Journal of Artificial Intelligence Research 81 (2024) 203-262 Submitted 01/2024; published 09/2024

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

Michael Bernreiter MBERNREI@DBAI.TUWIEN.AC.AT
Wolfgang Dvořák DVORAK@DBAI.TUWIEN.AC.AT
Institute of Logic and Computation, TU Wien, Austria

Anna Rapberger A.RAPBERGER@IMPERIAL.AC.UK
Department of Computing, Imperial College London, UK

Stefan Woltran WOLTRAN@DBAI.TUWIEN.AC.AT
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 four new classes of claim-augmented argumentation frameworks. These classes behave differently from each other with respect to semantic properties and computational complexity, but also in connection with structured argumentation formalisms such as assumption-based argumentation. 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 (Li, Oren, & Norman, 2011; Atkinson, Baroni, Giacomin, Hunter, Prakken, Reed, Simari, Thimm, & Villata, 2017). Indeed, preferences are nowadays a standard feature of many structured argumentation formalisms (Modgil & Prakken, 2013; Cyras & Toni, 2016). At the same time, there are numerous generalizations of abstract Argumentation Frameworks (AFs) (Dung, 1995) that consider the impact of preferences on the abstract level, be it in terms of argument strength (Kaci, van der Torre, Vesic, & Villata, 2021; Modgil, 2009; Bernreiter, Dvořák, & Woltran, 2024), preferences between values (Atkinson & Bench-Capon, 2021), or weighted arguments/attacks (Bistarelli & Santini, 2021). In 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. Deciding acceptability of arguments via argumentation semantics is thus reflected in terms of the modified attack relation (Kaci et al., 2021).

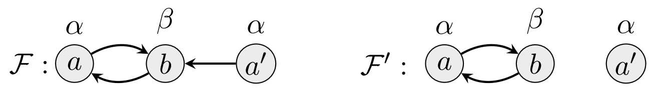
The difference in argument strength and the resulting modification of the attack relation naturally influences the acceptability of the arguments' conclusion (the *claim* of the argument). Claim acceptance in argumentation systems, i.e., the evaluation of commonly acceptable statements while disregarding their particular justifications, is an integral part of many structured argumentation formalisms (Modgil & Prakken, 2018; Dung, Kowalski, & Toni, 2009) and has received increasing attention in the literature (Horty, 2002; Baroni & Riveret, 2019; Dvořák & Woltran, 2020; Rocha & Cozman, 2022b). A simple yet powerful generalization of AFs that allow for claim-based evaluation are Claim-augmented AFs (CAFs) (Dvořák & Woltran, 2020), where each argument is assigned 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+ (Modgil & Prakken, 2018) and other formalisms which utilize abstract argumentation semantics whilst also considering the claims of the arguments in the evaluation. Moreover, CAF semantics capture semantics of logic programs without the need of additional mappings (Rapberger, 2020), in contrast to classical AF-instantiations (Caminada, Sá, Alcântara, & Dvořák, 2015). Thus, we obtain a direct correspondence between 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 (Dvořák & Woltran, 2020; Prakken & Vreeswijk, 2002; Modgil & Prakken, 2018) stemming from the fact that claims can appear as conclusion of several different arguments. As a consequence, several properties of AF semantics cannot be taken for granted when considered in terms of the arguments' claims. For instance, the property of *I-maximality*, i.e., $\subseteq$-maximality of extensions, which gives insights into the expressiveness of semantics (Dunne, Dvořák, Linsbichler, & Woltran, 2015) and skeptical argument justification (Baroni & Giacomin, 2007) is not satisfied by most CAF semantics (Dvořák, Rapberger, & Woltran, 2023). 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 (Dvořák, Greßler, Rapberger, & Woltran, 2023). 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*. This class satisfies that all arguments with the same claim attack the same arguments; thus modeling a very natural behavior of arguments that is common to all leading structured argumentation formalisms and instantiations. Well-formed CAFs have the main advantage that most of the semantics behave 'as expected', e.g., they retain I-maximality, and their computational complexity is located at the same level of the polynomial hierarchy as for 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: $\mathcal{F}$ represents the original situation while $\mathcal{F}'$ is the CAF resulting from deleting the unsuccessful attack from a' on the argument b.



Note that $\mathcal{F}$ 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 $\mathcal{F}'$, 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 cannot 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, transitive) over the set of arguments; on the other hand, it is evident that the deletion, reversion, and other types of attack manipulation impose 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) (Amgoud & Cayrol, 1998) 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 (Amgoud & Vesic, 2014); two other reductions have been introduced in (Kaci, van der Torre, & Villata, 2018). We apply these four preference reductions to well-formed CAFs. 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 not retaining full expressiveness of general CAFs. We investigate the relationship between these classes.

- We study semantic properties of the novel CAF classes. Our results highlight a significant advantage of a particular reduction when it comes to admissible based semantics: under this modification, subset-maximization (as used in preferred semantics for example) on the argument-level coincides with subset-maximization on the claim-level. Moreover, this modification preserves I-maximality. The other reductions fail to preserve these properties in most cases; moreover, for the conflict-free-based naive and stage semantics, I-maximality cannot be guaranteed for any of the four reductions.

- We investigate the complexity of reasoning for CAFs with preferences. 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.

- Finally, we examine the relationship between CAFs with preferences and assumption-based argumentation with preferences (ABA^+). Specifically, we show that if preferences in well-formed CAFs are handled via attack reversion, we can fully capture ABA^+ frameworks in which the axiom of weak contraposition is satisfied.

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 AFs yield insights for formalisms that can be captured by AFs.

This paper is organized as follows. In Section 2, we recall necessary background. In Section 3, we introduce Preference-based CAFs (PCAFs) which combine PAFs with well-formed CAFs. We characterize the novel CAF classes based on the preference reductions in Section 4, study the I-maximality of the semantics in Section 5, and their computational complexity in Section 6. We then investigate relationship between PCAFs and ABA^+ in Section 7 and conclude in Section 8.

This is an extended version of a paper published at AAAI 2023 (Bernreiter, Dvořák, Rapberger, & Woltran, 2023). The following contributions are new in this version: in addition to inherited CAF-semantics, we now also consider hybrid CAF-semantics (see Definition 9) and investigate them with respect to their semantic properties (in Section 5) and their computational complexity (in Section 6). Section 7, where we investigate the relationship between PCAFs and ABA^+, is entirely new. Moreover, this version contains full proofs for our results, as well as additional figures and explanations.

2. Preliminaries

In this section, we recall the necessary preliminaries needed for this paper. We start with Dung-style AFs in Subsection 2.1. In Subsection 2.2 we recall PAFs, which provide the foundation needed to deal with preferences. Finally, in Subsection 2.3 we recall CAFs, which allow us to take not only arguments but also their claims into account.

2.1 Abstract Argumentation Frameworks (AFs)

Abstract Argumentation Frameworks (AFs) (Dung, 1995) are a simple yet powerful formalism that allows us to model discussions. AFs contain abstract arguments, which are abstract in the sense that we are not concerned with the internal structure of the argument themselves. Rather, we are interested in the relationship between arguments, which is modeled via *attacks* between arguments. If there is an attack between two arguments, then the arguments are in conflict and cannot be jointly accepted. Moreover, usually we require that an attacked argument must be *defended* against all its attackers in order to be accepted.

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

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 (Baroni, Caminada, & Giacomin, 2018). We consider for σ the functions *cf* (conflict-free), *adm* (admissible), *com* (complete), *grd* (grounded), *naive* (naive), *stb* (stable), *prf* (preferred), *sem* (semi-stable), and *stg* (stage).

