Argument & Computation -1 (2023) 1–46
DOI 10.3233/AAC-220012
IOS Press

Argumentation with justified preferences

Sung-Jun Pyon
Faculty of Philosophy, Kim Il Sung University, Pyongyang, The Democratic People's Republic of Korea
E-mail: phil2@ryongnamsan.edu.kp

Abstract. It is often necessary and reasonable to justify preferences before reasoning from them. Moreover, justifying a preference ordering is reduced to justifying the criterion that produces the ordering. This paper builds on the well-known ASPIC+ formalism to develop a model that integrates justifying qualitative preferences with reasoning from the justified preferences. We first introduce a notion of preference criterion in order to model the way in which preferences are justified by an argumentation framework. We also adapt the notion of argumentation theory to build a sequence of argumentation frameworks, in which an argumentation framework justifies preferences that are to underlie the next framework. That is, in our formalism, preferences become not only an input of an argumentation framework, but also an output of it. This kind of input-output process can be applied in the further steps of argumentation. We also explore some interesting properties of our formalism.

Keywords: Defeasible reasoning, ASPIC+, justifying preferences, reasoning from preferences, reasoning about preferences, rule-based systems

1. Introduction

Argumentation is one of the mainstream approaches dealing with inconsistent information in intelligent systems. Dung's argumentation framework (AF) consists of a set of arguments and a binary attack relation between them [11,31]. A *semantics* is used for identifying acceptable arguments and drawing plausible conclusions. A set of arguments identified by a semantics as acceptable is called an *extension*. The attack relation in an AF can represent inconsistency of information and an extension identified by a semantics represents the ability of an AF to model inferences to plausible conclusion under inconsistent information.

In the field of argumentation theory, there has been a general consensus that arguments may not have the same strength [3,17] and the preferences should be considered in the evaluation of the successfulness of argument attacks [56,59]. Preferences are so important in evaluating arguments that preference-based (and value-based) AFs (PAFs) cover a crucial part of the spectrum of the existing argumentation formalisms. In a PAF, a preference ordering over the set of arguments is set to filter out the attack relation between arguments. A semantics is applied to the remained attacks. When we set a preference ordering over the set of arguments of an AF, some decision problems that may be intractable in the standard AF become very easier to solve. Preference-based (and value-based) AFs can be employed in the wide range of applications and domains such as merging conflicting knowledge bases [5], modeling dialogues [4,49], practical reasoning [13,57], legal reasoning [14,16] and even moral reasoning [10,63].

In this paper, our primary concern is to develop a model that enables not only reasoning from preferences but also justifying the preferences before applying them to reasoning. In practice, different people have different preferences and often a different context calls for different preferences. Therefore, any given preference ordering is not a 'universally accepted assumption' and may be questioned. This advises a rational agent to justify preferences before applying them to reasoning.

Indeed, every preference ordering comes from the criteria behind it. Any preference ordering is based on comparative evaluations, which "cannot begin until you come up with one or more classes or categories to which the objects of comparison can belong." Then, the class or category will provide the criterion for the comparative evaluation. Such a criterion "may amount to an ideal definition of the class" [16, p.244]. Therefore, justifying a preference ordering is reduced to justifying the criteria. Several papers have emphasized the need to support a given preference ordering in an AF [15,38,43]. Furthermore, there may be more than one criterion behind a preference ordering. That is why multi-criteria decision-making (MCDM) is dominant in the field. Multiple criteria may cooperatively produce a single preference ordering over different objects.

Although justifying multiple preference criteria before applying them to reasoning follows a general pattern by which we think, such formalisms have not been widely studied. Most of the existing preference-based AFs take a preference ordering over the set of arguments as its input, but make no justifications for the preference ordering or criteria behind it. The audience-dependency of selecting a preference (value order) was addressed by Bench-Capon and his colleagues in their value-based AF [13,15]. Thus, in their framework, a specific audience, which is defined as a total ordering over a set of values, can be regarded as a criterion from which a preference ordering over arguments is produced. Modgil has integrated reasoning about preferences with reasoning from the preferences mainly in his so-called extended argumentation framework [43,44]. He employed recursive attacks as the means of reasoning about preferences, thus his framework enables not only reasoning from preferences, but also reasoning about preferences. Sedki [58] suggests the use of an AF for preference elicitation with an approach based on the association between a given *PQCL* (Prioritized Qualitative Choice Logic [18]) theory and a value-based AF.

Integrating justification for preferences with reasoning from the justified preferences has been done in the field of recommender systems. Teze et al. proposed a recommender system with dynamic multiple criteria by extending DeLP-server [60]. Their system embeds multiple preference criteria in a DeLP-query and justifies them by generating a derivation from a DeLP-program. However, the selection of an appropriate criterion is derived from facts and strict rules. In other words, the selection of a criterion in their system is based on perfect information, thus can never be doubted. However, in everyday and legal argumentation, the selection of a criterion can often be doubted and be in conflict. Like all human judgments and actions, the selection of a criterion may be based on imperfect, especially inconsistent information rather than perfect and consistent information. Since argumentation is an effective approach dealing with inconsistent information, we can justify a preference criterion by means of an argumentation-based approach.

The legal reasoning, where a lawyer must demonstrate the priority of a legal norm to other conflicting norms [38], backs up our idea of this integration. In our proposal, an AF is not only used for drawing plausible conclusions, but also used for justifying preferences. Therefore, a preference ordering over a set of arguments becomes both output and input of an AF. In order to represent the condition under which a preference criterion is justified, we borrow the notion of *guard* from [60]. The guard of a preference criterion is defined as the information which must be drawn from an AF in order to justify the criterion. If the guard of a criterion is empty, the preference criteria can be said to be always justified, thus, applicable, under any condition.

To develop a model that enables not only reasoning from preferences but also justifying preferences, we choose ASPIC+, which is structured and preference-based, as the basic formal framework [39,46,55]. ASPIC+ has been proposed on the basis of several former works on rule-based argumentation [54,56,64] cognitive science [52,53], classical logics and contemporary argumentation schemes [62,65–67].

It has also been proven that this formalism satisfies rationality postulates stipulated in [27] even when applying preferences under some assumptions.[1] We modify the notion of argumentation theory with the concept of *preference criterion* to develop a model which enables not only reasoning from preferences, but also justifying the preferences. Then we devise a sequence of PAFs built over the theory, in which a PAF justifies some preference criteria that are to be also taken as an input of the next PAF.

Traditional preference-based argumentation formalisms including the ASPIC+ framework have only one repairing step where the attack relation is filtered through preferences. But our proposal has two or more repairing steps. One, where the attack relation is filtered through preferences whose criteria have empty guard, is for selecting appropriate preference criteria. The other, where the attack relation is filtered through justified preferences, is for drawing plausible conclusions taking into account the justified preferences. Then, the justified preferences are the output of the first step and the input of the second step. If repairing an AF twice through preferences is not enough for resolving conflicts between extensions, then our framework enables further repairing steps.

The remainder of this paper is organized as follows: Section 2 briefly reviews Dung's abstract AF and PAF and some basic definitions of the ASPIC+ formalism are together with it interesting property. In Section 3, we give some motivations for developing the model that is able to not only reason from preferences but also justify the preferences. Section 4 is dedicated to integrating justification for preference criteria with reasoning from the justified preferences. In Section 5, the results and properties of our formalism are deeply discussed with some decision-making examples. Section 6 shows some related works, and finally we conclude the paper in Section 7.

2. Preliminaries

In this section, we briefly review preference-based AF that is an extended version of Dung's abstract AF and the well-known ASPIC+ framework.

2.1. Preference-based argumentation framework

Let us first recall Dung's notion of abstract argumentation framework. An abstract argumentation framework (AF) is a pair $\mathcal{H} = \langle \mathcal{A}, \mathcal{R} \rangle$ where $\mathcal{A}$ is a set of arguments and $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is a binary attack relation among the arguments [31]. We say that an argument A attacks an argument B iff $\langle A, B \rangle \in \mathcal{R}$ (denoted by $A\mathcal{R}B$). A set of arguments $\mathcal{S} \subseteq \mathcal{A}$ is said to be *conflict-free* wrt. $\mathcal{R}$ iff $\nexists A, B \in \mathcal{S}$, such that $\langle A, B \rangle \in \mathcal{R}$. A set of arguments $\mathcal{S} \subseteq \mathcal{A}$ is also said to *defend* an argument $A \in \mathcal{A}$ iff for all $B \in \mathcal{A}$, such that $\langle B, A \rangle \in \mathcal{R}$, there exists a $C \in \mathcal{S}$, such that $\langle C, B \rangle \in \mathcal{R}$. We write $\mathtt{F}(\mathcal{S})$ for the set of all arguments that $\mathcal{S}$ defends. Then, the acceptability semantics over a set of arguments are defined as follows:

- A set $\mathcal{S} \subseteq \mathcal{A}$ of arguments is an *admissible* set iff it is conflict-free and defends all its elements;
- A set $\mathcal{S} \subseteq \mathcal{A}$ of arguments is a *complete* extension iff $\mathcal{S}$ is conflict-free and $\mathcal{S} = \mathtt{F}(\mathcal{S})$;
- A set $\mathcal{S} \subseteq \mathcal{A}$ of arguments is a *preferred* extension iff it is a maximal complete extension (wrt. set inclusion);
- A set $\mathcal{S} \subseteq \mathcal{A}$ of arguments is a *grounded* extension iff it is the minimal complete extension (wrt. set inclusion);
- A set $\mathcal{S} \subseteq \mathcal{A}$ of arguments is a *stable* extension iff it is conflict-free and for all $A \in \mathcal{A} \backslash \mathcal{S}$, there exists a $B \in \mathcal{S}$ such that $B\mathcal{R}A$;

[1] The former ASPIC framework satisfies rationality postulates only when preferences are not taken into account.

Remember that given an AF $\mathcal{H} = \langle \mathcal{A}, \mathcal{R} \rangle$, function $\mathtt{Ext}_{\mathtt{y}}(\mathcal{H})$ returns the set of all extensions under $\mathtt{y}$ semantics, where $\mathtt{y} \in \{\mathtt{pre}, \mathtt{sta}, \mathtt{gro}, \mathtt{com}, \mathtt{adm}\}$ and $\mathtt{pre}$ (respectively $\mathtt{sta}$, $\mathtt{gro}$, $\mathtt{com}$, $\mathtt{adm}$) means preferred (respectively stable, ground, complete, admissible) semantics.

In several works, the abstract AF has been extended into preference-based AFs (PAF) by adding a preference ordering over a set of arguments in order to model the generally accepted idea that arguments may not be equally preferred [3,17] and that argument preferences should be taken into account while determining whether an attack is successful or failed [56,59].

Formally, a PAF is a triple $\mathcal{H} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq \rangle$ where $\mathcal{A}$ is a set of arguments, $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ is an attack relation and $\succcurlyeq$ is a (partial or total) preference ordering over $\mathcal{A}$ [1]. Let $A, B \in \mathcal{A}$ be arguments. $A \succcurlyeq B$ means that argument A is at least as preferred as B. The strict counterpart of $\succcurlyeq$ is $\succ$. Note that AF $\langle \mathcal{A}, \mathcal{R} \rangle$ is a special case of PAF $\langle \mathcal{A}, \mathcal{R}, \succcurlyeq \rangle$ where $\succcurlyeq = \emptyset$. Thus, we will use the notion of PAF instead of AF, throughout this paper.

A preference ordering over the set of arguments of an AF is useful for *filtering* the attack relation, that is, removing the (preference-dependent) attacks where the attackee is preferred to the attacker.[2] Then, for the sake of applying the acceptability semantics to the filtered attack relation, it is necessary to define the notion of the *repaired* AF of a PAF. The repaired AF of a PAF $\langle \mathcal{A}, \mathcal{R}, \succcurlyeq \rangle$ is $\langle \mathcal{A}, \mathcal{R}' \rangle$ with $\mathcal{R}' = \{(a, b) | (a, b) \in \mathcal{R} \text{ and not } b \succ a\}$. The extensions of $\langle \mathcal{A}, \mathcal{R}, \succcurlyeq \rangle$ under a given semantics are Dung extensions of the repaired version $\langle \mathcal{A}, \mathcal{R}' \rangle$ under the same semantics. Furthermore, $\langle \mathcal{A}, \mathcal{R} \rangle$ is called the *original* version of a PAF $\langle \mathcal{A}, \mathcal{R}, \succcurlyeq \rangle$.

Although, by means of PAFs, we are able to not only model real-world argumentation, but also make many decision problems that have been proved to be intractable in Dung-style AFs much easier, such an approach gives rise to an unintuitive result. According to this approach, if one argument asymmetrically attacks another, but fails, then these two arguments become conflict-free. This is problematic since whether an attack is successful or failed is irrelevant to determining a conflict. A conflict between arguments is only the matter of their incompatiblity [46]. As a result, some formalisms that takes preferences or values into account lead to violating rationality postulates stipulated in [27].

To address this shortcoming, three main solutions have been proposed. These proposals aim to guarantee conflict-freeness of a PAF extension whether or not an attack is successful. One solution is to add preferences at semantics level, not attack level [7]. Another suggestion is to inverse the direction of failed and asymmetric attacks [9]. The ASPIC+ formalism has found the way to avoid losing the conflict-freeness of an extension in distinguishing between preference-dependent and preference-independent attacks [45,46,55]. Amgoud & Vesic [7] argue that their approach is more general than ASPIC+. However, everything has merits and demerits and we have found that, ASPIC+'s way of distinguishing between preference-dependent and preference-independent attacks has a very useful property which we will proceed to study in the next section.

2.2. *ASPIC+ framework*

In ASPIC+ framework, arguments are defined as inference trees formed by applying two kinds of inference rules: strict rules representing generalizations which do not allow exceptional cases like "A mammal is an animal" and defeasible rules representing commonly-hold generalizations with exceptional cases like "Mammals generally live on land". This naturally leads to three ways of attacking an

[2] In ASPIC+, only preference-dependent attacks can be removed (see the next subsection). However, in other formalisms such as deductive argumentation [9] or assumption-based argumentation with preferences (ABA+) [28], any attack should be removed from the framework if the attackee is preferred to the attacker and then the reversed attack is added.

argument: attacking a premise, a conclusion and an inference, which are respectively called *undermining*, *rebutting* and *undercutting*. In order to characterize the structure of arguments and the nature of the attack relation, we need to make a minimal assumption on the logical language that certain well-formed formulae are a contrary or contradictory of certain other well-formed formulae. Now, let us present some basic definitions of the framework.

Definition 1 (Argumentation system [46])**.** An argumentation system (AS) is a tuple $\langle \mathcal{L}, \bar{}, \mathcal{R}, \leqslant, \mathtt{n} \rangle$ where

- $\mathcal{L}$ is a logical language.
- $\bar{}$ is a contrariness function from $\mathcal{L}$ to $2^{\mathcal{L}}$, such that:
  - if $a \in \overline{b}$, and $b \notin \overline{a}$, then a is called a contrary of b (usually denoted by $a =\sim b$), and
  - else if $a \in \overline{b}$, and $b \in \overline{a}$, a and b are called contradictory (usually denoted by $a = \neg b$), and
  - each $a \in \mathcal{L}$ has at least one contradictory.
- $\mathcal{R} = \mathcal{R}s \cup \mathcal{R}d$ is a set of strict ($\mathcal{R}s$) and defeasible ($\mathcal{R}d$) inference rules such that $\mathcal{R}s \cap \mathcal{R}d = \emptyset$.
- $\leqslant$ is a partial preorder on $\mathcal{R}d$. The strict counterpart of $\leqslant$ is $<$.
- $\mathtt{n}$: $\mathcal{R}d \to \mathcal{L}$ is naming convention for defeasible rules. Informally, $\mathtt{n}(r)$ is a well formed formula in $\mathcal{L}$ which says that the defeasible rule $r \in \mathcal{R}$ is applicable. Note that if a rule is of the form $x_1, \ldots, x_n \to \overline{n(r)}$, such that $r \in \mathcal{R}d$ (or $x_1, \ldots, x_n \Rightarrow \overline{n(r)}$), then it should be read as follows: If $x_1, \ldots, x_n$ hold, then r is not (generally) applicable.

A set $\mathcal{S} \subseteq \mathcal{L}$ is said to be *consistent* iff $\nexists a, b \in$S such that $a = \overline{b}$, otherwise it is called *inconsistent*. Let $\mathcal{R}s$ be a set of strict rules. A set $\mathcal{S} \subseteq \mathcal{L}$ is said to be indirectly consistent (wrt. $\mathcal{R}s$) iff the closure of $\mathcal{S}$ under $\mathcal{R}s$ is consistent.

The framework also defines the notion of knowledge base from which arguments can be constructed, inspired by [37] where they classified premises of an argument or an argumentation scheme into several categories.

Definition 2 (Knowledge base [46])**.** A knowledge base of an AS $\langle \mathcal{L}, \bar{}, \mathcal{R}, \leqslant, \mathtt{n} \rangle$ is a pair $\langle \mathcal{K}, \leqslant' \rangle$, where $\mathcal{K} \subseteq \mathcal{L}$ consists of two disjoint subsets, i.e., $\mathcal{K} = \mathcal{K}n \cup \mathcal{K}p$ and $\leqslant'$ is a partial preorder on $\mathcal{K}p$

- $\mathcal{K}n$ is a set of (necessarily true) axioms. Intuitively, arguments cannot be attacked on their axiom premises.
- $\mathcal{K}p$ is a set of ordinary premises. Intuitively, arguments can be attacked on their ordinary premises and whether this results in defeat must be determined by comparing the attacker and the attacked premise.

The ASPIC+ framework defines the notion of an argument as an inference tree formed by applying strict and defeasible rules to premises that are well-formed formulae. Below, the ASPIC+'s definition of an argument can be found. For any argument A, $\mathtt{Prem}(A)$ returns all the formulas of $\mathcal{K}$ (called premises) used to build A, $\mathtt{Conc}(A)$ returns A's conclusion, $\mathtt{Sub}(A)$ returns all of A's subarguments, $\mathtt{DefRules}(A)$ and $\mathtt{StRules}(A)$ respectively return all of defeasible and strict rules in A and $\mathtt{TopRule}(A)$ returns the lastly applied rule in A.

Definition 3 (Arguments [55])**.** An argument A on the basis of a knowledge base $\langle \mathcal{K}, \leqslant' \rangle$ in an AS $\langle \mathcal{L}, \bar{}, \mathcal{R}, \leqslant, \mathtt{n} \rangle$ is

(1) φ if $\varphi \in \mathcal{K}$ with $\mathtt{Prem}(A) = \{\varphi\}$, $\mathtt{Conc}(A) = \varphi$, $\mathtt{Sub}(A) = \{\varphi\}$, $\mathtt{DefRules}(A) = \emptyset$, $\mathtt{StRules}(A) = \emptyset$, $\mathtt{TopRule}(A) =$ undefined.

(2) $A_1, \ldots, A_n \to \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a strict rule $\mathtt{Conc}(A_1), \ldots, \mathtt{Conc}(A_n) \to \psi$ in $\mathcal{R}s$,
$\mathtt{Prem}(A) = \mathtt{Prem}(A_1) \cup \cdots \cup \mathtt{Prem}(A_n)$,
$\mathtt{Conc}(A) = \psi$
$\mathtt{Sub}(A) = \mathtt{Sub}(A_1) \cup \cdots \cup \mathtt{Sub}(A_n) \cup \{A\}$,
$\mathtt{DefRules}(A) = \mathtt{DefRules}(A_1) \cup \cdots \cup \mathtt{DefRules}(A_n)$,
$\mathtt{StRules}(A) = \mathtt{StRules}(A_1) \cup \cdots \cup \mathtt{StRules}(A_n) \cup \{\mathtt{Conc}(A_1), \ldots, \mathtt{Conc}(A_n) \to \psi\}$,
$\mathtt{TopRule}(A) = \mathtt{Conc}(A_1), \ldots, \mathtt{Conc}(A_n) \to \psi$.

(3) $A_1, \ldots, A_n \Rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a defeasible rule $\mathtt{Conc}(A_1), \ldots, \mathtt{Conc}(A_n) \Rightarrow \psi$ in $\mathcal{R}d$,
$\mathtt{Prem}(A) = \mathtt{Prem}(A_1) \cup \cdots \cup \mathtt{Prem}(A_n)$,
$\mathtt{Conc}(A) = \psi$
$\mathtt{Sub}(A) = \mathtt{Sub}(A_1) \cup \cdots \cup \mathtt{Sub}(A_n) \cup \{A\}$,
$\mathtt{DefRules}(A) = \mathtt{DefRules}(A_1) \cup \cdots \cup \mathtt{DefRules}(A_n) \cup \{\mathtt{Conc}(A_1), \ldots, \mathtt{Conc}(A_n) \Rightarrow \psi\}$,
$\mathtt{StRules}(A) = \mathtt{StRules}(A_1) \cup \cdots \cup \mathtt{StRules}(A_n)$,
$\mathtt{TopRule}(A) = \mathtt{Conc}(A_1), \ldots, \mathtt{Conc}(A_n) \Rightarrow \psi$.

In addition, an argument is called *strict* iff $\mathtt{DefRules}(A) = \emptyset$; *defeasible* iff $\mathtt{DefRules}(A) \neq \emptyset$; *firm* iff $\mathtt{Prem}(A) \subseteq \mathcal{K}n$; *plausible* if $\mathtt{Prem}(A) \not\subseteq \mathcal{K}n$. Notice also that for any argument A, $\mathtt{Prem}_{\mathtt{n}}(A) = \mathtt{Prem}(A) \cap \mathcal{K}n$, and $\mathtt{Prem}_{\mathtt{p}}(A) = \mathtt{Prem}(A) \cap \mathcal{K}p$.

Example 1. Consider a knowledge base $\mathcal{KB}_1 = \langle \mathcal{K}n_1 \cup \mathcal{K}p_1, \leqslant'_1 \rangle$ in an argumentation system $\mathcal{AS}_1 = \langle \mathcal{L}, \bar{}, \mathcal{R}s_1 \cup \mathcal{R}d_1, \leqslant_1, \mathtt{n} \rangle$ such that:

- $\mathcal{L}$ is a propositional language which consists of a set of propositional atoms $\{p, q, r, \ldots\}$ and the symbols $\neg$ and $\sim$ respectively denoting strong and weak negation (i.e., negation as failure). α is a strong literal if α is a propositional atom or of the form $\neg\beta$ where β is a propositional atom. α is a wff. of $\mathcal{L}$, if α is a strong literal or of the form $\sim \beta$ where β is a strong literal.
- $\alpha \in \overline{\beta}$ iff (1) α is of the form $\neg\beta$ or β is of the form $\neg\alpha$; or (2) β is of the form $\sim \alpha$ (i.e., for any wff. α, α and $\neg\alpha$ are contradictories and α is a contrary of $\sim \alpha$).
- $\mathcal{R}s_1 = \{t \to \neg\mathtt{n}(p \Rightarrow s)\}$
- $\mathcal{R}d_1 = \{p \Rightarrow s; q \Rightarrow t; r \Rightarrow u; u \Rightarrow \neg t; s \Rightarrow v; t \Rightarrow \neg v\}$
- $\mathcal{K}p_1 = \{q, r, \}$
- $\mathcal{K}n_1 = \{p\}$
- $\leqslant'_1 = \leqslant_1 = \emptyset$

We construct 10 arguments from $\mathcal{KB}_1$: $A_1 = p$; $A_2 = q$; $A_3 = r$; $A_4 = A_1 \Rightarrow s$; $A_5 = A_2 \Rightarrow t$; $A_6 = A_3 \Rightarrow u$; $A_7 = A_4 \Rightarrow v$; $A_8 = A_5 \Rightarrow \neg v$; $A_9 := A_6 \Rightarrow \neg t$; $A_{10} = A_5 \Rightarrow \neg\mathtt{n}\,(p \Rightarrow s)$.

