The Effect of Preferences in Abstract Argumentation Under a Claim-Centric View

Michael Bernreiter*, Wolfgang Dvořák, Anna Rapberger and Stefan Woltran

Institute of Logic and Computation, TU Wien, Austria

Abstract

In this paper, we study the effect of preferences in abstract argumentation under a claim-centric perspective. Recent work has revealed that semantical and computational properties can change when reasoning is performed on claim-level rather than on the argument-level, while under certain natural restrictions (arguments with the same claims have the same outgoing attacks) these properties are conserved. We now investigate these effects when, in addition, preferences have to be taken into account and consider four prominent reductions to handle preferences between arguments. As we shall see, these reductions give rise to different classes of claim-augmented argumentation frameworks, and behave differently in terms of semantic properties and computational complexity. This strengthens the view that the actual choice for handling preferences has to be taken with care.

1. Introduction

Arguments vary in their plausibility. Research in formal argumentation has taken up this aspect in both quantitative and qualitative terms [1, 2]. Indeed, preferences are nowadays a standard feature of many structured argumentation formalisms [3, 4]. At the same time, there are numerous generalizations of abstract Argumentation Frameworks (AFs) [5] that consider the impact of preferences on the abstract level, be it in terms of argument strength [6, 7] or preferences between values [8]. In Dung AFs in which conflicts are expressed as a binary relation between arguments (*attack relation*), the incorporation of preferences typically results in the deletion or reversion of attacks between arguments of different strength—deciding acceptability of arguments via argumentation semantics is thus reflected in terms of the modified attack relation [9].

The difference in argument strength and the resulting modification of the attack relation naturally influences the acceptability of the arguments' conclusion (the *claim* of the argument). Claim acceptance in argumentation systems, i.e., the evaluation of commonly acceptable statements while disregarding their particular justifications, is an integral part of many structured argumentation formalisms [10, 11] and has received increasing attention in the literature [12, 13, 14]. A simple yet powerful generalization of Dung AFs that allow for claim-based evaluation are Claim-augmented AFs (CAFs) [14]. They extend AFs by assigning to each argument a claim. Semantics for CAFs can be obtained by evaluating the underlying AF before inspecting the claims of the acceptable arguments in the final step. CAFs serve as an ideal target formalism for ASPIC+ [10] and other knowledge representation formalisms which utilize abstract argumentation semantics whilst also considering the claims of the arguments in the evaluation. Thus, CAFs help to streamline the instantiation process by avoiding additional mappings to obtain semantical correspondence; e.g., in contrast to classical AF-instantiations of logic programs [15] where additional mappings are needed, claim-based semantics of CAFs capture logic programming semantics without detours [16]. In this way, we obtain a direct correspondence between the claim-extensions in the CAF and conclusion-extensions in the original formalism.

Although the acceptance of claims is closely related to argument-acceptance, there are subtle differences as observed in [14, 17, 10] stemming from the fact that claims can appear as conclusion of several different arguments. As a consequence, several properties of AF semantics such as *I-maximality*, i.e., $\subseteq$-maximality of extensions, cannot be taken for granted when considered in terms of the arguments' claims [18]. Furthermore, the additional level of claims causes a rise in the computational complexity of standard decision problems (in particular, verification is one level higher in the polynomial hierarchy as for standard AFs), see [14, 19]. Luckily, these drawbacks can be alleviated by taking fundamental properties of the attack relation into account: the basic observation that attacks typically depend on the claim of the attacking arguments gives rise to the central class of *well-formed CAFs*. CAFs from this class require that arguments with the same claim attack the same arguments, thus modeling – on the abstract level – a very natural behavior of arguments that is common to all leading structured argumentation formalisms and instantiations. Well-formed

NMR 2022: 20th International Workshop on Non-Monotonic Reasoning, August 07–09, 2022, Haifa, Israel

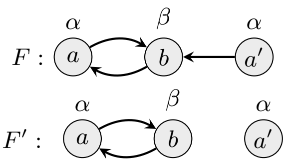
*Corresponding author.
mbernrei@dbai.tuwien.ac.at (M. Bernreiter);
dvorak@dbai.tuwien.ac.at (W. Dvořák);
arapberg@dbai.tuwien.ac.at (A. Rapberger);
woltran@dbai.tuwien.ac.at (S. Woltran)

CAFs have the main advantage that most of the semantics behave ‘as expected’. For instance, they retain the fundamental property of I-maximality, and their computational complexity is located at the same level of the polynomial hierarchy as for Dung AFs.

Unfortunately, it turns out that well-formedness cannot be assumed if one deals with preferences in argumentation, as arguments with the same claim are not necessarily equally plausible. The following example demonstrates this.

Example 1. *Consider two arguments a, a' with claim α, and another argument b having claim β. Moreover, both a and a' attack b, while b attacks a. Furthermore assume that we are given the additional information that b is preferred over a' (for example, if assumptions in the support of b are stronger than assumptions made by a'). A common method to integrate such information on argument rankings is to* delete *attacks from arguments that attack preferred arguments. In this case, we delete the attack from a' to b.*

Both frameworks are depicted below: F represents the original situation while F' is the CAF resulting from deleting the unsuccessful attack from a' on the argument b.



Note that F is well-formed since all arguments with the same claims attack the same arguments. The unique acceptable argument-set w.r.t. stable semantics (cf. Definition 2) is $\{a, a'\}$ which translates to $\{\alpha\}$ on the claim-level.

The CAF F', on the other hand, is no longer well-formed since a' does not attack b. In F', the argument-sets $\{a, a'\}$ and $\{a, b\}$ are both acceptable w.r.t. to stable-semantics. In terms of claims this translates to $\{\alpha\}$ and $\{\alpha, \beta\}$, which shows that I-maximality is violated on the claim-level.

Although well-formedness can not be guaranteed in view of preferences, this does not imply arbitrary behavior of the resulting CAF: on the one hand, preferences conform to a certain type of ordering (e.g., asymmetric, strict, partial, or total orders) over the set of arguments; on the other hand, it is evident that the deletion, reversion, and other types of attack manipulation impose certain restrictions on the structure of the resulting CAF. Combining both aspects, we obtain that, assuming well-formedness of the initial framework, it is unlikely that preference incorporation results in arbitrary behavior. The key motivation of this paper is to identify and exploit structural properties of preferential argumentation in the scope of claim acceptance. The aforementioned restrictions suggest beneficial impact on both the computational complexity and on desired semantical properties such as I-maximality.

In this paper, we tackle this issue by considering four commonly used methods, so-called reductions, to integrate preference orderings into the attack relation: the most common modification is the deletion of attacks in case the attacking argument is less preferred than its target. This method is typically utilized to transform preference-based argumentation frameworks (PAFs) [20] into AFs but is also used in many structured argumentation formalisms such as ASPIC+. This reduction has been criticized due to several problematic side-effects, e.g., it can be the case that two conflicting arguments are jointly acceptable, and has been accordingly adapted in [21]; two other reductions have been introduced in [9]. We apply these four preference reductions to well-formed CAFs with preferences. In particular, our main contributions are as follows:

- For each of the four reductions, we characterize the possible structure of CAFs that are obtained by applying the reduction to a well-formed CAF and a preference relation. This results in four novel CAF classes, each of which constitutes a proper extension of well-formed CAFs but does not retain the full expressiveness of general CAFs. We investigate the relationship between these classes.
- We study I-maximality of stable, preferred, semi-stable, stage, and naive semantics of the novel CAF classes. Our results highlight a significant advantage of a particular reduction: we show that, for admissible-based semantics, this modification preserves I-maximality. The other reductions fail to preserve I-maximality; moreover, for naive and stage semantics, I-maximality cannot be guaranteed for any of the four reductions.
- Finally, we investigate the complexity of reasoning for CAFs with preferences with respect to conflict-free, admissible, complete, and all of the aforementioned semantics. We show that for three of the four reductions, the verification problem drops by one level in the polynomial hierarchy for all except complete semantics and is thus not harder than for well-formed CAFs (which in turn has the same complexity as the corresponding AF problems). Complete semantics remain hard for all but one preference reduction. Moreover, it turns out that verification for the reduction which deletes attacks from weaker arguments remains as hard as for general CAFs.

Our results constitute a systematic study of the structural and computational effect of preferences on claim acceptance. Since we use CAFs as our base formalism, our investigations extend to large classes of formalisms that can be represented as CAFs, just like results on Dung AFs yield insights for formalisms that can be captured by AFs.

2. Preliminaries

We first define (abstract) argumentation frameworks [5]. U denotes a countable infinite domain of arguments.

