Determining Preferences over Extensions: A Cautious Approach to Preference-Based Argumentation Frameworks

Saul Gebhardt(✉) and Dragan Doder

Utrecht University, Utrecht, The Netherlands
saul.gebhardt@tiscali.nl, d.doder@uu.nl

Abstract. Preferences in abstract argumentation frameworks allow to represent the comparative strength of arguments, or preferences between values that arguments promote. In this paper, we reconsider the approach by Amgoud and Vesic, which computes the extensions of a preference-based argumentation framework by aggregating preferences and attacks into a new attack relation in a way that it favors preferred arguments in conflicts, and then simply applying Dung's semantics to the resulting graph. We argue that this approach is too rigid in some situations, as it discards other sensible (even if less preferred) alternatives. We propose a more cautious approach to preference-based argumentation, which favors preferred arguments in attacks, but also does not discard feasible alternatives. Our semantics returns a set of extensions and a preference relation between them. It generalizes the approach by Amgoud and Vesic, in the sense that the extensions identified by their semantics will be more preferred than other extensions.

Keywords: Abstract Argumentation · Preferences · Dung's Semantics

1 Introduction

In the last couple of decades, argumentation has emerged as an increasingly important field of artificial intelligence research [7,9,16]. It was used as a formalism for solving problems in various fields, like non-monotonic reasoning [14], decision making [1,18], paraconsistent logics [12,17], and in the domains of law and medicine [7]. The simplest, and in the same time the most popular formal models are so called Dung's (abstract) argumentation frameworks [11]. They are just directed graphs where vertices represent the arguments and the edges represent conflict between the arguments. Dung [11] proposed several *semantics* for evaluating the arguments, whose goal is to identify jointly acceptable sets of arguments (called *extensions*).

For some applications, Dung's argumentation frameworks appear too simple for proper modelling all aspects of an argumentation problem. One such shortcoming is the lack of ability to represent comparative strength of arguments, an

Z. Bouraoui and S. Vesic (Eds.): ECSQARU 2023, LNAI 14294, pp. 109–120, 2024.
https://doi.org/10.1007/978-3-031-45608-4_9

aspect which typically occurs if an argument relies on certain information while another argument relies on less certain ones [4], or when different arguments promote values of different importance [8]. This calls for augmenting argumentation frameworks with *preferences* over arguments [2,4,5,8,10,13,15]. Whenever argumentation frameworks are extended with preferences, the central question still remains how arguments are evaluated.

In early papers on preference-based argumentation [2,8], an attack is ignored if the target argument is preferred to its attacker. The extensions are then identified by applying Dung's semantics to the reduced argumentation framework with remaining attacks. This approach has been criticised in [4,6], as the resulting extensions are not necessarily *conflict-free*. Consider the following example, essentially taken from [6].

Consider an individual who wants to buy a violin. An expert says that the violin is made by Stradivari, which is why it's an expensive violin (we represent this with argument a*). Suppose that the individual has brought their child along to the store. This child then states that the violin was not made by Stradivari (argument* b*). It is clear that* b *attac* a*. On the other hand,* a *is preferred to* b*, since the expert should be a lot better at determining whether a violin is a proper Stradivarius or not than a child.*

Since b is preferred to a, the above mentioned method [2,8] will ignore the attack from b to a, so every Dung's semantics will accept both arguments, while there is clearly a conflict between them. To overcome this issue, Amgoud and Vesic [5] proposed a technical solution: to invert the direction of an attack in the case that its target is more preferred than the attacker. This approach preserves conflicts between pairs of arguments, thus ensuring conflict-freeness of extensions. Moreover, in any conflict it favors preferred arguments. In the violin example, any Dung's semantics will accept a and discard b, which is sensible given the disbalance between expertise levels.