Given two argumentation theories $\langle \mathcal{AS}_1, \mathcal{KB}_1 \rangle$ and $\langle \mathcal{AS}_2, \mathcal{KB}_2 \rangle$, $\langle \mathcal{AS}_1, \mathcal{KB}_1 \rangle \subseteq \langle \mathcal{AS}_2, \mathcal{KB}_2 \rangle$ iff $\mathcal{AS}_1 \subseteq \mathcal{AS}_2$ and $\mathcal{KB}_1 \subseteq \mathcal{KB}_2$.

As we can see in Definition 3, since arguments are inference trees, three kinds of argument attacks are possible: *undermining*, *rebutting* and *undercutting* attack.

Definition 4 (Attack [46])**.**

- Argument A undercuts argument B (on B') iff $\texttt{Conc}(A) \in \overline{n(r)}$ for some $B' \in \texttt{Sub}(B)$ such that the top rule of B' is defeasible.
- Argument A rebuts argument B (on B') iff $\texttt{Conc}(A) \in \overline{\varphi}$, for some $B' \in \texttt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \varphi$. In such a case, A contrary-rebuts B iff $\texttt{Conc}(A)$ is a contrary of φ and contradictory-rebuts B iff $\texttt{Conc}(A)$ is a contradictory of φ.
- Argument A undermines argument B (on φ) iff $\texttt{Conc}(A) \in \overline{\varphi}$ for some $\varphi \in \texttt{Prem}(B) \backslash \mathcal{K}n$. In such a case, argument A contrary-undermines B iff $\texttt{Conc}(A)$ is a contrary of φ. In addition, it is said that A contradictory-undermines B iff $\texttt{Conc}(A)$ is a contradictory of φ.

Example 1 (cont.)**.** A_{10} undercuts A_4 and A_7. And A_7 rebuts A_8 on A_8, thus they rebut each other (symmetric rebut). Exactly the same happens between A_9 and A_5. In addition, A_9 rebuts A_8 on A_5, but it is not the case that A_8 rebuts A_9. Therefore, A_8 and A_9 are in asymmetric contradictory-rebut.

In the above definition, argument attacks are divided into three categories according to what part of the attackee the attacker attacks on. Moreover, the ASPIC+ framework distinguishes preference-dependent attacks from preference-independent attacks. If an attack is one of undercutting, contrary-rebutting or contrary-undermining attack, it is a preference-independent attack, and otherwise it is a preference-dependent attack. Therefore, it seems likely to us that Kaci et al.'s famous formula "defeat = conflict + preference"[3] [41] holds only for contradictory-rebuts and contradictory-undermines and does not hold for undercut, contrary-rebut and contrary-undermine.

Example 1 (cont.)**.** As one can notice, A_{10}'s attacks on A_4 and A_7 are preference-independent, whereas rebuts between A_7 and A_8 and between A_5 and A_9 are preference-dependent. It is remarkable that A_9's attack on A_8 is preference-dependent although it is asymmetric.

The notion of argument defeat can be defined on the basis of the definition of preference-dependent and preference-independent attack as follows:

Definition 5 (Defeat [46])**.** Let A and B be arguments. Then A defeats B on B' iff

(1) A undercuts, contray-undermines, or contrary rebuts B on B', or
(2) A rebuts B on B' and $A \not\prec B'$, or
(3) A undermines B on φ and $A \not\prec \varphi$.

A strictly defeats B iff A defeats B, but B does not defeat A.

Below, we define the notion of argumentation theory that is the basis for constructing arguments and determining attack relations among the arguments.

Definition 6 (Argumentation theory [46])**.** An argumentation theory is a pair $\langle \mathcal{AS}, \mathcal{KB} \rangle$, where $\mathcal{AS}$ is an argumentation system, $\mathcal{KB}$ a knowledge base in $\mathcal{AS}$.

Finally, argumentation theories can be linked to PAFs.

[3] In [41], actually, "attack = conflict + preference" appears, but in this paper, we replace the term "attack" with "defeat", because in the original paper, the term "attack" stands for "successful attack", not "failed attack".

Definition 7 (PAF corresponding to an argumentation theory [46])**.** A PAF corresponding to an argumentation theory $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ is a triple $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle$ where $\mathcal{A}$ is the set of arguments on the basis of $\mathcal{T}$ as defined by Definition 3, $\mathcal{R}$ is the binary attack relation on $\mathcal{A}$ defined by Definition 4 and $\preccurlyeq$ is a preference ordering on $\mathcal{A}$.

Given an argumentation theory $\mathcal{T}$, remember that $\texttt{Arg}(\mathcal{T})$ returns the set of all finite arguments built from $\mathcal{T}$.

An argument ordering is a partial preorder $\preccurlyeq$ on arguments (whose strict counterpart is $\prec$) and is *admissible* iff firm and strict arguments are strictly preferred to defeasible or plausible ones and a strict inference rule cannot make an argument weaker or stronger. In this paper, we do not include an argument ordering $\preccurlyeq$ in the notion of argumentation theory as in [55] because it always follows from the underlying argumentation system and knowledge base. We rather include the argument ordering in the notion of PAF corresponding to an argumentation theory as in [46]. Such an inclusion is more appropriate to develop a model which integrates justifying preferences with reasoning from the justified preferences.

An AF $\langle \mathcal{A}, \mathcal{R}' \rangle$ where $\mathcal{A}$ is the set of arguments on the basis of an argumentation theory as defined by Definition 3 and $\mathcal{R}'$ is the binary defeat relation on $\mathcal{A}$ as defined by Definition 5 is the *repaired* version of $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle$. Moreover, $\langle \mathcal{A}, \mathcal{R} \rangle$ is called the *original* AF corresponding to the theory.

Generally, two ways of deriving argument orderings from orderings on rules or ordinary premises (last-link and weakest-link principles) have been recognized. Those two principles employ a general definition of a partial order $\trianglelefteq_{\texttt{s}}$ on sets in terms of a partial preorder $\leqslant_e$ on their elements as follows [46]:

(1) if $\mathcal{S}_1 = \emptyset$, then $\mathcal{S}_1 \trianglelefteq_{\texttt{s}} \mathcal{S}_2$;
(2) if $\mathcal{S}_1 = \emptyset$ and $\mathcal{S}_2 \neq \emptyset$, then $\mathcal{S}_1 \trianglelefteq_{\texttt{s}} \mathcal{S}_2$;
(3) if $\texttt{s} = \texttt{Eli}$, then $\mathcal{S}_1 \trianglelefteq_{\texttt{Eli}} \mathcal{S}_2$ iff there exists a $e_1 \in \mathcal{S}_1$ such that for all $e_2 \in \mathcal{S}_2$, it holds that $e_1 \leqslant_e e_2$;
(4) if $\texttt{s} = \texttt{Dem}$, then $\mathcal{S}_1 \trianglelefteq_{\texttt{Dem}} \mathcal{S}_2$ iff there exists a $e_2 \in \mathcal{S}_2$ such that for all $e_1 \in \mathcal{S}_1$, it holds that $e_1 \leqslant_e e_2$.

The last-link principle prefers an argument A over another argument B if the last defeasible rules used in B are less preferred (wrt. $\prec_{\texttt{s}}$) than those in A or, in case both arguments are strict, if the premises of B are less preferred (wrt. $\prec_{\texttt{s}}$) than the premises of A. The weakest-link principle considers all uncertain elements in an argument unlike the last-link principle. The weakest-link principle prefers argument A to B if A is preferred to B with respect to $\trianglelefteq_{\texttt{s}}$ on both their premises and defeasible rules. Below, the function $\texttt{LastDefRules}(A)$ returns last defeasible rules of argument A as follows: (1) $\texttt{LastDefRules}(A) = \emptyset$ iff $\texttt{DefRules}(A) = \emptyset$; (2) If A is of the form $A_1, \ldots, A_n \Rightarrow \varphi$, then $\texttt{LastDefRules}(A) = \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \varphi\}$, otherwise $\texttt{LastDefRules}(A) = \texttt{LastDefRules}(A_1) \cup \ldots \texttt{LastDefRules}(A_n)$.

Definition 8 (Last-link and weakest-link principles [46])**.**

- **Last-link principle.** Let A and B be two arguments. Then $A \prec B$ iff either:

  (1) $\texttt{LastDefRules}(A) \trianglelefteq_{\texttt{s}} \texttt{LastDefRules}(B)$ or
  (2) $\texttt{LastDefRules}(A)$ and $\texttt{LastDefRules}(B)$ are empty and $\texttt{Prem}(A) \trianglelefteq_{\texttt{s}} \texttt{Prem}(B)$.

- **Weakest-link principle.** Let A and B be two arguments. Then $A \prec B$ iff:

  (1) if both A and B are strict, then $\texttt{Prem}(A) \trianglelefteq_{\texttt{s}} \texttt{Prem}(B)$, else;

(2) if both A and B are firm, then $\texttt{DefRules}(A) \trianglelefteq_{\texttt{s}} \texttt{DefRules}(B)$, else;
(3) $\texttt{Prem}(A) \trianglelefteq_{\texttt{s}} \texttt{Prem}(B)$ and $\texttt{DefRules}(A) \trianglelefteq_{\texttt{s}} \texttt{DefRules}(B)$.

Modgil and Prakken also define the notion of maximal fallible subarguments and strict continuations of arguments that are useful for proving some propositions (An argument is fallible if it is plausible or defeasible). The maximal fallible subarguments of an argument are those with the 'last' defeasible inferences in that argument or else (if the argument is strict) they are the argument's ordinary premises [46].

Definition 9 (Maximal fallible subargument [46])**.** The set $\texttt{M}(A)$ of the maximal fallible subarguments of an argument A is defined such that for any $A' \in \texttt{Sub}(A)$, $A' \in \texttt{M}(A)$ iff:

(1) the top rule of A' is defeasible or A' is an ordinary premise, and;
(2) there is no $A'' \in \texttt{Sub}(A)$ such that $A'' \neq A$ and $A' \in \texttt{Sub}(A'')$, and A' satisfies the (1).

Definition 10 (Strict continuations of arguments [46])**.** For any set of arguments $\{A_1, \ldots, A_n\}$, the argument A is a strict continuation of $\{A_1, \ldots, A_n\}$ iff:

(1) $\bigcup_{i=1}^{n} \texttt{Prem}_{\texttt{p}}(A_i) = \texttt{Prem}_{\texttt{p}}(A)$ (i.e., the ordinary premises in A are exactly those in $\{A_1, \ldots, A_n\}$);
(2) $\bigcup_{i=1}^{n} \texttt{DefRules}(A_i) = \texttt{DefRules}(A)$ (i.e., the defeasible rules in A are exactly those in $\{A_1, \ldots, A_n\}$);
(3) $\bigcup_{i=1}^{n} \texttt{StRules}(A_i) \subseteq \texttt{StRules}(A)$ and $\bigcup_{i=1}^{n} \texttt{Prem}_{\texttt{n}}(A_i) \subseteq \texttt{Prem}_{\texttt{n}}(A)$ (i.e., the strict rules and axiom premises of A are a superset of the strict rules and axiom premises in $\{A_1, \ldots, A_n\}$).

It has been shown in [46] that ASPIC+ satisfies the postulates of *Closure* under subarguments and strict rule application unconditionally. However, the postulates of *Direct Consistency* and *Indirect Consistency* hold only under the assumption of *reasonable* argument ordering. An argument ordering is reasonable if it satisfies properties that one might expect to hold of orderings over arguments composed from fallible and infallible elements [55]. Formally:

Definition 11 (Reasonable argument ordering [46])**.** An argument ordering $\preccurlyeq$ is *reasonable* iff:

(1) for all A and B such that A is strict and firm and B is plausible or defeasible, it holds that $B \prec A$;
(2) for all A and B such that B is strict and firm, it holds that $B \not\prec A$.
(3) for all A, A' and B such that A' is a strict continuation of $\{A\}$, if $A \not\prec B$, then $A' \not\prec B$, and if $B \not\prec A$ then $B \not\prec A'$ (i.e. applying strict rules to a single argument's conclusion and possibly adding new axiom premises does not weaken, respectively, strengthen, arguments).
(4) Let $\{C_1, \ldots, C_n\}$ be a finite subset of $\mathcal{A}$, and for $i = 1, \ldots, n$, let $C^{+\setminus i}$ be some strict continuation of $\{C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n\}$. Then, it is not the case that for all i, $C^{+\setminus i} \prec C^i$.

An argumentation theory is said to be (*directly*) *consistent* if the set of conclusions of all arguments in an arbitrary extension of the PAF built over the theory is consistent; an argumentation theory is *indirectly consistent* if the closure of the set of conclusions of all arguments in an arbitrary extension of the PAF built over the theory under strict rule application is consistent [46,55].

Now, it is time to define the notion of the output of a PAF whose input is an argumentation theory. For any argumentation theory and the PAF corresponding to the theory, let $\texttt{Ext}_{\texttt{y}}(\mathcal{H}) = \{\mathcal{E}_1, \ldots, \mathcal{E}_n\}$ ($n \geqslant 1$). The output of the PAF (or the output of the argumentation theory) is the set of acceptable conclusions under the given semantics. Credulous and skeptical viewpoints are possible.

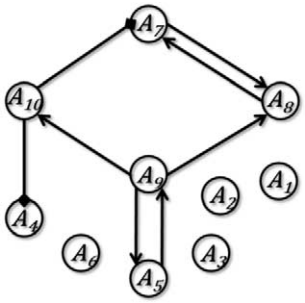

Fig. 1. Preference-dependent and preference-independent attacks.

- $\texttt{Concs}(\mathcal{E}i) = \{\texttt{Conc}(A) | A \in \mathcal{E}i\}$ $(i \in \{1, \ldots, n\})$.
- The set of credulously justified conclusions is $\texttt{COutput}_{\texttt{Y}}(\mathcal{H}) = \bigcup_{i=1,\ldots,n} \texttt{Concs}(\mathcal{E}i)$.
- The set of skeptically justified conclusions is $\texttt{SOutput}_{\texttt{Y}}(\mathcal{H}) = \bigcap_{i=1,\ldots,n} \texttt{Concs}(\mathcal{E}i)$.

Remember that $\texttt{Output}_{\texttt{Y}}(\mathcal{H})$ can be either $\texttt{SOutput}_{\texttt{Y}}(\mathcal{H})$ or $\texttt{COutput}_{\texttt{Y}}(\mathcal{H})$ according to the adopted viewpoint.

Interestingly, a preferred/stable extension of the PAF built over a standard ASPIC+ argumentation theory is also a preferred/stable extension of its original version or its subset. This property, which is very useful for developing a model which is capable of justifying preferences, can be formalized as follows:

Proposition 1. *Let $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle$ be a PAF corresponding to an argumentation theory which has a reasonable argument ordering. Then, under preferred semantics, it holds that for all $\mathcal{E}' \in \texttt{Ext}_{pre}(\langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle)$, there exists an extension $\mathcal{E} \in \texttt{Ext}_{pre}(\langle \mathcal{A}, \mathcal{R} \rangle)$ such that $\mathcal{E}' \subseteq \mathcal{E}$. Under stable semantics, it just holds that $\texttt{Ext}_{sta}(\langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle) \subseteq \texttt{Ext}_{sta}(\langle \mathcal{A}, \mathcal{R} \rangle)$.*

Example 1 (cont.)**.** In Fig. 1, graphical representation of the PAF built over the argumentation theory $\langle \mathcal{KB}_1, \mathcal{AS}_1 \rangle$ is given. Open arrows represent preference-dependent attacks and diamond arrows represent preference-independent attacks. Note that $\leqslant'_1 = \leqslant_1 = \emptyset$.

Now, let us extend $\leqslant'_1$, $\leqslant_1$ with preference information $r \Rightarrow u \prec q \Rightarrow t \prec u \Rightarrow \neg t \prec t \Rightarrow \neg v \langle p \Rightarrow s \prec s \Rightarrow v$, and $q \rangle r$. From the weakest-link principle we can infer that $A_9 \prec A_{10}$, $A_9 \prec A_5$, $A_9 \prec A_8$, $A_8 \prec A_7$. Therefore, the result of filtering the attack relation with the argument ordering can be depicted as follows (see Fig. 2).

The original AF has two preferred/stable extensions: $\{A_1, A_2, A_3, A_5, A_6, A_8, A_{10}\}$ and $\{A_1, A_2, A_3, A_4, A_6, A_7, A_9\}$, while the PAF has a preferred/stable extension $\{A_1, A_2, A_3, A_5, A_6, A_8, A_{10}\}$. This fact illustrates Proposition 1.

The above proposition shows that the preferences in a PAF corresponding to an ASPIC+ argumentation theory not only filter the attack relation but also filter the set of extensions under preferred/stable semantics. According to Amgoud & Vesic, preferences play two roles in an AF [8,9]. They may be used for handling an attack where the attackee is preferred to the attacker or may be used for refining the result of a PAF. Interestingly, in the ASPIC+ framework, preferences handle failed attacks and simultaneously, refine the set of extensions of the original AF under preferred/stable semantics.

Note that the refinement role which preferences play in ASPIC+ has more or less different meaning from that in Amgoud & Vesic's deductive argumentation. In [9], preferences are used for refining the

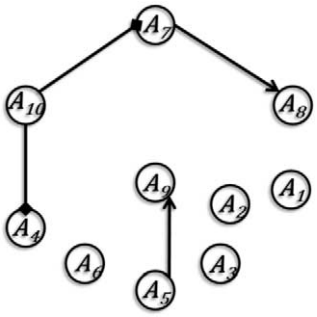

Fig. 2. Filtering extensions through preferences.

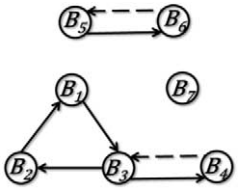

Fig. 3. Reduction in an extension.

result of a repaired AF, while the above proposition shows that preferences have the ability of refining the result of an original framework. Moreover, in Amgoud & Vesic's formalism, filtered extensions are exactly of those of a PAF. But, in our formalism, filtered extensions may get smaller, that is, lose some arguments as the result of filtering (under preferred semantics). Some extensions pass the *filter bed* of preferences without any *loss* of arguments and some fail to pass. We can also see that some extensions that pass the *filter bed* lose some arguments, and as a result, get smaller.

Example 2. Let us consider a PAF $\mathcal{H}_2$ depicted in Fig. 3 where $B_4 \prec B_3$ (the dashed arrows represent failed attack). The original AF has two preferred extensions $\{B_2, B_4, B_5, B_7\}$ and $\{B_2, B_4, B_6, B_7\}$, while the repaired AF also has two preferred extensions $\{B_5, B_7\}$ and $\{B_6, B_7\}$. So, both extensions of the original AF, which pass the filter bed of $B_4 \prec B_3$, lose two arguments B_2 and B_4.

One might think that preferences play the role of refining the extensions of an original framework in any preference-based argument formalism. However, several preferences-based formalisms [1,2,8,9,13, 28,43] fail to make the preference ordering over arguments play the role of refining the extensions of an original framework. For example, consider an AF formalized in [1] where argument A attacks B and B is preferred to A. The original framework has one preferred extension $\{A\}$, but the repaired framework that takes preferences into account has one preferred extension $\{A, B\}$. Thus, the preference $B \succ A$ has failed to refine the result of the original framework.

3. Preferences, preference criteria and justifying preferences

In a PAF, a preference ordering over the set of arguments is used for filtering the attack relation, in turn, producing a defeat relation and defining plausible conclusions. In most cases, such preferences over arguments come from the defeasible rule priorities those arguments are based on, as in ASPIC+

framework. However, why should we apply such orderings? Are all those orderings undeniable facts or axioms? Probably, most of them are not. Therefore, a rational agent may not only reason from preferences, but also have to justify those preferences. There are few preference orderings that are taken for granted. Most preference orderings should be the outcome of justification in order to be applied.

In practice, preference orderings over a set of objects vary from person to person. This audience-dependency of selecting a preference (value order) was already explored by Bench-Capon in their value-based AF [13,15]. Furthermore, preferences may vary from context to context even for a single agent. An agent has several preference orderings over a set of objects in its mind and applies them to reasoning. Then, which preference ordering to apply depends on the particularity of a given context. Therefore, in order to persuade others to do something or prove that something is right or good, we should first try to convince them of the preferences behind an advice or a claim that we propose. The context-dependency of preference application leads us into the topic of justifying preferences in an AF.

First of all, justifying a preference ordering is justifying the criteria behind it. Any preference ordering is based on comparative evaluations, which "cannot begin until you come up with one or more classes or categories to which the objects of comparison can belong." Then, the class or category will provide the criterion for the comparative evaluation. Such a criterion "may amount to an ideal definition of the class" [33, p. 244].

Regarding preference criteria, several points should be made clear. Not only one, but several preference criteria are applicable to a set of objects. For example, when we produce a preference ordering over houses, criteria such as size, distance to working place and location are applicable. And when we produce a preference ordering over the beauty of some things (aesthetic evaluation), multiple criteria such as proportion, slight distortion, contrast, harmony, craftsmanship, association and so on can be applied [33]. More importantly, preference orderings on the same set of objects differ according to the criteria. The preference ordering over a set of houses according to their size may be different from that according to the distance to working place. Incidentally, different context recommends different criteria. For this reason, even a single person's preferences may vary from context to context.

Second, two preference criteria may be incompatible over a set of objects. The criteria size and distance to working is said to be *incompatible* over a set of houses $\{h_1, h_2\}$ if h_1 is preferred to h_2 according to size, and h_2 is preferred to h_1 according to distance to working place. The possibility of several preference orderings (or criteria) being incompatible is one of the main reasons why we should justify preferences in PAFs. In fact, among several preference criteria applicable to a set of objects, which criterion an agent should select depends on the particularity of a given context. An agent in a dangerous place should select *safeness* as its preference criterion, while an agent in a safe place should select *comfort* as its criterion[4] [60]. Moreover, as noted above, it cannot be said that there is only one criterion behind a preference ordering.

Third, multiple criteria may cooperatively produce a single preference ordering. That is why the approach of multi-criteria is dominant in the field of quantitative decision-making. For example, if the *size* criterion produces preference ordering $h_1 \succ h_2$ and the *price* criterion produces $h_2 \succ h_3$, then these two criteria cooperatively produces a total preference ordering $h_1 \succ h_2 \succ h_3$ over the set of houses

[4]Note that, for a rational agent, selecting a criterion is justifying the criterion. A selected criterion in a context is one that is justified with respect to the information available in the context. For instance, suppose, among two criteria size and distance, an agent prefers size to distance and has decided to live in h_1. Then the size is selected because it is justified with respect to the information available in the context (for example, because of the large family, the size is important), the distance is not selected because it fails to be justified with respect to the information (for example, as having a car and no traffic jam, the distance from work is less important than size).

$\{h_1, h_2, h_3\}$.[5] However, only compatible criteria can cooperatively produce a preference ordering over a set of objects, but incompatible ones cannot. Therefore, usually, an agent cannot select two incompatible criteria in a context.