Definition 1. *An argumentation framework (AF) is a tuple $F = (A, R)$ where $A \subseteq U$ is a finite set of arguments and $R \subseteq A \times A$ is an attack relation between arguments. Let $E \subseteq A$. We say E* attacks *b (in F) if $(a, b) \in R$ for some $a \in E$; $E^+_F = \{b \in A \mid \exists a \in E : (a, b) \in R\}$ denotes the set of arguments attacked by E. $E^\oplus_F = E \cup E^+_F$ is the* range *of E in F. An argument $a \in A$ is* defended *(in F) by E if $b \in E^+_F$ for each b with $(b, a) \in R$.*

Given an AF $F = (A, R)$ it can be convenient to write $a \in F$ for $a \in A$ and $(a, b) \in F$ for $(a, b) \in R$. Semantics for AFs are defined as functions σ which assign to each AF $F = (A, R)$ a set $\sigma(F) \subseteq 2^A$ of extensions. We consider for σ the functions cf, adm, com, $naive$, stb, prf, sem, and stg which stand for conflict-free, admissible, complete, naive, stable, preferred, semi-stable, and stage, respectively [22].

Definition 2. *Let $F = (A, R)$ be an AF. A set $S \subseteq A$ is* conflict-free (in F)*, if there are no $a, b \in S$, such that $(a, b) \in R$. $cf(F)$ denotes the collection of conflict-free sets of F. For a conflict-free set $S \in cf(F)$, it holds that*

- *$S \in adm(F)$ if each $a \in S$ is defended by S in F;*
- *$S \in com(F)$ if $S \in adm(F)$ and each $a \in A$ defended by S in F is contained in S;*
- *$S \in naive(F)$ if there is no $T \in cf(F)$ with $S \subset T$;*
- *$S \in stb(F)$ if each $a \in A \setminus S$ is attacked by S in F;*
- *$S \in prf(F)$ if $S \in adm(F)$ and there is no $T \in adm(F)$ with $S \subset T$;*
- *$S \in sem(F)$ if $S \in adm(F)$ and there is no $T \in adm(F)$ with $S^\oplus_F \subset T^\oplus_F$;*
- *$S \in stg(F)$ if there is no $T \in cf(F)$ with $S^\oplus_F \subset T^\oplus_F$.*

Example 2. *Consider the AF $F = (\{a, a', b\}, \{(a, b), (a', b), (b, a)\})$ from Example 1, ignoring claims α and β. Then $cf(F) = \{\emptyset, \{a\}, \{a'\}, \{b\}, \{a, a'\}\}$, $adm(F) = \{\emptyset, \{a\}, \{a'\}, \{a, a'\}\}$, $naive(F) = \{\{b\}, \{a, a'\}\}$, and $\sigma(F) = \{\{a, a'\}\}$ for $\sigma \in \{com, stb, prf, sem, stg\}$.*

CAFs are AFs in which each argument is assigned a claim, and thus constitute a straightforward generalization of AFs [14].

Table 1
I-maximality of CAFs.

	$naive_c$	stb_c	prf_c	sem_c	stg_c
CAF	x	x	x	x	x
$wfCAF$	x	✓	✓	✓	✓

Table 2
Complexity of CAFs ($\Delta \in \{CAF, wfCAF\}$).

σ	$Cred^\Delta_\sigma$	$Skept^\Delta_\sigma$	Ver^{CAF}_σ	Ver^{wfCAF}_σ
cf	in P	trivial	NP-c	in P
adm	NP-c	trivial	NP-c	in P
com	NP-c	P-c	NP-c	in P
$naive$	in P	coNP-c	NP-c	in P
stb	NP-c	coNP-c	NP-c	in P
prf	NP-c	Π^P_2-c	Σ^P_2-c	coNP-c
sem/stg	Σ^P_2-c	Π^P_2-c	Σ^P_2-c	coNP-c

Definition 3. *A claim-augmented argumentation framework (CAF) is a triple $(A, R, claim)$ where (A, R) is an AF and $claim\colon A \to Claims$ is a function that maps arguments to claims. The claim-function can be extended to sets of arguments, i.e., $claim(E) = \{claim(a) \mid a \in E\}$. A well-formed CAF (wfCAF) is a CAF $(A, R, claim)$ in which all arguments with the same claim attack the same arguments, i.e., for all $a, b \in A$ with $claim(a) = claim(b)$ we have $\{c \mid (a, c) \in R\} = \{c \mid (b, c) \in R\}$.*

The semantics of CAFs are based on those of AFs.

Definition 4. *Let $F = (A, R, claim)$ be a CAF. The claim-based variant of a semantics σ is defined as $\sigma_c(F) = \{claim(S) \mid S \in \sigma((A, R))\}$.*

Example 3. *Consider the CAF F from Example 1. Formally, $F = (A, R, claim)$ with $A = \{a, a', b\}$, $R = \{(a, b), (a', b), (b, a)\}$, $claim(a) = claim(a') = \alpha$, and $claim(b) = \beta$. F is well-formed and the underlying AF of F was investigated in Example 2. From there we can infer that, e.g., $cf_c(F) = \{\emptyset, \{\alpha\}, \{\beta\}\}$, $adm_c(F) = \{\emptyset, \{\alpha\}\}$, $naive_c(F) = \{\{\alpha\}, \{\beta\}\}$, and $stb_c(F) = \{\{\alpha\}\}$.*

Well-known basic relations between different AF semantics σ also hold for σ_c: $stb_c(F) \subseteq sem_c(F) \subseteq prf_c(F) \subseteq adm_c(F)$ as well as $stb_c(F) \subseteq stg_c(F) \subseteq naive_c(F) \subseteq cf_c(F)$ [18].

Note that the semantics $\sigma \in \{naive, stb, prf, sem, stg\}$ employ argument maximization and result in incomparable extensions on regular AFs: for all $S, T \in \sigma(F)$, $S \subseteq T$ implies $S = T$. This property is referred to as I-maximality, and is defined analogously for CAFs:

Definition 5. *σ_c is I-maximal for a class $\mathcal{C}$ of CAFs if, for all CAFs $F \in \mathcal{C}$ and all $S, T \in \sigma_c(F)$, $S \subseteq T$ implies $S = T$.*

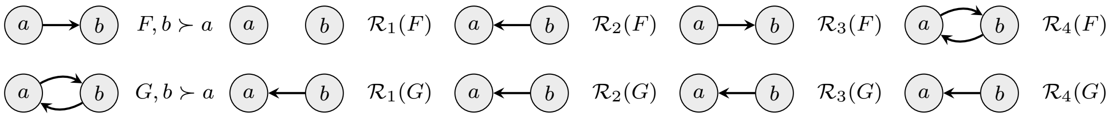


Figure 1: Effect of the four reductions on the attack relation between two arguments.

Table 1 shows I-maximality properties of CAFs [18], revealing an important property of wfCAFs compared to general CAFs: I-maximality is preserved in all semantics except $naive_c$, implying natural behavior of these maximization-based semantics analogous to regular AFs; see, e.g., [23] for a general discussion of such properties.

Regarding computational complexity, we consider the following decision problems pertaining to CAF-semantics σ_c:

- *Credulous Acceptance* ($Cred_\sigma^{CAF}$): Given a CAF F and claim α, is α contained in some $S \in \sigma_c(F)$?
- *Skeptical Acceptance* ($Skept_\sigma^{CAF}$): Given a CAF F and claim α, is α contained in each $S \in \sigma_c(F)$?
- *Verification* (Ver_σ^{CAF}): Given a CAF F and a set of claims S, is $S \in \sigma_c(F)$?

We furthermore consider these reasoning problems restricted to wfCAFs and denote them by $Cred_\sigma^{wfCAF}$, $Skept_\sigma^{wfCAF}$, and Ver_σ^{wfCAF}. Table 2 shows the complexity of these problems [14, 19]. Here we see that the complexity of the verification problem drops by one level in the polynomial hierarchy when comparing general CAFs to wfCAFs. This is an important advantage of wfCAFs, as a lower complexity in the verification problem allows for a more efficient enumeration of claim-extensions (cf. [14]).

3. Preference-based CAFs

As discussed in the previous sections, wfCAFs are a natural subclass of CAFs with advantageous properties in terms of I-maximality and computational complexity. However, when resolving preferences among arguments the resulting CAFs are typically no longer well-formed (cf. Example 1). In order to study preferences under a claim-centric view we introduce preference-based CAFs. These frameworks enrich the notion of wfCAFs with the concept of argument strength in terms of preferences. Our main goals are then to understand the effect of resolved preferences on the structure of the underlying wfCAF on the one hand, and to determine whether the advantages of wfCAFs are maintained on the other hand. Given this motivation, it is reasonable to consider the impact of preferences on *well-formed* CAFs only.

Definition 6. *A preference-based claim-augmented argumentation framework (PCAF) is a quadruple $F = (A, R, claim, \succ)$ where $(A, R, claim)$ is a well-formed CAF and $\succ$ is an asymmetric preference relation over A.*

Notice that preferences in PCAFs are not required to be transitive. While transitivity of preferences is often assumed in argumentation [21, 9], it cannot always be guaranteed in practice [6]. However, we will consider the effect of transitive orderings when applicable.