Kaci et al. [13] argued that the proposal of Amgoud and Vesic [5] contains an implicit strong constraint that an argument never never able to attack a preferred argument. While we in general agree with the idea of Amgoud and Vesic that the preferred arguments should be favored when involved in an attack, regardless of its direction, we also agree with Kaci et al. that in some situation original direction of the attack should also be considered. To illustrate our position, let us slightly modify the above violin example by replacing the child with another expert, just slightly less reputed. In this situation, the argumentation graph doesn't change, but intuitively acceptance of $\{b\}$ becomes a sensible alternative - even if less preferred than acceptance of $\{a\}$. While ideally we would like distinguish between two scenarios by saying "how strongly" is one argument preferred to another one, that is not possible due to purely qualitative nature of preference relations. This calls for more cautions approach, which does not automatically discard possibly sensible alternatives.

In this paper, we propose a more cautious approach to preference-based argumentation, which favors preferred arguments in attacks, but also does not discard feasible alternatives. In the violin example, it will return two extensions, $\{a\}$ and $\{b\}$, with the first one being more preferred to the second one.

In general, it returns a set of extensions and a preference relation between them. The extensions that fully favor preferred arguments in attacks (i.e., the extensions identified by Amgoud and Vesic [5]) will be more preferred than any other, and the remaining extensions which correspond to feasible, but less likely alternative scenarios will also be ordered according to their to feasibility.

Technically, we propose a two-step procedure for generating possible extensions and preferences over them. In the *first* step, we extract multiple argumentation graphs from the same preference-based argumentation framework, where each of them corresponds to a feasible scenario, and we define an order over them induced by the given preference order over arguments. In the example from above, we will extract two graphs: one in which b attacks a, and one in which the attack is inverted. The latter graph will be more preferred than the former. We noted that in more complex graphs there is more than one sensible way to define a preference order over the extracted graphs, so we first proposed some guiding principles that each such order should satisfy. We then provided two concrete orders that satisfy the principles. In the *second* step, we define a preference relation over extensions of extracted graphs, using previously defined order over the graphs as a starting point. We first deal with the case when the preference order over graphs is total, in which case we employ a variant of lexicographic order. Then we show that we can properly generalize that idea to the case when the order over extracted graphs is partial.

2 Preference-Based Argumentation Frameworks

Dung [11] defined an abstract argumentation frameworks (AFs) as a pair consisting of a set of arguments and attacks, a binary relation between arguments.

Definition 1 (Dung's Argumentation Framework). *A Dung's Argumentation Framework (AF) is a tuple:* $G = \langle A, def \rangle$*, where* A *represents a set of arguments and* $def \subseteq A \times A$ *is a set of attacks between arguments.*

Note that we denote the attack relation by def (from defeat), in order to distinguish this relation from the "attack" relation in preference-based argumentation.

Definition 2. *Let* $G = \langle A, def \rangle$ *be an AF.*

- *A set of arguments* $S \subset A$ *is said to be conflict-free if and only if there are no* $a, b \in S$ *such that* $(a, b) \in def$*.*
- *A set of arguments* $S \subset A$ *is said to attack an argument* b *if and only if there exists some* $a \in S$ *such that* $(a, b) \in def$*.*
- *A set of arguments* $S \subset A$ *defends an argument* a *if and only if* $\forall b \in A$ *if* $(b, a) \in def$*, then* S *attacks* b*.*

Acceptable sets of arguments, called *extensions* are defined by *acceptability semantics* proposed by Dung in [11].

Definition 3. *Let* $G = \langle A, def \rangle$ *be an AF and* $B \subseteq A$ *a set of arguments.*

- *An argument $a \in A$ is acceptable with respect to B if and only if for $\forall a' \in A$: if $(a', a) \in def$, then B attacks a'.*
- *B is admissible if and only if it is conflict-free and each element in B is acceptable with respect to B.*
- *B is a preferred extension if and only if it is a maximal (with respect to $\subseteq$) admissible set*
- *B is a stable extension if and only if B is conflict-free and $\forall a \in A \backslash B$, $\exists b \in B$ such that $(b, a) \in def$.*
- *B is a complete extension if and only if B is admissible and $\forall a \in A$, if B defends a, then $a \in B$.*
- *B is the grounded extension if and only if B is the minimal (with respect to $\subseteq$) complete extension.*

For $\sigma \in \{preferred, stable, complete, grounded\}$, we write $E \in \sigma(G)$ to denote that E is a σ extension of G.