One way to adapt to multiple preference criteria is to use a meta-preference ordering over a set of criteria. When evaluating a house, we may prefer size to distance to work. However, where does this meta-preference come from? The meta-preference and the criteria behind it should also be justified.

Like all human judgments and actions, justifying criteria to generate a preference ordering must be based on imperfect, especially inconsistent knowledge. Therefore, the selection of a criterion may be doubted, uncertain or even be in conflict. Actually, we can see that there are debates on whether it is reasonable to apply a certain criterion to evaluating something. For instance, perfect proportion and slight distortion have been billed as two of key criteria in aesthetic evaluation, but artists and aestheticians may dispute on whether they should apply perfect proportion or slight distortion.

In this paper, we are interested in implementing an elaborate mechanism that allows to justify preference criteria in a PAF. Moreover, we know that argumentation is an effective approach for dealing with inconsistent information. Therefore, the conflict among available preference criteria can be resolved through an argumentation. The mechanism that is capable of justifying preference criteria may be the AF itself in which the mechanism should be implemented. An AF can be used for justifying preferences as well as defining plausible conclusions. The core of this section is to introduce the idea that the preference ordering embedded in a PAF should also be justified by another PAF with the same arguments and attack relations.

We find that especially legal reasoning, where a lawyer may have to justify the priority of a certain norm to its conflicting norms, resonates with our idea of using AF for justifying preferences. As mentioned in [38], a conflict between legal norms from different sources or promulgated at different times may arise in legal reasoning. However, most lawyers neither have power to change any of the conflicting norms, nor have control to change established evidence in a case. The only way to resolve the conflict is to introduce a preference ordering over the legal norms. In the legal literature, three major principles which are used to set a preference ordering over conflicting norms are identified. Those are *Lex Superior* which prefers a norm whose legislative source is higher in the legislative source hierarchy, *Lex Posterior* which prefers a norm promulgated more recently and *Lex Specialis* which prefers a more specific norm. Thus, *Lex Superior*, *Lex Posterior* and *Lex Specialis* can be regarded as three main criteria, by which a preference ordering over legal norms can be generated. However, the application of one of these criteria should also be justified. If you would like to resolve a conflict between legal norms by applying *Lex Superior*, then you must justify why we should apply only *Lex Superior*, not *Lex Posterior* nor *Lex Specialis*. That is because if we apply another criterion, the opposite judicial decision may be drawn.

Along the everyday life, it seems that a human has a number of criteria in his mind and attaches a condition to every criterion, under which the application of the criterion is justified. That is, a human justifies a preference criterion by proving the satisfaction of the condition attached to the criterion. For example, a person not only knows that both perfect proportion and slight distortion can be the criteria against which we assess the beauty of something like clothes, but also knows that he should promote either of them based on looking into what is all the rage this season. A lawyer also knows that *Lex Superior, Lex Posterior* and *Lex Specilis* are the criteria from which a preference ordering over legal norms may be produced. He applies some of those criteria for resolving conflicts among legal norms

[5]Criterion may not define total preference ordering over a set of objects. The size criterion and the price criterion in this example define a partial order over $\{h_1, h_2, h_3\}$.

and knows under which condition application of one of those criteria can be justified. On the basis on such an intuition, diverse user's preference handling models have been proposed [24,25,29]. Our formalism will conform to this intuition of a preference-based reasoning that every preference ordering comes from criteria and the adoption of criteria should be justified with regard to a certain context.

4. Justifying preferences and reasoning from the justified preferences in ASPIC+

In this section, we give a definition of a preference criterion on the basis of [60]'s notion of guard and modify the notion of ASPIC+ argumentation theory in order to model the way in which preferences are justified by an AF.

4.1. Justifying preferences

Criteria behind a preference ordering is so essential that justifying the preference ordering boils down to justifying the criteria. Teze et al.'s notion of *guard* is useful for modeling the way in which a criterion is justified with regard to a certain context since a *context* can always be modeled by terms of a set of literals which are true in the context [60]. Our model needs to select an appropriate preference criteria depending on certain conditions, thus, the notion of guard, which offers a special way of associating these conditions to a preference criterion, plays an important role. A guard could be viewed as a way of guiding the choice of a criterion [60]. We define a guard as a set of literals that should be justified by a given AF to apply the associated criterion. Therefore, we can define a preference criterion as follows:

Definition 12 (Preference Criterion)**.** A preference criterion is a pair $c = \langle \mathcal{G}c, \mathcal{S}c \rangle$, where $\mathcal{G}c \subseteq \mathcal{L}$ (called *guard*) is a set of literals such that each literal must be justified by an AF in order to allow the criterion to be applied and $\mathcal{S}c$ stores all the preference information[6] attached to the criterion.

Example 1 (cont.)**.** To apply preferences to our example, we take four criteria $c_1 = \langle \{p, q, r, s\}, \{(q \Rightarrow t > r \Rightarrow u), (t \Rightarrow \neg v > r \Rightarrow u)\} \rangle$, $c_2 = \langle \{q, \neg s\}, \{(s \Rightarrow v > t \Rightarrow \neg v), (r \Rightarrow u > t \Rightarrow \neg v)\} \rangle$, $c_3 = \langle \emptyset, \{(p \Rightarrow s > q \Rightarrow t)\} \rangle$ and $c_4 = \langle \{p, q, r, s\}, \{(q > r)\} \rangle$.

In the above example, criterion c_1 together with c_3 can produce a preference ordering $p \Rightarrow s > q \Rightarrow t > r \Rightarrow u$ and this shows that there may be more than two criteria behind an ordering.

Let $c = \langle \mathcal{G}c, \mathcal{S}c \rangle$ be a preference criterion, then $\texttt{Guard}(c)$ returns $\mathcal{G}c$ and $\texttt{Stored}(c)$ returns $\mathcal{S}c$. As mentioned above, in this paper, we justify a criterion by adopting the argumentation-based approach. Justified criterion implies that a PAF built over an available argumentation theory justifies all literals involved in the guard. In other words, if the PAF identifies all literals involved in the guard of a criterion as *acceptable* under a given semantics, then the criterion is also identified as justified one.

Definition 13 (Justifying a Criterion)**.** Let $c = \langle \mathcal{G}c, \mathcal{S}c \rangle$ be a preference criterion and $\mathcal{H} = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle$ a PAF. The criterion c is justified by $\mathcal{H}$ under a given semantics $\mathtt{y}$ iff $\texttt{Guard}(c) \subseteq \texttt{Output}_{\mathtt{y}}(\mathcal{H})$.

While justifying preference criteria by a PAF corresponding to APSIC+ argumentation theory, we usually adopt preferred or stable semantics because under such semantics, our model has some desirable properties (see Section 4.3)

[6]The term "*preference information*" means that $\mathcal{S}c$ is a partial order over the ordinary premises or defeasible rules of a knowledge base.

A criterion should usually be justified from the skeptical viewpoint, since, as we will show in what follows, such viewpoint does not allow inconsistent criteria to be justified simultaneously. We also say that a criterion is justified *by an extension* under a given semantics. Let $c = \langle \mathcal{G}c, \mathcal{S}c \rangle$ be a preference criterion, $\mathcal{H} = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle$ a PAF and $\mathcal{E}$ one of the extensions of $\mathcal{H}$ under a given semantics. If $\texttt{Guard}(c) \subseteq \texttt{Concs}(\mathcal{E})$, we say that the criterion c is justified by the extension $\mathcal{E}$.

Remark. A criterion whose guard is $\emptyset$ is justified by any PAF under any semantics.

We call a criterion whose guard is $\emptyset$ an *absolute* criterion in the sense that this criterion can be applied without justification. In a certain domain of reasoning, we can think of an absolute criterion such as specificity of defeasible rules.

Example 1 (cont.)**.** The criteria c_1, c_3 and c_4 are skeptically justified by the PAF, while c_2 is not justified under the preferred/stable semantics.

A pair of criteria may or may not be compatible. The incompatiblity of criteria can be defined as follows:

Definition 14 (Incompatible criteria)**.** Let $c_1 = \langle \mathcal{G}c_1, \mathcal{S}c_1 \rangle$ and $c_2 = \langle \mathcal{G}c_2, \mathcal{S}c_2 \rangle$ be preference criteria. We say that c_1 and c_2 are incompatible iff there exists a preference $(p > q) \in \texttt{Stored}(c_1)$ such that $(p < q) \in \texttt{Stored}(c_2)$.[7]

A set of preference criteria is *inconsistent* if and only if it includes at least two incompatible criteria. Otherwise, it is *consistent*. Note also that a set of preference criteria $\mathcal{C}_1$ is said to be *inconsistent* with another criteria set $\mathcal{C}_2$ if and only if $\mathcal{C}_1 \cup \mathcal{C}_2$ is inconsistent.

We cannot apply two incompatible preference criteria in the same context. Thus, the guards of incompatible criteria should also be incompatible, namely, should not be justified, with respect to the argumentation theory available in the context. Otherwise, a PAF built over an argumentation theory may justify two incompatible criteria simultaneously. In such a way, we can resolve conflicts among preference orderings, since any set of incompatible criteria cannot belong to a single extension. The notion of valid criteria set reflects this idea.

Definition 15 (Valid criteria set)**.** Let $\mathcal{C}$ be a set of preference criteria and $\mathcal{T}$ an argumentation theory that satisfies indirect consistency. Then, $\mathcal{C}$ is valid wrt. $\mathcal{T}$ iff for all c_1 and c_2 in $\mathcal{C}$ that are incompatible it holds that $\mathcal{K}n \cup \texttt{Guard}(c_1) \cup \texttt{Guard}(c_2)$, is indirectly inconsistent under the strict rules of $\mathcal{T}$ (here, $\mathcal{K}n$ is the set of axioms of $\mathcal{T}$).

Proposition 2. *Let $\mathcal{T}$ be an argumentation theory which is indirectly consistent, $\mathcal{C}$ a set of valid preference criteria wrt. $\mathcal{T}$, and $\mathcal{H} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq \rangle$ a PAF built over $\mathcal{T}$. For all c_1, c_2 in $\mathcal{C}$, that are incompatible, it is impossible for both of c_1 and c_2 to be justified by an extension of $\mathcal{H}$ under a given semantics (admissible, ground, complete, preferred and stable) and thus be skeptically justified by $\mathcal{H}$.*

The requirement reflected in Definition 15 may seem too strong since the difficulty of reasoning preferences is prominent in the fact that rational agents cannot avoid dealing with inconsistent preference criteria that are justifiable in the very same context. For example, in moral or aesthetical reasoning, a

[7] Here, $p > q$ and $p < q$ represents opposite rule priorities or premise orderings.

rational agent may face dilemmas in which there are no clear-cut solutions to eliminate incompatible preference alternatives. Then, we can adopt credulous standpoint for modeling such dilemmas.

The following corollary directly comes from the above proposition.

Corollary. *A preference criterion that is incompatible with a criterion whose guard is* $\emptyset$ *cannot be justified by any AF.*

Example 1 (cont.)**.** One can easily see that c_1 and c_2 are incompatible and the guards of this two criteria themselves contains inconsistent literals ($s \in \texttt{Guard}(c_1)$ and $\neg s \in \texttt{Guard}(c_2)$), so both cannot be simultaneously justified. That is, the criteria set $\{c_1, c_2, c_3, c_4\}$ is a valid criteria set.

4.2. Reasoning from justified preferences

The selection of a criterion may be doubted or even be in conflict, and, in turn, be in need of justification. Since argumentation is an effective approach dealing with imperfect information, the selection of a certain criterion can also be justified by an argumentation.

Traditionally, a PAF filters the attack relation through its preferences. The result of this step of argumentation is called repaired framework. Nevertheless, most of the existing PAFs including ASPIC+ have only one repairing step because they provide a mechanism only for reasoning from preferences, i.e. the preferences are only the input of the framework. Our proposal is to have two (or more than two) repairing steps because an intelligent agent should not only reason from preferences, but also justify the preferences before reasoning from them. One repairing step is for justifying preferences, the other is for reasoning from those preferences. We consider both justifying preferences and reasoning from the justified preferences in an integrated way, as it accords with human-style argumentation. For the sake of justifying preferences with regard to a certain context, we first need to revise the notion of argumentation theory should be revised in terms of preference criteria.

Definition 16 (Argumentation theory with preference criteria)**.** Argumentation theory with preference criteria (ATPC) is a pair $\langle \mathcal{AS}, \mathcal{KB} \rangle$ where

- $\mathcal{AS} = \langle \mathcal{L}, \bar{}, \mathcal{R}, \mathcal{C}, \mathtt{n} \rangle$ is an argumentation system, where $\mathcal{L}, \bar{}, \mathcal{R} = \mathcal{R}s \cup \mathcal{R}d$ and $\mathtt{n}$ are respectively logical language, contrariness function, a set of inference rules and a naming function defined by Definition 1 and $\mathcal{C}$ is a set of preference criteria which store defeasible rule (in $\mathcal{R}d$) priority information,
- $\mathcal{KB} = \langle \mathcal{K}, \mathcal{C}' \rangle$ a knowledge base in $\mathcal{AS}$, where $\mathcal{K} = \mathcal{K}n \cup \mathcal{K}p$ is defined by Definition 2 and $\mathcal{C}'$ is a set of preference criteria which store information about orderings on $\mathcal{K}p$.

Below, we elaborate our proposal where justifying preferences and reasoning from the justified preferences are integrated. An AF with justified preferences can be seen as having those two steps. Notice that there are two repairing steps.

(1) Building the *primary* PAF over an ATPC, with preferences whose criteria guard is $\emptyset$[8] and determining justified preferences.
(2) Building the *advanced* PAF, with justified preferences and concluding or defining the justified conclusions.

[8] As a criterion whose guard is the empty set is justified by any AF, we can take it as input for determining justified preference criteria.

In the routine above, the *primary* PAF is used for determining justified preferences and the *advanced* PAF is used for reasoning from those justified preferences. The justified preferences are the output of the primary PAF and simultaneously the input of the advanced framework. Furthermore, the step (2) could be an iterative step in this procedure, that is, the preferences justified by the advanced PAF can also be adopted for building further PAFs. It bears a resemblance to a complex argument structure where the conclusion of a subargument becomes a premise of another subargument. The notion of the *primary* and *advanced* PAFs can be defined as follows:

Definition 17 (*Primary* and *advanced* PAF corresponding to an ATPC)**.**

- Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ with $\mathcal{AS} = \langle \mathcal{L}, \bar{}, \mathcal{R}, \mathcal{C}, \mathtt{n} \rangle$ and $\mathcal{KB} = \langle \mathcal{K}, \mathcal{C}' \rangle$ be an ATPC with valid criteria sets. The *primary* PAF built over $\mathcal{T}$ is $\mathcal{H}_{pri} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{pri} \rangle$, where $\mathcal{A} = \mathtt{Arg}(\mathcal{T})$, $\mathcal{R}$ is a binary attack relation on $\mathcal{A}$ defined by Definition 4 and $\succcurlyeq_{pri}$ is the preference ordering over $\mathcal{A}$ produced from the set of criteria $\mathcal{C}_{pri} = \{c | c \in \mathcal{C} \cup \mathcal{C}', \mathtt{Guard}(c) = \emptyset\}$.
- Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ with $\mathcal{AS} = \langle \mathcal{L}, \bar{}, \mathcal{R}, \mathcal{C}, \mathtt{n} \rangle$ and $\mathcal{KB} = \langle \mathcal{K}, \mathcal{C}' \rangle$ be an ATPC with valid criteria sets and $\mathcal{H}_{pri} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{pri} \rangle$ be the primary PAF built over $\mathcal{T}$. The *advanced* PAF built over $\mathcal{T}$ is $\mathcal{H}_{adv} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{adv} \rangle$ where $\mathcal{A} = \mathtt{Arg}(\mathcal{T})$, $\mathcal{R}$ is a binary attack relation on $\mathcal{A}$ defined by Definition 4 and $\succcurlyeq_{adv}$ is the preference ordering over $\mathcal{A}$ produced from the set of criteria $\mathcal{C}_{adv} = \{c | c \in \mathcal{C} \cup \mathcal{C}', c \text{ is skeptically justified by } \mathcal{H}_{pri}\}$.

In determining justified preference criteria, the skeptical standpoint should usually be adopted because incompatible criteria cannot be skeptically justified by any PAF under valid criteria set as shown in Proposition 2. When reasoning from justified preferences, the original AF should be repaired twice by *primary preferences* whose guard is $\emptyset$ and *advanced preferences* whose criteria are justified by the primary framework. Note that primary preferences are justified by any AF.

Remark. Let $\mathcal{C}_{pri}$ be the set of criteria adopted by a primary PAF corresponding to an argumentation theory and $\mathcal{C}_{adv}$ the set of criteria adopted by its advanced version. Then, $\mathcal{C}_{pri} \subseteq \mathcal{C}_{adv}$, thus $\succcurlyeq_{pri} \subseteq \succcurlyeq_{adv}$. Let $\mathcal{H}'_{pri} = \langle \mathcal{A}, \mathcal{R}'_{pri} \rangle$ be the repaired version of $\mathcal{H}_{pri}$ and $\mathcal{H}'_{adv} = \langle \mathcal{A}, \mathcal{R}'_{adv} \rangle$ the repaired version of $\mathcal{H}_{adv}$, it holds that $\mathcal{R}'_{pri} \supseteq \mathcal{R}_{adv}$'

To generalize, in our formalism, the *primary* PAF is the result of the first filtering with preference criteria whose guards are empty set, while the *advanced* PAF is the result of the second filtering with preference criteria whose guards belong to the output of the former primary PAF. However, what can we do if the conflict between extensions still remains unresolved even after the second filtering, that is, if the advanced PAF has two or more conflicting (preferred/stable) extensions? In such a case, if possible, we can do further third, fourth, ..., and the nth filtering and thus make a sequence of PAFs, where every PAF shares the same set of arguments and the attack relation, but differs only in their preferences. The concept of '*result of the nth filtering* (*or the nth PAF*)' can be defined inductively as follows:

Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ with $\mathcal{AS} = \langle \mathcal{L}, \bar{}, \mathcal{R}, \mathcal{C}, \mathtt{n} \rangle$ and $\mathcal{KB} = \langle \mathcal{K}, \mathcal{C}' \rangle$ be an ATPC.

- The result of the first filtering is the primary PAF corresponding to $\mathcal{T}$.
- Let $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq_{n-1} \rangle$ $(n = 2, 3, \dots)$ be the result of the $n-1$th filtering, then the result of the nth filtering is PAF $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq_n \rangle$ where $\preccurlyeq_n$ is a preference ordering over $\mathcal{A}$ defined by the set of criteria $\mathcal{C}_n = \{c | c \in \mathcal{C} \cup \mathcal{C}', c \text{ is skeptically justified by } \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_{n-1} \rangle\}$

We call the preference criteria whose guards are empty set the first preference criteria. The preference criteria that are justified by the result of the nth filtering are called the $n+1$th preference criteria in the sense that the $n+1$th PAF is to be based on the criteria.

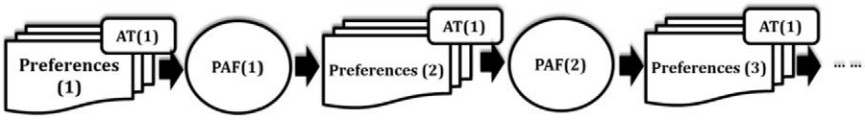

Fig. 4. The nth filtering.

4.3. Properties of the sequence of PAFs

We suggest using a PAF for the sake of justifying preferences, by which the attack relation is filtered, in turn, the output of the PAF is also changed. In the above sequence of PAFs, the output of a PAF determines which preferences to select and the selected preferences affect the output of the next PAF (see Fig. 4). The first preference criteria determine the output of the primary PAF and the primary PAF determines preferences that are to be adopted by the advanced PAF. In the same way, the nth preference criteria determine the output of the result of the nth filtering, which, in turn, determines the $n+1$th preference criteria. Here, it is fundamental to ensure the consistency between the nth and the $n+1$th preference criteria. The nth preference criteria are an input of the nth PAF, whose output also includes the $n+1$th preference criteria. Therefore, it is unintuitive to allow inconsistency between the nth and $n+1$th preference criteria, in the same way as it is regarded as irrational to allow inconsistency between the premises and conclusions of a single argument.

From Proposition 2, we can see that inconsistent preference criteria set cannot be justified by a PAF, if the argumentation theory over which the PAF is built has valid criteria set. Since a preference criterion that is incompatible with a criterion whose guard is empty set cannot be justified by any PAF, the second preference criteria cannot include a criterion incompatible with one of the first criteria. However, if $n > 1$, Proposition 2 may not ensure the consistency between the nth and $n+1$th preference criteria, since they are justified by different PAFs (the $n-1$th and nth PAF). For a simple example, let us consider the advanced PAF $\mathcal{H}_{adv} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{adv} \rangle$ built over an ATPC. The advanced PAF is based on the second preference criteria that are justified by the primary PAF $\mathcal{H}_{pri} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{pri} \rangle$. From the viewpoint of *dynamics* of AFs, the advanced PAF is an *attack abstraction* of the primary version because $\succcurlyeq_{pri} \subseteq \succcurlyeq_{adv}$ [19,20].[9] However, an attack abstraction preserves the set of accepted arguments only under some strict assumptions (with respect to the status of the attacker and attackee of the removed attack relation and the given semantics). Hence, unless the argumentation formalism is carefully defined, it may lead to very unintuitive result where the advanced PAF justifies some preference criteria that are inconsistent with the second criteria (that are justified by the primary PAF). The same may also happen in the further steps of such a sequence of PAFs.

Fortunately, ASPIC+ bears a desirable property that other structured argumentation formalisms do not have (Proposition 1). As aforementioned, we modify ASPIC+ argumentation theory with the notion of preference criteria, so as to build a sequence of PAFs, by which we can not only justify preferences but also reasoning from the justified preferences. Then, thanks to the desirable property of ASPIC+, we can ensure consistency between preferences from which we reason and those that we justify.

The following proposition, which reveals the relation between the extensions of a primary PAF and those of its advanced version, is useful for showing that the sequence of PAFs built over an ATPC does not allow inconsistency between input and output preferences of a PAF in it.

[9] Let $\mathcal{H} = \langle \mathcal{A}, \mathcal{R} \rangle$ and $\mathcal{H}' = \langle \mathcal{A}', \mathcal{R}' \rangle$ be two AFs. Then, $\mathcal{H}'$ is an *attack abstraction* of $\mathcal{H}$ iff $\mathcal{A} = \mathcal{A}'$ and $\mathcal{R}' \subseteq \mathcal{R}$.