If a and b are arguments and $a \succ b$ holds then we say that a is stronger than b. But what effect should this ordering have? How should this influence, e.g., the set of admissible arguments? One possibility is to remove all attacks from weaker to stronger arguments in our PCAF, and to then determine the set of admissible arguments in the resulting CAF. This altering of attacks in a PCAF based on its preference-ordering is called a reduction. The literature describes four such reductions for regular AFs [9, 24, 21]. Following [9] we next recall these reductions.

Definition 7. *Given a PCAF $F = (A, R, claim, \succ)$, a corresponding CAF $\mathcal{R}_i(F) = (A, R', claim)$ is constructed via Reduction i, where $i \in \{1, 2, 3, 4\}$, as follows:*

- $i = 1$: $\forall a, b \in A: (a, b) \in R' \Leftrightarrow (a, b) \in R, b \not\succ a$
- $i = 2$: $\forall a, b \in A: (a, b) \in R' \Leftrightarrow ((a, b) \in R, b \not\succ a) \vee ((b, a) \in R, (a, b) \notin R, a \succ b)$
- $i = 3$: $\forall a, b \in A: (a, b) \in R' \Leftrightarrow ((a, b) \in R, b \not\succ a) \vee ((a, b) \in R, (b, a) \notin R)$
- $i = 4$: $\forall a, b \in A: (a, b) \in R' \Leftrightarrow ((a, b) \in R, b \not\succ a) \vee ((b, a) \in R, (a, b) \notin R, a \succ b) \vee ((a, b) \in R, (b, a) \notin R)$

Figure 1 visualizes the above reductions. Intuitively, Reduction 1 removes attacks that contradict the preference ordering while Reduction 2 reverts such attacks. Reduction 3 removes attacks that contradict the preference ordering, but only if the weaker argument is attacked by the stronger argument also. Reduction 4 can be seen as a combination of Reductions 2 and 3. Observe that all four reductions are polynomial time computable with respect to the input PCAF.

Note that many structured argumentation formalisms make use of preference-reductions as well. For instance, ABA+ [4] employs attack reversal similar to Reduction 2 while some instances of ASPIC [3] delete attacks from weaker arguments in the spirit of Reduction 1.

The semantics for PCAFs can now be defined in a straightforward way: first, one of the four reductions is applied to the given PCAF; then, CAF-semantics are applied to the resulting CAF.

Definition 8. *Let F be a PCAF and let $i \in \{1, 2, 3, 4\}$. The preference-claim-based variant of a semantics σ relative to Reduction i is defined as $\sigma^i_p(F) = \sigma_c(\mathcal{R}_i(F))$.*

Example 4. *Let $F = (A, R, claim, \succ)$ be the PCAF where $A = \{a, a', b\}$, $R = \{(a, b), (a', b), (b, a)\}$, $claim(a) = claim(a') = \alpha$, $claim(b) = \beta$, and $b \succ a'$. The underlying CAF $(A, R, claim)$ of F was examined in Example 3.*

$\mathcal{R}_1(F) = (A, R', claim)$ with $R' = \{(a, b), (b, a)\}$, which is the same CAF as F' in Example 1. It can be verified that, e.g., $adm^1_p(F) = adm_c(\mathcal{R}_1(F)) = \{\{\emptyset, \{\alpha\}, \{\beta\}, \{\alpha, \beta\}\}$ and $stb^1_p(F) = \{\{\alpha\}, \{\alpha, \beta\}\}$.

Indeed, the choice of reduction can influence the extensions of a PCAF. For example, $\mathcal{R}_2(F) = (A, R'', claim)$ with $R'' = \{(a, b), (b, a), (b, a)\}$, $adm^2_p(F) = \{\emptyset, \{\alpha\}, \{\beta\}\}$, and $stb^2_p(F) = \{\{\alpha\}, \{\beta\}\}$.

It is easy to see that basic relations between semantics carry over from CAFs, as, if we have $\sigma_c(F) \subseteq \tau_c(F)$ for two semantics σ, τ and all CAFs F, then also $\sigma^i_p(F) \subseteq \tau^i_p(F)$ for all PCAFs F. It thus holds that for all $i \in \{1, 2, 3, 4\}$, $stb^i_p(F) \subseteq sem^i_p(F) \subseteq prf^i_p(F) \subseteq adm^i_p(F)$ as well as $stb^i_p(F) \subseteq stg^i_p(F) \subseteq naive^i_p(F) \subseteq cf^i_p(F)$.

Remark 1. *In this paper we require the underlying CAF of a PCAF to be well-formed. The reason for this is that we are interested in whether the benefits of well-formed CAFs are preserved when preferences have to be taken into account. Even from a technical perspective, admitting PCAFs with a non-well-formed underlying CAF is not very interesting with respect to the questions addressed in this paper. Indeed, any CAF could be obtained from such general PCAFs, regardless of which preference reduction we are using, by simply specifying the desired CAF and an empty preference relation. Thus, such general PCAFs have the same properties regarding I-maximality and complexity as general CAFs.*

4. Characterization & Expressiveness

Our first step towards understanding the effect of preferences on wfCAFs is to examine the impact of resolving preferences on the structure of the underlying CAF. To this end, we consider four new CAF classes which are obtained from applying the reductions of Definition 7 to PCAFs.

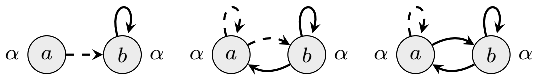


Figure 2: CAFs contained only in $\mathcal{R}_1$-$\mathbf{CAF}$, $\mathcal{R}_2$-$\mathbf{CAF}$, and $\mathcal{R}_4$-$\mathbf{CAF}$ respectively. Solid arrows are attacks, dashed arrows indicate where attacks are missing for the CAF to be well-formed.

Definition 9. *$\mathcal{R}_i$-$\mathbf{CAF}$ denotes the set of CAFs that can be obtained by applying Reduction i to PCAFs, i.e., $\mathcal{R}_i$-$\mathbf{CAF} = \{\mathcal{R}_i(F) \mid F \text{ is a PCAF}\}$.*

It is easy to see that $\mathcal{R}_i$-$\mathbf{CAF}$, with $i \in \{1, 2, 3, 4\}$, contains all wfCAFs (we can simply specify the desired wfCAF and an empty preference relation). However, not all CAFs are contained in $\mathcal{R}_i$-$\mathbf{CAF}$. For example, $F = (\{a, b\}, \{(a, b), (b, a)\}, claim)$ with $claim(a) = claim(b)$ can not be obtained from a PCAF F': such F' would need to contain either (a, b) or (b, a). But then, since the underlying CAF of a PCAF must be well-formed, F' would have to contain a self-attack which can not be removed by any of the reductions. This is enough to conclude[1] that the four new classes are located in-between wfCAFs and general CAFs:

Proposition 1. *Let $\mathbf{CAF}$ be the set of all CAFs and $\mathbf{wfCAF}$ the set of all wfCAFs. For all $i \in \{1, 2, 3, 4\}$ it holds that $\mathbf{wfCAF} \subset \mathcal{R}_i$-$\mathbf{CAF} \subset \mathbf{CAF}$.*

Furthermore, the new classes are all distinct from each other, i.e., we are indeed dealing with *four* new CAF classes:

Proposition 2. *For all $i \in \{1, 2, 4\}$ and all $j \in \{1, 2, 3, 4\}$ such that $i \neq j$ it holds that $\mathcal{R}_i$-$\mathbf{CAF} \not\subseteq \mathcal{R}_j$-$\mathbf{CAF}$ and $\mathcal{R}_3$-$\mathbf{CAF} \subset \mathcal{R}_i$-$\mathbf{CAF}$.*

Proof sketch. Figure 2 shows CAFs that are in only one of $\mathcal{R}_1$-$\mathbf{CAF}$, $\mathcal{R}_2$-$\mathbf{CAF}$, and $\mathcal{R}_4$-$\mathbf{CAF}$. Consider the PCAF $F = (\{a, b\}, \{(a, b), (b, b)\}, claim, \succ)$ with $claim(a) = claim(b) = \alpha$ and $b \succ a$. Then $\mathcal{R}_1(F)$, $\mathcal{R}_2(F)$, and $\mathcal{R}_4(F)$ are the CAFs of Figure 2. Since self-attacks are not removed or introduced by any reduction, and the underlying CAF must be well-formed, F is the only PCAF from which $\mathcal{R}_1(F)$, $\mathcal{R}_2(F)$, and $\mathcal{R}_4(F)$ can be obtained. Note that $\mathcal{R}_3(F)$ is simply the underlying CAF of F. $\mathcal{R}_3$-$\mathbf{CAF} \subset \mathcal{R}_i$-$\mathbf{CAF}$ follows by the fact that if an attack (a, b) is removed by Reduction 3 from some PCAF G, then $(b, a) \in G$. In this case, all reductions behave in the same way (cf. Definition 7 or Figure 1). □

[1] Although many proof sketches are provided in this text, a preprint of this paper with full proofs in the appendix can be accessed at https://arxiv.org/abs/2204.13305.