In [2], Dung's AFs were extended with preferences over arguments.

Definition 4 (Preference Argumentation Framework). *A Preference Argumentation Framework (PAF) is a tuple $\mathcal{F} = \langle A, att, \geq \rangle$, where A is a set of arguments, $att \subseteq A \times A$ is an attack relation between arguments and $\geq$ is the preference relation.*

Following [2], in this paper we use term *preference relation* for a (partial or total) *preorder* (i.e., a reflexive and transitive binary relation). We use $a > b$ as an abbreviation for $a \geq b \wedge \neg(b \geq a)$.

In [5] attacks are referred to as *critical* attacks if it is an attack from a less preferred to a more preferred argument.

Definition 5 (Critical Attack). *The set of all critical attacks in any PAF $\mathcal{F} = \langle A, att, \geq \rangle$ is defined as follows:*

$$Critical(\mathcal{F}) = \{(a, b) \mid \forall a, b \in A : (a, b) \in att \wedge b > a\}.$$

Amgoud and Vesic [5] proposed a method which takes a PAF and uses the attacks and preferences over arguments to identify which arguments *defeat* which other arguments. The method entails inverting the direction of critical attacks.

Definition 6. *Let $\mathcal{F} = \langle A, att, \geq \rangle$ be a PAF. Then, $G = \langle A, def \rangle$ is the reduced AF of $\mathcal{F}$ if $def = \{(a, b) \mid (a, b) \in att \wedge b \not> a\} \cup \{(b, a) \mid (a, b) \in att \wedge b > a\}$.*

After applying this method to a PAF, Dung's acceptability semantics are applied to the reduced AF to identify the extensions of the PAF.

3 Extracting Multiple AFs from a Single PAF

In this section, we propose a method for extracting multiple AFs from a single PAF, and determining preferences between these AFs. In the case of violin and example, two different AFs will be created: one in which the direction of the critical attack has been inverted, and the other in which its direction is maintained. The former graph will be preferred to the latter one. Our method for extracting AFs from a PAF based on the reduction method proposed in [5], but differs from it in that not necessarily the direction of all critical attacks will be inverted.

Definition 7 (Extracting AFs from a PAF). *Let $\mathcal{F} = \langle A, att, \geq\rangle$ be a PAF. If $R \subseteq Critical(\mathcal{F})$ is a set of critical attacks, then $G = \langle A, def\rangle$ is an AF* extracted from $\mathcal{F}$, *where*

$$def = \{(a, b) \mid (b, a) \in R\} \cup (att \backslash R).$$

Let I_G denote the set of attacks in AF G which are obtained by inverting the direction of attacks in PAF $\mathcal{F}$, i.e., $I_G = \{(a, b) \mid (b, a) \in R\}$. Let $S(\mathcal{F})$ denote the set of all of the AFs that can be extracted from PAF $\mathcal{F}$.

Next we develop a method to define preferences over $S(\mathcal{F})$. This method, called an AF preference method, is a function which maps each PAF $\mathcal{F}$ to a preference order over AFs extracted from $\mathcal{F}$. While multiple of these methods can be defined, some basic conditions state that any AF preference method should be a transitive and reflexive order.

Definition 8. *An AF preference method (AFpm) is a function m that maps each PAF $\mathcal{F}$ into $\succeq^{\mathcal{F}}_m$, where $\succeq^{\mathcal{F}}_m \subseteq S(\mathcal{F}) \times S(\mathcal{F})$, and $\succeq^{\mathcal{F}}_m$ is a preference relation.*

We use the following abbreviations: $G_1 \succ^{\mathcal{F}}_m G_2$ is the conjunction $G_1 \succeq^{\mathcal{F}}_m G_2$ and $G_2 \not\succeq^{\mathcal{F}}_m G_1$, while $G_1 \approx^{\mathcal{F}}_m G_2$ denotes $G_1 \succeq^{\mathcal{F}}_m G_2$ and $G_2 \succeq^{\mathcal{F}}_m G_1$. If $\mathcal{F}$ is clear from context, we will write $G_1 \succeq_m G_2$ instead of $G_1 \succeq^{\mathcal{F}}_m G_2$. We will also omit m whenever it is clear from context or it is irrelevant.