Definition 2 (AF-semantics)**.** *Let $F = (A, R)$ be an AF. A set $S \subseteq A$ is* conflict-free (in F)*, iff 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*

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

Let us provide a small example. AFs will be depicted as directed graphs, where the nodes are arguments and the edges are attacks between arguments.

Example 2. *Let $F = (A, R)$ be the AF depicted in Figure 1, ignoring claims α, β, and γ, i.e.,*

$$\begin{aligned} A &= \{a, b, c, d, e, f\}, \\ R &= \{(a, b), (b, a), (b, c), (c, f), (d, c), (d, e), (e, d), (f, e), (f, f)\}. \end{aligned}$$

Regarding conflict-free semantics, observe that, e.g., $\{a, b\} \notin cf(F)$ since $(a, b) \in R$. On the other hand, $\{a, c\} \in cf(F)$ since $(a, c) \notin R$, $(c, a) \notin R$, $(a, a) \notin R$, and $(c, c) \notin R$. Moreover, note that $\{f\} \notin cf(F)$ since the argument f is self-attacking, i.e., since $(f, f) \in R$.

$$\begin{aligned} cf(F) = \{&\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{e\}, \\ &\{a, c\}, \{a, d\}, \{a, e\}, \{b, d\}, \{b, e\}, \{c, e\}, \{a, c, e\}\}. \end{aligned}$$

Since naive extensions are the subset-maximal conflict-free sets, we have

$$naive(F) = \{\{a, d\}, \{b, d\}, \{b, e\}, \{a, c, e\}\}.$$

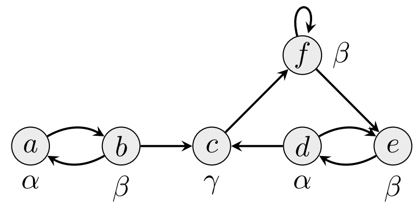


Figure 1: Example AF where each argument is associated with a claim α, β, or γ.

Note that there is only one conflict-free set S such that $S_F^{\oplus} = A$, namely $S = \{a, c, e\}$. Thus, $\{a, c, e\}_F^{\oplus} \supset T_F^{\oplus}$ for all $T \in cf(F)$ such that $T \neq \{a, c, e\}$. Therefore,

$$stg(F) = stb(F) = \{\{a, c, e\}\}.$$

Regarding admissible semantics we have, e.g., $\{c\} \notin adm(F)$ since c does not defend itself against the attacks from b and d. However, $\{a, c, e\} \in adm(F)$ since a defends c and since c defends e. Overall we have

$$adm(F) = \{\emptyset, \{a\}, \{b\}, \{d\}, \{a, d\}, \{b, d\}, \{a, c, e\}\}.$$

The preferred semantics are the subset-maximal admissible sets, i.e.,

$$prf(F) = \{\{a, d\}, \{b, d\}, \{a, c, e\}\}.$$

Analogously to conflict-free sets, there is only one admissible set S such that $S_F^{\oplus} = A$, namely $S = \{a, c, e\}$. Thus,

$$sem(F) = stb(F) = \{\{a, c, e\}\}.$$

As for complete semantics, we have $com(F) = adm(F)$ since no admissible set defends an argument outside of the set. Thus,

$$com(F) = \{\emptyset, \{a\}, \{b\}, \{d\}, \{a, d\}, \{b, d\}, \{a, c, e\}\}.$$

Lastly, the subset-minimal complete extensions is $\emptyset$, i.e.,

$$grd(F) = \{\emptyset\}.$$

2.2 Preference-based Argumentation Frameworks (PAFs)

Preference-based AFs generalize standard Dung-style AFs by introducing preferences between arguments (Kaci et al., 2021).

Definition 3 (PAF)**.** *A Preference-based AF (PAF) is a triple $P = (A, R, \succ)$ where (A, R) is an AF and $\succ$ is an asymmetric preference relation over A.*

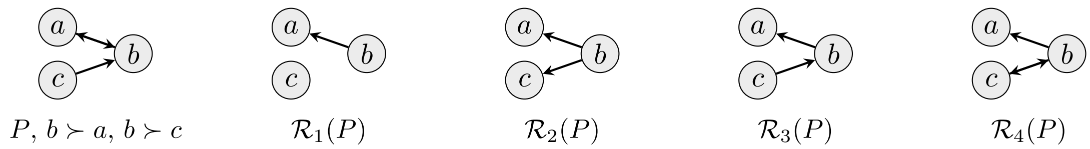


Figure 2: Effect of the four preference reductions on the attack relation.

Notice that preferences in PAFs are not required to be transitive. While transitivity of preferences is often assumed in argumentation (Amgoud & Vesic, 2014; Kaci et al., 2018), it cannot always be guaranteed in practice (Kaci et al., 2021). In this paper, 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 (and that b is weaker than a). But what effect should this ordering have? How should this influence, e.g., the admissible extensions of the framework? One possibility is to remove all attacks from weaker to stronger arguments, and to then determine the admissible extensions in the resulting AF. This altering of attacks in a PAF based on its preference-ordering is called a reduction. The literature describes four such reductions for regular AFs (Kaci et al., 2021), which we now adapt.

Definition 4 (Preference reduction)**.** *Given a PAF $P = (A, R, \succ)$, the corresponding AF $\mathcal{R}_i(P) = (A, R')$ 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 2 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: if a weak argument attacks a stronger argument, and there is no reverse attack, add a reverse attack but do not remove the attack from the weak to the strong argument; if a weak argument attacks a stronger argument, but there is a reverse attack, remove the attack from the weaker argument.

The semantics for PAFs are defined in a straightforward way: first, one of the four reductions is applied to the given PAF; then, AF-semantics are applied to the resulting AF.

Definition 5 (PAF-semantics)**.** *Let P be a PAF and let $i \in \{1, 2, 3, 4\}$. The preference-based variant of an AF-semantics σ relative to Reduction i is defined as $\sigma^i(P) = \sigma(\mathcal{R}_i(P))$.*

Example 3. *Consider the PAF $P = (\{a, b, c\}, \{(a, b), (b, a), (c, b)\}, \succ)$ with $b \succ a$ and $b \succ c$ depicted in Figure 2. The AFs resulting from applying the various preference reductions, i.e., $\mathcal{R}_1(P)$, $\mathcal{R}_2(P)$, $\mathcal{R}_3(P)$, and $\mathcal{R}_4(P)$, are also depicted in Figure 2.*

For Reduction 1 we have $adm^1(P) = adm(\mathcal{R}_1(P)) = \{\emptyset, \{b\}, \{c\}, \{b, c\}\}$ *and therefore* $prf^1(P) = prf(\mathcal{R}_1(P)) = \{\{b, c\}\}$. *If we use Reduction 2 we get* $adm^2(P) = adm(\mathcal{R}_2(P)) = \{\emptyset, \{b\}\}$ *and therefore* $prf^2(P) = prf(\mathcal{R}_2(P)) = \{\{b\}\}$.

2.3 Claim-augmented Argumentation Frameworks (CAFs)

CAFs generalize standard AFs by assigning a claim to each argument (Dvořák & Woltran, 2020). The notion of enriching arguments with claims/conclusions appears often and under various names in the literature. For instance, Conclusion-based AF (Rocha & Cozman, 2022b, 2022a) are equivalent to CAFs as we consider them, while Argument-Conclusion Structures (Baroni, Governatori, & Riveret, 2016) are not technically equivalent but strongly related to CAFs.