Proposition 3. *Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ with $\mathcal{AS} = \langle \mathcal{L}, \bar{}, \mathcal{R}, \mathcal{C}, \mathtt{n} \rangle$ and $\mathcal{KB} = \langle \mathcal{K}, \mathcal{C}' \rangle$ be an ATPC with valid criteria sets and $\mathcal{H}_{pri} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{pri} \rangle$ and $\mathcal{H}_{adv} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{adv} \rangle$ respectively the primary and advanced PAF corresponding to $\mathcal{T}$ which have reasonable argument orderings. Then, under preferred semantics, for all $\mathcal{E}' \in \mathtt{Ext}_{pre}(\mathcal{H}_{adv})$, there exists an extension $\mathcal{E} \in \mathtt{Ext}_{pre}(\mathcal{H}_{pri})$ such that $\mathcal{E}' \subseteq \mathcal{E}$. Under stable semantics, it holds that $\mathtt{Ext}_{sta}(\mathcal{H}_{pri}) \supseteq \mathtt{Ext}_{sta}(\mathcal{H}_{adv})$.*

Example 1 (cont.)**.** Let us consider an ATPC $\langle \mathcal{AS}_1, \mathcal{KB}_1 \rangle$ with $\mathcal{AS}_1 = \langle \mathcal{L}, \bar{}, \mathcal{R}, \mathcal{C}_1, \mathtt{n} \rangle$ and $\mathcal{KB}_1 = \langle \mathcal{K}, \mathcal{C}'_1 \rangle$, where $\mathcal{C}_1 = \{c_1, c_2, c_3\}, \mathcal{C}'_1 = \{c_4\}$ and the other elements of the theory can be found in Section 2. The criterion whose guard is $\emptyset$ is c_3 which stores preference information $\{(p \Rightarrow s > q \Rightarrow t)\}$. Then, from the weakest-link principle, the argument A_7 is preferred to A_8 and A_{10}, and A_4 is preferred to A_{10}. Therefore, the primary PAF $\langle \mathcal{A}_1, \mathcal{R}_1, \succcurlyeq_1 \rangle$, where $\succcurlyeq_1 = \{(A_7 \succ A_8), (A_7 \succ A_{10}), (A_4 \succ A_{10})\}$, is built. The result of repairing $\mathcal{H}_1 = \langle \mathcal{A}_1, \mathcal{R}_1 \rangle$ with $\succcurlyeq_1$ is as follows (see Fig. 5).

The attack $\langle A_8, A_7 \rangle$ is removed from the framework because it is a preference-dependent attack and $A_7 \succ A_8$. However, as we can notice in the diagram, the attacks $\langle A_{10}, A_7 \rangle$ and $\langle A_{10}, A_4 \rangle$ are not removed from the framework despite $A_7 \succ A_{10}$, $A_4 \succ A_{10}$ because they are preference-independent undercuts. Preferences over arguments in undercut have no effect on the attack. Now, if we reckon extensions of $\langle \mathcal{A}_1, \mathcal{R}_1, \succcurlyeq_1 \rangle$, it still has two preferred extensions: $\mathcal{E}_1 = \{A_1, A_2, A_3, A_6, A_{10}, A_8, A_5\}$ and $\mathcal{E}_2 = \{A_1, A_2, A_3, A_6, A_4, A_7, A_9\}$. Since $\mathtt{Guard}(c_1) = \mathtt{Guard}(c_4) \subset \mathtt{Concs}(\mathcal{E}_1 \cap \mathcal{E}_2)$ and $\mathtt{Guard}(c_2) \nsubseteq \mathtt{Concs}(\mathcal{E}_1 \cup \mathcal{E}_2)$, c_1 and c_4 are the criteria skeptically justified by the primary PAF. Thus, we should build an advanced framework based on c_1, c_3 and c_4. Since c_1 stores $\{(q \Rightarrow t > r \Rightarrow u), (t \Rightarrow \neg v > r \Rightarrow u)\}$ and c_4 stores $\{(q > r)\}$ the advanced version of $\langle \mathcal{A}_1, \mathcal{R}_1, \succcurlyeq_1 \rangle$ is $\langle \mathcal{A}_1, \mathcal{R}_1, \succcurlyeq_2 \rangle$, where $\succcurlyeq_2 = \{(A_2 \succ A_3), (A_5 \succ A_6), (A_7 \succ A_8), (A_7 \succ A_{10}), (A_4 \succ A_{10}), (A_8 \succ A_9), (A_{10} \succ A_9), (A_5 \succ A_9\}$. Then, the original AF $\langle \mathcal{A}_1, \mathcal{R}_1 \rangle$ is repaired once again by $\succcurlyeq_2$ and the following framework is produced (see Fig. 6).

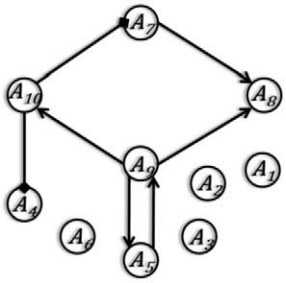

Fig. 5. A PAF filtered through *primary* preferences.

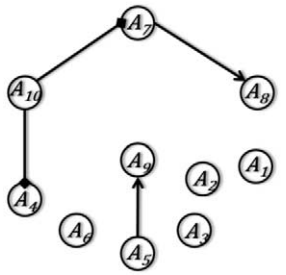

Fig. 6. An AF filtered through *advanced* preferences.

The PAF $\langle \mathcal{A}_1, \mathcal{R}_1, \succcurlyeq_2 \rangle$ has only one preferred extension: $\mathcal{E}_1 = \{A_1, A_2, A_3, A_6, A_{10}, A_8, A_5\}$. Unlike other argumentation formalisms, the preference ordering $\succcurlyeq_2$ and the criteria $\{c_1, c_3, c_4\}$ behind it are justified with respect to an argumentation theory.

The following proposition shows that any inconsistency cannot be found between advanced preference criteria and those justified by the advanced PAF built over an ATPC.

Proposition 4. *Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ with $\mathcal{AS} = \langle \mathcal{L}, \bar{}, \mathcal{R}, \mathcal{C}, n \rangle$ and $\mathcal{KB} = \langle \mathcal{K}, \mathcal{C}' \rangle$ be an ATPC with valid criteria sets and $\mathcal{H}_{pri} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{pri} \rangle$ and $\mathcal{H}_{adv} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{adv} \rangle$ respectively the primary and advanced PAF corresponding to $\mathcal{T}$ which have reasonable argument orderings. In addition, let $\mathcal{C}_{adv}$ and $\mathcal{C}_3$ be respectively be sets of preference criteria that produce $\succcurlyeq_{adv}$ and that is justified by $\mathcal{H}_{adv}$. Then there exists no $c \in \mathcal{C}_{adv}$ and $c' \in \mathcal{C}_3$ such that c and c' are incompatible.*

Some other argumentation formalisms, namely, deductive argumentation [9] and assumption-based argumentation with preferences (ABA+ for short) [28], which take preferences take into account, inverse the direction of a failed attack in order to guarantee conflict-freeness of extensions with respect to the attack relation. For this reason, in such formalisms, the extension of a PAF is not that of the original AF under stable semantics (or a subset of an extension of the original AF under preferred semantics). As mentioned in Section 2, Amgoud and Vesic also made preferences refine extensions [9]. However, in their formalism, the extensions refined are those of the repaired framework, and thus the extensions of the PAF may deviate from those of the original AF. In a word, Proposition 1 does not hold in deductive argumentation or ABA+. It may give rise to inconsistency between the outputs of the PAF and the original AF. Therefore, if we built such a sequence of PAFs based on deductive argumentation or ABA+, then we would not guarantee consistency between the preferences that we reason from and those that we justify.

Now, it is time to generalize Proposition 3.

Proposition 5. *Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ be an ATPC with valid criteria sets, $\mathcal{H}_n = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_n \rangle$ and $\mathcal{H}_{n+1} = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_{n+1} \rangle$ respectively the resulting PAFs of the nth and $n+1$th filtering which have reasonable argument orderings. Then, if $\preccurlyeq_{n+1} \supseteq \preccurlyeq_n$, then under preferred semantics, for all $\mathcal{E}' \in \mathtt{Ext}_{pre}(\mathcal{H}_{n+1})$, there exists an extension $\mathcal{E} \in \mathtt{Ext}_{pre}(\mathcal{H}_n)$ such that $\mathcal{E}' \subseteq \mathcal{E}$. Under stable semantics, it holds that $\mathtt{Ext}_{sta}(\mathcal{H}_n) \supseteq \mathtt{Ext}_{sta}(\mathcal{H}_{n+1})$.*

The above proposition, differently from Proposition 3, is conditioned on the antecedent $\preccurlyeq_{n+1} \supseteq \preccurlyeq_n$ (under preferred semantics). Sometimes, it may hold that $\preccurlyeq_n \supseteq \preccurlyeq_{n+1}$ under preferred semantics. This may make the sequence of PAFs with justified preferences fall into an endless loop.

Example 2 (cont.)**.** Now, let us revise Example 2 with some modifications in order to illustrate it. Suppose that the argument preference $B_4 \prec B_3$ comes from the criteria whose guards belong to the conclusions of B_2 and B_4. Assuming that the argumentation theory has no absolute criteria, the primary PAF has two preferred extensions $\{B_2, B_4, B_5, B_7\}$ and $\{B_2, B_4, B_6, B_7\}$. Then, the preference criteria from which $B_4 \prec B_3$ comes are justified by the primary PAF because $\{B_2, B_4\}$ is the intersection of those two preferred extensions. As a result, the preference ordering $B_4 \prec B_3$ is activated and thus the advanced PAF is constructed as follows (see Fig. 7).

The advanced PAF also has two preferred extensions $\{B_5, B_7\}$ and $\{B_6, B_7\}$. Arguments B_2 and B_4 are not justified by the advanced PAF and the preference criteria is now deactivated. Then, the next (third)

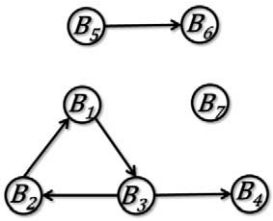

Fig. 7. How can a sequence of PAFs fall into an endless loop?

PAF coincides with the primary PAF. Once again, the next (fourth) PAF coincides with the advanced PAF. Consequently, the sequence of PAFs falls into undesirable loop. In such a case, it is more prudent to stop at the advanced PAF.

The above example shows that if the framework filtered through a preference may not justify the preference over which it is built, it may lead us to unprofitable and endless loop.

However, under stable semantics (or even under preferred semantics if the PAFs built over an ATPC do not contain any odd length cycles of attack),[10] such undesirable outcomes are impossible. Under stable semantics, it holds that $\texttt{Ext}_{sta}(\mathcal{H}_n) \supseteq \texttt{Ext}_{sta}(\mathcal{H}_{n+1})$ from Proposition 5. Let $\texttt{Ext}_{sta}(\mathcal{H}_n) = \{\mathcal{E}_1, \ldots, \mathcal{E}_n\}$ and $\texttt{Ext}_{sta}(\mathcal{H}_{n+1}) = \{\mathcal{E}'_1, \ldots, \mathcal{E}'_m\}$ $(m \leqslant n)$. One can notice that $\bigcap_{i=1}^{n} \mathcal{E}_i \subseteq \bigcap_{i=1}^{m} \mathcal{E}'_i$. Because skeptical viewpoint is usually adopted when we justify preferences, any preference criteria justified by the nth PAF is also justified by the $n+1$th PAF. Therefore, a monotonic increase in justified preference information brings about another monotonic increase in sceptical view and monotonic decrease in credulous view. If it were to hold that $\texttt{Ext}_{sta}(\mathcal{H}_n) \not\supseteq \texttt{Ext}_{sta}(\mathcal{H}_{n+1})$ (such an assumption is possible under preferred semantics), then it would be possible to have $\bigcap_{i=1}^{n} \mathcal{E}_i \not\subseteq \bigcap_{i=1}^{m} \mathcal{E}'_i$, thus there may exist a criterion c whose guard belongs to $\bigcap_{i=1}^{n} \mathcal{E}_i \setminus \bigcap_{i=1}^{m} \mathcal{E}'_i$. Then, we can recognize that the criterion c is justified by the nth PAF, but fails to be justified by the $n+1$th PAF. It is a very unintuitive anomaly since c is one of the underlying criteria, over which, the $n+1$th PAF is built, but the $n+1$th PAF which is based on c does not justify c.

The following proposition (generalization of Proposition 4) shows that the standard ASPIC+ prohibits inconsistency between the input and output preference criteria of a PAF in a sequence of PAFs.

Proposition 6. *Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ be an ATPC with valid criteria sets, $\mathcal{H}_n = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_n \rangle$ and $\mathcal{H}_{n+1} = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_{n+1} \rangle$ respectively the resulting PAFs of the nth and $n+1$th filtering which have reasonable argument orderings. In addition, let $\mathcal{C}_n$ and $\mathcal{C}_{n+1}$ be respectively sets of preference criteria that produce $\preccurlyeq_n$ and $\preccurlyeq_{n+1}$. Then there exists no $c \in \mathcal{C}_n$ and $c' \in \mathcal{C}_{n+1}$ such that c and c' are incompatible.*

In the above proposition, $\mathcal{C}_n$ is taken as an input of $\mathcal{H}_n$, while $\mathcal{C}_{n+1}$ is an output of $\mathcal{H}_n$ and as one can see, no inconsistency is observed. Note that Proposition 6 does not bear the antecedent $\preccurlyeq_{n+1} \supseteq \preccurlyeq_n$.

An important issue that arises here is when we should stop the filtering in a sequence of PAFs? The proposition below will be the answer to this question.

Proposition 7. *Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ with $\mathcal{AS} = \langle \mathcal{L}, \bar{}, \mathcal{R}, \mathcal{C}, n \rangle$ and $\mathcal{KB} = \langle \mathcal{K}, \mathcal{C}' \rangle$ be an ATPC with valid criteria sets, $\mathcal{H}_{n-1}$, $\mathcal{H}_n$ and $\mathcal{H}_{n+1}$ $(n = 2, 3, \ldots)$ respectively results of the $n-1$th, nth and $n+1$th*

[10] It is shown that the stable and the preferred semantics coincide when the AF does not contain any odd length cycles of attacks [31].

filtering with reasonable argument orderings. Then, it holds that if $\mathtt{Ext}_y(\mathcal{H}_{n-1}) = \mathtt{Ext}_y(\mathcal{H}_n)$*, then* $\mathtt{Ext}_y(\mathcal{H}_n) = \mathtt{Ext}_y(\mathcal{H}_{n+1})$ *under preferred or stable semantics (that is,* $y \in \{\mathtt{pre}, \mathtt{sta}\}$*).*

The above proposition teaches us that if the result of a present filtering is the same as the previous filtering, then the next filtering must also be the same, thus, there is no need of further filtering. Therefore, under such circumstances, it will be more helpful to enrich the underlying argumentation theory with more preference criteria rather than repeating the unprofitable filtering, since the added criteria may contribute to resolving conflicts among the extensions.

5. Discussions

It is reasonable to justify preferences before reasoning from them in everyday argumentation. Thus, we propose using the ASPIC+ framework to integrate justifying preferences with reasoning from those preferences. Our proposal includes a somewhat meta-perspective on arguments within AFs themselves, which is novel in the literature. In this section, we investigate the formalism more closely.

An argumentation, as a mechanism for reasoning from inconsistent information, can be used not only for defining plausible conclusions, but also for selecting appropriate preference criteria. If a preference criterion is justified by the argumentation built over the information available in a context, the criterion is selected as an appropriate one to the context. Then, the selected criteria are used for repairing the previous argumentation and finally defining conclusions.

Using argumentation for justifying preferences conforms the way in which we ordinarily reason and argue. The reason why argumentation becomes a powerful paradigm of AI is that it is capable of not only modeling non-monotonic reasoning, but also providing rational explanations identified plausible conclusions. For example, argumentation-based decision support systems can explain why the recommended choices are desirable [6,26,68]. Nonetheless, preferences may underlie such decisions or belief. In everyday discussions and debates, people may have to justify the desirability of a course of action or the acceptability of a judgment by appealing to preferences such as value orders or rule priorities. However, if a preference that underlies one's argument is not self-evident to everyone, it should also be justified like other statements. Our proposal makes it possible to employ multiple preference criteria and justify some of them as appropriate for a context by adopting an argumentation-based approach. The notion of *valid criteria set* is used to resolve conflicts between incompatible preferences, since an arbitrary pair of incompatible criteria in such a set cannot belong to a single extension. Consider the following decision-making example that was described in [34].

Example 3. A robotic agent performs a cleaning task (Fig. 8). The robot should decide which boxes to carry first to the specified place called store (grey area in the figure). There are four boxes (*box1*, *box2*, *box3*, *box4*) in the environment, which are of different sizes and in different locations.

Because of the difference in their size (*box3* is the biggest and *box4* is the smallest), the agent cannot carry any two of *box1*, *box2*, *box3* at the same time, but can carry *box1* and *box4* or *box2* and *box4* at the same time. The agent cannot carry *box3* and *box4* together. Several preference criteria for selecting boxes may be applicable, for example, the robot may prefer boxes nearer to it or prefer boxes nearer to the store. However, the robot applies these criteria only when some conditions are satisfied. When the robot is near to the store, then it will prefer boxes nearer to the store. Once the robot load itself with *box4* (recall that it can carry *box1* or *box2* with *box4* at the same time), it will prefer the box nearer to *box4* (to

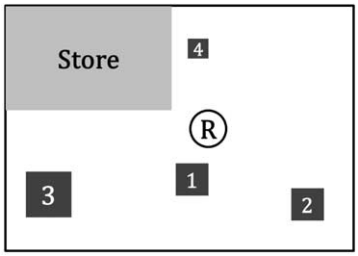

Fig. 8. The robotic environment: scenario 1.

save energy). Here, we will use first-order predicate language. Now, consider an ATPC $\mathcal{T}_3 = \langle \mathcal{AS}_3, \mathcal{KB}_3 \rangle$ with $\mathcal{AS}_3 = \langle \mathcal{L}, \bar{}, \mathcal{R}s_3 \cup \mathcal{R}d_3, \mathcal{C}_3, \mathrm{n} \rangle$ and $\mathcal{KB}_3 = \langle \mathcal{K}n_3 \cup \mathcal{K}p_3, \mathcal{C}'_3 \rangle$ where:

$$\begin{aligned}
\mathcal{R}s_3 = \big\{ & Select(box1) \to \neg Seclect(box2), Select(box2) \to \neg Select(box1), \\
& Select(box2) \to \neg Select(box3), Select(box3) \to \neg Select(box2), \\
& Select(box3) \to \neg Select(box1), Select(box1) \to \neg Select(box3), \\
& Select(box3) \to \neg Select(box4), Select(box4) \to \neg Select(box3) \big\}, \\
\mathcal{R}d_3 = \emptyset, \quad & \mathcal{C}_3 = \emptyset, \\
\mathcal{K}p_3 = \big\{ & Select(box1), Select(box2), Select(box3), Select(box4), NearStore(Robot) \big\}, \\
\mathcal{K}n_3 = \emptyset, \quad & \mathcal{C}'_3 = \{c_1, c_2\}, \\
c_1 = \big\langle & \big\{ NearStore(Robot) \big\}, \big\{ Select(box4) > Select(box3) \big\} \big\rangle, \\
c_2 = \big\langle & \big\{ Select(box4) \big\}, \big\{ Select(box1) > Select(box2) \big\} \big\rangle.
\end{aligned}$$

We could construct 13 arguments:

$$\begin{aligned}
& A = Select(box1), \quad B = Select(box2), \\
& C = Select(box3), \quad D = Select(box4), \\
& A_1 = A \to \neg Seclect(box2), \quad B_1 = B \to \neg Select(box3), \\
& C_1 = C \to \neg Select(box1), \quad D_1 = D \to \neg Select(box3), \\
& A_2 = A \to \neg Select(box3), \quad B_2 = B \to \neg Select(box1), \\
& C_2 = C \to \neg Select(box2), \quad C_3 = C \to \neg Select(box4), \\
& E = NearStore(Robot).
\end{aligned}$$

Then, the primary PAF (Fig. 9) has three preferred/stable extensions: $\{A, A_1, A_2, D, D_1, E\}$, $\{B, B_1, B_2, D, D_1, E\}$, $\{C, C_1, C_2, C_3, E\}$. In the primary PAF, only E is skeptically accepted, and thus c_1 is justified. As a result, the advanced PAF (Fig. 10) has two preferred/stable extensions: $\{A, A_1, A_2, D, D_1, E\}$, $\{B, B_1, B_2, D, D_1, E\}$. Then, c_2 is skeptically justified by the advanced PAF.

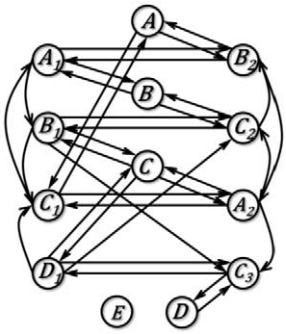

Fig. 9. Primary PAF.

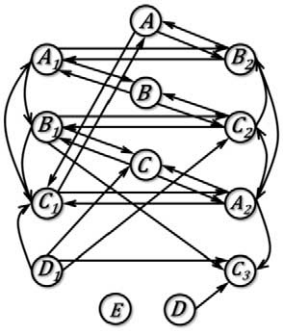

Fig. 10. Advanced PAF.

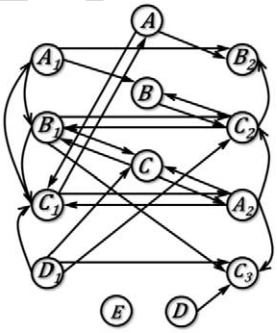

Fig. 11. Third PAF.

The third PAF based on c_1 and c_2 (thus also based on premise ordering $Select(box1) > Select(box2)$) has only one preferred/stable extension: $\{A, A_1, A_2, D, D_1, E\}$. Therefore, the recommended decision is to carry *box1* and *box4* at the same time (Fig. 11).

In this example, we built a sequence of three PAFs based on an ATPC. As we can see, the multiple extensions are subsequently filtered through justified preferences. As it shows, the proposed model makes it possible to provide justifications for the preferences that underlie established belief or selected decisions.

Fig. 12. The robotic environment: scenario 2.

Nevertheless, the credulous standpoint allows incompatible criteria to be simultaneously justified in our formalism. In such a case, another preference criterion which stores meta-preference information may be useful. For example, given two incompatible preference criteria c_1 and c_2, we can think of another criteria $\langle\{a_1, \ldots, a_n\}, \{c_1 > c_2\}\rangle$, where $\{a_1, \ldots, a_n\}$ is a set of literals which should be identified as acceptable by a given AF to apply the meta-preference $c_1 > c_2$. Now, we can extend an ATPC with meta-preference criteria.

Definition 16 (Extended argumentation theory with preference criteria)**.** An extended argumentation theory with preference criteria (EATPC) is a tuple $\langle \mathcal{AS}, \mathcal{KB}, \mathcal{C}_{\text{meta}}\rangle$, where $\langle \mathcal{AS}, \mathcal{KB}\rangle$ is an ATPC with $\mathcal{AS} = \langle \mathcal{L}, \bar{}, \mathcal{R}, \mathcal{C}, \mathtt{n}\rangle$ and $\mathcal{KB} = \langle \mathcal{K}, \mathcal{C}'\rangle$ and $\mathcal{C}_{\text{meta}}$ is a set of meta-preference criteria which store information about ordering on $\mathcal{C}$ and $\mathcal{C}'$.

Example 4. We present another scenario featuring a cleaning robot reasoning about its environment. In this scenario, there are four boxes too, but we should also take their weights into account as well as sizes and locations. We have the ordering in weight: *box2*, *box3*, *box1*, *box4* (that is, *box2* is the biggest and *box4* is the smallest, see Fig. 12).