The above definition is not very restrictive, as the only constraint is that any AF preference method has to be a reflexive and transitive relation. Following the general approach of Amgoud and Vesic [4] which favors preferred arguments in conflicts, we use principle based approach to describe the subclass of AF preference methods which enforces that it is preferable to invert the direction of critical attacks rather than maintain their direction.

Our first principle formalizes that idea in the case of simplest setting that roughly correspond to the violin example: there are only two arguments, which means that they can only contain one critical attack. However, we show that our two principles already ensure that the idea will be respected by all extracted graphs (see Theorem 1).

Principle 1 (Inversion Preference). *Let $\mathcal{F} = \langle\{a, b\}, att, \geq\rangle$ be a PAF such that $(a, b) \in Critical(\mathcal{F})$. Let $G_1, G_2 \in S(\mathcal{F})$ such that $I_{G_1} = \{(b, a)\}$ and $I_{G_2} = \emptyset$. Then $G_1 \succ^{\mathcal{F}}_m G_2$.*

Our second principle ensures that an AF preference method orders AFs in a consistent way in different PAFs,m i.e., that its strategy does not change when we switch from one framework to another one. Consider two AFs G_1 and G_2

that are extracted from the same PAF $\mathcal{F}$ such that $G_1 \succeq_m^{\mathcal{F}} G_2$. By adding an argument a to both G_1 and G_2 and to add the same attacks from and to a, two new AFs are created, G_1' and G_2'. The second principle then enforces that $G_1' \succeq_m G_2'$. However, since G_1' and G_2' no longer contain the same arguments as $\mathcal{F}$, they can no longer be extracted from $\mathcal{F}$. Instead, a new PAF, $\mathcal{F}'$ needs to be defined which contains all the arguments of $\mathcal{F}$ and a. To go towards a formal definition of the second principle, a notion of reducing a PAF or AF with respect to a set of arguments is required.

Definition 9 (Reduction with respect to a set of arguments). *Let $G = \langle A, def\rangle$ be an AF. The reduced AF of G with respect to arguments $A' \subset A$ is $G|_{A'} = \langle A', def'\rangle$, where $def' = def \cap (A' \times A')$.*

Let $\mathcal{F} = \langle A, att, \geq\rangle$ be a PAF. The reduced PAF of $\mathcal{F}$ with respect to arguments $A' \subset A$ is $\mathcal{F}|_{A'} = \langle A', att', \geq'\rangle$, where $att' = att \cap (A' \times A')$ and $\geq' = \geq \cap (A' \times A')$.

With the notation of a reduction with respect to a set of arguments present, it is possible to define the second principle formally.

Principle 2 (Expansion). *Let PAF $\mathcal{F} = \langle A, att, \geq\rangle$. For any $A' \subset A$, let $\mathcal{Q}$ be a PAF such that $\mathcal{Q} = \mathcal{F}|_{A'}$. Let $Q_1, Q_2 \in S(\mathcal{Q})$ and let $G_1, G_2 \in S(\mathcal{F})$ such that $Q_1 = G_1|_{A'}$, $Q_2 = G_2|_{A'}$ and $I_{G_1} \backslash I_{Q_1} = I_{G_2} \backslash I_{Q_2}$. If $Q_1 \succeq_m^{\mathcal{Q}} Q_2$, then $G_1 \succeq_m^{\mathcal{F}} G_2$.*

Now that these two principles have been defined, we are ready to propose a class of AF preference methods that all capture the intuition properly that it is preferable to invert critical attacks rather than maintaining the critical attacks in extracted AFs.