Definition 6 (CAF)**.** *A Claim-augmented AF (CAF) is a triple* $\mathcal{F} = (A, R, cl)$ *where* (A, R) *is an AF and* $cl\colon A \to \mathcal{C}$ *is a function that maps arguments to an infinite domain of claims* $\mathcal{C}$. *The claim-function is extended to sets of arguments* $S \subseteq A$ *via* $cl(S) = \{cl(a) \mid a \in S\}$. *A well-formed CAF (wfCAF) is a CAF* (A, R, cl) *in which all arguments with the same claim attack the same arguments, i.e., for all* $a, b \in A$ *with* $cl(a) = cl(b)$ *we have that* $\{a\}^+_{(A,R)} = \{b\}^+_{(A,R)}$.

There are two types of semantics for CAFs, inherited and hybrid. Inherited semantics apply AF-semantics to the underlying AF of a given CAF, and then collect the claims of arguments contained in an extension.

Definition 7 (Inherited semantics)**.** *Let* $\mathcal{F} = (A, R, cl)$ *be a CAF. The inherited CAF-variant of an AF-semantics* σ *is defined as* $\sigma_{inh}(\mathcal{F}) = \{cl(S) \mid S \in \sigma((A, R))\}$.

Example 4. *Let* $\mathcal{F} = (A, R, cl)$ *be the CAF depicted in Figure 1, i.e.,*

$$\begin{aligned}
&A = \{a, b, c, d, e, f\},\\
&R = \{(a, b), (b, a), (b, c), (c, f), (d, c), (d, e), (e, d), (f, e), (f, f)\},\\
&cl(a) = cl(d) = \alpha,\\
&cl(b) = cl(e) = cl(f) = \beta,\\
&cl(c) = \gamma.
\end{aligned}$$

Note that $\mathcal{F}$ *is not well-formed, since, e.g.,* $(a, b) \in R$ *but* $(d, b) \notin R$ *despite* $cl(a) = cl(d)$.

The underlying AF (A, R) *of* $\mathcal{F}$ *is the AF we examined in Example 2. The extensions of* $\mathcal{F}$ *on the claim-level can be inferred from the extensions of* (A, R) *on the argument-level (see Example 2). Thus, we have*

$$\begin{aligned}
cf_{inh}(\mathcal{F}) &= \{\emptyset, \{\alpha\}, \{\beta\}, \{\gamma\}, \{\alpha, \gamma\}, \{\alpha, \beta\}, \{\beta, \gamma\}, \{\alpha, \beta, \gamma\}\},\\
naive_{inh}(\mathcal{F}) &= \{\{\alpha\}, \{\beta\}, \{\alpha, \beta\}, \{\alpha, \beta, \gamma\}\},\\
adm_{inh}(\mathcal{F}) = com_{inh}(\mathcal{F}) &= \{\{\emptyset, \{\alpha\}, \{\beta\}, \{\alpha, \beta\}, \{\alpha, \beta, \gamma\}\},\\
prf_{inh}(\mathcal{F}) &= \{\{\alpha\}, \{\alpha, \beta\}, \{\alpha, \beta, \gamma\}\},\\
stg_{inh}(\mathcal{F}) = sem_{inh}(\mathcal{F}) = stb_{inh} &= \{\{\alpha, \beta, \gamma\}\},\\
grd(\mathcal{F}) &= \{\emptyset\}.
\end{aligned}$$

Hybrid semantics (Rapberger, 2020; Dvořák et al., 2023) employ subset-maximization (such as in preferred semantics) on the claim-level rather than the argument level.

Definition 8 (Claim-defeat & claim-range)**.** *Let $\mathcal{F} = (A, R, cl)$ be a CAF. A set of arguments $S \subseteq A$* defeats *a claim $\alpha \in cl(A)$ in $\mathcal{F}$ iff S attacks every $a \in A$ with $cl(a) = \alpha$ (in $\mathcal{F}$). $S^*_{\mathcal{F}} = \{\alpha \in cl(A) \mid S$ defeats α in $\mathcal{F}\}$ denotes the set of all claims which are defeated by S in $\mathcal{F}$. The claim-range of a set $S \subseteq A$ of arguments is denoted by $S^{\circledast}_{\mathcal{F}} = cl(S) \cup S^*_{\mathcal{F}}$.*

Definition 9 (Hybrid semantics)**.** *Let $\mathcal{F} = (A, R, cl)$ be a CAF with underlying AF $F = (A, R)$. Consider a set of claims $C \subseteq cl(A)$. We call $S \subseteq A$ a σ_{inh}-realization of C in $\mathcal{F}$ iff $S \in \sigma(A, R)$ and $cl(S) = C$.*

- *$C \in prf_{hyb}(\mathcal{F})$ if C is $\subseteq$-maximal in $adm_{inh}(\mathcal{F})$;*
- *$C \in naive_{hyb}(\mathcal{F})$ if C is $\subseteq$-maximal in $cf_{inh}(\mathcal{F})$;*
- *$C \in stb\text{-}adm_{hyb}(\mathcal{F})$ if there is a adm_{inh}-realization S of C which defeats any $\alpha \in cl(A) \setminus C$ (i.e., $S^{\circledast}_{\mathcal{F}} = cl(A)$);*
- *$C \in stb\text{-}cf_{hyb}(\mathcal{F})$ if there is a cf_{inh}-realization S of C which defeats any $\alpha \in cl(A) \setminus C$ (i.e., $S^{\circledast}_{\mathcal{F}} = cl(A)$);*
- *$C \in sem_{hyb}(\mathcal{F})$ if there is an adm_{inh}-realization S of C in $\mathcal{F}$ such that there is no $T \in adm(F)$ with $S^{\circledast}_{\mathcal{F}} \subset T^{\circledast}_{\mathcal{F}}$;*
- *$C \in stg_{hyb}(\mathcal{F})$ if there is an cf_{inh}-realization S of C in $\mathcal{F}$ such that there is no $T \in cf(F)$ with $S^{\circledast}_{\mathcal{F}} \subset T^{\circledast}_{\mathcal{F}}$.*

In the remainder of the paper, we refer to an arbitrary CAF-semantics via σ_μ or τ_ν, i.e., $\sigma_\mu, \tau_\nu \in \{cf_{inh}, adm_{inh}, com_{inh}, grd_{inh}, naive_{inh}, naive_{hyb}, stb_{inh}, stb\text{-}adm_{hyb}, stb\text{-}cf_{hyb}, prf_{inh}, prf_{hyb}, sem_{inh}, sem_{hyb}, stg_{inh}, stg_{hyb}\}$.

Example 5. *Consider again the CAF $\mathcal{F} = (A, R, cl)$ depicted in Figure 1. Recall that we already investigated this CAF with regards to inherited semantics in Example 4. In contrast to inherited semantics, for hybrid naive and preferred semantics we have*

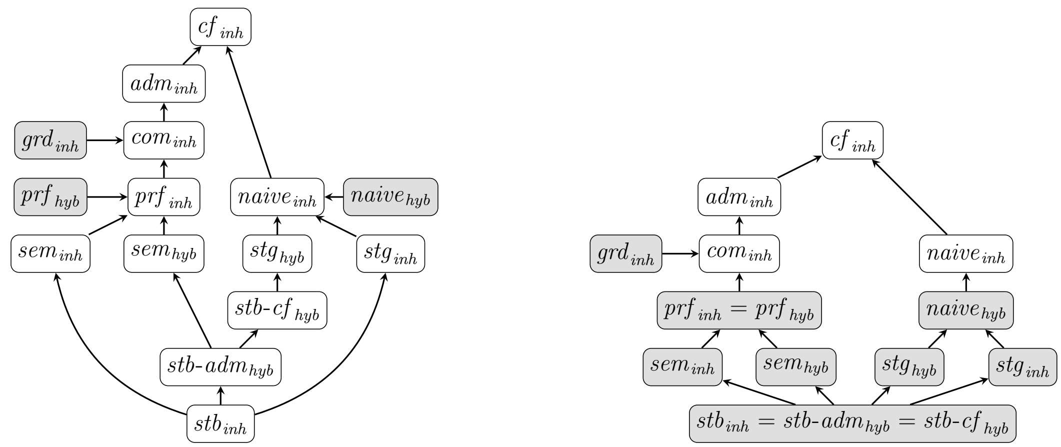
$$naive_{hyb}(\mathcal{F}) = prf_{hyb}(\mathcal{F}) = \{\{\alpha, \beta, \gamma\}\}.$$