While the classes $\mathcal{R}_1$-$\mathbf{CAF}$, $\mathcal{R}_2$-$\mathbf{CAF}$, and $\mathcal{R}_4$-$\mathbf{CAF}$ are incomparable we observe $\mathcal{R}_3$-$\mathbf{CAF} \subset \mathcal{R}_i$-$\mathbf{CAF}$ which reflects that Reduction 3 is the most conservative of the four reductions, removing attacks only when there is a counter-attack from the stronger argument.

We now know that applying preferences to wfCAFs results in four distinct CAF-classes that lie in-between wfCAFs and general CAFs. It is still unclear, however, how to determine whether some CAF belongs to one of these classes or not. Especially for $\mathcal{R}_2$-$\mathbf{CAF}$ and $\mathcal{R}_4$-$\mathbf{CAF}$ this is not straightforward, since Reductions 2 and 4 not only remove but also introduce attacks and therefore allow for many possibilities to obtain a particular CAF as result. We tackle this problem by characterizing the new classes via the so-called wf-problematic part of a CAF.

Definition 10. *A pair of arguments (a, b) is wf-problematic in a CAF $F = (A, R, claim)$ if $a, b \in A$, $(a, b) \notin R$, and there is $a' \in A$ with $claim(a') = claim(a)$ and $(a', b) \in R$. The set $wfp(F) = \{(a, b) \mid (a, b)$ is wf-problematic in $F\}$ is called the wf-problematic part of F.*

Intuitively, the wf-problematic part of a CAF F consists of those attacks that are missing for F to be well-formed (cf. Figure 2). Indeed, F is a wfCAF if and only if $wfp(F) = \emptyset$. The four new classes can be characterized as follows:

Proposition 3. *Let $F = (A, R, claim)$ be a CAF. Then*

- *$F \in \mathcal{R}_1$-$\mathbf{CAF}$ iff $(a, b) \in wfp(F)$ implies $(b, a) \notin wfp(F)$;*
- *$F \in \mathcal{R}_2$-$\mathbf{CAF}$ iff there are no arguments a, a', b, b' in F with $claim(a) = claim(a')$ and $claim(b) = claim(b')$ such that $(a, b) \in wfp(F)$, $(b, a) \notin R$, $(a', b) \in R$, and either $(b, a') \in R$ or $((a', b') \notin R$ and $(b', a') \notin R)$;*
- *$F \in \mathcal{R}_3$-$\mathbf{CAF}$ iff $(a, b) \in wfp(F)$ implies $(b, a) \in R$;*
- *$F \in \mathcal{R}_4$-$\mathbf{CAF}$ iff there are no arguments a, a', b, b' in F with $claim(a) = claim(a')$ and $claim(b) = claim(b')$ such that $(a, b) \in wfp(F)$, $(b, a) \notin R$, $(a', b) \in R$, and either $(b, a') \notin R$ or $((a', b') \notin R$ and $(b', a') \notin R)$.*

Proof sketch. Regarding $\mathcal{R}_1$-$\mathbf{CAF}$, observe that Reduction 1 can only delete but not introduce attacks. If $(a, b) \in wfp(F)$ implies $(b, a) \notin wfp(F)$ then we can construct a PCAF F' with $R' = R \cup \{(a, b) \mid (a, b) \in wfp(F)\}$ and $b \succ a$ iff $(a, b) \in R' \setminus R$. Observe that $\succ$ is asymmetric. Conversely, a CAF G with arguments a, b such that $(a, b) \in wfp(G)$ and $(b, a) \in wfp(G)$ can not be obtained via Reduction 1 from a PCAF G', since G' would have to contain both the attacks $(a, b), (b, a)$ as well as the preferences $b \succ a, a \succ b$. The argument for $\mathcal{R}_3$-$\mathbf{CAF}$ is similar.

For $\mathcal{R}_2$-$\mathbf{CAF}$, suppose there are a, a', b with $claim(a) = claim(a')$, $(a, b) \in wfp(F)$, $(b, a) \notin R$, and $(a', b) \in R$. Assume there is a PCAF $F' = (A, R', claim, \succ)$ such that $\mathcal{R}_2(F') = F$. Since Reduction 2 can not completely remove conflicts, $(a, b) \notin R'$ and $(b, a) \notin R'$. If $(b, a') \in R$, then $(a', b) \in R'$ and $(b, a') \in R'$ since Reduction 2 can not introduce symmetric attacks. But then $(A, R', claim)$ is not well-formed. Now suppose $(b, a') \notin R$, but there is some b' with $claim(b) = claim(b')$, $(a', b') \notin R$, and $(b', a') \notin R$. Then also $(a', b') \notin R'$ and $(b', a') \notin R'$. But since $(a', b) \in R$ we have either $(a', b) \in R'$ or $(b, a') \in R'$, which means that $(A, R', claim)$ is not well-formed. In all other cases we can construct a PCAF $F'' = (A, R'', claim, \succ)$ such that $\mathcal{R}_2(F'') = F$: first revert all attacks (a', b) in F for which there is some a with $claim(a) = claim(a')$ and $(a, b) \notin R$, $(b, a) \notin R$; then, add all remaining pairs (a, b) that are still wf-problematic as attacks. Define $b \succ a$ iff $(a, b) \in R'' \setminus R$. It can be verified that $(A, R'', claim)$ is well-formed, $\succ$ is asymmetric, and $\mathcal{R}_2(F'') = F$. The argument for $\mathcal{R}_4$-$\mathbf{CAF}$ is similar. □

The above characterizations give us some insights into the effect of the various reductions on wfCAFs. Indeed, the similarity between the characterizations of $\mathcal{R}_1$-$\mathbf{CAF}$ and $\mathcal{R}_3$-$\mathbf{CAF}$, resp. $\mathcal{R}_2$-$\mathbf{CAF}$ and $\mathcal{R}_4$-$\mathbf{CAF}$, can intuitively be explained by the fact that Reductions 1 and 3 only remove attacks, while Reductions 2 and 4 can also introduce attacks. Furthermore, Proposition 3 allows us to decide in polynomial time whether a given CAF F can be obtained by applying one of the four preference reductions to a PCAF.

But what happens if we restrict ourselves to transitive preferences? Analogously to $\mathcal{R}_i$-$\mathbf{CAF}$, by $\mathcal{R}_i$-$\mathbf{CAF}_{tr}$ we denote the set of CAFs obtained by applying Reduction i to PCAFs with a transitive preference relation. It is clear that $\mathcal{R}_i$-$\mathbf{CAF}_{tr} \subseteq \mathcal{R}_i$-$\mathbf{CAF}$ for all $i \in \{1, 2, 3, 4\}$. Interestingly, the relationship between the classes $\mathcal{R}_i$-$\mathbf{CAF}_{tr}$ is different to that between $\mathcal{R}_i$-$\mathbf{CAF}$ (Proposition 2). Specifically, $\mathcal{R}_3$-$\mathbf{CAF}_{tr}$ is not contained in the other classes. Intuitively, this is because, in certain PCAFs F, transitivity can force $a_1 \succ a_n$ via $a_1 \succ a_2 \succ \ldots \succ a_n$ such that $(a_n, a_1) \in F$ but $(a_1, a_n) \notin F$. In this case, only Reduction 3 leaves the attacks between a_1 and a_n unchanged.

Proposition 4. *For all $i, j \in \{1, 2, 3, 4\}$ such that $i \neq j$ it holds that $\mathcal{R}_i$-$\mathbf{CAF}_{tr} \not\subseteq \mathcal{R}_j$-$\mathbf{CAF}_{tr}$.*

We will not characterize all four classes $\mathcal{R}_i$-$\mathbf{CAF}_{tr}$. However, capturing $\mathcal{R}_1$-$\mathbf{CAF}_{tr}$ will prove useful when analyzing the computational complexity of PCAFs using

Reduction 1 (see Section 6). Note that $wfp(F)$ can be seen as a directed graph, with an edge between vertices a and b whenever $(a, b) \in wfp(F)$. Thus, we may use notions such as paths and cycles in the wf-problematic part of a CAF.