Definition 10 (Inversion-based AFpm). *An Inversion-based AFpm is an AFpm that respects Principle 1 and Principle 2.*

We now show that for any two AFs that are extracted from the same PAF, if they are ordered using an inversion-based AFpm, it is preferred to invert the direction of any critical attack rather than to maintain its direction.

Theorem 1. *Let m be an inversion-based AFpm. Let $\mathcal{F}$ be a PAF, where $(a, b) \in Critical(\mathcal{F})$. Let $G_1, G_2 \in S(\mathcal{F})$ such that $(b, a) \notin I_{G_2}$ and $I_{G_1} = I_{G_2} \cup \{(b, a)\}$. Then, $G_1 \succ_m^{\mathcal{F}} G_2$.*

The previous result provides the most and the last preferred extracted graph of a PAF, regardless of the choice of the inversion-based AFpm.

Corollary 1. *Let m be an inversion-based AFpm and let $\mathcal{F} = \langle A, att, \geq\rangle$ be a PAF.*

1. *If G_1 is the reduced graph of $\mathcal{F}$ (according to Definition 6) and $G \in S(\mathcal{F})$ such that $G_1 \neq G$, then, $G_1 \succ_m^{\mathcal{F}} G$.*
2. *For every $G \in S(\mathcal{F})$ if $G \neq \langle A, att\rangle$ then $G \succ_m^{\mathcal{F}} \langle A, att\rangle$.*

Definition 10 does not determine a unique inversion-based AFpms. Now we propose two different methods, both of them guided by the idea that we prefer graphs in which we invert "more", but with two different interpretations of "more": first using the subset relation, and then using cardinality.

Definition 11. *Let $\mathcal{F}$ be a PAF. Then s maps $\mathcal{F}$ to a preference relation $\succeq_s^{\mathcal{F}}$ such that for any AFs $G_1, G_2 \in S(\mathcal{F})$ $G_1 \succeq_s^{\mathcal{F}} G_2$ if and only if $I_{G_1} \supseteq I_{G_2}$.*

This method obviously gives birth to preference relations that are partial preorders. On the other hand, the following method defines a total preorder over $S(\mathcal{F})$ for every PAF $\mathcal{F}$.

Definition 12. *Let $\mathcal{F}$ be a PAF. Then c maps $\mathcal{F}$ to a preference relation $\succeq_c^{\mathcal{F}}$ such that for any AFs $G_1, G_2 \in S(\mathcal{F})$, $G_1 \succeq_c^{\mathcal{F}} G_2$ if and only if $|I_{G_1}| \geq |I_{G_2}|$.*

Both mappings defined above satisfy Definition 10.

Theorem 2. *Both s and c are inversion-based AF preference methods.*

Note that the strategy of c is to "weight" every critical attack equally: only the number of inversions matters. We might also search for other methods which violates that assumption; for example one might prefer to invert the attack from the weakest argument in a framework to the strongest one, than to invert some other critical attack.

On the other hand, the strategy of s is more cautious. It is easy to see that the total order defined by c extends the partial order defined by s. In fact, we can show that s is the most cautious inversion-based AFpm, in the sense that any other inversion-based AFpm is a further refinement of s.

Theorem 3. *For any inversion-based AFpm m and any PAF $\mathcal{F}$, if $G_1 \succeq_s^{\mathcal{F}} G_2$, then $G_1 \succeq_m^{\mathcal{F}} G_2$.*

4 Preferences over Extensions of a PAF

In the previous section a single PAF $\mathcal{F}$ was used to generate multiple different AFs and to generate preferences between these different AFs based on an inversion-based AFpm. In this section, preferences over extensions of elements of $S(\mathcal{F})$ will be defined. In the following two subsections, different definitions will be given to determine the preferences over extensions depending on whether the preference order over $S(\mathcal{F})$ is a total order or not. Throughout this section, we write $E_1 \succcurlyeq^{\mathcal{F}} E_2$ to denote that E_1 is an extension that is at least as preferred as E_2 (according to some considered semantics σ). If $\mathcal{F}$ is clear from context, it may be omitted for convenience, which means $E_1 \succcurlyeq E_2$ will be used.