Regarding claim-range, notice that the admissible argument-set $\{a, c, e\}$ already contains every claim in $\mathcal{F}$, i.e., $cl(\{a, c, e\}) = cl(A)$. Thus, $\{a, c, e\}^{\circledast}_{\mathcal{F}} = cl(A)$. For the admissible argument set $\{b, d\}$ we have $cl(\{b, d\}) = \{\alpha, \beta\}$ and $\{b, d\}^*_{\mathcal{F}} = \{\gamma\}$, i.e., $\{b, d\}^{\circledast}_{\mathcal{F}} = cl(A)$. There is no other admissible argument set $S \in adm(A, R)$ such that $S^{\circledast}_{\mathcal{F}} = cl(A)$. Thus,

$$sem_{hyb}(\mathcal{F}) = stb\text{-}adm_{hyb}(\mathcal{F}) = \{\{\alpha, \beta\}, \{\alpha, \beta, \gamma\}\}.$$

For the conflict-free (but not admissible) argument set $\{b, e\}$ we have $cl(\{b, e\}) = \{\beta\}$ and $\{b, e\}^*_{\mathcal{F}} = \{\alpha, \gamma\}$, i.e., $\{b, e\}^{\circledast}_{\mathcal{F}} = cl(A)$. There is no other conflict-free argument set $S \in cf((A, R))$ such that $S^{\circledast}_{\mathcal{F}} = cl(A)$. Thus,

$$stg_{hyb}(\mathcal{F}) = stb\text{-}cf_{hyb}(\mathcal{F}) = \{\{\beta\}, \{\alpha, \beta\}, \{\alpha, \beta, \gamma\}\}.$$



(a) General CAFs. (b) Well-formed CAFs.

Figure 3: Relations between semantics on (well-formed) CAFs. If there is an arrow from σ_μ to τ_ν, then $\sigma_\mu(\mathcal{F}) \subseteq \tau_\nu(\mathcal{F})$ for all CAFs $\mathcal{F}$ of the respective CAF-class. Semantics highlighted in gray are I-maximal.

The relationship between the various CAF-semantics has been investigated for both general and well-formed CAFs (Dvořák et al., 2023). See Figure 3 for a summary of these results. It can be seen that many inherited and hybrid semantics coincide on wfCAFs, but not on general CAFs.

Many argumentation semantics employ argument maximization (e.g. preferred or naive) and therefore deliver incomparable extensions on standard AFs: for all $S, T \in prf(F)$, $S \subseteq T$ implies $S = T$. This fundamental property is called I-maximality (Baroni & Giacomin, 2007), and is defined analogously for CAFs:

Definition 10 (I-maximality)**.** *A CAF-semantics σ_μ is I-maximal for a class $\mathfrak{F}$ of CAFs if, for all CAFs $\mathcal{F} \in \mathfrak{F}$ and all $C, D \in \sigma_\mu(\mathcal{F})$, $C \subseteq D$ implies $C = D$.*

Figure 3 shows I-maximality properties of semantics for CAFs (Dvořák et al., 2023), with the semantics highlighted in gray being I-maximal. I-maximality gives insights into the expressiveness of semantics both on the argument-level (Dunne et al., 2015) and on the claim-level (Dvořák et al., 2023), and therefore has been one of the first properties in claim-based argumentation to be investigated (Dvořák, Rapberger, & Woltran, 2020). For wfCAFs, I-maximality is preserved in all maximization-based semantics except $naive_{inh}$, implying natural behavior analogous to standard AFs. On the other hand, I-maximality is not preserved for general CAFs, revealing a fundamental difference regarding how the various semantics behave on wfCAFs versus on general CAFs.

The computational complexity of CAFs has been investigated before (Dvořák & Woltran, 2020; Dvořák et al., 2023), revealing more differences between general CAFs and wfCAFs. We assume familiarity with the complexity classes P, NP, and coNP. Moreover, Σ_2^{P} is the class of decision problems that can be decided in nondeterministic polynomial time by an algorithm with access to an NP-oracle (Arora & Barak, 2009). Π_2^{P} is the complementary class of Σ_2^{P}. DP is the class of languages that can be expressed as an intersection of a language in NP and a language in coNP.

Table 1: Computational Complexity of CAFs.

<table>
<tr><th rowspan="2">σ_μ</th><th>$Cred^{\Delta}_{\sigma_\mu}$</th><th colspan="2">$Skept^{\Delta}_{\sigma_\mu}$</th><th colspan="2">$Ver^{\Delta}_{\sigma_\mu}$</th></tr>
<tr><th>$\Delta \in \{CAF, wfCAF\}$</th><th>$\Delta = CAF$</th><th>$\Delta = wfCAF$</th><th>$\Delta = CAF$</th><th>$\Delta = wfCAF$</th></tr>
<tr><td>cf_{inh}</td><td>in P</td><td colspan="2">trivial</td><td>NP-c</td><td>in P</td></tr>
<tr><td>adm_{inh}</td><td>NP-c</td><td colspan="2">trivial</td><td>NP-c</td><td>in P</td></tr>
<tr><td>com_{inh}</td><td>NP-c</td><td colspan="2">P-c</td><td>NP-c</td><td>in P</td></tr>
<tr><td>grd_{inh}</td><td>in P</td><td colspan="2">in P</td><td colspan="2">in P</td></tr>
<tr><td>stb_{inh}
$stb\text{-}adm_{hyb}$
$stb\text{-}cf_{hyb}$</td><td>NP-c</td><td colspan="2">coNP-c</td><td>NP-c</td><td>in P</td></tr>
<tr><td>$naive_{inh}$</td><td rowspan="2">in P</td><td colspan="2">coNP-c</td><td>NP-c</td><td rowspan="2">in P</td></tr>
<tr><td>$naive_{hyb}$</td><td>Π_2^{P}-c</td><td>coNP-c</td><td>DP-c</td></tr>
<tr><td>prf_{inh}</td><td rowspan="2">NP-c</td><td colspan="2" rowspan="2">Π_2^{P}-c</td><td>Σ_2^{P}-c</td><td rowspan="2">coNP-c</td></tr>
<tr><td>prf_{hyb}</td><td>DP-c</td></tr>
<tr><td>sem_{inh}
sem_{hyb}</td><td>Σ_2^{P}-c</td><td colspan="2">Π_2^{P}-c</td><td>Σ_2^{P}-c</td><td>coNP-c</td></tr>
<tr><td>stg_{inh}
stg_{hyb}</td><td>Σ_2^{P}-c</td><td colspan="2">Π_2^{P}-c</td><td>Σ_2^{P}-c</td><td>coNP-c</td></tr>
</table>

Definition 11 (Decision problems for CAFs)**.** *We consider the following decision problems pertaining to a CAF-semantics σ_μ:*

- Credulous Acceptance *($Cred^{CAF}_{\sigma_\mu}$): Given a CAF $\mathcal{F}$ and claim α, is α contained in some $C \in \sigma_\mu(\mathcal{F})$?*
- Skeptical Acceptance *($Skept^{CAF}_{\sigma_\mu}$): Given a CAF $\mathcal{F}$ and claim α, is α contained in each $C \in \sigma_\mu(\mathcal{F})$?*
- Verification *($Ver^{CAF}_{\sigma_\mu}$): Given a CAF $\mathcal{F}$ and a set of claims C, is $C \in \sigma_\mu(\mathcal{F})$?*

We furthermore consider these reasoning problems restricted to wfCAFs and denote them by $Cred^{wfCAF}_{\sigma_\mu}$, $Skept^{wfCAF}_{\sigma_\mu}$, and $Ver^{wfCAF}_{\sigma_\mu}$.