Because of their size, the robot cannot carry the pair of *box1* and *box2* or the pair of *box2* and *box3* at the same time. Moreover, the robot cannot carry *box3* and *box4* together because they are too heavy (in Fig. 12, the dark grey color represents that the box is heavy). The robotic agent's context-sensitive preference criteria are as follows: if the robot selects *box1*, then it will prefer *box3* to *box2* or *box4* and if it selects *box3*, then it will select *box1* rather than *box2* or *box4* (because both *box1* and *box3* are nearer the store); if the robot chooses *box2*, then it will select *box4* next rather than *box1* or *box3* and if it chooses *box4*, it will also prefer *box4* to *box1* or *box3* (because both *box2* and *box4* are farther from the store). The robot also has mete-preference criteria that is also context-sensitive: if the robot is near the store, then it will prefer c_1 and c_2; if the robot is far from the store, then it will prefer c_3 and c_4. Consider an EATPC $\mathcal{T}_4 = \langle \mathcal{AS}_4, \mathcal{KB}_4, \mathcal{C}_{\text{meta4}}\rangle$ with $\mathcal{AS}_4 = \langle \mathcal{L}, \bar{}, \mathcal{R}s_4 \cup \mathcal{R}d_4, \mathcal{C}_4, \mathtt{n}\rangle$ and $\mathcal{KB}_4 = \langle \mathcal{K}n_4 \cup \mathcal{K}p_4, \mathcal{C}'_4\rangle$ where:

$$\mathcal{R}s_4 = \big\{Select(box1) \rightarrow \neg Select(box2), Select(box2) \rightarrow \neg Select(box1),$$
$$Select(box2) \rightarrow \neg Select(box3), Select(box3) \rightarrow \neg Select(box2),$$
$$Select(box3) \rightarrow \neg Select(box4), Select(box4) \rightarrow \neg Select(box3)\big\},$$
$$\mathcal{R}d_3 = \emptyset, \qquad \mathcal{C}_4 = \emptyset,$$
$$\mathcal{K}p_3 = \big\{Select(box1), Select(box2), Select(box3), Select(box4), FarfromStore(Robot)\big\},$$

$$\mathcal{K}n_4 = \emptyset, \qquad \mathcal{C}'_4 = \{c_1, c_2, c_3, c_4\}, \qquad \mathcal{C}_{\text{meta4}} = \{c_5, c_6\},$$
$$c_1 = \langle \{Select(box1)\}, \{Select(box3) > Select(box2), Select(box3) > Select(box4)\} \rangle,$$
$$c_2 = \langle \{Seclect(box3)\}, \{Select(box1) > Select(box2), Select(box1) > Select(box4)\} \rangle,$$
$$c_3 = \langle \{Select(box2)\}, \{Select(box4) > Select(box1), Select(box4) > Select(box3)\} \rangle,$$
$$c_4 = \langle \{Select(box4)\}, \{Select(box2) > Select(box1), Select(box2) > Select(box3)\} \rangle,$$
$$c_5 = \langle \{NearStore(Robot)\}, \{c_1 > c_3, c_2 > c_4\} \rangle \quad \text{and}$$
$$c_6 = \langle \{FarformStore(Robot)\}, \{c_1 < c_3, c_2 < c_4\} \rangle.$$

We construct 11 arguments:

$$A = Select(box1), \qquad B = Select(box2),$$
$$C = Select(box3), \qquad D = Select(box4),$$
$$A_1 = A \rightarrow \neg Select(box2), \qquad B_1 = B \rightarrow \neg Select(box3),$$
$$C_1 = C \rightarrow \neg Select(box4), \qquad B_2 = B \rightarrow \neg Select(box1),$$
$$C_2 = C \rightarrow \neg Select(box2), \qquad D_1 = D \rightarrow \neg Select(box3),$$
$$E = FarfromStore(Robot).$$

The primary PAF (Fig. 13) has two preferred/stable extensions: $\{A, A_1, C, C_1, C_2, E\}$ and $\{B, B_1, B_2, D, D_1, E\}$. Thus, the meta-preference criterion c_5 is justified. Then, we can adopt credulous standpoint with the justified meta-preferences. As one can notice, c_1, c_2, c_3, c_4 are credulously justified by the primary PAF. That is, from the credulous viewpoint, all of c_1, c_2, c_3, c_4 are applicable. But from the justified meta-preference criterion c_6, we should adopt $Select(box1) < Select(box2)$ and $Select(box3) < Select(box4)$ since both c_3 and c_4 are preferred. The subsequent PAF is built as Fig. 14. The PAF with the justified meta-preferences has the single preferred/stable extension: $\{A, A_1, C, C_1, C_2, E\}$. What the sequence of PAFs built over $\mathcal{T}_4$ recommends is to carry *box2* and *box4* first.

As the above example shows, when the credulous standpoint allows inconsistent preference criteria, we may make use of meta-preferences, but they should also be justified. Several argumentation-based decision-making systems trade on a single meta-preference criterion. For example, the decision-making

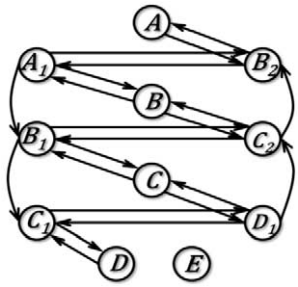

Fig. 13. PAF without meta-preference.

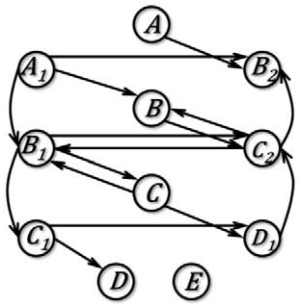

Fig. 14. PAF with meta-preferences.

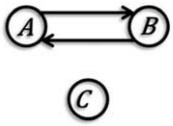

Fig. 15. *Self-preferring* and *self-discarding* extensions.

system based on dynamic argumentation system includes a meta-preference (a strict total order over the preferences) as a component of their so-called *abstract decision framework* [35].

Furthermore, our formalism enables an extension to have a support for preferring itself. In Dung-style AFs, although the principle of argument acceptability and the concept of an admissible set of arguments seem straightforward enough, it turns out that intricate formal puzzles loom-it can happen that an argument is both admissibly provable and refutable [61]. This formal puzzle is due to an AF having two or more conflicting preferred/stable extensions. The symmetric attack often makes it the case that an AF has more than two conflicting preferred/stable extensions, thus, violate rationality postulates when the credulous viewpoint is adopted. In the literature, to resolve such conflicts among extensions, a preference ordering over a set of arguments has been introduced. However, there was no justification for given preferences. Preferences over a set of arguments are neither undeniable facts nor axioms that are taken for granted. In our proposal, the AF built over a theory is used not only for defining plausible conclusions, but also for justifying appropriate preferences as in Example 1. As we have mentioned in Section 4.2, some of the preference criteria are selected and justified by the primary PAF. Then, the advanced PAF is built based on the justified criteria and used for drawing plausible conclusions. In the traditional PAFs, the preferences are used for calculating extensions, but the extensions are not used for justifying the preferences. In a word, the relation between a preference and an extension has been mono-directional. However, in our proposal the relation between a preference ordering and an extension may be bi-directional, that is, the preferences are used for calculating extensions, and simultaneously, the extensions can be used for justifying the underlying preferences. When two preferred/stable extensions are in conflict, one of them may provide a support for preferring itself to the opposite one. Let us consider the following example.

Example 5. Let us consider an AF $\mathcal{H}_5$ depicted in Fig. 15. The theory over which the AF is built has no criterion whose guard is $\emptyset$, thus the primary PAF coincides with $\mathcal{H}_5$.

$\mathcal{H}_5$ has two conflicting preferred/stable extensions: $\{A, C\}$ and $\{B, C\}$. Then let us assume that there is a criterion c with $\texttt{Guard}(c) \subseteq \texttt{Concs}(C)$, such that c stores preference information from which

the ordering ($A \succ B$) comes. In this case, it is obvious that the criterion c is skeptically justified, so we should choose $\{A, C\}$ and reject $\{B, C\}$ in the advanced version of the framework. Now, if we note that $\texttt{Guard}(c) \subseteq \texttt{Concs}(\{A, C\})$, it can be said that the extension $\{A, C\}$ provides a support for preferring itself to its rival extension $\{B, C\}$. Also, since $\texttt{Guard}(c) \subseteq \texttt{Concs}(\{B, C\})$, we note that $\{B, C\}$ provides a support for preferring its rival extension to itself. Therefore, $\{B, C\}$ can be called a *self-discarding* extension, while $\{A, C\}$ can be called a *self-preferring* extension.

In the above example, the extension of the advanced PAF is calculated on the basis of the preference ($A \succ B$), but provides a support for the preference ordering. This is neither circular reasoning nor begging the question because an extension of an AF can consist of more than one argument and draw multiple conclusions. A self-preferring extension provides a support for preferring itself to its rivals, while a self-discarding extension provides a support for preferring its rival to itself. Formally:

Definition 17 (*Self-preferring* and *self-discarding* preferred/stable extensions)**.** Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ be an ATPC, $\mathcal{H}_{pri} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{pri} \rangle$ the primary PAF corresponding to $\mathcal{T}$ and $\{\mathcal{E}_1, \ldots, \mathcal{E}_n\}$ a set of conflicting preferred/stable extensions of $\mathcal{H}_{pri}$.

- $\mathcal{E}_i$ is *self-preferring* iff there exists a set of criteria $\mathcal{C}_{sp}$ such that for all $c \in \mathcal{C}_{sp}$, it holds that $\texttt{Guard}(c) \subseteq \texttt{Concs}(\mathcal{E}_i)$ and the PAF with the preferences defined from $\mathcal{C}_{sp}$ has only one preferred/stable extension $\mathcal{E}_i$.
- $\mathcal{E}_i$ is *self-discarding* iff there exists a set of criteria $\mathcal{C}_{sd}$ such that for all $c \in \mathcal{C}_{sd}$, it holds that $\texttt{Guard}(c) \subseteq \texttt{Concs}(\mathcal{E}_i)$ and $\mathcal{E}_i$ is not a preferred/stable extension of the PAF with the preferences defined from $\mathcal{C}_{sd}$.[11]

In Example 1, the extension $\mathcal{E}_1$ is self-preferring, while $\mathcal{E}_2$ is self-discarding. The above definition suggests that we can classify extensions into three categories after introducing preferences: self-preferring extensions, self-discarding extensions and the extensions that are neither self-preferring nor self-discarding. Based on this classification of extensions, we can also set a preference ordering over a set of extensions, namely, a powerset of arguments.[12] A self-preferring extension should be preferred to an extension which is neither self-preferring nor self-discarding. An extension which is neither self-preferring nor self-discarding is preferred to a self-discarding extension. It is also self-evident that a self-preferring extension is preferred to a self-discarding one and that is why we choose $\{A, C\}$ in Example 2. In particular, conflicting extensions of an AF without preferences are at the same preference level because they all belong to the same category: extensions which are neither self-preferring nor self-discarding.

When we justify preferences in an AF, we usually adopt the skeptical viewpoint (Example 1 and 3), since it ensures that incompatible criteria are not justified simultaneously. However, what should we do if the skeptically justified criteria do not have adequate preference information for resolving the conflict between extensions? Meta-preference criteria may be useful. Then, what if we have no meta-preference criteria? In such a case, an alternative approach may be to look at the extensions of an AF and to consider that each extension give rise to different justified preferences, which in turn give rise to different PAFs. Then, instead of a sequence of PAFs, a (two-level) tree structure of PAFs comes into being, where every leaf represents different possibilities.

[11] This definition can also be extended for the resulting PAF of the nth filtering.

[12] The extensions over which we set a preference ordering are those of an original framework. In other words, let $\mathcal{H} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq \rangle$ be a PAF, then we can set a preference ordering over the extensions of $\langle \mathcal{A}, \mathcal{R} \rangle$, not $\langle \mathcal{A}, \mathcal{R}, \succcurlyeq \rangle$ as in [9].

Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ be an ATPC, $\mathcal{H}_{pri} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{pri} \rangle$ the primary PAF corresponding to $\mathcal{T}$ and $\{\mathcal{E}_1, \ldots, \mathcal{E}_n\}$ be a set of conflicting preferred/stable extensions of $\mathcal{H}_{pri}$. If no preference criteria are justified by $\mathcal{H}_{pri}$, then a (two-level) PAF tree corresponding to $\mathcal{T}$ would be defined as follows:

- the root node holds $\mathcal{H}_{pri}$;
- for every leaf $\mathbf{L}$, $\mathbf{L}$ is the PAF $\langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{\mathbf{L}} \rangle$, where $\succcurlyeq_{\mathbf{L}}$ is the preference ordering on $\mathcal{A}$ produced from the set of criteria that are justified by $\mathcal{E}i$ $(1 \leqslant i \leqslant n)$.

In the above definition, it can be easily noticed that an extension of a leaf PAF coincides with or belongs to an extension of the primary PAF (root PAF) from Proposition 1 under preferred/stable semantics. Hence, every set of preference criteria justified by an extension of a leaf PAF is also justified by an extension of the root PAF. As a result, it seems needless to further branch the tree structure with preferences justified by an extension of a leaf PAF.

Sometimes, the preferences justified by an extension of a leaf PAF may be enough for resolving the conflict. Or sometimes, the leaf PAFs may share one or more extensions. If the leaf PAFs share only one extension, the extension may be chosen by the user. Consider the following examples.

Example 6. Let us consider a scenario where an agent should decide whether to buy a laptop computer or not. It is clear that he should buy a computer of which CPU speed is high, RAM capacity is large and battery is good. However, the computer that he is thinking of buying does not seem to have all of these three properties. The salesperson who is a computer expert says that the CPU speed of the computer is incredibly high, while a window-shopper, who represents himself as another computer expert, says that CPU speed of the computer is not high. Then, the agent thinks that the window-shopper is not trustworthy and pretends to be an expert. The agent also believes that the RAM capacity is large and the battery is not good. He does not have any total preference ordering over the three attributes (CPU speed, RAM capacity, battery), but he knows that the high CPU speed should be backed up by large RAM capacity. Therefore, in the context where the actual CPU speed is high, the agent prefers RAM capacity to battery, while in the context where the CPU speed is not high, the agent prefers battery to RAM capacity.

Then, let $\mathcal{T}_6 = \langle \mathcal{AS}_6, \mathcal{KB}_6 \rangle$ be an ATPC with $\mathcal{AS}_6 = \langle \mathcal{L}, \bar{}, \mathcal{R}_6, \mathcal{C}_6, \mathtt{n} \rangle$ and $\mathcal{KB}_6 = \langle \mathcal{K}_6, \mathcal{C}'_6 \rangle$ where:

$\mathcal{K}n_6 = \{Expert1\_says(high\_CPU\_speed(com1)), Expert2\_says(\neg high\_CPU\_speed(com1))\}$;

$\mathcal{K}p_6 = \{Untrustworthy(expert2), large\_RAM\_capacity(com1), \neg good\_battery(com1)\}$;

$\mathcal{R}s_6 = \{r_1\}$; $\mathcal{R}d_6 = \{r_2, r_3, r_4, r_5, r_6, r_7\}$; $\mathcal{C}_6 = \{c_1, c_2\}$ and

$r_1 : Untrustworthy(expert2) \rightarrow \neg \mathtt{n}(r_3)$;

$r_2 : Expert1\_says(high\_CPU\_speed(com1)) \Rightarrow high\_CPU\_speed(com1)$;

$r_3 : Expert2\_says(\neg high\_CPU\_speed(com1)) \Rightarrow \neg high\_CPU\_speed(com1)$;

$r_4 : high\_CPU\_speed(com1) \Rightarrow buy(com1)$;

$r_5 : \neg high\_CPU\_speed(com1) \Rightarrow \neg buy(com1)$;

$r_6 : large\_RAM\_capacity(com1) \Rightarrow buy(com1)$;

$r_7 : \neg good\_battery(com1) \Rightarrow \neg buy(com1)$;

$r_8 : \neg large\_RAM\_capacity(com1) \Rightarrow \neg buy(com1)$;

$r_9 : \mathit{good\_battery}(\mathit{com1}) \Rightarrow \mathit{buy}(\mathit{com1})$;

$c_1 = \langle \{\mathit{high\_CPU\ \_speed}(\mathit{com1})\}, \{(r_6 \succ r_7), (r_8 \succ r_9), (r_4 \succ r_7), (r_5 \succ r_9)\} \rangle$;

$c_2 = \langle \{\neg \mathit{high\_CPU\ \_speed}(\mathit{com1})\}, \{(r_7 \succ r_6), (r_9 \succ r_8), (r_7 \succ r_4), (r_9 \succ r_5)\} \rangle$.

The following arguments can be built over the theory.

$A = \mathit{Expert1\_says}\big(\mathit{high\_CPU\ \_speed}(\mathit{com1})\big)$;

$B = \mathit{Expert2\_says}\big(\neg \mathit{high\_CPU\ \_speed}(\mathit{com1})\big)$;

$C = \mathit{Untrustworthy}(\mathit{expert2})$;

$D = \mathit{large\_RAM\_capacity}(\mathit{com1})$;

$E = \neg \mathit{good\_battery}(\mathit{com1})$;

$A_1 = A \Rightarrow \mathit{high\_CPU\ \_speed}(\mathit{com1})$;

$B_1 = B \Rightarrow \neg \mathit{high\_CPU\ \_speed}(\mathit{com1})$;

$C_1 = C \rightarrow \neg \mathtt{n}(r_3)$;

$D_1 = D \Rightarrow \mathit{buy}(\mathit{com1})$;

$E_1 = E \Rightarrow \neg \mathit{buy}(\mathit{com1})$;

$A_2 = A_1 \Rightarrow \mathit{buy}(\mathit{com1})$;

$B_2 = B_1 \Rightarrow \neg \mathit{buy}(\mathit{com1})$.

The theory has no criterion whose guard is $\emptyset$. Fig. 16 depicts the primary PAF $\mathcal{H}_6$ built over $\mathcal{T}_6$ (diamond arrows represent undercut).

Note that the AF above has two conflicting preferred/stable extensions: $\mathcal{E}_1 = \{A, B, C, D, E, C_1, A_1, A_2, D_1\}$ and $\mathcal{E}_2 = \{A, B, C, D, E, C_1, E_1\}$. The former extension recommends buying the laptop, while the latter recommends not buying it. Since $\mathtt{Guard}(c_1) \nsubseteq \mathtt{Concs}(\mathcal{E}_1) \cap \mathtt{Concs}(\mathcal{E}_2)$, $\mathtt{Guard}(c_2) \nsubseteq \mathtt{Concs}(\mathcal{E}_1) \cap \mathtt{Concs}(\mathcal{E}_2)$, neither c_1 nor c_2 is skeptically justified by the primary framework. Then, the conflict between $\mathcal{E}_1$ and $\mathcal{E}_2$ remains unresolved. Then, what should we do if the skeptical standpoint does not justify enough preference criteria for resolving a conflict? Then, we can construct a two-level tree structure of PAFs, where the root holds $\mathcal{H}_6$ and the leaves hold PAFs with preferences justified by

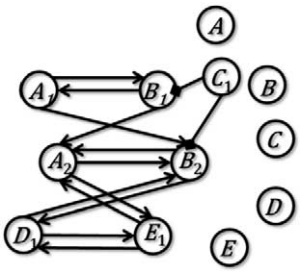

Fig. 16. Making decision without preference criteria.

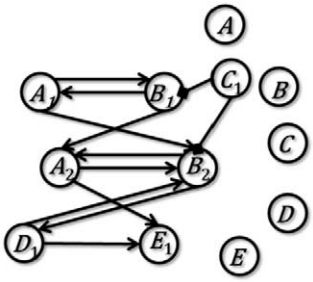

Fig. 17. Making decision with justified preference criteria by an extension.

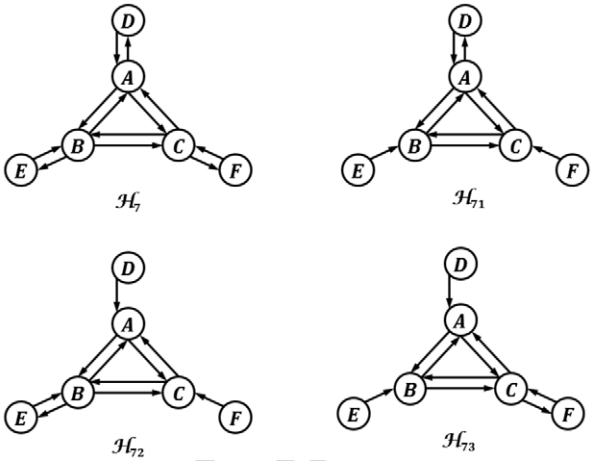

Fig. 18. What if the primary PAF skeptically justifies no preference criteria?

$\mathcal{E}_1$ and $\mathcal{E}_2$. From the fact that $\texttt{Guard}(c_1) \subseteq \texttt{Concs}(\mathcal{E}_1)$ (since we have $\texttt{Conc}(A_1) = \texttt{Guard}(c_1)$), $\mathcal{E}_1$ justifies the preference criterion c_1, while $\mathcal{E}_2$ does not justify any criteria. Now, the primary PAF should be repaired by means of the preference information which the criterion c_1 stores and the leaf PAF $\mathcal{H}'_6$ is built as Fig. 17 depicts. As one can notice, $E \prec A_2$ and $D_1 \succ E$, thus the attacks $\langle E, A_2\rangle$ and $\langle E, D_1\rangle$ should be removed from the framework. The leaf PAF has only one preferred extension $\mathcal{E}_1$, thus $\mathcal{E}_2$ is rejected and the conflict is resolved. The other leaf is the same as $\mathcal{H}_6$ because $\mathcal{E}_2$ justifies no preference criteria. Then, the extension that is common to two leaves is $\mathcal{E}_1$. Therefore, it seems to be more desirable to choose $\mathcal{E}_1$, which recommends buying the laptop. In fact, as we can see, $\mathcal{E}_1$ provides a support for preferring itself to its opposite extension $\mathcal{E}_2$, thus $\mathcal{E}_1$ is a *self-preferring* extension. As mentioned above, a self-preferring extension ($\mathcal{E}_1$) is preferred to an extension that is neither self-preferring nor self-discarding ($\mathcal{E}_2$).

The examples show that our proposal sometimes produces interesting results where an extension provides a support for preferring itself and rejecting its rival extensions.

Example 7. Consider an ATPC, over which the primary PAF $\mathcal{H}_7$ is built as in Fig. 18 (the theory contains no preference criteria whose guards are $\emptyset$).

The primary PAF has four preferred/stable extensions: $\{A, E, F\}$, $\{B, D, F\}$, $\{C, D, E\}$ and $\{D, E, F\}$. Now, assume that we have three preference criteria c_1, c_2 and c_3 with $\texttt{Guard}(c_1) \subseteq$

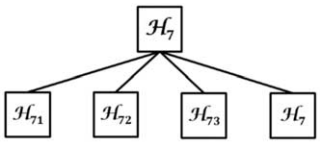

Fig. 19. A tree structure of PAFs.

$\texttt{Concs}(A)$, $\texttt{Guard}(c_2) \subseteq \texttt{Concs}(B)$ and $\texttt{Guard}(c_3) \subseteq \texttt{Concs}(C)$, such that c_1, c_2 and c_3 store preference information from which the orderings ($E \succ B$, $F \succ C$), ($D \succ A$, $F \succ C$) and ($D \succ A$, $E \succ B$) come, respectively. The PAFs $\mathcal{H}_{71}$, $\mathcal{H}_{72}$, $\mathcal{H}_{73}$, depicted in Fig. 18, are those with these argument orderings. The primary PAF skeptically justifies no preference criteria. Therefore, we can construct a tree structure of PAFs as Fig. 19.