4.1 Preferences over Extensions When $\succeq$ over $S(\mathcal{F})$ It Total

In this subsection we define preferences over extensions when there exists a total preference order over $S(\mathcal{F})$, using a variant of lexicographical order. To illustrate the idea, imagine that for some PAF $\mathcal{F}$, $G_1, \dots, G_4 \in S(\mathcal{F})$ are all extracted graphs of $\mathcal{F}$ that contain some of $E_1, \dots, E_4$ as an extension of given semantics σ. Let the total preference order over $S(\mathcal{F})$ rank $G_1, \dots, G_4$ as represented by Fig. 1. Below each of the AFs, extensions of that specific AF according to σ are written.

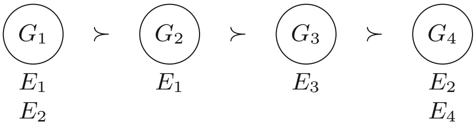

Fig. 1. Representation of total order

For instance, since $E_1 \in \sigma(G_1)$ and $E_2 \in \sigma(G_1)$, we need to look at the extensions of the next most preferred AF, G_2. Since $E_1 \in \sigma(G_2)$ but $E_2 \notin \sigma(G_2)$, it must be the case that E_1 is preferred to E_2. In Table 1, the extensions are ordered according to lexicographical order of the sequences of numbers, where 1 means that E_i is an extension of G_j, and 0 that it is not.

Table 1. Lexicographical order of Fig. 1.

	G_1	G_2	G_3	G_4
E_1	1	1	0	0
E_2	1	0	0	1
E_3	0	0	1	0
E_4	0	0	0	1

Since any extension may only appear once in an AF, all of the values in Table 1 are either 1 or 0. This is the case because there are only strict preferences between AFs in Fig. 1. However, it is possible for AFs extracted from the same PAF to be equally preferred. In that case we will "group" some AFs together.

Definition 13 (Class of all equally preferred AFs). *For any PAF $\mathcal{F}$ and AF $G \in S(\mathcal{F})$, let $[G]^{\mathcal{F}}_m$ represent the class of all AFs in $S(\mathcal{F})$ that are equally preferred to G according to AFpm m, i.e., $[G]^{\mathcal{F}}_m = \{G' \in S(\mathcal{F}) \mid G \approx^{\mathcal{F}}_m G'\}$.*

Informally, in such case we would replace the individual graphs in the first row of Table 1 with the classes defined above, and ordered according to the assumed total order. Then the remaining rows would contain positive integers that indicate how many AFs from a class contains E_i as an extension. Lexicographic order of those rows will define preference order over E_i's.

We now introduce a notion that will also be useful in the rest of this section.

Definition 14. *Let σ be a Dung's semantics. For two sets of arguments E and E' and for any set of AFs S, we say that E is preferred to E' wrt. S, and we write $Pr_\sigma(E, E', S)$, if and only if $|\{G \in S \mid E \in \sigma(G)\}| > |\{G' \in S \mid E' \in \sigma(G')\}|$.*

In other words, if the amount of times E is an extension of elements of S is higher than the amount of times E' is an extension of the elements of S under acceptability semantics σ, then $Pr^\sigma(E, E', S)$.

As has been shown in this section, whenever a preference order over AFs is a total order, we can use those AFs to create a table counting the amount of times an extension is an extension of AFs that are equally preferred to each other. The extension that is counted the most often in the most preferred AF then becomes the most preferred extension. In the case of a tie, the second most preferred class of AFs is counted, until a difference has been observed. If there are no differences in any of the classes of equally preferred AFs between 2 extensions E_1 and E_2, they are equally preferred to each other. ?In terms of a formal definition, Definition 14 can be used to help express this.