Table 1 shows the complexity of these problems as established in (Dvořák et al., 2023).[1] The complexity of the verification problem drops by one level in the polynomial hierarchy when comparing general CAFs to wfCAFs (except for grd_{inh}). This is an important advantage of wfCAFs, as a lower complexity in the verification problem allows for a more efficient enumeration of claim-extensions (Dvořák & Woltran, 2020).

1. Note that the complexity of grounded semantics has not been investigated explicitly in (Dvořák et al., 2023), but it is easy to see that $Cred^{CAF}_{grd_{inh}}$, $Skept^{CAF}_{grd_{inh}}$, and $Ver^{CAF}_{grd_{inh}}$ are in P since the unique grounded argument-extension can be computed in polynomial time (Dvořák & Dunne, 2018).

3. Preference-based Claim-augmented AFs (PCAFs)

As discussed in the previous sections, wfCAFs are a natural subclass of CAFs with advantageous semantic and computational properties. 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 12 (PCAF)**.** *A Preference-based Claim-augmented AF (PCAF) is a quadruple* $\mathcal{P} = (A, R, cl, \succ)$ *where* (A, R, cl) *is a wfCAF and* $(A, R, \succ)$ *is a PAF.*

Preferences in PCAFs are resolved via one of the four preference reductions, analogously to how they are resolved in PAFs (cf. Definition 4). Observe that all four reductions are polynomial time computable with respect to the input PCAF.

Definition 13 (Preference reductions applied to PCAFs)**.** *Given a PCAF* $\mathcal{P} = (A, R, cl, \succ)$*, the corresponding CAF* $\mathcal{R}_i(\mathcal{P}) = (A, R', cl)$ *is obtained by applying Reduction* i*, where* $i \in \{1, 2, 3, 4\}$*, to the underlying PAF* $P = (A, R, \succ)$ *of* $\mathcal{P}$*, i.e.,* $(A, R') = \mathcal{R}_i(P)$*.*

The semantics of PCAFs work by first resolving preferences between arguments, and then applying CAF-semantics to the resulting CAF.

Definition 14 (PCAF-semantics)**.** *Let* $\mathcal{P}$ *be a PCAF and let* $i \in \{1, 2, 3, 4\}$*. The PCAF-variant of a CAF-semantics* σ_μ *relative to Reduction* i *is defined as* $\sigma_\mu^i(\mathcal{P}) = \sigma_\mu(\mathcal{R}_i(\mathcal{P}))$*.*

Note that many structured argumentation formalisms use preference reductions. For instance, ABA+ (Cyras & Toni, 2016) employs attack reversal similar to Reduction 2 while some instances of ASPIC (Modgil & Prakken, 2013) delete attacks from weaker arguments in the spirit of Reduction 1.

Example 6. *Let* $\mathcal{P} = (A, R, cl, \succ)$ *be the PCAF with arguments* $A = \{a, a', b\}$*, attacks* $R = \{(a, b), (a', b), (b, a)\}$*, claims* $cl(a) = cl(a') = \alpha$ *and* $cl(b) = \beta$*, and the preference* $b \succ a'$*. The underlying CAF* (A, R, cl) *of* $\mathcal{P}$ *is the same CAF as* $\mathcal{F}$ *in Example 1.*

Note that $\mathcal{R}_1(\mathcal{P}) = (A, R', cl)$ *with* $R' = \{(a, b), (b, a)\}$*, which is the same CAF as* $\mathcal{F}'$ *in Example 1. It can be verified that, e.g.,* $adm_{inh}^1(\mathcal{P}) = adm_{inh}(\mathcal{R}_1(\mathcal{P})) = \{\{\emptyset, \{\alpha\}, \{\beta\}, \{\alpha, \beta\}\}$ *and* $stb_{inh}^1(\mathcal{P}) = \{\{\alpha\}, \{\alpha, \beta\}\}$*.*

As in PAFs (cf. Example 3) the choice of reduction can influence the extensions of a PCAF. For example, $\mathcal{R}_2(\mathcal{P}) = (A, R'', cl)$ *with* $R'' = \{(a, b), (b, a), (b, a')\}$*,* $adm_{inh}^2(\mathcal{P}) = \{\emptyset, \{\alpha\}, \{\beta\}\}$*, and* $stb_{inh}^2(\mathcal{P}) = \{\{\alpha\}, \{\beta\}\}$*.*

Remark. *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. Syntactic 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 4 to PCAFs.

Definition 15 (CAF-classes)**.** $\mathcal{R}_i$-**CAF** *denotes the set of CAFs that can be obtained by applying Reduction i to PCAFs, i.e.,* $\mathcal{R}_i$-**CAF** $= \{\mathcal{R}_i(\mathcal{P}) \mid \mathcal{P}$ *is a PCAF*$\}$.

It is easy to see that $\mathcal{R}_i$-**CAF**, where $i \in \{1, 2, 3, 4\}$, contains all wfCAFs (we can simply specify the desired wfCAF and an empty preference relation). Moreover, not all CAFs are contained in $\mathcal{R}_i$-**CAF**, i.e., the four new classes are located in-between wfCAFs and general CAFs:

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

Proof. Let $i \in \{1, 2, 3, 4\}$. **wfCAF** $\subseteq \mathcal{R}_i$-**CAF** follows from the fact that any $(A, R, cl) \in$ **wfCAF** can be obtained via Reduction i from the PCAF $(A, R, cl, \emptyset)$.

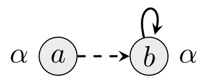
wfCAF $\subset \mathcal{R}_i$-**CAF**: consider the PCAF $\mathcal{P} = (\{a, b\}, \{(a, a), (a, b), (b, a), (b, b)\}, cl, \succ)$ with $cl(a) = cl(b)$, and $b \succ a$. For all $i \in \{1, 2, 3, 4\}$ we have $\mathcal{R}_i(\mathcal{P}) = (\{a, b\}, \{(a, a), (b, a), (b, b)\}, cl)$, i.e., the resulting CAF $\mathcal{R}_i(\mathcal{P})$ is not well-formed.

$\mathcal{R}_i$-**CAF** $\subset$ **CAF**: since **CAF** contains all CAFs, we have $\mathcal{R}_i$-**CAF** $\subseteq$ **CAF**. It remains to show that $\mathcal{R}_i$-**CAF** $\neq$ **CAF**. Towards a contradiction, assume there is a PCAF $\mathcal{P} = (A, R, cl, \succ)$ such that $\mathcal{R}_i(\mathcal{P}) = (A, R', cl)$ with $(a, b), (b, a) \in R'$ but $(a, a), (b, b) \notin R'$ for some $a, b \in A$ with $cl(a) = cl(b)$. This means that either $(a, b) \in R$ or $(b, a) \in R$, since none of four reductions can introduce the attacks (a, b) and (b, a) at the same time. By symmetry, we only look at the case that $(a, b) \in R$. Then, since (A, R, cl) is well-formed and since $cl(a) = cl(b)$, $(b, b) \in R$. But $\succ$ is non-reflexive, i.e., (b, b) is not removed by Reduction i and therefore $(b, b) \in R'$. Contradiction. □

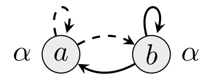
Furthermore, the new classes are all distinct from each other, i.e., we are indeed dealing with *four* new CAF classes. Specifically, $\mathcal{R}_1$-**CAF**, $\mathcal{R}_2$-**CAF**, and $\mathcal{R}_4$-**CAF** are incomparable while $\mathcal{R}_3$-**CAF** is strictly contained in the other three classes. This reflects the fact that Reduction 3 is the most conservative of the four preference reductions, removing attacks from weak to strong arguments only when there is a counter-attack from the strong argument.