Proposition 5. *$F \in \mathcal{R}_1$-$\mathbf{CAF}_{tr}$ for a CAF F if and only if (1) $wfp(F)$ is acyclic and (2) $(a, b) \in F$ implies that there is no path from a to b in $wfp(F)$.*

Proof sketch. Assume there is a cycle $(a_1, \dots, a_n, a_1)$ in $wfp(F)$. Then, since Reduction 1 can not introduce attacks, if there is a PCAF F' such that $\mathcal{R}_1(F') = F$, we have $(a_1, a_2), \dots, (a_n, a_1) \in F'$. This implies $a_1 \succ a_n \succ a_{n-1} \succ \dots \succ a_1$, i.e., $\succ$ is not asymmetric. Similarly, if there is a path $(a_1, \dots, a_n)$ in $wfp(F)$ we have to define $a_n \succ \dots \succ a_1$ in F'. But then $(a_1, a_n) \notin \mathcal{R}_1(F')$.

If $wfp(F)$ is acyclic and there is no path from a to b in $wfp(F)$ such that $(a, b) \in F$, then we can construct a PCAF F' such that $\mathcal{R}_1(F') = F$ in the same way as when $\succ$ is not transitive (cf. proof of Proposition 3). ☐

From the high-level point of view, our characterization results yield insights into the expressiveness of argumentation formalisms that allow for preferences. Propositions 3 and 5 show which situations can be captured by formalisms which (i) constructs attacks based on the claim of the attacking argument (i.e., formalisms with well-formed attack relation) and (ii) incorporate asymmetric or transitive preference relations on arguments using one of the four reductions.

5. I-Maximality

One of the advantages of wfCAFs over general CAFs is that they preserve I-maximality under most maximization-based semantics (cf. Table 1), which leads to more intuitive behavior of these semantics when considering extensions on the claim-level. We now investigate whether these advantages are preserved when preferences are introduced.

Definition 11. *σ_p^i is I-maximal for a class $\mathcal{C}$ of PCAFs if, for all F in $\mathcal{C}$ and all $S, T \in \sigma_p^i(F)$, $S \subseteq T$ implies $S = T$.*

From known properties of wfCAFs (cf. Table 1) it follows directly that $naive_p^i$ is not I-maximal for PCAFs. It remains to investigate the I-maximality of prf_p^i, stb_p^i, sem_p^i, and stg_p^i for PCAFs. For convenience, given a CAF $F = (A, R, claim)$ and $E \subseteq A$, we sometimes write $E \in \sigma(F)$ for $E \in \sigma((A, R))$.

Lemma 6. *Let $F = (A, R, claim, \succ)$ be a PCAF and $E \subseteq A$. $E \in cf(\mathcal{R}_i(F))$ if and only if $E \in cf((A, R, claim))$ for $i \in \{2, 3, 4\}$.*

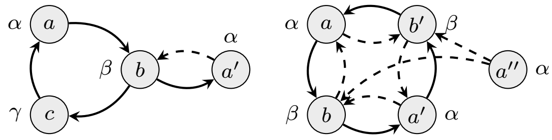


Figure 3: CAFs used as counter examples for I-maximality (cf. Proposition 8 and 9). Dashed arrows are edges in $wfp(F)$.

In other words, Reductions 2, 3, and 4 preserve conflict-freeness. It is easy to see that this is not the case for Reduction 1. In fact, Reduction 1 has been deemed problematic for exactly this reason when applied to regular AFs [21], although it is still discussed and considered in the literature alongside the other reductions [6]. We first consider Reduction 3, and show that it preserves I-maximality for some, but not all, semantics.

Proposition 7. *prf_p^3, stb_p^3, and sem_p^3 are I-maximal for PCAFs.*

Proof. By $stb_p^3(F) \subseteq sem_p^3(F) \subseteq prf_p^3(F)$ it suffices to consider prf_p^3. Towards a contradiction, assume there is a PCAF $F = (A, R, claim, \succ)$ such that $S \subset T$ for some $S, T \in prf_p^3(F)$. Then there are $S', T' \subseteq A$ such that $S' \in prf(\mathcal{R}_3(F))$, $claim(S') = S$, $T' \in prf(\mathcal{R}_3(F))$, and $claim(T') = T$. $S' \not\subseteq T'$ since $S' \in prf(\mathcal{R}_3(F))$. Thus, there is $x \in S'$ such that $x \notin T'$. But $claim(x) \in T$, i.e., there is $x' \in T'$ with $claim(x') = claim(x)$. There are two possibilities for why $x \notin T'$.

Case 1: $T' \cup \{x\} \notin cf(\mathcal{R}_3(F))$, i.e., there exists $y \in T'$ such that $y \notin S'$ and either $(x, y) \in F$ or $(y, x) \in F$. In fact, $(x, y) \notin F$: otherwise, by the well-formedness of $(A, R, claim)$, we have $(x', y) \in F$ and, by Lemma 6, $T' \notin cf(\mathcal{R}_3(F))$. Thus, $(y, x) \in F$. By the definition of Reduction 3, $(y, x) \in \mathcal{R}_3(F)$. S' must defend x in $\mathcal{R}_3(F)$, i.e., there exists $z \in S'$ such that $(z, y) \in \mathcal{R}_3(F)$. Then $(z, y) \in F$. Since $S \subset T$ there exists $z' \in T'$ such that $claim(z') = claim(z)$. $(z', y) \in F$ by the well-formedness of $(A, R, claim)$. But then $T' \notin cf(\mathcal{R}_3(F))$. Contradiction.

Case 2: x is not defended by T', i.e., there exists $y \in A$ that is not attacked by T' and such that $(y, x) \in \mathcal{R}_3(F)$. By the same argument as above, there is $z' \in T'$ with $(z', y) \in F$. It cannot be that $(z', y) \in \mathcal{R}_3(F)$, i.e., $y \succ z'$. By the definition of Reduction 3, $(y, z') \in F$ and thus $(y, z') \in \mathcal{R}_3(F)$. But then $T' \notin adm(\mathcal{R}_3(F))$. Contradiction. ☐

Of course, positive results regarding the I-maximality of PCAFs with arbitrary preferences, such as in the above proposition, still hold for PCAFs with transitive preference orderings. Conversely, for negative results, it suffices to show that I-maximality is not preserved on transitive orderings to obtain results for the more general case.

Table 3
I-maximality of PCAFs. Results also hold when considering only PCAFs with transitive preferences.

	$naive_p^i$	stb_p^i	prf_p^i	sem_p^i	stg_p^i
$i \in \{1,2,4\}$	x	x	x	x	x
$i = 3$	x	✓	✓	✓	x

Proposition 8. *stg_p^3 is not I-maximal for PCAFs, even when considering only transitive preferences.*

Proof sketch. Let F be the CAF shown on the left in Figure 3. Observe that $F \in \mathcal{R}_3\text{-}\mathbf{CAF}_{tr}$ since $\mathcal{R}_3(F') = F$ for the PCAF F' with the same arguments as F, attacks $\{(a,b),(b,c),(c,a),(a',b),(b,a')\}$ and $b \succ a'$. Moreover, it can be verified that $stg_p^3(F') = \{\{\alpha\},\{\alpha,\gamma\},\{\beta\}\}$. □

In contrast to Reduction 3, under Reductions 1, 2, and 4 we lose I-maximality for *all* semantics.

Proposition 9. *σ_p^i, with $\sigma \in \{prf, stb, sem, stg\}$ and $i \in \{1,2,4\}$, is not I-maximal for PCAFs, even when considering only transitive preferences.*

Proof sketch. We only need to show this for stb_p^i since $stb_p^i(F) \subseteq sem_p^i(F) \subseteq prf_p^i(F)$ and $stb_p^i(F) \subseteq stg_p^i(F)$.

For $i \in \{1,4\}$, consider F' from Example 1. $F' \in \mathcal{R}_1\text{-}\mathbf{CAF}_{tr}$ by Proposition 5, and $F' \in \mathcal{R}_4\text{-}\mathbf{CAF}_{tr}$ since $\mathcal{R}_4(F) = F$ for $F = (\{a,a',b\},\{(b,a)\}, claim, \succ)$ with $a \succ b$. It can be verified that $stb_c(F') = \{\{\alpha\},\{\alpha,\beta\}\}$.

For $i = 2$, let G be the CAF shown on the right in Figure 3. $G \in \mathcal{R}_2\text{-}\mathbf{CAF}_{tr}$ since $\mathcal{R}_2(G') = G$ for the PCAF G' with attacks $\{(b,a),(b,a'),(b',a),(b',a')\}$ and preferences $a \succ b$ and $a' \succ b'$. $stb_c(G) = \{\{\alpha\},\{\alpha,\beta\}\}$. □

Table 3 summarizes our I-maximality results. Reduction 3 manages to preserve I-maximality in most cases. It is also the most conservative of the reductions, preserving conflict-freeness and not adding new attacks. Interestingly, the other three reductions lose I-maximality for *all* semantics.