In Fig. 19, the primary PAF $\mathcal{H}_7$ is set once again as a leaf PAF, because the extension $\{D, E, F\}$ does not justify any preference criteria, and thus the PAF that the extension gives rise to is the primary PAF itself. Then, the extension common to four leaf PAFs is $\{D, E, F\}$. Here, notice that the extensions of $\mathcal{H}_{71}$, $\mathcal{H}_{72}$ and $\mathcal{H}_{73}$ are of the extensions $\mathcal{H}_7$. Therefore, it is needless to further branch the tree structure.

As shown through the examples, if a sequence of PAFs is impossible due to lack of skeptically justified preference criteria, we can enumerate possible alternative conclusions by means of a tree structure of PAFs.

6. Related works

Preference is a common topic in the field of artificial intelligence (for more details, see [51]). Preferences can also be embedded into an AF that can serve as a core engine for reasoning under imperfect and inconsistent information in intelligent systems. In a PAF, a preference ordering over the set of arguments is embedded to filter the attack relation between arguments. However, some early PAFs [1,2,13,15,43] do not guarantee conflict-freeness of extensions. Since if one argument asymmetrically attacks another and the attackee is preferred to the attacker, then this attack is counted as failed one, and thus should be removed from the AF. This gives rise to a very unintuitive result where two conflicting arguments are in the same extension if the attack is failed. Therefore, several works [7,9,46] have been done to guarantee conflict-freeness in a PAF and make PAFs satisfy rationality postulates presented in [27].

One proposal is to define the notion of conflict-freeness in terms of attack (including failed attacks) rather than defeat (excluding failed attacks) in ASPIC+ framework. Modgil & Prakken argued that since attacks indicate the mutual incompatibility of the information contained in the attacking and attacked arguments, then intuitively one should continue to define conflict-free sets in terms of those that do not contain mutually attacking arguments [46]. Defining conflict-freeness of an extension in terms of attack, not defeat, make the framework satisfy rationality postulates, but violate Dung's fundamental lemma. Thus, they defined a '*reasonable*' argument ordering to make their framework satisfy Dung's fundamental lemma. Another proposal which was made by [7] is to introduce preferences into the semantics level, not attack level. Instead of changing original attacks, they take into account preferences when determining the acceptability of arguments, i.e. at the semantics level. They define a semantics as a dominance relation on the powerset of the set $\mathcal{A}$ of arguments [7]. Then, they pin down three postulates that such a relation should satisfy and generalize Dung's semantics. In another paper, Amgoud & Vesic also suggested inverting the direction of failed attacks to guarantee the conflict-freeness of extensions in a PAF

[9]. However, if it is possible to invert the direction of an asymmetric attack, then should the attack be counted as a symmetric one from the beginning? Since the ASPIC+ framework gurantees conflict-freeness of extensions (Proposition 16 of [46]), the sequence of PAFs built over an ATPC bears desirable properties that are shown in Section 4.3.

Amgoud & Vesic [8,9] also identified two roles that preferences may play in an AF: (1) handling failed (critical) attacks and (2) refining the result of a PAF. To make the preferences in their system play the second role, they defined two preference relations (called *democratic* and *elicit*) over the powerset of arguments. These relations are used to return the best among the extensions of the repaired AF. In this paper, we classify extensions into three categories: self-preferring extensions, self-discarding extensions and extensions that are neither self-preferring nor self-discarding. We can also set a preference ordering over these three categories. Our classification and ordering can be compared to Amgoud & Vesic's preference relation over a powerset of arguments. Furthermore, Proposition 3 shows that our formalism makes preferences play two roles simultaneously.

The need to justify preferences before introducing them to reasoning has recently been emphasized in [38]. According to them, we do not have control over legal norms and their modification, but we can rather argue that one norm instead of another should be applied to a specific case. Governatori et al. also gave an example where a conflict between norms occurs and different criteria prefer different norms [38]. In legal reasoning, a lawyer should not only make an appeal to some of three criteria (*Lex Superior, Lex Posterior* and *Lex Specialis* which are appeared in Section 4.1), but also justify the criteria if a conflict between those criteria occurs as in their example.

For the sake of justifying preferences before reasoning from them, Booth et al. employ the model of so-called *property-based preferences*, where a preference ordering over arguments is derived from preferences over properties of the arguments. [23] Therefore, it can be regarded that, in [23], preferences are justified by the information on properties of arguments and change as the result of moving to different motivational states, which also bring about some change in argument properties. In ASPIC+, argument orderings are derived from the preferences over defeasible rules and ordinary premises, which are *components* of arguments. In our approach, those preferences over defeasible rules and ordinary premises are also justified by AFs. Furthermore, while Booth et al. deal with the justification of preferences in abstract argumentation [36], we integrate justifying preferences with reasoning from preferences in structured argumentation.

A proposal to use AFs for preference elicitation has appeared in [58]. Sedki explores the correlation between a given *PQCL* theory and a value-based AF and discusses using value-based AF for preference elicitation. Recently, Oguego et al. has proposed to use argumentation to manage user preferences [48]. They explore a generalized framework that can be used to handle conflicts among user preferences in ambient intelligence.

Preference learning has been a substantial topic in the field of artificial intelligence [50,51]. Learning or eliciting preferences means to acquire preference information in either direct or indirect way, from preference statements, critiques to examples, observation of user's clicking behavior, etc. [51] However, learning preferences is different from justifying (or reasoning about) preferences. An agent rarely has a preference that holds under any condition. The agent rather has a preference that holds under a certain condition (for example, Yong Chol prefers red wine to white wine, given that the second course is fish). Therefore, learning preference not only means acquiring preference information (preference for red wine to white wine), but also stipulating the condition (given that the second course is fish) under which such preference information is justified. Nevertheless, a justified preference means that the condition attached to the preference has been satisfied. That is, justifying preference means proving that such condition

attached to a preference has already been satisfied. Therefore, learning preferences is prerequisite to justifying preferences.

Modgil has acknowledged the non-prespecificity of the preference ordering of an AF [43]. As a result, he has proposed an extended argumentation framework (EAF) that enables even *reasoning about* preferences. He adopted recursive attack as the means of producing a preference ordering over two conflicting arguments. He extended Dung's AF by adding a recursive attack relation which ranges from an argument to an attack between two arguments. His framework is powerful especially in resolving conflicts between incompatible preferences because it regards two arguments as being in a symmetric attack if they respectively attack an attack and the reversal of the attack. Later, the EAF has been extended to structured extended argumentation framework (SEAF), which satisfies rationality postulates for *bounded hierarchical* EAFs [44].

SEAFs allow a special kind of rule whose head is a rule (or ordinary premise) priority, and thus make it possible to construct arguments expressing preferences over other arguments. As a result, we can reason about preferences with SEAFs. On the other hand, SEAFs determine the recursive attack relation based on the rule priorities, which are conclusions of arguments. Therefore, we can also reason from preferences with SEAFs.

Here, four main points are worth remarking to compare our model with SEAFs. First, as already remarked, failed attacks are recursively attacked in SEAFs. This leads to finding a special set of defeats called a *reinstatement set* to determine whether an argument is acceptable or not with respect to a certain set of arguments. A reinstatement set for a defeat ensures that the defeat succeeded in surviving the recursive attack on it (for the definition of reinstatement set, see [43]). In our formalism, failed attacks are removed from a framework as in some others such as deductive argumentation and ABA+.

The second difference concerns the fact that our formalism adopts the notion of preference criteria, while SEAFs use the special kind of rule in order to model the way in which we justify or reason about preferences. In fact, every preference criterion can be converted into one or more rules whose heads are rule (or ordinary premise) priorities.[13] For example, the preference criterion $\langle\{p, q\}, \{r_1 > r_2, r_3 > r_4\}\rangle$ can be converted into two rules: $p, q \Rightarrow r_1 > r_2$ and $p, q \Rightarrow r_3 > r_4$. Apart from this formal correspondence, it should be remarked that our model is different from SEAFs in how to deal with inconsistent preferences.

Intuitively, it should be avoided that inconsistent preferences are justified together and thus applied to an AF. If an argumentation formalism allows rules whose heads are rule priorities, it should also include the special kind of strict rules regarding rule priorities. For example, Modgil and Prakken's extended argumentation theory contains strict rules axiomatising partial orders such as $y > x, z > y \rightarrow z > x$ (here, x, y, z are meta-variables ranging over rule names) [44]. Instead of containing such spcial kind of strict rules, we define the notion of a *valid criteria set*, which ensures that no pair of incompatible preference criteria are justified simultaneously. As noted in Section 4, the notion of valid criteria set is based on the intuition that every preference criterion bears a certain condition under which its application is justified and, as a result, incompatible criteria should also bear inconsistent guards. In other words, preference criteria are incompatible because their guards are inconsistent. Consider, for example, that a rule priority $r_1 > r_2$ (here r_1 refers to "what the source **A** said is true.", while r_2 refers to "what the source **B** said is true.") can reference a statement a (which refers to "**A** is more trustworthy than **B**")

[13] Since every preference criterion can be converted into the special rules, justifying preferences and reasoning about preferences, we think, can be seen as having the same meaning if they are broadly understood. They all mean giving support or reason for applying the preferences or explaining why the preferences should be adopted. In this paper, we use the term "justifying" instead of "reasoning about" simply for differentiating our model from SEAFs.

made by a source $\mathbf{S_1}$, while $r_2 > r_1$ can reference a statement b (which refers to "**B** is more trustworthy than **A**") made by another source $\mathbf{S_2}$. Hence, the condition under which the application of $r_1 > r_2$ and $r_2 > r_1$ are justified can be represented by two statements a and b, respectively. Here, intuitively, there must be, at least, a strict rule $a \rightarrow \neg b$.

In our formalism, since preference criteria are employed, the conflicts among arguments due to preferences are usually concealed. In Example 5, it looks like that, intuitively, there is a conflict between B and C because C demotes B, but this is not a part of this AF. Nevertheless, SEAFs makes such a kind of conflicts come to the fore by means of recursive attack (if we had built EAF, we could produce a recursive attack from C to $\langle B, A\rangle$ in Example 5). Even though conflicts due to preferences (e.g. that between B and C) do not come to the fore in PAFs built over an ATPC, a set of arguments containing such kind of conflicts cannot be an extension of the advanced PAF under Dung's standard semantics. In Example 5, the set of arguments $\{B, C\}$ is not the preferred/stable extension of the advanced version and it is *self-discarding*. Therefore, we can say that our model successfully deals with the problem related with such hidden conflicts.

Third, SEAFs should allow *collective* attacks (or *joint* attacks as coined by the others [36,47]), while our model does not. As noted earlier, an argument ordering is derived from the preferences over defeasible rules or ordinary premises in both SEAFs and our framework. Then, single argument may contain more than one defeasible rules or ordinary premises. Therefore, a defeasible rule (or ordinary premise) priority that is the conclusion of a SEAF argument may not be enough for determining that an argument is preferred to another. Therefore, it becomes possible for two or more arguments to *collectively* (and recursively) attack an attack in order to undermine the success of the latter as defeats [44].

Fourth, we can contrast our model with Modgil and Prakken's SEAFs, in which all reasoning about preferences is catered for in a *single* EAF. In contrast, an ATPC is used for building a *sequence* of PAFs, where a PAF plays the role of the reasoning mechanism for justifying preferences that the next PAF is to rest on.

In fact, it has shown that some EAFs can be converted into hierarchical AFs $\langle \mathcal{A}_n, \mathcal{R}_n\rangle, \ldots, \langle \mathcal{A}_1, \mathcal{R}_1\rangle$ where each $\langle \mathcal{A}_i, R_i\rangle$ outputs justified claims that yield preferences applied to attacks amongst arguments in $\langle \mathcal{A}_{i-1}, \mathcal{R}_{i-1}\rangle$ $(i > 2)$. [42,43] This seems to be very similar to the sequence of PAFs built over an ATPC. However, in the sequence of PAFs, all PAFs share the same set of arguments and attack relation. In such a sequence, the only thing that makes a PAF different from another is its argument ordering. In contrast, individual AFs involved in a hierarchical AF have different set of arguments and thus attack relations.

Here, it is remarkable that our approach can be closely related with dynamics of AFs [21,22]. Let $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq_1\rangle, \ldots, \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_n\rangle$ be the sequence of PAFs built over an ATPC and $\langle \mathcal{A}, \mathcal{R}'_1\rangle, \ldots, \langle \mathcal{A}, \mathcal{R}'_n\rangle$ the repaired AFs of $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq_1\rangle, \ldots, \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_n\rangle$, respectively. Then, it can be easily noticed that every repaired version of the PAFs involved in the sequence is an *attack abstraction*[14] from the original framework $\langle \mathcal{A}, \mathcal{R}\rangle$. In addition, the repaired framework of the ith PAF is an attack abstraction from that of the $i-1$th PAF $(1 < i \leqslant n)$ under stable semantics (and even under preferred semantics if the repaired framework of the ith PAF does not contain any odd-length cycles of attacks).

It is central issue in dynamics of AFs to stipulate some principles, by which we can expect the outcome of a changed AF. Such principles are usually of the form "If an argument (or an attack) is removed (or added), such that a given property P_1 is satisfied, then the outcome of the argumentation framework satisfies P'_2" [30]. Here, P_1 expresses some constraint on the addition or removal of arguments or attacks

[14]Given two AFs $\langle \mathcal{A}, \mathcal{R}\rangle$ and $\langle \mathcal{A}', R'\rangle$, $\langle A', R'\rangle$ is an *attack abstraction* from $\langle \mathcal{A}, \mathcal{R}\rangle$ iff $\mathcal{A} = \mathcal{A}'$ and $\mathcal{R}' \subseteq \mathcal{R}$ [12].

of an AF, while P_2 usually expresses the relation between the original and changed AFs. As shown in Section 4.3, in the sequence of PAFs, where only the defeat relations subsequently change, each of skeptically accepted arguments by a PAF is also skeptically accepted by the subsequent PAFs. It is also impossible for every credulously rejected arguments by a PAF to be credulously accepted by the subsequent PAFs. Therefore, we can find two sets of arguments whose status does not change throughout the sequence of PAFs: skeptically accepted and credulously rejected arguments by the primary PAF (under stable semantics and under preferred semantics when the primary PAF does not contain any odd-length cycles of attacks). Such interesting results cannot be recognized in hierarchical AFs.

One of the important problems of AF dynamics is a *semantical defect* of an agent's AF which prevents her from drawing any plausible conclusion in the sense that no argument is accepted, or prevents her from drawing enough conclusions in the sense that the accepted arguments are not enough for giving answers she wants [12].[15] Such undesirable situations should be avoided. Therefore, the agent should change the AF so as to *cure the defect*. If an agent counters a semantical defect, she wants to know what are minimal diagnoses of the given knowledge base, i.e. which parts are causing the semantical defect. For instance, a certain minimal diagnosis may consist of arguments which are somehow out of date or not as significant in comparison to the others and thus should be discarded [12]. An efficient minimal diagnosis may also consist of attacks where the attackee is preferred to the attacker because of some reason (e.g. preference criteria in our approach). Consequently, an agent may tend to remove these attacks. An agent can find such diagnosis by introducing some preference criteria and constructing a sequence of PAFs. In the sequence of PAFs $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq_1 \rangle, \ldots, \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_n \rangle$ built over an ATPC, where the original AF $\langle \mathcal{A}, \mathcal{R} \rangle$ suffers from a semantical defect, if $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq_n \rangle$ has a single extension under a given semantics, then, the attacks removed by $\preccurlyeq_n$ can be regarded as an *attack-based diagnosis*. Furthermore, the repaired framework of $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq_n \rangle$ becomes an *attack-based repair* of $\langle \mathcal{A}, \mathcal{R} \rangle$ (for formal definitions of diagnosis and repair, see Definition 3.2 of [12]). Here, we should note that preference criteria whose guards are $\emptyset$ play a crucial role for finding diagnosis of a semantical defect where no argument is accepted. If the primary PAF based on such preferences still has multiple extensions, we may also build subsequent PAFs with justified preferences.

EAFs and SEAFs have also been carefully studied under the grounded semantics. It is shown that the grounded extension of a special kind of EAF called *bounded hierarchical* EAF is the least fixed point of its characteristic function.[16] In the sequence of PAFs, the grounded semantics can also play an important role, not just because, the repaired framework of a PAF, as a standard Dung AF, has the least fixed point of its characteristic function as its grounded extension, but also because every PAF has a single grounded extension that monotonically changes. Formally:

[15] In [40], Baumannn and Ulbricht define a semantical defect of an AF as a situation where it is impossible to draw any plausible conclusion because no argument is accepted. In this paper, I have broadened the meaning of semantical defect. A semantical defect of an agent's AF may also include a situation where it is impossible to draw *enough* conclusions for giving needed answers even though the AF includes arguments which gives such answers. This kind of situations usually comes into being when an AF has multiple extensions and thus skeptically accepted arguments are not enough for giving the answers an agent wants. For instance, in Example 3, only E is skeptically accepted, which concludes that the robot is near the store. However, this is not an answer that the agent wants. The robotic agent should decide which boxes to carry first.

[16] Generally, given an AF $\mathcal{H}$ with a set of arguments $\mathcal{A}$, the characteristic function $\mathtt{F}_{\mathcal{H}}$ is defined as follows:

$$\mathtt{F}_{\mathcal{H}} : 2^{\mathcal{A}} \mapsto 2^{\mathcal{A}},$$

$$\mathtt{F}_{\mathcal{H}}(\mathcal{S}) = \{A | A \text{ is acceptable wrt. } \mathcal{S} \text{ in } \mathcal{H}\}.$$

Proposition 8. *Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ be an ATPC with valid criteria sets, $\mathcal{H}_n = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_n \rangle$ and $\mathcal{H}_{n+1} = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_{n+1} \rangle$ respectively the resulting PAFs of the nth and $n+1$th filtering which have reasonable argument orderings. If $\mathcal{E}_n$ and $\mathcal{E}_{n+1}$ are respectively grounded extensions of $\mathcal{H}_n$ and $\mathcal{H}_{n+1}$, it holds that $\mathcal{E}_n \subseteq \mathcal{E}_{n+1}$.*

The above proposition shows that the sequence of PAFs built over an ATPC brings about a monotonic increase in grounded extension as the justified preferences change. As a result, in the sequence, an accepted argument by a PAF under grounded semantics cannot be rejected by the subsequent PAFs.

Bench-Capon et al. has also recognized the non-prespecificity of a value order and proposed a novel solution to the problem of producing value orders [15]. According to them, 'we cannot assume that the parties to a debate will come with a clear ranking of values: rather these rankings appear to emerge during the course of the debate.' They defined a dialogue process for evaluating the status of arguments in a value-based AF. The dialogue process can be used to construct positions by which the orderings of values will be determined. They adopted the dialogue framework that was developed to prove the acceptability of arguments in AFs [15, Section 6].

The audience-dependency of preferences (or value orders) in an AF was addressed in value-based AFs [13,15]. Perrussel et al. also defined *multiple* PAF to model the intuition that different agents have different preferences [49]. However, even a single agent may promote different preferences in different contexts. An agent will select a preference criterion only if he notices that the condition, under which the criterion is justified, is satisfied in a specific context. Our formalism reflects this intuition by borrowing the notion of guard from [60]. Teze et al. proposed a recommender system which uses justified preference criteria [60]. They introduced the notion of conditional-preference expressions (somewhat like the notion of a conditional preference network [29]) that represents IF... THEN... ELSE IF... structure of criteria selection. In their system, a single criterion is justified in a context. Furthermore, other structured argumentation formalisms that trade on the mathematical notion of conditional preference network like have recently been studied and applied [26,32]. However, in our formalism, we do not adopt the notion of conditional-preference expression, because we thought that multiple compatible criteria may be available behind a preference ordering in a context. Furthermore, their recommender systems justify a criterion by means of strict derivations. This implies that a selection of a criterion is based on perfect information and, therefore, it could never be doubted. However, in practice, a selection of a criterion is rarely taken for granted, rather it is based on imperfect information, and thus may be questioned or even be in conflict with some other available piece of information. From this viewpoint, we adopt the argumentation-approach for the sake of justifying preferences. In our formalism, a PAF is used for not only drawing plausible conclusions, but also justifying preferences.

7. Conclusions

Many successful preference-based and value-based AFs prove the usefulness of preferences or value orders taken as an input of the AF. Nonetheless, preferences in an argumentation or value orders in a practical reasoning are not 'universal presuppositions', thus, they should also be an output of an argumentation or reasoning [15]. This leads us to building a model where both justifying preferences and determining acceptable conclusions by taking the preferences into account are possible. In this paper, we have argued that a PAF built over an argumentation theory should adopt preferences that have been justified by another framework with the same arguments and attack relations. Hence, we propose to build

a sequence of PAFs over an argumentation theory, where a PAF justifies preferences that the next PAF is to be based on.

On the other hand, justifying preferences is reduced to justifying criteria behind them with available information in a context. Accordingly, we modify the notion of APSIC+ argumentation theory with preference criteria and as a result, an ATPC is defined

It is very interesting that in a standard ASPIC+ framework, preferences may not only be used for filtering the attack relation, but also for filtering the extensions of the original AF (Proposition 1). It also makes it possible for us to build such a sequence of PAFs over an ATPC.

The *sequence* of PAFs built over an argumentation theory with preference criteria (ATPC) involves two or more individual PAFs: primary PAF, advanced PAF, third PAF and so on. The primary PAF is for justifying preferences and the advanced PAF is for concluding or determining acceptable statements. The primary PAF takes into account only the preference criteria whose guards are empty set, while the advanced PAF takes into account the preference criteria justified by the primary PAF. Our formalism has also been thoroughly discussed through some examples. Especially, Example 6 shows that our proposal accords well with practical wisdom when some attributes of an object that is under consideration depend on each other. Let us consider Example 6 again.

Example 6 (cont.)**.** Let us replace the theory $\mathcal{T}_6 = \langle \mathcal{AS}_6, \mathcal{KB}_6 \rangle$ with $\mathcal{T}_7 = \langle \mathcal{AS}_7, \mathcal{KB}_7 \rangle$ where $\mathcal{AS}_7 = \langle \mathcal{L}, ^-, \mathcal{R}_7, \mathcal{C}_7 \rangle$ and $\mathcal{KB}_7 = \langle \mathcal{K}_7, \mathcal{C}'_7 \rangle$ such that $\mathcal{K}n_7 = \{$*Expert1_says*(*high_CPU_speed*(*com1*)), *Expert2_says*(¬*high_CPU_speed*(*com1*))$\}$, $\mathcal{K}p_7 = \{$*Untrustworthy*(*Expert2*), ¬*large_RAM_capacity* (*com1*), *good_battery*(*com1*)$\}$, $\mathcal{R}s_7 = \mathcal{R}s_6$, $\mathcal{R}d_7 = \mathcal{R}d_6$, $\mathcal{C}_7 = \mathcal{C}_6$ and $\mathcal{C}'_7 = \mathcal{C}'_6$. Then, the PAF corresponding to $\mathcal{T}_7$ will recommend not buying *com1*.