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* $\mathcal{R}_i$-**CAF** $\not\subseteq \mathcal{R}_j$-**CAF** *and* $\mathcal{R}_3$-**CAF** $\subset \mathcal{R}_i$-**CAF**.

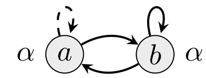
Proof. We show the various statements separately.



(a) only in $\mathcal{R}_1$-**CAF**



(b) only in $\mathcal{R}_2$-**CAF**



(c) only in $\mathcal{R}_4$-**CAF**

Figure 4: CAFs that are contained only in $\mathcal{R}_1$-**CAF**, $\mathcal{R}_2$-**CAF**, and $\mathcal{R}_4$-**CAF** respectively. Dashed arrows are attacks that are missing for the CAF to be well-formed.

- $\mathcal{R}_1$-**CAF** $\not\subseteq \mathcal{R}_j$-**CAF** with $j \in \{2,3,4\}$: let $\mathcal{F}$ be the CAF shown in Figure 4a. $\mathcal{F}$ is in $\mathcal{R}_1$-**CAF** as it can be obtained by applying Reduction 1 to the PCAF $(A, R, cl, \succ)$ with $A = \{a, b\}$, $R = \{(a, b), (b, b)\}$, $cl(a) = cl(b) = \alpha$, and $b \succ a$. Note that (a, b) is deleted and is therefore not in $\mathcal{R}_1(\mathcal{P}) = \mathcal{F}$. Towards a contradiction, assume there is a PCAF $\mathcal{P}$ such that $\mathcal{R}_j(\mathcal{P}) = \mathcal{F}$. Since self-attacks cannot be removed by any of the four reductions, $(b, b) \in \mathcal{P}$. Since the underlying CAF of $\mathcal{P}$ must be well-formed, also $(a, b) \in \mathcal{P}$. But then, by the definition of Reduction j, either $(a, b) \in \mathcal{R}_j(\mathcal{P})$ or $(b, a) \in \mathcal{R}_j(\mathcal{P})$. Contradiction.

- $\mathcal{R}_2$-**CAF** $\not\subseteq \mathcal{R}_j$-**CAF** with $j \in \{1,3,4\}$: let $\mathcal{F}$ be the CAF shown in Figure 4b. $\mathcal{F}$ is in $\mathcal{R}_2$-**CAF** as it can be obtained by applying Reduction 2 to the PCAF $(A, R, cl, \succ)$ with $A = \{a, b\}$, $R = \{(a, b), (b, b)\}$, $cl(a) = cl(b) = \alpha$, and $b \succ a$. Towards a contradiction, assume there is a PCAF $\mathcal{P}$ such that $\mathcal{R}_j(\mathcal{P}) = \mathcal{F}$. Then $(b, b) \in \mathcal{P}$ and therefore also $(a, b) \in \mathcal{P}$. But $(b, a) \notin \mathcal{P}$, since $(a, a) \notin \mathcal{F}$ and therefore also $(a, a) \notin \mathcal{P}$. But Reductions 1 and 3 cannot introduce (b, a) in this case, while Reduction 4 cannot introduce (b, a) without retaining (a, b).

- $\mathcal{R}_4$-**CAF** $\not\subseteq \mathcal{R}_j$-**CAF** with $j \in \{1,2,3\}$: let $\mathcal{F}$ be the CAF shown in Figure 4c. $\mathcal{F}$ is in $\mathcal{R}_4$-**CAF** as it can be obtained by applying Reduction 4 to the PCAF $(A, R, cl, \succ)$ with $A = \{a, b\}$, $R = \{(a, b), (b, b)\}$, $cl(a) = cl(b) = \alpha$, and $b \succ a$. Towards a contradiction, assume there is a PCAF $\mathcal{P}$ such that $\mathcal{R}_j(\mathcal{P}) = \mathcal{F}$. Then $(b, b) \in \mathcal{P}$ and therefore also $(a, b) \in \mathcal{P}$. But $(b, a) \notin \mathcal{P}$, since $(a, a) \notin \mathcal{P}$. But Reduction 1, 2 and 3 cannot introduce (b, a), at least not without deleting (a, b).

- $\mathcal{R}_3$-**CAF** $\subset \mathcal{R}_j$-**CAF** with $j \in \{1,2,4\}$: let $\mathcal{F}$ be any CAF in $\mathcal{R}_3$-**CAF**. Then there is a PCAF $\mathcal{P} = (A, R', cl, \succ)$ such that $\mathcal{R}_3(\mathcal{P}) = \mathcal{F}$. If $(a, b) \in \mathcal{P}$ and $(a, b) \in \mathcal{F}$ we can assume that $b \not\succ a$ without loss of generality. If $(a, b) \in \mathcal{P}$ but $(a, b) \notin \mathcal{F}$, then, by definition of Reduction 3, $(b, a) \in \mathcal{P}$ and $b \succ a$. In this case, Reduction j functions in the same way as Reduction 3 (cf. Definition 4 and Figure 2), i.e., $\mathcal{R}_j(\mathcal{P}) = \mathcal{F}$. This proves $\mathcal{R}_3$-**CAF** $\subseteq \mathcal{R}_j$-**CAF**. $\mathcal{R}_3$-**CAF** $\subset \mathcal{R}_j$-**CAF** follows from $\mathcal{R}_j$-**CAF** $\not\subseteq$ $\mathcal{R}_3$-**CAF**. □

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 $\mathcal{R}_2$-**CAF** and $\mathcal{R}_4$-**CAF** this is not straightforward, since Reductions 2 and 4 not only remove but also introduce attacks and therefore allow for several possibilities by which a particular CAF can be obtained. We tackle this problem by characterizing the new classes via the so-called wf-problematic part of a CAF.

Definition 16 (wf-problematic part)**.** *A pair of arguments (a, b) is wf-problematic in a CAF $\mathcal{F} = (A, R, cl)$ iff $a, b \in A$, $(a, b) \notin R$, and there is $a' \in A$ with $cl(a') = cl(a)$ and $(a', b) \in R$. The set $wfp(\mathcal{F}) = \{(a, b) \mid (a, b)$ is wf-problematic in $\mathcal{F}\}$ is called the wf-problematic part of $\mathcal{F}$.*

Intuitively, the wf-problematic part of a CAF $\mathcal{F}$ consists of those attacks that are missing for $\mathcal{F}$ to be well-formed (cf. Figure 4). Indeed, $\mathcal{F}$ is a wfCAF if and only if $wfp(\mathcal{F}) = \emptyset$. The four new classes can be characterized as follows:

Proposition 3. *Let $\mathcal{F} = (A, R, cl)$ be a CAF. Then*

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

Proof. Here we consider $\mathcal{R}_1$-**CAF**. The remaining cases can be found in Appendix A (Lemma 40 for $\mathcal{R}_2$-**CAF**, Lemma 41 for $\mathcal{R}_3$-**CAF**, and Lemma 42 for $\mathcal{R}_4$-**CAF**).