Definition 15. *Extension E is preferred to extension E', $E \succapprox^{\mathcal{F}} E'$ if and only if $\exists G : Pr^\sigma\{E, E', [G]^{\mathcal{F}}_m\}$, $\forall G'$ if $G' \succ G$, then it is cannot be the case that $Pr^\sigma\{E', E, [G']^{\mathcal{F}}_m\}$.*

In other words, if $E \succapprox^{\mathcal{F}} E'$, there needs to exist a group of equally preferred AFs, $[G]$, where E is an extension more often than E' ($Pr^\sigma\{E, E', [G]\}$). Note that the preference relation over extensions is a transitive relation. Moreover, for any AF G' which is preferred to G, in that class of equally preferred AFs, $[G']$, E' is not allowed to be an extension more often than E.

4.2 Preferences over Extensions When $\succeq$ over $S(\mathcal{F})$ Is Partial

The previous method is only applicable if the underlying AFpm always provides a total order. In particular, it can't be applied to s (Definition 11).

Let $\mathcal{F}$ be any random PAF such that $Critical(\mathcal{F}) = \{(a, b), (c, d)\}$. Let $I_{G_1} = \{(b, a), (d, c)\}$, $I_{G_2} = \{(b, a)\}$, $I_{G_3} = \{(d, c)\}$ and $I_{G_4} = \emptyset$. We use inversion-based AFpm s to determine the preferences between these four different AFs. It is clear that G_2 and G_3 are incomparable. The AFs are represented in Fig. 2 with preferences between them.

Since G_2 and G_3 are incomparable AFs, the preference order 'branches'.

Similarly to preferences over extensions when $S(\mathcal{F})$ is totally ordered, we would like to prefer extension E_1 over extension E_2 if we cannot find a reason why E_2 is preferred to E_1. In other words, if for any AF G_2 such that $E_2 \in \sigma(G_2)$, there exists an AF G_1 such that $G_1 \succ G_2$ and $E_1 \in \sigma(G_1)$, then E_1 would be preferred to E_2. Compared to the case where $S(\mathcal{F})$ is totally ordered by some

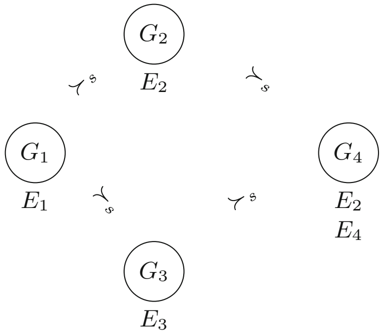

Fig. 2. Representation of partial order

$\succeq^{\mathcal{F}}$, AFs being incomparable makes the method a bit more complex, as different 'branches' can exist (such as in Fig. 2). To be able to express that an extension E is preferred to another extension E', it needs to be checked that whenever E' is an extension of an AF G', that E is an extension of an AF G such that $G \succ G'$.

Since that it could be possible that some AFs extracted from the PAF $\mathcal{F}$ are equally preferred under some AF preference method, we employ Definition 14.

Definition 16. *For any PAF $\mathcal{F}$ and all AFs extracted from $\mathcal{F}$, $S(\mathcal{F})$, E is preferred to E', denoted by $E \succcurlyeq^{\mathcal{F}} E'$, if and only if $\forall G' : Pr^{\sigma}(E', E, [G']^{\mathcal{F}}_{m}) \rightarrow \exists G : (Pr^{\sigma}(E, E', [G]^{\mathcal{F}}_{m}) \wedge G \succ G')$.*

Theorem 4 (Transitivity). *If $E_1 \succcurlyeq E_2$ and $E_2 \succcurlyeq E_3$ then $E_1 \succcurlyeq E_3$.*

At the end of the section, we prove that the method proposed for partial order properly generalize the method for the total order proposed in the previous subsection. In other words, if $S(\mathcal{F})$ is totally ordered, both methods will give the same preferences over extensions.

Theorem 5. *For any PAF $\mathcal{F}$ such that $S(\mathcal{F})$ is totally ordered, the order of extensions found by using Definition 16 is exactly the same as found by using Definition 15.*

Finally, from Corollary 1 it follows that the extensions identified by Amgoud and Vesic [4] will be more preferred than other extensions.