When we decide which object or course of action to choose, their attributes should be considered. The selection of an object or a course of action derives from an ordered set of their attributes. For example, if a house is close to work but not large and an agent prefers size to distance to work, then the agent will not choose the house, although it is close to work. In contrast, if the agent prefers distance to work to size, then he will choose the house, although, it is not large. Deriving preferences over objects from ordered sets of their attributes was deeply studied in [40]. However, sometimes, some attributes of an object or course of an action may be dependent each other. Here, dependence of attributes means that one attribute is helpless without the presence of the other attribute. For example, the highness of a computer CPU is dependent on the largeness of its RAM. The high CPU of a computer should be backed up by a large RAM, that is, if the RAM capacity of a computer is not large, its CPU whose speed is high is helpless.[17] In such a case, it is hard to set a unified ranking over the attributes. One cannot easily say that he prefers CPU speed to battery capacity of a computer since if the computer does not have large RAM, the CPU whose speed is high is helpless. Then, our formalism will be useful for dealing with such problems. If some of the desired attributes (high CPU speed, large RAM capacity) depend on each other and at least one of them is not equipped, it is wise to prefer an independent attribute (battery capacity) to the dependent attributes. The results of Example 6 show that our model accords well with this wisdom. Furthermore, our formalism that integrates justifying preferences with reasoning from the justified preferences brings about interesting results where extensions may be *self-preferring* or *self-discarding*. Thus, we can classify extensions into three categories and set an ordering over extensions based on this classification.

[17] Here, the term "helpless" does not mean that the CPU is useless, but means that the high CPU speed is useless (i.e. even low CPU speed is OK). That is, when the RAM capacity is not large enough, it will be rather better off using a cheaper CPU whose speed is low, instead of high and expensive CPU.

Acknowledgements

I would like to thank the anonymous reviewers for their comments and suggestions that helped me to improve the paper.

Appendix. Proof of propositions

Proposition 1. *Let $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle$ be a PAF corresponding to an argumentation theory which has a reasonable argument ordering. Then, under preferred semantics, it holds that for all $\mathcal{E}' \in \mathtt{Ext}_{pre}(\langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle)$, there exists an extension $\mathcal{E} \in \mathtt{Ext}_{pre}(\langle \mathcal{A}, \mathcal{R} \rangle)$ such that $\mathcal{E}' \subseteq \mathcal{E}$. Under stable semantics, it just holds that $\mathtt{Ext}_{sta}(\langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle) \subseteq \mathtt{Ext}_{sta}(\langle \mathcal{A}, \mathcal{R} \rangle)$.*

Proof.

– Under preferred semantics

Let $\langle \mathcal{A}, \mathcal{R}' \rangle$ be the repaired AF of $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle$. From Proposition 16 of [46], $\mathcal{E}'$ is also conflict-free wrt. $\mathcal{R}$. Now, we are to prove that $\mathcal{E}'$ defends all of its elements wrt. $\mathcal{R}$. Let A and B be respectively arguments that belong to $\mathcal{E}'$ and $\mathcal{A} \backslash \mathcal{E}'$ such that B attacks A on A' (wrt. $\mathcal{R}$). Then, the following three cases are possible:

First, if $\langle B, A \rangle \in \mathcal{R}'$, then there exists a $C \in \mathcal{E}'$, such that $\langle C, B \rangle \in \mathcal{R}'$ since $\mathcal{E}'$ defends all of its elements wrt. $\mathcal{R}'$. Because it holds that $\mathcal{R}' \subseteq \mathcal{R}$, there exists a $C \in \mathcal{E}'$, such that $\langle C, B \rangle \in \mathcal{R}$. Thus, $\mathcal{E}'$ defends all of its elements wrt. $\mathcal{R}$.

Second, if $\langle B, A \rangle \notin \mathcal{R}'$ because $B \prec A$, and $\langle B, A \rangle$ is an asymmetric preference-dependent attack, then from Lemma 36 of [46], A' defeats B (when the top rule of B is defeasible) or there exists a strict continuation A^+ of A such that A^+ defeats B (when the top rule of B is strict). Furthermore, because $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle$ satisfies the postulate of closure under subarguments and strict rules, it holds that $B' \in \mathcal{E}'$ and $A^+ \in \mathcal{E}'$. Then since it holds that $\mathcal{R}' \subseteq \mathcal{R}\mathcal{E}'$ defends all of its elements wrt. $\mathcal{R}$ in this case.

Third, if $\langle B, A \rangle \notin \mathcal{R}'$ because $B \prec A$, and $\langle B, A \rangle$ is a symmetric preference-dependent attack, then $\mathcal{E}'$ also defends all of its elements since $\langle A, B \rangle \in \mathcal{R}$.

Now suppose that $\mathcal{E}'$ is not maximal admissible set, that is, there exists a $A \in \mathcal{A} \backslash \mathcal{E}'$ such that $\mathcal{E}' \cup \{A\}$ is admissible wrt. $\mathcal{R}$. Since $\mathcal{E}'$ is a preferred extension wrt. $\mathcal{R}'$, $\mathcal{E}' \cup \{A\}$ is not admissible wrt. $\mathcal{R}'$. $\mathcal{E}' \cup \{A\}$ is also conflict-free wrt. $\mathcal{R}'$ because the conflict-freeness is preserved from Proposition 16 of [46]. Therefore, $\mathcal{E}' \cup \{A\}$ fails to defend its element A (it defends all elements that belongs to $\mathcal{E}'$ because of the preferred nature of $\mathcal{E}'$ wrt. $\mathcal{R}'$) wrt. $\mathcal{R}'$. This implies that there exists a $B \in \mathcal{A} \backslash \mathcal{E}'$ such that $(B, A) \in \mathcal{R}'$, and for all $C \in \mathcal{E}' \cup \{A\}$ it holds that $(C, B) \notin \mathcal{R}'$. From the admissibility of $\mathcal{E}' \cup \{A\}$ wrt. $\mathcal{R}$, if $(B, A) \in \mathcal{R}' \subseteq \mathcal{R}$, then there exists a $C \in \mathcal{E}' \cup \{A\}$ such that C attacks B on B' wrt. $\mathcal{R}$. That is, $\langle C, B \rangle \in \mathcal{R}$, but $\langle C, B \rangle \notin \mathcal{R}'$.

Now, two choices are possible. First, assume that $C \neq A$. This implies that $\langle C, B \rangle$ is a preference-dependent attack. If $\langle C, B \rangle$ is a symmetric attack (Fig. 20), then it holds that $\langle B, C \rangle \in \mathcal{R}'$. Because of admissibility of $\mathcal{E}'$ wrt. $\mathcal{R}'$, there exists a $D \in \mathcal{E}'$ such that $\langle D, B \rangle \in \mathcal{R}'$. Contradiction with the fact that $\mathcal{E}' \cup \{A\}$ fail to defend its element A wrt $\mathcal{R}'$. If $\langle C, B \rangle$ is an asymmetric attack (Fig. 21), then from then from Lemma 36 of [46], B' defeats C (when the top rule of C is defeasible) or there exists a strict continuation B^+ of B such that B^+ defeats C (when the top rule of B is strict). $\mathcal{E}'$ wrt. $\mathcal{R}'$, there exists a $D \in \mathcal{E}'$ such that $\langle D, B' \rangle \in \mathcal{R}'$ or $\langle D, B^+ \rangle \in \mathcal{R}'$. If $\langle D, B' \rangle \in \mathcal{R}'$, then D also defeats B on B'. And if $\langle D, B^+ \rangle \in \mathcal{R}'$, there exists a strict continuation of D^+ of D such that D^+ defeats B (under reasonable

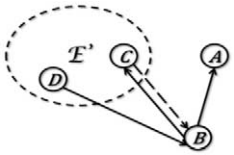

Fig. 20. When $\langle C, B\rangle$ is symmetric.

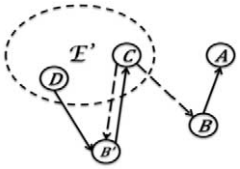

Fig. 21. When $\langle C, B\rangle$ is asymmetric.

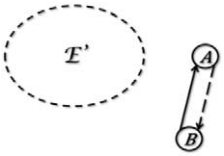

Fig. 22. When $C = A$.

argument ordering). Also, it holds that $D^+ \in \mathcal{E}'$. Both cases lead us to the contradiction with the fact that $\mathcal{E}' \cup \{A\}$ fail to defend its element A wrt. $\mathcal{R}'$. Thus, $\mathcal{E}' \cup \{A\}$ is not admissible wrt $\mathcal{R}$, that is, $\mathcal{E}'$ is also a preferred extension wrt. $\mathcal{R}$.

Next, if $C = A$, then we know that there exists a failed attack from A to B (Fig. 22). Then, if we define A as an argument which is acceptable wrt. $\mathcal{E}' \cup \{A\}$ in $\langle \mathcal{A}, \mathcal{R}\rangle$, then it is self-evident that $\mathcal{E}' \cup \{A\}$ is also admissible. Otherwise, $\mathcal{E}' \cup \{A\}$ is not admissible, therefore, $\mathcal{E}'$ becomes a preferred extension of $\langle \mathcal{A}, \mathcal{R}\rangle$. As a result, a conflict-free set $\mathcal{E} = \mathcal{E}' \cup \{A | A$ is acceptable wrt. $\mathcal{E}' \cup \{A\}$ in $\langle \mathcal{A}, \mathcal{R}\rangle\}$ is a preferred extension of $\langle \mathcal{A}, \mathcal{R}\rangle$. Hence, $\mathcal{E}' \subset \mathcal{E} \in \mathtt{Ext}_{pre}(\langle \mathcal{A}, \mathcal{R}\rangle)$.

In the *above* figures, solid arrows represent attacks that belongs to $\mathcal{R}'$, and the dashed arrow represents the attack that belongs to $\mathcal{R}$, but fails to belong to $\mathcal{R}'$. Note that every attack that belongs to $\mathcal{R}'$ also belongs to $\mathcal{R}$. In addition, if $\emptyset \in \mathtt{Ext}_{pre}(\langle \mathcal{A}, \mathcal{R}, \preccurlyeq\rangle)$, then the proposition self-evidently holds, since the empty set is a subset of any set.

– Under stable semantics

If the *repaired* framework has no stable extension ($\mathtt{Ext}_{sta}(\langle \mathcal{A}, \mathcal{R}, \preccurlyeq\rangle) = \emptyset$), then the proposition holds, since an empty set is a subset of any set.

If $Ext_{sta}(\langle \mathcal{A}, \mathcal{R}, \preccurlyeq\rangle) \neq \emptyset$, we should prove that for all $\mathcal{E}' \in \mathtt{Ext}_{sta}(\langle \mathcal{A}, \mathcal{R}, \preccurlyeq\rangle)$, it also holds that $\mathcal{E}' \in \mathtt{Ext}_{sta}(\langle \mathcal{A}, \mathcal{R}\rangle)$. Since the conflict-freeness is guaranteed by Proposition 16 of [46], $\mathcal{E}'$ is also conflict-free wrt. $\mathcal{R}$. From the definition of stable semantics, $\mathcal{E}'$ attacks all arguments that do not belong to $\mathcal{E}'$, that is, for all $B \in \mathcal{A}\backslash\mathcal{E}'$, there exists a $C \in \mathcal{E}'$ such that $(C, B) \in \mathcal{R}'$. Since $\mathcal{R}' \subseteq \mathcal{R}$, it holds that for all $B \in \mathcal{A}\backslash\mathcal{E}'$, there exists a $C \in \mathcal{E}'$ such that $(C, B) \in \mathcal{R}$. It implies that $\mathcal{E}'$ also attacks all the arguments that is not belong to $\mathcal{E}'$ wrt. $\mathcal{R}$. Furthermore, $\nexists C \in \mathcal{A}\backslash\mathcal{E}'$ such that $\mathcal{E}' \cup \{C\}$ is conflict-free (wrt. $\mathcal{R}$ or $\mathcal{R}'$). Therefore, $\mathcal{E}'$ is the maximal set that is conflict-free and attacks all the arguments that do not belong to $\mathcal{E}'$ wrt $\mathcal{R}$, i.e. $\mathcal{E}'$ is a stable extension wrt. $\mathcal{R}$. □

Proposition 2. *Let $\mathcal{T}$ be an argumentation theory which is indirectly consistent, $\mathcal{C}$ a set of valid preference criteria wrt. $\mathcal{T}$, and $\mathcal{H} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq \rangle$ a PAF built over $\mathcal{T}$. For all c_1, c_2 in $\mathcal{C}$, that are incompatible, it is impossible for both of c_1 and c_2 to be justified by an extension of $\mathcal{H}$ under a given semantics thus be skeptically justified by $\mathcal{H}$.*

Proof. From Definition 15, for all c_1 and c_2 in $\mathcal{C}$ that are incompatible it holds that $\mathcal{K}n \cup \texttt{Guard}(c_1) \cup \texttt{Guard}(c_2)$, is indirectly inconsistent under the strict rules of $\mathcal{T}$ (here, $\mathcal{K}n$ is the set of axioms of $\mathcal{T}$).

Suppose that both of c_1 and c_2 are justified by an extension of $\mathcal{H}$ under a given semantics. This means that $\texttt{Guard}(c_1) \cup \texttt{Guard}(c_2) \subseteq \texttt{Concs}(\mathcal{E})$. Furthermore, if we take into account that strict arguments cannot be attacked in any manner, then it holds that $\mathcal{K}n \subseteq \texttt{Concs}(\mathcal{E})$. Therefore, it holds that $\mathcal{K}n \cup \texttt{Guard}(c_1) \cup \texttt{Guard}(c_2) \subseteq \texttt{Concs}(\mathcal{E})$. However, from the indirect consistency of $\mathcal{T}$, $\texttt{Concs}(\mathcal{E})$ is also indirectly consistent. Contradiction with the fact that $\mathcal{K}n \cup \texttt{Guard}(c_1) \cup \texttt{Guard}(c_2)$, is indirectly inconsistent under the strict rules of $\mathcal{T}$. Therefore, for any extension $\mathcal{E}$ under a given semantics (*admissible*, *ground*, *complete*, *preferred*, *stable*), $\texttt{Guard}(c_1) \cup \texttt{Guard}(c_2) \nsubseteq \texttt{Concs}(\mathcal{E})$. Eventually, it is impossible for both c_1 and c_2 to be justified by any extension of the PAF corresponding to $\mathcal{T}$. This also implies that both c_1 and c_2 cannot be skeptically justified under a given semantics. □

Proposition 3. *Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ with $\mathcal{AS} = \langle \mathcal{L}, \bar{}, \mathcal{R}, \mathcal{C}, n \rangle$ and $\mathcal{KB} = \langle \mathcal{K}, \mathcal{C}' \rangle$ be an ATPC with valid criteria sets and $\mathcal{H}_{pri} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{pri} \rangle$ and $\mathcal{H}_{adv} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{adv} \rangle$ respectively the primary and advanced PAF corresponding to $\mathcal{T}$ which have reasonable argument orderings. Then, under preferred semantics, for all $\mathcal{E}' \in \texttt{Ext}_{pre}(\mathcal{H}_{adv})$, there exists an extension $\mathcal{E} \in \texttt{Ext}_{pre}(\mathcal{H}_{pri})$ such that $\mathcal{E}' \subseteq \mathcal{E}$. Under stable semantics, it holds that $\texttt{Ext}_{sta}(\mathcal{H}_{pri}) \supseteq \texttt{Ext}_{sta}(\mathcal{H}_{adv})$.*

Proof. Let $\mathcal{H}'_{pri} = \langle \mathcal{A}, \mathcal{R}'_{pri} \rangle$ be the repaired AF of $\mathcal{H}_{pri}$ and $\langle \mathcal{A}, \mathcal{R}'_{adv} \rangle$ that of $\mathcal{H}_{adv}$. Here, $\mathcal{R}'_{pri}$ is the result of filtering $\mathcal{R}$ through $\succcurlyeq_{pri}$ and $\mathcal{R}'_{adv}$ is the result of filtering $\mathcal{R}$ through $\succcurlyeq_{adv}$. Then, we also know that $\succcurlyeq_{pri} \subseteq \succcurlyeq_{adv}$, because $\succcurlyeq_{pri}$ comes from preference criteria that can be applied without any justification. Therefore, it can be seen that $\mathcal{H}_{adv} = \langle \mathcal{A}, \mathcal{R}'_{pri}, \succcurlyeq_{adv} \setminus \succcurlyeq_{pri} \rangle$. Then, from Proposition 1, under preferred semantics, it holds that for all $\mathcal{E}' \in \texttt{Ext}_{pre}(\mathcal{H}_{adv})$, there exists $\mathcal{E} \in \texttt{Ext}_{pre}(\mathcal{H}_{adv})$ such that $\mathcal{E}' \subseteq \mathcal{E}$. Under stable semantics, it holds that $\texttt{Ext}_{sta}(\mathcal{H}_{pri}) \supseteq \texttt{Ext}_{sta}(\mathcal{H}_{adv})$. □

Proposition 4. *Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ with $\mathcal{AS} = \langle \mathcal{L}, \bar{}, \mathcal{R}, \mathcal{C}, n \rangle$ and $\mathcal{KB} = \langle \mathcal{K}, \mathcal{C}' \rangle$ be an ATPC with valid criteria sets and $\mathcal{H}_{pri} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{pri} \rangle$ and $\mathcal{H}_{adv} = \langle \mathcal{A}, \mathcal{R}, \succcurlyeq_{adv} \rangle$ respectively the primary and advanced PAF corresponding to $\mathcal{T}$ which have reasonable argument orderings. In addition, let $\mathcal{C}_{adv}$ and $\mathcal{C}_3$ be respectively be sets of preference criteria that produce $\succcurlyeq_{adv}$ and that is justified by $\mathcal{H}_{adv}$. Then there exists no $c \in \mathcal{C}_{adv}$ and $c' \in \mathcal{C}_3$ such that c and c' are incompatible.*

Proof. For all c in $\mathcal{C}_3$, c's guard belongs to the intersection of all extensions of $\mathcal{H}_{adv}$. In the same way, the guard of any criterion of $\mathcal{C}_{adv}$ also belongs to the intersection of all extensions of $\mathcal{H}_{pri}$. Then, under stable semantics, since it holds that $\texttt{Ext}_{sta}(\mathcal{H}_{pri}) \supseteq \texttt{Ext}_{sta}(\mathcal{H}_{adv})$ from Proposition 3, the intersection of extensions of $\mathcal{H}_{adv}$ is a superset of that of $\mathcal{H}_{pri}$. Hence, it holds that $\mathcal{C}_{adv} \subseteq \mathcal{C}_3$. From Proposition 2, there exists no $c \in \mathcal{C}_{adv}$ and $c' \in \mathcal{C}_3$ such that c and c' are incompatible. On the other hand, under preferred semantics, since skeptical standpoint is adopted, for all $\mathcal{E}' \in \texttt{Ext}_{pre}(\mathcal{H}_{adv})$, all criteria in $\mathcal{C}_3$ are justified by $\mathcal{E}'$. From Proposition 3, there exists $\mathcal{E} \in \texttt{Ext}_{pre}(\mathcal{H}_{adv})$ such that $\mathcal{E}' \subseteq \mathcal{E}$. We here note that all criteria in $\mathcal{C}_{adv}$ are justified by $\mathcal{E}$. The proposition holds since every couple of incompatible criteria cannot justified by a single extension (Proposition 2). □

Proposition 5. *Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ be an ATPC with valid criteria sets, $\mathcal{H}_n = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_n \rangle$ and $\mathcal{H}_{n+1} = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_{n+1} \rangle$ respectively the resulting PAFs of the nth and n + 1th filtering which have reasonable argument orderings. Then, if $\preccurlyeq_{n+1} \supseteq \preccurlyeq_n$, then under preferred semantics, for all $\mathcal{E}' \in \mathtt{Ext}_{pre}(\mathcal{H}_{n+1})$, there exists an extension $\mathcal{E} \in \mathtt{Ext}_{pre}(\mathcal{H}_n)$ such that $\mathcal{E}' \subseteq \mathcal{E}$. Under stable semantics, it holds that $\mathtt{Ext}_{sta}(\mathcal{H}_n) \supseteq \mathtt{Ext}_{sta}(\mathcal{H}_{n+1})$.*

Proof. Let $\langle \mathcal{A}, \mathcal{R}'_n \rangle$ be the repaired AF of $\mathcal{H}_n$. Then, we can rewrite $\mathcal{H}_{n+1}$ as $\langle \mathcal{A}, \mathcal{R}'_n, \succcurlyeq_{n+1} \setminus \succcurlyeq_n \rangle$. From Proposition 1, it holds that for all $\mathcal{E}' \in \mathtt{Ext}_{pre}(\mathcal{H}_{n+1})$, there exists $\mathcal{E} \in \mathtt{Ext}_{pre}(\mathcal{H}_n)$ such that $\mathcal{E}' \subseteq \mathcal{E}$. It also holds that $\mathtt{Ext}_{sta}(\mathcal{H}_n) \supseteq \mathtt{Ext}_{sta}(\mathcal{H}_{n+1})$. □

Proposition 6. *Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ be an ATPC with valid criteria sets, $\mathcal{H}_n = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_n \rangle$ and $\mathcal{H}_{n+1} = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_{n+1} \rangle$ respectively the resulting PAFs of the nth and n + 1th filtering which have reasonable argument orderings. In addition, let $\mathcal{C}_n$ and $\mathcal{C}_{n+1}$ be respectively sets of preference criteria that produce $\preccurlyeq_n$ and $\preccurlyeq_{n+1}$. Then there exists no $c \in \mathcal{C}_n$ and $c' \in \mathcal{C}_{n+1}$ such that c and c′ are incompatible.*

Proof. Let $H_{n-1} = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_{n-1} \rangle$ be the resulting PAF of the $n-1$th filtering. If $\preccurlyeq_n \supseteq \preccurlyeq_{n-1}$, then from Proposition 5, for all $\mathcal{E}' \in \mathtt{Ext}_{pre}(\mathcal{H}_n)$, it holds that there exists an extension $\mathcal{E} \in \mathtt{Ext}_{pre}(\mathcal{H}_{n-1})$ such that $\mathcal{E}' \subseteq \mathcal{E}$. It also holds that $\mathtt{Ext}_{sta}(\mathcal{H}_{n-1}) \supseteq \mathtt{Ext}_{sta}(\mathcal{H}_n)$. Since $\mathcal{C}_n$ and $\mathcal{C}_{n+1}$ are respectively justified by $\mathcal{H}_{n-1}$ and $\mathcal{H}_n$, there exists no $c \in \mathcal{C}_n$ and $c' \in \mathcal{C}_{n+1}$ such that c and c' are incompatible (see the proof procedure of Proposition 4). In contrast, if $\preccurlyeq_{n-1} \supseteq \preccurlyeq_n$, we can rewrite $\mathcal{H}_{n-1}$ and $\mathcal{H}_n$ as $\langle \mathcal{A}, \mathcal{R}'_n, \preccurlyeq_{n-1} \setminus \preccurlyeq_n \rangle$ and $\langle \mathcal{A}, \mathcal{R}'_n \rangle$ respectively (here $\langle \mathcal{A}, \mathcal{R}'_n \rangle$ is the repaired version of $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq_n \rangle$. From Proposition 3, for all $\mathcal{E}' \in \mathtt{Ext}_{pre}(\mathcal{H}_{n-1})$, it holds that there exists an extension $\mathcal{E} \in \mathtt{Ext}_{pre}(\mathcal{H}_n)$ such that $\mathcal{E}' \subseteq \mathcal{E}$ under preferred semantics. It also holds that $\mathtt{Ext}_{sta}(\mathcal{H}_n) \supseteq \mathtt{Ext}_{sta}(\mathcal{H}_{n-1})$. Since $\mathcal{C}_{n+1}$ and $\mathcal{C}_n$ are respectively justified by $\mathcal{H}_n$ and $\mathcal{H}_{n-1}$, there exists no $c \in \mathcal{C}_n$ and $c' \in \mathcal{C}_{n+1}$ such that c and c' are incompatible (see the proof procedure of Proposition 4 once again). □