"$\Longrightarrow$": By contrapositive. Suppose there is $(a, b) \in wfp(\mathcal{F})$ such that $(b, a) \in wfp(\mathcal{F})$. Towards a contradiction, assume $\mathcal{F} \in \mathcal{R}_1$-**CAF**. Then there is a PCAF $\mathcal{P} = (A, R', cl, \succ)$ such that $\mathcal{R}_1(\mathcal{P}) = \mathcal{F}$. Since Reduction 1 can only delete but not introduce attacks, and since the underlying CAF of $\mathcal{P}$ must be well-formed, $(a, b) \in R'$ and $(b, a) \in R'$. However, then also $(b \succ a)$ and $(a \succ b)$ which means that $\mathcal{P}$ is not asymmetric. Contradiction.

"$\Longleftarrow$": Suppose that $(a, b) \in wfp(\mathcal{F})$ implies $(b, a) \notin wfp(\mathcal{F})$. Then $\mathcal{R}_1(\mathcal{P}) = \mathcal{F}$ for the PCAF $\mathcal{P} = (A, R', cl, \succ)$ with $R' = R \cup \{(a, b) \mid (a, b) \in wfp(\mathcal{F})\}$ as well as $a \succ b$ iff $(b, a) \in R' \setminus R$. The underlying CAF of $\mathcal{P}$ is well-formed since $wfp((A, R', cl)) = \emptyset$. Furthermore, $\succ$ is asymmetric since $(a, b) \in wfp(\mathcal{F})$ implies $(b, a) \notin wfp(\mathcal{F})$ and by construction of $\mathcal{P}$. □

The above characterizations give us some insights into the effect of the various reductions on wfCAFs. Indeed, the similarity between the characterizations of $\mathcal{R}_1$-**CAF** and $\mathcal{R}_3$-**CAF**, resp. $\mathcal{R}_2$-**CAF** and $\mathcal{R}_4$-**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. Proposition 3 allows us to decide in polynomial time whether a given CAF $\mathcal{F}$ can be obtained by applying one of the four preference reductions to a PCAF. Moreover, in the proof of Proposition 3 we see how, given $\mathcal{F} \in \mathcal{R}_i$-**CAF**, we can construct a PCAF $\mathcal{P}$ such that $\mathcal{R}_i(\mathcal{P}) = \mathcal{F}$ in polynomial time.

But what happens if we restrict ourselves to transitive preferences? Analogously to $\mathcal{R}_i$-**CAF** (cf. Definition 15), 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 $\mathcal{R}_i$-$\mathbf{CAF}_{tr} \subseteq \mathcal{R}_i$-**CAF** for all $i \in \{1, 2, 3, 4\}$. Moreover, in the proof of Proposition 1 we actually made use

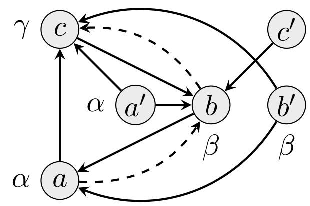


Figure 5: A CAF which shows that $\mathcal{R}_3$-$\mathbf{CAF}_{tr} \not\subseteq \mathcal{R}_j$-$\mathbf{CAF}_{tr}$ for $j \in \{1, 2, 4\}$. Dashed arrows are edges in the wf-problematic part.

of transitive preferences, i.e., $\mathbf{wfCAF} \subset \mathcal{R}_i$-$\mathbf{CAF}_{tr}$ for all $i \in \{1, 2, 3, 4\}$. Interestingly, however, 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. The reason for this is that, in certain PCAFs $\mathcal{P}$, transitivity can force $a_1 \succ a_n$ via $a_1 \succ a_2 \succ \ldots \succ a_n$ such that $(a_n, a_1) \in \mathcal{P}$ but $(a_1, a_n) \notin \mathcal{P}$. In this case, only Reduction 3 leaves the attacks between a_1 and a_n unchanged.

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

Proof. Note that the preference relations of the PCAFs used in the proof of Proposition 2 are transitive. We therefore have $\mathcal{R}_i$-$\mathbf{CAF}_{tr} \not\subseteq \mathcal{R}_j$-$\mathbf{CAF}_{tr}$ for every $i \in \{1, 2, 4\}$ and $j \in \{1, 2, 3, 4\}$ such that $i \neq j$.

It remains to show $\mathcal{R}_3$-$\mathbf{CAF}_{tr} \not\subseteq \mathcal{R}_j$-$\mathbf{CAF}_{tr}$ for $j \in \{1, 2, 4\}$. Let $\mathcal{F}$ be the CAF shown in Figure 5. $\mathcal{F}$ is in $\mathcal{R}_3$-$\mathbf{CAF}_{tr}$: to see this, let $\mathcal{P}$ be the PCAF with the same arguments and attacks as $\mathcal{F}$, and additionally attacks (a, b) and (b, c); Moreover, let $c \succ b$, $b \succ a$, and $c \succ a$; the attack (a, c) is not deleted by Reduction 3 if there is no attack (c, a); Thus, $\mathcal{R}_3(\mathcal{P}) = \mathcal{F}$. We show that $\mathcal{F} \notin \mathcal{R}_j$-$\mathbf{CAF}_{tr}$ for $j \in \{1, 2, 4\}$.

- $\mathcal{F}$ is not in $\mathcal{R}_1$-$\mathbf{CAF}_{tr}$ since a PCAF that reduces to $\mathcal{F}$ would need to have $c \succ b$, $b \succ a$, and therefore also $c \succ a$. But Reduction 1 would delete the attack (a, c).
- Towards a contradiction, assume there is a PCAF $\mathcal{P}$ such that $\mathcal{R}_2(\mathcal{P}) = \mathcal{F}$. First, we show that $(a, b) \in \mathcal{P}$, $(b, a) \in \mathcal{P}$, $(b, c) \in \mathcal{P}$, and $(c, b) \in \mathcal{P}$.
  - Assume $(a, b) \notin \mathcal{P}$. Then two things most hold. Firstly, it must be that $(b, a) \in \mathcal{P}$, otherwise $(b, a) \notin \mathcal{F}$. Secondly, $(a', b) \notin \mathcal{P}$, otherwise the underlying CAF of $\mathcal{P}$ would not be well-formed. This means that (a', b) must have been introduced into $\mathcal{F}$ by applying Reduction 2, i.e., by reversing (b, a'). Therefore, $(b, a') \in \mathcal{P}$. But then also $(b', a') \in \mathcal{P}$, otherwise the underlying CAF of $\mathcal{P}$ is not well-formed. But then, by the definition of Reduction 2, either $(b', a') \in \mathcal{F}$ or $(a', b') \in \mathcal{F}$, which is not the case. Contradiction.
  - Assume $(b, a) \notin \mathcal{P}$. Then, since the underlying CAF of $\mathcal{P}$ must be well-formed, $(b', a) \notin \mathcal{P}$. This means $(a, b') \in \mathcal{P}$, otherwise we cannot obtain $\mathcal{F}$ from $\mathcal{P}$ via Reduction 2. This means that $(a', b') \in \mathcal{P}$, which is not possible since neither $(a', b') \in \mathcal{F}$ nor $(b', a') \in \mathcal{P}$.

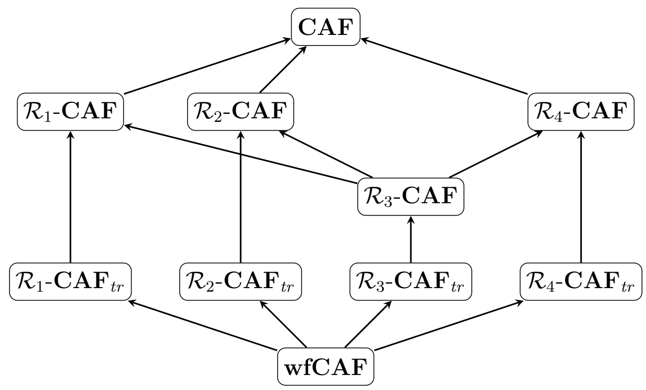


Figure 6: Relations between the various CAF-classes. An arrow indicates that a class is a strict subset of the other, e.g. $\mathcal{R}_3$-**CAF** $\subset$ $\mathcal{R}_4$-**CAF**.