5 Conclusion

This paper proposes a cautious approach to preference-based argumentation, which favors preferred arguments in attacks, but also does not discard feasible alternatives. Our semantics returns a set of extensions and a preference relation between them. We generalize the proposal by Amgoud and Vesic [4], which avoid the problem of conflicting extensions present in early approaches to preference-based argumentation [3,8,15]. There are two more reduction approaches in the literature [13]. Similarly as [4], those approaches reduce a PAF to an AF and return the extensions of that AF, therefore they discard all other possible AFs and they do not define preferences over extensions.

References

1. Amgoud, L.: Argumentation for decision making. In: Simari, G., Rahwan, I. (eds.) Argumentation in Artificial Intelligence, pp. 301–320. Springer, Boston (2009). https://doi.org/10.1007/978-0-387-98197-0_15
2. Amgoud, L., Cayrol, C.: A reasoning model based on the production of acceptable arguments. Anna. Math. Artif. Intell. **34**, 197–215 (2002)
3. Amgoud, L., Cayrol, C.: Inferring from inconsistency in preference-based argumentation frameworks. J. Autom. Reason. **29**(2), 125–169 (2002)
4. Amgoud, L., Vesic, S.: A new approach for preference-based argumentation frameworks. Ann. Math. Artif. Intell. **63**(2), 149–183 (2011)
5. Amgoud, L., Vesic, S.: On the role of preferences in argumentation frameworks. In: 2010 22nd IEEE International Conference on Tools with Artificial Intelligence. vol. 1, pp. 219–222 (2010)
6. Amgoud, L.B., Vesic, S.: Repairing preference-based argumentation frameworks. In: Twenty-First International Joint Conference on Artificial Intelligence. Citeseer (2009)
7. Atkinson, K., et al.: Towards artificial argumentation. AI Mag. **38**(3), 25–36 (2017). https://doi.org/10.1609/aimag.v38i3.2704
8. Bench-Capon, T.J.: Persuasion in practical argument using value-based argumentation frameworks. J. Logic Comput. **13**(3), 429–448 (2003)
9. Bench-Capon, T.J., Dunne, P.E.: Argumentation in artificial intelligence. Artif. Intell. **171**(10–15), 619–641 (2007)
10. Bernreiter, M., Dvorák, W., Woltran, S.: Abstract argumentation with conditional preferences. In: Toni, F., et al Computational Models of Argument - Proceedings of COMMA 2022, Cardiff, Wales, UK, 14–16 September 2022. Frontiers in Artificial Intelligence and Applications. IOS Press, vol. 353. pp. 92–103 (2022). https://doi.org/10.3233/FAIA220144
11. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif. Intell. **77**(2), 321–357 (1995)
12. Grooters, D., Prakken, H.: Combining paraconsistent logic with argumentation. In: COMMA, pp. 301–312 (2014)
13. Kaci, S., van der Torre, L., Villata, S.: Preference in abstract argumentation. In: 7th International Conference on Computational Models of Argument (COMMA). vol. 305, pp. 405–412. IOS Press (2018)
14. Lin, F., Shoham, Y.: Argument systems: a uniform basis for nonmonotonic reasoning. In: KR, vol. 89, pp. 245–255 (1989)
15. Modgil, S.: Reasoning about preferences in argumentation frameworks. Artif. Intell. **173**(9–10), 901–934 (2009)
16. Rahwan, I., Simari, G.R.: Argumentation in Artificial Intelligence, vol. 47. Springer, New York (2009). https://doi.org/10.1007/978-0-387-98197-0

17. Simari, G.R., Loui, R.P.: A mathematical treatment of defeasible reasoning and its implementation. Artif. Intell. **53**(2–3), 125–157 (1992)
18. Zhong, Q., et al.: An explainable multi-attribute decision model based on argumentation. Expert Syst. Appl. **117**, 42–61 (2019)