Proposition 7. *Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ with $\mathcal{AS} = \langle \mathcal{L}, \bar{}, \mathcal{R}, \mathcal{C}, n \rangle$ and $\mathcal{KB} = \langle \mathcal{K}, \mathcal{C}' \rangle$ be an ATPC with valid criteria sets, $\mathcal{H}_{n-1}$, $\mathcal{H}_n$ and $\mathcal{H}_{n+1}$ ($n = 2, 3, \ldots$) respectively results of the $n-1$th, nth and $n+1$th filtering with reasonable argument orderings. Then, it also holds that if $\mathtt{Ext}_y(\mathcal{H}_{n-1}) = \mathtt{Ext}_y(\mathcal{H}_n)$, then $\mathtt{Ext}_y(\mathcal{H}_n) = \mathtt{Ext}_y(\mathcal{H}_{n+1})$ under preferred or stable semantics (that is, $y \in \{\mathtt{pre}, \mathtt{sta}\}$).*

Proof. Let $\succcurlyeq_{n-1}$, $\succcurlyeq_n$ and $\succcurlyeq_{n+1}$ be respectively argument orderings of $\mathcal{H}_{n-1}$, $\mathcal{H}_n$ and $\mathcal{H}_{n+1}$ and $\mathcal{C}_{n-1}$, $\mathcal{C}_n$ and $\mathcal{C}_{n+1}$ sets of preference criteria which produce $\succcurlyeq_{n-1}$, $\succcurlyeq_n$ and $\succcurlyeq_{n+1}$, respectively. It means that $\bigcup_{ci \in \mathcal{C}_n} \mathtt{Guard}(c_i) \subseteq \mathtt{SOutput}_y(\mathcal{H}_{n-1})$ and $\bigcup_{ci \in \mathcal{C}n+1} \mathtt{Guard}(c_i) \subseteq \mathtt{SOutput}_y(\mathcal{H}_n)$. Then, since $\mathcal{C}_n$ and $\mathcal{C}_{n+1}$ are derived from a single preference criteria set $\mathcal{C} \cup \mathcal{C}'$ and $\mathtt{SOutput}_y(\mathcal{H}_{n-1}) = \mathtt{SOutput}_y(\mathcal{H}_n)$, $\mathcal{C}_n = \mathcal{C}_{n+1}$. Therefore, it holds that $\succcurlyeq_n = \succcurlyeq_{n+1}$ and $\mathcal{H}_n$ and $\mathcal{H}_{n+1}$ are identical. As a result, $\mathtt{Ext}_y(\mathcal{H}_n) = \mathtt{Ext}_y(\mathcal{H}_{n+1})$. □

Proposition 8. *Let $\mathcal{T} = \langle \mathcal{AS}, \mathcal{KB} \rangle$ be an ATPC with valid criteria sets, $\mathcal{H}_n = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_n \rangle$ and $\mathcal{H}_{n+1} = \langle \mathcal{A}, \mathcal{R}, \preccurlyeq_{n+1} \rangle$ respectively the resulting PAFs of the nth and n + 1th filtering which have reasonable argument orderings. If $\mathcal{E}_n$ and $\mathcal{E}_{n+1}$ are respectively grounded extensions of $\mathcal{H}_n$ and $\mathcal{H}_{n+1}$, it holds that $\mathcal{E}_n \subseteq \mathcal{E}_{n+1}$.*

Proof. In order to prove the proposition, the following lemma should be proved. □

Lemma 1. *Let* $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle$ *be a PAF corresponding to an argumentation theory which has a reasonable argument ordering. Then, under complete semantics, it holds that* $\texttt{Ext}_{com}(\langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle) \subseteq \texttt{Ext}_{com}(\langle \mathcal{A}, \mathcal{R} \rangle)$.

Proof. We should prove that for all $\mathcal{E}' \in \texttt{Ext}_{com}(\langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle)$, it also holds that $\mathcal{E}' \in \texttt{Ext}_{com}(\langle \mathcal{A}, \mathcal{R} \rangle)$. Let $\langle \mathcal{A}, \mathcal{R}' \rangle$ be the repaired AF of $\langle \mathcal{A}, \mathcal{R}, \preccurlyeq \rangle$. As the proof of Proposition 1 shows, $\mathcal{E}'$ is also admissible (conflict-free and defends all of its elements) in $\langle \mathcal{A}, \mathcal{R} \rangle$. Now, let us assume that $\mathcal{E}'$ does not contain an argument A which is acceptable wrt. $\mathcal{E}'$ in $\langle \mathcal{A}, \mathcal{R} \rangle$. From Proposition 16 of [46], $\mathcal{E}' \cup \{A\}$ is also conflict-free in $\langle \mathcal{A}, \mathcal{R}' \rangle$. Since $\mathcal{E}'$ is a complete extension of $\langle \mathcal{A}, \mathcal{R}' \rangle$, A is not acceptable wrt. $\mathcal{E}'$ in $\langle \mathcal{A}, \mathcal{R}' \rangle$. It means that there exists an argument $B \in \mathcal{A} \backslash \mathcal{E}'$ such that B attacks A, but any argument in $\mathcal{E}'$ does not attack B in $\langle \mathcal{A}, \mathcal{R}' \rangle$. Because $\mathcal{R}' \subseteq \mathcal{R}$, B also attacks A in $\langle \mathcal{A}, \mathcal{R} \rangle$ and thus there is an argument $C \in \mathcal{E}'$ such that C attacks B in $\langle \mathcal{A}, \mathcal{R} \rangle$. Therefore, we have $C \prec B$. If $\langle C, B \rangle$ is a symmetric attack, it holds that B attacks C in $\langle \mathcal{A}, \mathcal{R}' \rangle$. Then, we have a $D \in \mathcal{E}'$ such that D attacks B in $\langle \mathcal{A}, \mathcal{R}' \rangle$ (because $\mathcal{E}'$ defends all of its elements). Contradiction with the fact that A is not acceptable wrt. $\mathcal{E}'$ in $\langle \mathcal{A}, \mathcal{R}' \rangle$. Next, if $\langle C, B \rangle$ is an asymmetric attack, then assume that C attacks B on B'. Then from then from Lemma 36 of [46], B' attacks C (when the top rule of C is defeasible) or there exists a strict continuation B^+ of B such that B^+ attacks C (when the top rule of B is strict) in $\langle \mathcal{A}, \mathcal{R}' \rangle$. Therefore, we have a $D \in \mathcal{E}'$ such that D attacks B' or B^+ in $\langle \mathcal{A}, \mathcal{R}' \rangle$. It means that D attacks B (on B') or there exists a strict continuation of D, which attacks B in $\langle \mathcal{A}, \mathcal{R}' \rangle$. Contradiction! Therefore, $\mathcal{E}'$ contains any argument which is acceptable wrt. it in $\langle \mathcal{A}, \mathcal{R} \rangle$.

Let us prove by mathematical induction that $\preccurlyeq_n \subseteq \preccurlyeq_{n+1}$ under grounded semantics. This self-evidently holds for $\preccurlyeq_1$ and $\preccurlyeq_2$ since $\preccurlyeq_1$ comes from preference criteria whose guards are $\emptyset$. Assume that $\preccurlyeq_{k-1} \subseteq \preccurlyeq_k$ under grounded semantics. Then, from Lemma 1 and the fact that the grounded extension of an AF is the least complete extension, we have $\mathcal{E}_{k-1} \subseteq \mathcal{E}_k$ $(k > 1)$. Hence, it holds that $\preccurlyeq_k \subseteq \preccurlyeq_{k+1}$. Now, we have $\preccurlyeq_n \subseteq \preccurlyeq_{n+1}$ under grounded semantics $(1 \leqslant n)$. Now, let $\langle \mathcal{A}, \mathcal{R}'_n \rangle$ be the repaired AF of $\mathcal{H}_n$. Then, we can rewrite $\mathcal{H}_{n+1}$ as $\langle \mathcal{A}, \mathcal{R}'_n, \succcurlyeq_{n+1} \setminus \succcurlyeq_n \rangle$. Then, the proposition directly follows from Lemma 1 and the fact that the grounded extension is the least complete extension. □

References

[1] L. Amgoud and C. Cayrol, A reasoning model based on the production of acceptable arguments, *Ann. Math. Artif. Intell.* **34** (2002), 197–216. doi:10.1023/A:1014490210693.

[2] L. Amgoud and C. Cayrol, Inferring from inconsistency in preference-based argumentation frameworks, *Int. J. Approx. Reason.* **29**(2) (2002), 125–169.

[3] L. Amgoud, C. Cayrol and D. LeBerre, Comparing arguments using preference orderings for argument-based reasoning, in: *Proceedings of the 8th International Conference on Tools with Artificial Intelligence*, 1996, pp. 400–403.

[4] L. Amgoud, N. Maudet and S. Parsons, Modeling dialogues using argumentation, in: *Proceedings of the Fourth International Conference on Multi-Agent Systems*, E. Durfee, ed., AAAI Press, Menlo Park, 2000, pp. 31–38. doi:10.1109/ICMAS.2000.858428.

[5] L. Amgoud and S. Parsons, An argumentation framework for merging conflicting knowledge bases, in: *Lecture Notes in Artificial Intelligence*, S. Flesca, S. Greco, N. Leone and G. Ianni, eds, Vol. 2424, Springer-Verlag, Berlin Heidelberg, 2002, pp. 27–37.

[6] L. Amgoud and H. Prade, Using arguments for making and explaining decisions, *Artif. Intell.* **173** (2009), 413–436. doi:10.1016/j.artint.2008.11.006.

[7] L. Amgoud and S. Vesic, A new approach for preference-based argumentation frameworks, *Ann. Math. Artif. Intell.* **63** (2011), 149–183. doi:10.1007/s10472-011-9271-9.

[8] L. Amgoud and S. Vesic, Two roles of preferences in argumentation frameworks, in: *Proceedings of the 11th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, W. Liu, ed., Springer-Verlag, Berlin Heidelberg, 2011, pp. 86–97. doi:10.1007/978-3-642-22152-1_8.

[9] L. Amgoud and S. Vesic, Rich preference-based argumentation frameworks, *Int. J. Approx. Reason.* **55** (2014), 585–606. doi:10.1016/j.ijar.2013.10.010.
[10] K. Atkinson and T.J.M. Bench-Capon, Addressing moral problems through practical reasoning, *J. Appl. Log.* **6** (2008), 135–151. doi:10.1016/j.jal.2007.06.005.
[11] P. Baroni, F. Toni and B. Verheij, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games: 25 years later, *Argument. Comput.* **11** (2020), 1–14. doi:10.3233/AAC-200901.
[12] R. Baumann and M. Ulbricht, If nothing is accepted-repairing argumentation frameworks, *J. Artif. Intell. Research* **66** (2019), 1099–1145. doi:10.1613/jair.1.11791.
[13] T.J.M. Bench-Capon, Persuasion in practical argument using value-based argumentation frameworks, *J. Log. Comput.* **13**(3) (2003), 429–448. doi:10.1093/logcom/13.3.429.
[14] T.J.M. Bench-Capon, Before and after dung: Argumentation in AI and law, *Argument. Comput.* **11** (2020), 221–228.
[15] T.J.M. Bench-Capon, S. Doutre and P.E. Dunne, Audiences in argumentation frameworks, *Artif. Intell.* **171** (2007), 42–71. doi:10.1016/j.artint.2006.10.013.
[16] T.J.M. Bench-Capon and S. Modgil, Norms and value based reasoning: Justifying compliance and violation, *Artif. Intell. Law.* **25** (2017), 29–64. doi:10.1007/s10506-017-9194-9.
[17] S. Benferhat, D. Dubois and H. Prade, Argumentative inference in uncertain and inconsistent knowledge bases, in: *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, D. Heckerman and A. Mamdani, eds, Springer, Morgan-Kaufmann, Washington, DC, 1993, pp. 411–419. doi:10.1016/B978-1-4832-1451-1.50054-8.
[18] S. Benferhat and K. Sedki, Two alternatives for handling preferences in qualitative choice logic, *Fuzzy Sets Syst.* **159**(15) (2008), 1889–1912. doi:10.1016/j.fss.2008.02.014.
[19] P. Bisquert, C. Cayrol, F.D. de Saint-Cyr and M.-C. Lagasquie-Schiex, Enforcement in argumentation is a kind of update, in: *Proceedings of the Seventh International Conference on Scalable Uncertainty Management (SUM'13)*, 2013, pp. 30–43.
[20] G. Boella, S. Kaci and L. van der Torre, Dynamics in argumentation with single extensions: Abstraction principles and the grounded extension, in: *Proceedings of the Tenth European Conferences on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2009)*, 2009, pp. 107–118. doi:10.1007/978-3-642-02906-6_11.
[21] G. Boella, S. Kaci and L. van der Torre, Dynamics in argumentation with single extensions: Abstraction principles and the grounded extension, in: *Proceedings of the Tenth European Conferences on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2009)*, 2009, pp. 107–118. doi:10.1007/978-3-642-02906-6_11.
[22] G. Boella, S. Kaci and L. van der Torre, Dynamics in argumentation with single extensions: Attack refinement and the grounded extension, in: *Proceedings of the International Conference on Autonomous Agents and Multiagents Systems (AAMAS 2009)*, 2009, pp. 1213–1214.
[23] R. Booth, S. Kaci and T. Rienstra, Property-based preferences in abstract argumentation, in: *Algorithmic Decision Theory: Third International Conference, ADT 2013*, P. Perny, M. Pirlot and A. Tsoukiás, eds, Lecture Notes in Artificial Intelligence, Vol. 8176, Springer-Verlag, Berlin Heidelberg, 2013, pp. 86–100. doi:10.1007/978-3-642-41575-3_7.
[24] R. Brafman and Y. Dimopoulos, Extended semantics and optimization algorithms for CP-networks, *Comput. Intell.* **20** (2004), 219–245.
[25] R. Brafman and C. Domshlak, Graphically structured value-function compilation, *Artif. Intell.* **172** (2008), 325–349. doi:10.1016/j.artint.2007.07.002.
[26] M.E.B. Brarda, L.H. Tamargo and A.J. Garcia, An approach to enhance argument-based multi-criteria decision systems with conditional preferences and explainable answers, *Expert Syst. Appl.* **126** (2019), 171–186. doi:10.1016/j.eswa.2019.02.021.
[27] M. Caminada and L. Amgoud, On the evaluation of argumentation formalisms, *Artif. Intell.* **171** (2007), 286–310. doi:10.1016/j.artint.2007.02.003.
[28] K. Cyras and F. Toni, ABA+: Assumption-based argumentation with preferences, in: *Principles of Knowledge Representation and Reasoning, 15th International Conference, Cape Town*, C. Baral, J.P. Delgrande and F. Wolter, eds, AAAI Press, 2016, pp. 553–556.
[29] C. Domshlak and R. Brafman, CP-nets-reasoning and consistency testing, in: *Proceedings of KR'02*, D. Fensel and F. Giunchiglia, eds, 2002, pp. 121–132.
[30] S. Doutre and J.G. Mailly, Constraints and changes: A survey of abstract argumentation dynamics, *Argument. Comput.* **9** (2018), 223–248. doi:10.3233/AAC-180425.
[31] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games, *Artif. Intell.* **77**(2) (1995), 321–357. doi:10.1016/0004-3702(94)00041-X.
[32] P.M. Dung, P.M. Thang and T.C. Son, On structured argumentation with conditional preferences, in: *The Thirty-Third AAAI Conference on Artificial Intelligence*, AAAI Press, Honolulu, Hawaii, USA, 2019, pp. 2792–2800.
[33] J. Fahnestock and M. Secor, *A Rhetoric of Argument*, 2nd edn, McGraw Hill Publisher, Boston, 1982.
[34] E. Ferretti, M. Errecalde, A.J. Garcia and G.R. Simari, Decision rules and arguments in defeasible decision making, in: *Computational Models of Argument: Proceedings of COMMA*, Toulouse, 2008, pp. 171–182.

[35] E. Ferretti, L.H. Tamargo, A.J. Garcia, M.L. Errecalde and G.R. Simari, An approach to decision making based on dynamic argumentation systems, *Artif. Intell.* **242** (2017), 107–131. doi:10.1016/j.artint.2016.10.004.
[36] G. Flouris and A. Bikakis, A comprehensive study of argumentation frameworks with sets of attacking arguments, *Int. J. Approx. Reason.* **109** (2019), 55–86. doi:10.1016/j.ijar.2019.03.006.
[37] T.F. Gordon, H. Prakken and D.N. Walton, The Carneades model of argument and burden of proof, *Artif. Intell.* **171** (2007), 875–896. doi:10.1016/j.artint.2007.04.010.
[38] G. Governatori, F. Olivieri, M. Cristani and S. Scannapieco, Revision of defeasible preferences, *Int. J. Approx. Reason.* **104** (2019), 205–230. doi:10.1016/j.ijar.2018.10.020.
[39] J. Heyninck and C. Straßer, Rationality and maximal consistent sets for a fragment of ASPIC+ without undercut, *Argument Comput.* **12** (2021), 3–47. doi:10.3233/AAC-200903.
[40] K.V. Hindriks, W. Visser and C.M. Jonker, Multi-attribute preference logic, in: *Lecture Notes in Artificial Intelligence*, N. Desai, A. Liu and M. Winikoff, eds, Vol. 7057, Springer-Verlag, Berlin Heidelberg, 2009, pp. 181–195.
[41] S. Kaci, L. van der Torre and E. Weydert, Acyclic argumentation: Attack = conflict + preference, in: *Frontiers in Artificial Intelligence*, G. Brewka, S. Coradeschi, A. Perini and P. Traverso, eds, Vol. 141, IOS Press, Amsterdam, 2006, pp. 725–726.
[42] S. Modgil, Hierarchical argumentation, in: *Proc. 10th European Conference on Logics in Artificial Intelligence (JELIA 2006)*, 2006, pp. 319–332. doi:10.1007/11853886_27.
[43] S. Modgil, Reasoning about preferences in argumentation frameworks, *Artif. Intell.* **173** (2009), 901–934. doi:10.1016/j.artint.2009.02.001.
[44] S. Modgil and H. Prakken, Reasoning about preferences in structured extended argumentation frameworks, in: *3rd International Conference on Computational Models of Argument*, 2010, pp. 347–358.
[45] S. Modgil and H. Prakken, Revisiting preferences and argumentation, in: *International Joint Conference on Artificial Intelligence*, 2011, pp. 1021–1026.
[46] S. Modgil and H. Prakken, A general account of argumentation with preferences, *Artif. Intell.* **195** (2013), 361–397. doi:10.1016/j.artint.2012.10.008.
[47] S.H. Nielsen and S. Parsons, A generalization of Dung's abstract framework for argumentation: Arguing with sets of attacking arguments, in: *ArgMAS 2006*, N. Maudet, S. Parsons and I. Rahwan, eds, LNAI, Vol. 4766, Springer-Verlag, Berlin Heidelberg, 2006, pp. 54–73.
[48] C.L. Oguego, J.C. Augusto, A. Muñoz and M. Springett, Using argumentation to manage users' preferences, *Future Generation Computer Systems.* **81** (2018), 235–243. doi:10.1016/j.future.2017.09.040.
[49] L. Perrussel, S. Doutr, J. Thevenin and P. McBurney, A persuasion dialog for gaining access to information, argumentation in multi-agent systems, in: *Argumentation in Multi-Agent Systems 2007*, I. Rahwan, S. Parsons and C. Reed, eds, Lecture Notes in Artificial Intelligence, Vol. 4946, Springer-Verlag, Berlin Heidelberg, 2008, pp. 63–79. doi:10.1007/978-3-540-78915-4_5.
[50] K. Pfannschmidt, P. Gupta, B. Haddenhorst and E. Hüllermeier, Learing context-dependent choice functions, *Int. J. Approx. Reason.* **140** (2022), 116–155. doi:10.1016/j.ijar.2021.10.002.
[51] G. Pigozzi, A. Tsoukias and P. Viappiani, Preferences in artificial intelligence, *Ann. Math. Artif. Intell.* **77** (2016), 361–401. doi:10.1007/s10472-015-9475-5.
[52] J.L. Pollock, *Knowledge and Justification*, Princeton University Press, Princeton, 1974.
[53] J.L. Pollock, Defeasible reasoning, *Cog. Sci.* **11** (1987), 481–518. doi:10.1207/s15516709cog1104_4.
[54] J.L. Pollock, Justification and defeat, *Artif. Intell.* **67** (1994), 377–408. doi:10.1016/0004-3702(94)90057-4.
[55] H. Prakken, An abstract framework for argumentation with structured arguments, *Argument Comput.* **1**(2) (2010), 93–124. doi:10.1080/19462160903564592.
[56] H. Prakken and G. Sartor, Argument-based extended logic programming with defeasible priorities, *J. Appl. Non-class. Log.* **7** (1997), 25–75. doi:10.1080/11663081.1997.10510900.
[57] I. Rahwan and L. Amgoud, An argumentation-based approach for practical reasoning, in: *Argumentation in Multi-Agent Systems 2006*, N. Maudet, S. Parsons and I. Rahwan, eds, Lecture Notes in Artificial Intelligence, Vol. 4766, Springer-Verlag, Berlin Heidelberg, 2007, pp. 74–90. doi:10.1007/978-3-540-75526-5_5.
[58] K. Sedki, Value-based argumentation framework built from prioritized qualitative choice logic, *Int. J. Approx. Reason.* **64** (2015), 75–94. doi:10.1016/j.ijar.2015.07.001.
[59] G.R. Simari and R.P. Loui, A mathematical treatment of defeasible reasoning and its implementation, *Artif. Intell.* **53** (1992), 125–157. doi:10.1016/0004-3702(92)90069-A.
[60] J.C.L. Teze, S. Gottifredi, A.J. García and G.R. Simari, Improving argumentation-based recommender systems through context-adaptable selection criteria, *Expert Syst. Appl.* **42**(21) (2015), 8243–8258. doi:10.1016/j.eswa.2015.06.048.
[61] F.H. van Eemeren, B. Garssen, E.C.W. Krabbe, A.F.S. Henkemans, B. Verheij and J.H.M. Wagemans, *Handbook of Argumentation Theory*, Springer Reference, Dordrecht, 2014.
[62] B. Verheij, Dialectical argumentation with argumentation schemes: An approach to legal logic, *Artif. Intell. Law.* **11**(2–3) (2003), 167–195. doi:10.1023/B:ARTI.0000046008.49443.36.

[63] B. Verheij, Formalizing value-guided argumentation for ethical systems design, *Artif. Intell. Law.* **24** (2016), 387–407. doi:10.1007/s10506-016-9189-y.
[64] G.A.W. Vreeswijk, Abstract argumentation systems, *Artif. Intell.* **90** (1997), 225–279. doi:10.1016/S0004-3702(96)00041-0.
[65] D.N. Walton, *Methods of Argumentation*, Cambridge University Press, New York, 2013.
[66] D.N. Walton and D. Godden, The nature of critical questions in argumentation schemes, in: *The Uses of Argument*, D. Hitchcock, ed., Hamilton, ON, Canada, McMaster University, 2005, pp. 476–484.
[67] D.N. Walton, C. Reed and F. Macagno, *Argumentation Schemes*, Cambridge University Press, Cambridge, 2008.
[68] Q. Zhong, X. Fan, X. Luo and F. Toni, An explainable multi-attribute decision model based on argumentation, *Expert Syst. Appl.* **117** (2018), 42–61. doi:10.1016/j.eswa.2018.09.038.