6. Computational Complexity

In this section, we investigate the impact of preferences on the computational complexity of claim-based reasoning. To this end, we adapt the decision problems introduced in Section 2 to PCAFs as follows: given a preference Reduction $i \in \{1,2,3,4\}$, we define $Cred_{\sigma,i}^{PCAF}$, $Skept_{\sigma,i}^{PCAF}$, and $Ver_{\sigma,i}^{PCAF}$ analogously to $Cred_{\sigma}^{CAF}$, $Skept_{\sigma}^{CAF}$, and Ver_{σ}^{CAF}, except that we take

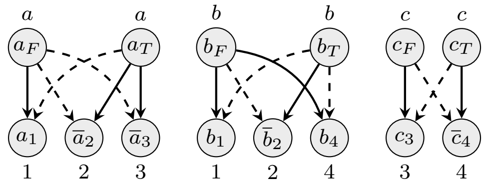


Figure 4: Reduction of 3-Sat-instance $C_1 = \{a,b,c\}$, $C_2 = \{\neg a, \neg b\}$, $C_3 = \{\neg a, c\}$, $C_4 = \{b, \neg c\}$, to an instance (F,S) of $Ver_{cf,1}^{PCAF}$ (cf. Proof of Proposition 10). Dashed arrows are attacks deleted in $\mathcal{R}_1(F)$, i.e., they are edges in $wfp(\mathcal{R}_1(F))$.

a PCAF instead of a CAF as input and appeal to the σ_p^i semantics instead of the σ_c semantics. Membership results for PCAFs can be inferred from results for general CAFs (recall that the preference reductions from PCAFs to the respective CAF class can be done in polynomial time), and hardness results from results for wfCAFs. Thus, the complexity of credulous and skeptical acceptance follows immediately from known results for CAFs and wfCAFs: given $i \in \{1,2,3,4\}$ and $\sigma \in \{cf, adm, com, naive, stb, prf, sem, stg\}$, the problems $Cred_{\sigma,i}^{PCAF}$ and $Skept_{\sigma,i}^{PCAF}$ have the same complexity as $Cred_{\sigma}^{wfCAF}$ and $Skept_{\sigma}^{wfCAF}$ respectively (cf. Table 2).

The computational complexity of the verification problem, on the other hand, is one level higher for general CAFs when compared to wfCAFs (cf. Table 2), i.e., the bounds that existing results yield for PCAFs are not tight. In the following, we examine the complexity of $Ver_{\sigma,i}^{PCAF}$ for each of the considered reductions and semantics. Let us first consider Reduction 1.

Proposition 10. *$Ver_{\sigma,1}^{PCAF}$ is* NP*-complete for $\sigma \in \{cf, naive\}$, even for transitive preferences.*

Proof sketch. NP-membership follows from known results for general CAFs. NP-hardness: let φ be an arbitrary instance of 3-Sat given as a set $\{C_1,\dots,C_m\}$ of clauses over variables X. We construct a PCAF $F = (A, R, claim, \succ)$ and a set of claims $S = \{1,\dots,m\} \cup X$ as follows:

- $A = V \cup \overline{V} \cup H$ where
  $V = \{x_i \mid x \in C_i, 1 \le i \le m\}$,
  $\overline{V} = \{\overline{x}_i \mid \neg x \in C_i, 1 \le i \le m\}$, and
  $H = \{x_T, x_F \mid x \in X\}$;
- $R = \{(x_T, x_i), (x_F, x_i) \mid x_i \in V\} \cup$
  $\{(x_T, \overline{x}_i), (x_F, \overline{x}_i) \mid \overline{x}_i \in \overline{V}\}$;
- $claim(x_i) = claim(\overline{x}_i) = i$ for $x_i, \overline{x}_i \in V \cup \overline{V}$,
  $claim(x_T) = claim(x_F) = x$ for $x \in X$;
- $x_i \succ x_T$ for all $x_i \in V$ and $\overline{x}_i \succ x_F$ for all $\overline{x}_i \in \overline{V}$.

Figure 4 illustrates the above construction. It can be verified that φ is satisfiable if and only if $S \in cf_p^1(F)$. The same can be shown for $naive_p^1$. Informally, the set S forces us to have, for each $x \in X$, x_T or x_F in S thus simulating a guess for an interpretation. Due to the removed attacks all corresponding occurrences x_i (resp. $\overline{x}_i$) can be added to S without conflict. Now it amounts to check whether these occurrences cover all i, i.e., make all clauses true under the actual guess. □

Note that the "trick" in above construction to guess an interpretation does not work for admissible-based semantics, since the occurrences of x_i resp. $\overline{x}_i$ in S would remain undefended. Indeed, we need a more involved construction next.

Proposition 11. *$Ver_{\sigma,1}^{PCAF}$ is* NP*-complete for $\sigma \in \{stb, adm, com\}$, even for transitive preferences.*

Proof sketch. We show NP-hardness. Let φ be a 3-Sat-instance given as a set $\{C_1, \dots, C_m\}$ of clauses over variables X. For convenience, we directly construct a CAF $F = (A, R, claim)$ with $F \in \mathcal{R}_1\text{-}\mathbf{CAF}_{tr}$ instead of providing a PCAF F' such that $\mathcal{R}_1(F') = F$. This is legitimate, as, by our characterization of $\mathcal{R}_1\text{-}\mathbf{CAF}_{tr}$ (see Proposition 5), we can obtain F' by simply adding all edges in $wfp(F)$ to R and defining $\succ$ accordingly.

- $A = V \cup \overline{V} \cup H$ where
  $V = \{x_i \mid x \in C_i, 1 \leq i \leq m\}$,
  $\overline{V} = \{\overline{x}_i \mid \neg x \in C_i, 1 \leq i \leq m\}$, and
  $H = \{x_{i,j}^k, \hat{x}_{i,j}^k \mid 1 \leq k \leq 4, x_i \in V, \overline{x}_j \in \overline{V}\}$;
- $R = \{(x_i, x_{i,j}^1), (x_{i,j}^1, x_{i,j}^2), (x_{i,j}^2, \overline{x}_j), (\overline{x}_j, x_{i,j}^3), (x_{i,j}^3, x_{i,j}^4), (x_{i,j}^4, x_i) \mid x_i \in V, \overline{x}_j \in \overline{V}\}$;
- $claim(x_i) = claim(\overline{x}_i) = i$ for all $x_i, \overline{x}_i$, $claim(x_{i,j}^k) = claim(\hat{x}_{i,j}^k) = x_{i,j}^k$ for all $x_{i,j}^k, \hat{x}_{i,j}^k$.

For verification consider the set $S = \{1, \dots, m\} \cup \{claim(a) \mid a \in H\}$. Figure 5 illustrates the above construction. It can be verified that (1) $F \in \mathcal{R}_1\text{-}\mathbf{CAF}_{tr}$; (2) φ is satisfiable iff $S \in stb_c(F)$. Likewise for adm_c and com_c. Intuitively, for each $x_i, \overline{x}_j$, the helper arguments $x_{i,j}^k$ and the corresponding cycle ensures that only one of $x_i, \overline{x}_j$ can be chosen. Note that x_i and $\overline{x}_j$ must not attack each other directly because of well-formedness in the original CAF. □

In fact, when applying Reduction 1, we lose the advantages of wfCAFs for *all* investigated semantics, since also for the remaining semantics verification remains harder than in the case of wfCAFs.

Proposition 12. *$Ver_{\sigma,1}^{PCAF}$ is Σ_2^{P}-complete for $\sigma \in \{prf, sem, stg\}$, even for transitive preferences.*

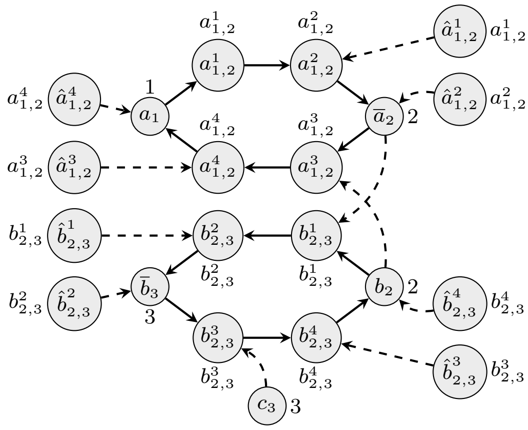


Figure 5: Reduction of 3-Sat-instance $C_1 = \{a\}$, $C_2 = \{\neg a, b\}$, $C_3 = \{\neg b, c\}$, to an instance (F', S) of $Ver_{stb,1}^{PCAF}$ (cf. Proof of Proposition 11). Dashed arrows are attacks deleted from F', i.e., they are edges in $wfp(\mathcal{R}_1(F'))$.

The proposition can be proven by adapting the standard translation for skeptical acceptance of preferred-semantics [25, Reduction 3.7].

We now turn our attention to Reductions 2, 3, and 4. Since these reductions do not remove conflicts between arguments, it is easy to see that verification for conflict-free and naive semantics remains tractable.

Proposition 13. *$Ver_{\sigma,i\in\{2,3,4\}}^{PCAF}$ is in* P *for $\sigma \in \{cf, naive\}$.*

Proof sketch. By Lemma 6, given a PCAF F, it suffices to test if C is conflict-free (resp. naive) in the underlying CAF of F. This problem is in P for wfCAFs (cf. Table 2). □

As it turns out, with Reductions 2, 3, and 4 we retain the benefits of wfCAFs over general CAFs for almost all investigated semantics with respect to computational complexity. In short, verification for wfCAFs is easier than on general CAFs because, given a wfCAF F and a set of claims C, a set of arguments S can be constructed in polynomial time such that S is the unique maximal admissible set in F with claim $claim(S) = C$ [14]. Making use of the fact that Reductions 2, 3, and 4 do not alter conflicts between arguments, we can construct such a maximal set of arguments also for PCAFs: given a PCAF F and set C of claims, we define the set $E_0(C)$ containing all arguments of F with a claim in C; the set $E_1^i(C)$ is obtained from $E_0(C)$ by removing all arguments attacked by $E_0(C)$ in the underlying CAF of F; finally, the set $E_*^i(C)$ is obtained by repeatedly removing all arguments not defended by $E_1^i(C)$ in $\mathcal{R}_i(F)$ until a fixed point is reached. Recall that $S_{(A,R)}^+ = \{a \mid (b,a) \in R, b \in S\}$ denotes the arguments attacked by S in (A, R).

Definition 12. *Given a PCAF $F = (A, R, claim, \succ)$, a set of claims C, and $i \in \{2, 3, 4\}$, we define*

$$E_0(C) = \{a \in A \mid claim(a) \in C\};$$
$$E_1^i(C) = E_0(C) \setminus E_0(C)^+_{(A,R)};$$
$$E_k^i(C) = \{x \in E_{k-1}^i(C) \mid x \text{ is defended by } E_{k-1}^i(C) \text{ in } \mathcal{R}_i(F)\} \text{ for } k \geq 2;$$
$$E_*^i(C) = E_k^i \text{ for } k \geq 2 \text{ such that } E_k^i(C) = E_{k-1}^i(C).$$

The above definition is based on [14, Definition 5], but with the crucial differences that undefended arguments are (a) computed w.r.t. $\mathcal{R}_i(F)$ and (b) are iteratively removed until a fixed point is reached.

For conflict-free based semantics we observe that the conflicts are not affected by the reductions and thus one can use existing results for well-formed CAFs [14, Lemma 1] to obtain that $E_1^i(C)$ is the unique candidate for the maximal conflict-free set of arguments that realizes the claim set C.

Lemma 14. *Let F be a PCAF, C be a set of claims and $i \in \{2, 3, 4\}$. We have that $C \in cf_p^i(F)$ iff $claim(E_1^i(C)) = C$. Moreover, if $C \in cf_p^i(F)$ then $E_1^i(C)$ is the unique maximal conflict-free set S in $\mathcal{R}_i(F)$ such that $claim(S) = C$.*

Regarding admissible semantics we are looking for a conflict-free set that defends all its arguments. Thus we start from the conflict-free set $E_1^i(C)$. Notice that arguments that are not in $E_1^i(C)$ cannot be contained in any admissible set S with $claim(S) = C$. We can then obtain the maximal admissible set realizing C in $\mathcal{R}_i(F)$ by iteratively removing arguments that are not defended by the current set of arguments. Once we reach a fixed-point we have an admissible set, but need to check whether we still cover all the claims of C.

Lemma 15. *Let F be a PCAF, C be a set of claims and $i \in \{2, 3, 4\}$. We have that $C \in adm_p^i(F)$ iff $claim(E_*^i(C)) = C$. Moreover, if $C \in adm_p^i(F)$ then $E_*^i(C)$ is the unique maximal admissible set S in $\mathcal{R}_i(F)$ such that $claim(S) = C$.*

By computing the maximal conflict-free (resp. admissible) extensions $E_1^i(C)$ (resp. $E_*^i(C)$) for a set of claims C, the verification problem becomes easier for most semantics.

Proposition 16. *$Ver_{\sigma, i \in \{2,3,4\}}^{PCAF}$ is in* P *for $\sigma \in \{adm, stb\}$. It is* coNP*-complete for $\sigma \in \{prf, sem, stg\}$, even when considering only transitive preferences.*

Proof sketch. Let $F = (A, R, claim, \succ)$ be a PCAF, let C be a set of claims, and let $i \in \{2, 3, 4\}$. We can compute $\mathcal{R}_i(F)$, $E_1^i(C)$, and $E_*^i(C)$ in polynomial time.

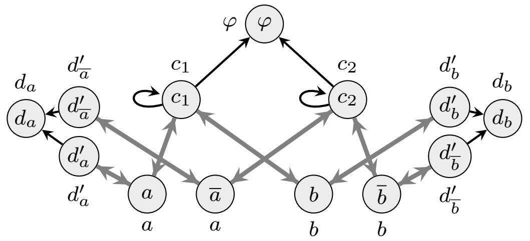


Figure 6: $\mathcal{R}_4(F)$ from the proof of Proposition 17, $\varphi = ((a \vee b) \wedge (\neg a \vee \neg b))$. Symmetric attacks drawn in gray and thick have been introduced by Reduction 4.

For adm, by Lemma 15, it suffices to test whether $claim(E_*^i(C)) = C$. For stb, we first check whether $C \in adm_p^i(F)$. If not, $C \notin stb_p^i(F)$. If yes, then, by Lemma 15, $claim(E_*^i(C)) = C$. We can check in polynomial time if $E_*^i(C) \in stb(\mathcal{R}_i(F))$. If yes, we are done. If no, then there is an argument x that is not in $E_*^i(C)$ but is also not attacked by $E_*^i(C)$ in $\mathcal{R}_i(F)$. Moreover, there can be no other $S \in stb(\mathcal{R}_i(F))$ with $claim(S) = C$ since for any such S we would have $S \subseteq E_*^i(C)$, which would imply that S does not attack x and that $x \notin S$.

The arguments for $\sigma \in \{prf, sem, stg\}$ are similar, but some checks require coNP-time. ☐

For complete semantics, only Reduction 3 retains the benefits of wfCAFs. Here, the fact that Reductions 2 and 4 can introduce new attacks leads to an increase in complexity.

Proposition 17. *$Ver_{com,3}^{PCAF}$ is in* P*. $Ver_{com, i \in \{2,4\}}^{PCAF}$ is* NP*-complete, even for transitive preferences.*

Proof sketch. P-membership for $Ver_{com,3}^{PCAF}$ is similar to the proof of Proposition 16. We demonstrate NP-hardness of $Ver_{com,4}^{PCAF}$. Let φ be an arbitrary instance of 3-Sat given as a set C of clauses over variables X and let $\overline{X} = \{\overline{x} \mid x \in X\}$. We construct a PCAF $F = (A, R, claim, \succ)$ as well as a set of claims $S = X \cup \{\varphi\}$:

- $A = \{\varphi\} \cup C \cup X \cup \overline{X} \cup \{d_x \mid x \in X\} \cup \{d'_x \mid x \in X \cup \overline{X}\}$;
- $R = \{(c, \varphi) \mid c \in C\} \cup \{(c, c) \mid c \in C\} \cup \{(c, x) \mid x \in c, c \in C\} \cup \{(c, \overline{x}) \mid \neg x \in c, c \in C\} \cup \{(d'_x, x) \mid x \in X \cup \overline{X}\} \cup \{(d'_x, d_x), (d'_{\overline{x}}, d_x) \mid x \in X\}$;
- $claim(x) = claim(\overline{x}) = x$ for $x \in X$, $claim(v) = v$ otherwise;
- $x \succ c$, $x \succ d'_x$ for all $x \in X \cup \overline{X}$ and all $c \in C$.

Figure 6 illustrates the above construction. It can be verified that φ is satisfiable iff $S \in com_c(\mathcal{R}_4(F))$. ☐

Table 4
Complexity of $Ver^{PCAF}_{\sigma,i}$. Results also hold when considering only PCAFs with transitive preferences.

σ	$i = 1$	$i \in \{2, 4\}$	$i = 3$
$cf/adm/naive/stb$	NP-c	in P	in P
com	NP-c	NP-c	in P
$prf/sem/stg$	Σ_2^P-c	coNP-c	coNP-c

Table 4 summarizes our complexity results. Reduction 3 preserves the lower complexity of wfCAFs for all investigated semantics, while Reductions 2 and 4 preserve the lower complexity for all but complete semantics. Reduction 1 does not preserve the advantages of wfCAFs, and rather exhibits the full complexity as general CAFs. Notice that the lower complexity of the verification problem is crucial for enumerating extensions. In particular, the improved enumeration algorithm for wfCAFs [14] is based on the polynomial time verification of claim-sets and thus extends to PCAFs under Reductions 2–4.

7. Conclusion

Many approaches to structured argumentation (i) assume that arguments with the same claims attack the same arguments and (ii) take preferences into account. Investigations on claim-augmented argumentation frameworks (CAFs) so far only consider (i), showing that the resulting subclass of well-formed CAFs (wfCAFs) has several desired properties. The research question of this paper is to analyze whether these properties carry over when preferences are taken into account, since the incorporation of preferences can violate the syntactical restriction of wfCAFs.

To this end, we introduced preference-based claim-augmented argumentation frameworks (PCAFs) and investigated the impact of the four preference reductions commonly used in abstract argumentation when applied to PCAFs. We examined and characterized CAF-classes that result from applying these reductions to PCAFs, yielding insights into the expressiveness of argumentation formalisms that can be instantiated as CAF and allow for preference incorporation. Furthermore, we investigated the fundamental properties of I-maximality and computational complexity for PCAFs. Preserving I-maximality is desirable since it implies intuitive behavior of maximization-based semantics, while the complexity of the verification problem is crucial for the enumeration of claim-extensions. Insights in terms of both semantical and computational properties provide necessary foundations towards a practical realization of this particular argumentation paradigm (we refer to, e.g., [26, 27], for a similar research endeavor in terms of incomplete AFs).

Our results show that (1) Reduction 3, the most conservative of the four reductions, exhibits the same properties as wfCAFs regarding computational complexity while also preserving I-maximality for most of the semantics; (2) Reductions 2 and 4 retain the advantages of wfCAFs regarding complexity for all but complete semantics, but do not preserve I-maximality for any investigated semantics; (3) under Reduction 1, neither complexity properties nor I-maximality are preserved. The above results hold even if we restrict ourselves to transitive preferences. It is worth noting that Reduction 3 behaves favorably on regular AFs as well, fulfilling many principles for preference-based semantics laid out by Kaci et al. (2018).

A possible direction for future work is to lift the preference ordering over arguments to sets of arguments and select extensions in this way. This has been investigated for regular AFs in combination with Reduction 2 [21]. Another direction is to extend our studies to alternative semantics for CAFs [18, 19], where subset-maximization is handled on the claim-level instead of on the argument-level.

Acknowledgments

We thank the anonymous reviewers for their valuable feedback. This work was funded by the Austrian Science Fund (FWF) under grants Y698 and W1255-N23 and by the Vienna Science and Technology Fund (WWTF) under grant ICT19-065.

References

[1] H. Li, N. Oren, T. J. Norman, Probabilistic argumentation frameworks, in: Proc. TAFA'11, volume 7132 of *LNCS*, Springer, 2011, pp. 1–16. URL: https://doi.org/10.1007/978-3-642-29184-5_1. doi:10.1007/978-3-642-29184-5_1.

[2] K. Atkinson, P. Baroni, M. Giacomin, A. Hunter, H. Prakken, C. Reed, G. R. Simari, M. Thimm, S. Villata, Towards artificial argumentation, AI Magazine 38 (2017) 25–36. doi:10.1609/aimag.v38i3.2704.

[3] S. Modgil, H. Prakken, A general account of argumentation with preferences, Artif. Intell. 195 (2013) 361–397. URL: https://doi.org/10.1016/j.artint.2012.10.008. doi:10.1016/j.artint.2012.10.008.

[4] K. Cyras, F. Toni, ABA+: assumption-based argumentation with preferences, in: Proc.KR'16, AAAI Press, 2016, pp. 553–556. URL: http://www.aaai.org/ocs/index.php/KR/KR16/paper/view/12877.

[5] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, Artif. Intell. 77 (1995) 321–358. doi:10.1016/0004-3702(94)00041-X.

[6] S. Kaci, L. W. N. van der Torre, S. Vesic, S. Villata, Preference in abstract argumentation, in: Handbook of Formal Argumentation, Volume 2, College Publications, 2021, pp. 211–248.

[7] S. Modgil, Reasoning about preferences in argumentation frameworks, Artificial Intelligence 173 (2009) 901–934.

[8] K. Atkinson, T. J. M. Bench-Capon, Value-based argumentation, IfCoLog Journal of Logic and its Applications 8 (2021) 1543–1588. URL: https://collegepublications.co.uk/ifcolog/?00048.

[9] S. Kaci, L. W. N. van der Torre, S. Villata, Preference in abstract argumentation, in: Proc. COMMA'18, volume 305 of *FAIA*, IOS Press, 2018, pp. 405–412. URL: https://doi.org/10.3233/978-1-61499-906-5-405. doi:10.3233/978-1-61499-906-5-405.

[10] S. Modgil, H. Prakken, Abstract rule-based argumentation, in: Handbook of Formal Argumentation, College Publications, 2018, pp. 287–364.

[11] P. M. Dung, R. A. Kowalski, F. Toni, Assumption-based argumentation, in: Argumentation in Artificial Intelligence, Springer, 2009, pp. 199–218. URL: https://doi.org/10.1007/978-0-387-98197-0_10. doi:10.1007/978-0-387-98197-0_10.

[12] J. F. Horty, Skepticism and floating conclusions, Artif. Intell. 135 (2002) 55–72. URL: https://doi.org/10.1016/S0004-3702(01)00160-6. doi:10.1016/S0004-3702(01)00160-6.

[13] P. Baroni, R. Riveret, Enhancing statement evaluation in argumentation via multi-labelling systems, J. Artif. Intell. Res. 66 (2019) 793–860. doi:10.1613/jair.1.11428.

[14] W. Dvořák, S. Woltran, Complexity of abstract argumentation under a claim-centric view, Artif. Intell. 285 (2020) 103290. doi:10.1016/j.artint.2020.103290.

[15] M. Caminada, S. Sá, J. Alcântara, W. Dvořák, On the equivalence between logic programming semantics and argumentation semantics, Int. J. Approx. Reasoning 58 (2015) 87–111. doi:10.1016/j.ijar.2014.12.004.

[16] A. Rapberger, Defining argumentation semantics under a claim-centric view, in: Proc. STAIRS'20, volume 2655 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020. URL: http://ceur-ws.org/Vol-2655/paper2.pdf.

[17] H. Prakken, G. A. Vreeswijk, Logics for defeasible argumentation, in: Handbook of Philosophical Logic, volume 4, Springer, 2002.

[18] W. Dvořák, A. Rapberger, S. Woltran, Argumentation semantics under a claim-centric view: Properties, expressiveness and relation to SETAFs, in: Proc. KR'20, IJCAI.org, 2020, pp. 341–350. doi:10.24963/kr.2020/35.

[19] W. Dvořák, A. Greßler, A. Rapberger, S. Woltran, The complexity landscape of claim-augmented argumentation frameworks, in: Proc. AAAI'21, AAAI Press, 2021, pp. 6296–6303. URL: https://ojs.aaai.org/index.php/AAAI/article/view/16782.

[20] L. Amgoud, C. Cayrol, On the acceptability of arguments in preference-based argumentation, in: Proc. UAI'98, Morgan Kaufmann, 1998, pp. 1–7. URL: https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=226&proceeding_id=14.

[21] L. Amgoud, S. Vesic, Rich preference-based argumentation frameworks, Int. J. Approx. Reason. 55 (2014) 585–606. URL: https://doi.org/10.1016/j.ijar.2013.10.010. doi:10.1016/j.ijar.2013.10.010.

[22] P. Baroni, M. Caminada, M. Giacomin, Abstract argumentation frameworks and their semantics, in: Handbook of Formal Argumentation, College Publications, 2018, pp. 159–236.

[23] L. van der Torre, S. Vesic, The principle-based approach to abstract argumentation semantics, IfCoLog Journal of Logic and its Applications 4 (2017) 2735–2778. URL: http://www.collegepublications.co.uk/downloads/ifcolog00017.pdf.

[24] L. Amgoud, C. Cayrol, A reasoning model based on the production of acceptable arguments, Ann. Math. Artif. Intell. 34 (2002) 197–215.

[25] W. Dvořák, P. E. Dunne, Computational problems in formal argumentation and their complexity, IfCoLog Journal of Logic and its Applications 4 (2017) 2557–2622.

[26] D. Baumeister, M. Järvisalo, D. Neugebauer, A. Niskanen, J. Rothe, Acceptance in incomplete argumentation frameworks, Artif. Intell. 295 (2021) 103470. URL: https://doi.org/10.1016/j.artint.2021.103470. doi:10.1016/j.artint.2021.103470.

[27] B. Fazzinga, S. Flesca, F. Furfaro, Revisiting the notion of extension over incomplete abstract argumentation frameworks., in: Proc. IJCAI'20, 2020, pp. 1712–1718.