- Assume $(b, c) \notin \mathcal{P}$. Then two things most hold. Firstly, it must be that $(c, b) \in \mathcal{P}$, otherwise $(c, b) \notin \mathcal{F}$. Secondly, $(b', c) \notin \mathcal{P}$, otherwise the underlying CAF of $\mathcal{P}$ would not be well-formed. This means that (b', c) must have been introduced into $\mathcal{F}$ by applying Reduction 2, i.e., by reversing (c, b'). Therefore, $(c, b') \in \mathcal{P}$. But then also $(c', b') \in \mathcal{P}$, otherwise the underlying CAF of $\mathcal{P}$ is not well-formed. But then, by the definition of Reduction 2, either $(c', b') \in \mathcal{F}$ or $(b', c') \in \mathcal{F}$, which is not the case. Contradiction.
- Assume $(c, b) \notin \mathcal{P}$. Then, since the underlying CAF of $\mathcal{P}$ must be well-formed, $(c', b) \notin \mathcal{P}$. This means $(b, c') \in \mathcal{P}$, otherwise we cannot obtain $\mathcal{F}$ from $\mathcal{P}$ via Reduction 2. This means that $(b', c') \in \mathcal{P}$, which is not possible since neither $(b', c') \in \mathcal{F}$ nor $(c', b') \in \mathcal{P}$.

Since $(a, b) \in \mathcal{P}$, $(b, a) \in \mathcal{P}$, $(b, c) \in \mathcal{P}$, and $(c, b) \in \mathcal{P}$, the only way to obtain $\mathcal{F}$ form $\mathcal{P}$ via Reduction 2 is to set $c \succ b$ and $b \succ a$. But then $c \succ a$ which means that $(a, c) \notin \mathcal{F}$. Contradiction, i.e., $\mathcal{F} \notin \mathcal{R}_2\text{-}\mathbf{CAF}_{tr}$.

- Now assume there is a PCAF $\mathcal{P}'$ such that $\mathcal{R}_4(\mathcal{P}') = \mathcal{F}$. It must be that $(a, b) \in \mathcal{P}'$ since we cannot obtain $(a', b) \in \mathcal{R}_4(\mathcal{P}')$ and $(b, a') \notin \mathcal{R}_4(\mathcal{P}')$ without $(a', b) \in \mathcal{P}'$. Analogously, it must be that $(b, c) \in \mathcal{P}'$. Then in order to have $\mathcal{R}_4(\mathcal{P}') = \mathcal{F}$ we need to set $c \succ b$ and $b \succ a$. But then $c \succ a$ which means that it cannot be that $(a, c) \in \mathcal{R}_4(\mathcal{P}')$ and $(c, a) \notin \mathcal{R}_4(\mathcal{P}')$. Contradiction, i.e., $\mathcal{F} \notin \mathcal{R}_4\text{-}\mathbf{CAF}_{tr}$. □

The above result also implies that $\mathcal{R}_i\text{-}\mathbf{CAF}_{tr} \subset \mathcal{R}_i\text{-}\mathbf{CAF}$ for $i \in \{1, 2, 4\}$ since we have $\mathcal{R}_3\text{-}\mathbf{CAF}_{tr} \subseteq \mathcal{R}_3\text{-}\mathbf{CAF} \subset \mathcal{R}_i\text{-}\mathbf{CAF}$ (cf. Proposition 2) and $\mathcal{R}_3\text{-}\mathbf{CAF}_{tr} \not\subset \mathcal{R}_i\text{-}\mathbf{CAF}_{tr}$ (cf. Proposition 4), which implies $\mathcal{R}_i\text{-}\mathbf{CAF}_{tr} \neq \mathcal{R}_i\text{-}\mathbf{CAF}$. It is also easy to see that $\mathcal{R}_3\text{-}\mathbf{CAF}_{tr} \subset \mathcal{R}_3\text{-}\mathbf{CAF}$: take the CAF from Figure 5 and add the additional attack (c, a). The resulting CAF is in $\mathcal{R}_3$-**CAF** since we do not need to set the preference $c \succ a$, whereas it is not in $\mathcal{R}_3\text{-}\mathbf{CAF}_{tr}$ since $c \succ a$ is enforced by $c \succ b \succ a$. Figure 6 summarizes the relationship between the CAF-classes.

We will not characterize all four classes $\mathcal{R}_i\text{-}\mathbf{CAF}_{tr}$ for transitive preferences. Indeed, while each $\mathcal{R}_i$-**CAF** and $\mathcal{R}_i\text{-}\mathbf{CAF}_{tr}$ are distinct syntactically, we will show that their semantic properties (cf. Section 5) and their computational complexity (cf. Section 6) are the same. However, we will characterize $\mathcal{R}_1$-$\mathbf{CAF}_{tr}$ as this will prove useful when analyzing the computational complexity of PCAFs using Reduction 1. Note that $wfp(\mathcal{F})$ can be seen as a directed graph, with an edge between vertices a and b whenever $(a, b) \in wfp(\mathcal{F})$. Thus, we may use notions such as paths and cycles in the wf-problematic part of a CAF.

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

Proof. Let $\mathcal{F} = (A, R, cl)$.

- Suppose $wfp(\mathcal{F})$ is acyclic and there is no $(a, b) \in \mathcal{F}$ with a path from a to b in $wfp(\mathcal{F})$. Construct the PCAF $\mathcal{P} = (A, R', cl, \succ)$ with $R' = R \cup \{(a, b) \mid (a, b) \in wfp(\mathcal{F})\}$ and $b \succ a$ iff there is a path from a to b in $wfp(\mathcal{F})$. (A, R', cl) is well-formed by construction. $\succ$ is transitive because if there is a path from a to b and from b to c, then there is also a path from a to c. $\succ$ is asymmetric because otherwise there would be a path from a to b and from b to a, which again would mean that there is a cycle. It remains to show that $\mathcal{R}_1(\mathcal{P}) = \mathcal{F}$. Let (a, b) be any attack in $\mathcal{P}$. We distinguish two cases:

  - $(a, b) \in \mathcal{F}$. Then, since there is no path from a to b in $wfp(\mathcal{F})$, $b \not\succ a$. Therefore, $(a, b) \in \mathcal{R}_1(\mathcal{P})$.
  - $(a, b) \notin \mathcal{F}$. Then, by construction, $(a, b) \in wfp(\mathcal{F})$ and therefore $b \succ a$. Thus, (a, b) is removed from $\mathcal{P}$ by Reduction 1, i.e., $(a, b) \notin \mathcal{R}_1(\mathcal{P})$.

  Note also that, by construction of $\mathcal{P}$, there can be no $(a, b) \in \mathcal{F}$ such that $(a, b) \notin \mathcal{P}$.

- Suppose $wfp(\mathcal{F})$ is cyclic. Then there are arguments $x_1, \ldots, x_n \in \mathcal{F}$ such that $x_1 = x_n$ and $(x_i, x_{i+1}) \in wfp(\mathcal{F})$ for all $1 \leq i < n$. Towards a contradiction, assume there is a PCAF $\mathcal{P} = (A, R', cl, \succ)$ such that $\mathcal{R}_1(\mathcal{P}) = \mathcal{F}$. Then $(x_i, x_{i+1}) \in \mathcal{P}$ for all $1 \leq i < n$, otherwise (A, R', cl) would not be well-formed. In order to have $\mathcal{R}_1(\mathcal{P}) = \mathcal{F}$ we must have $x_{i+1} \succ x_i$ for all $1 \leq i < n$. But then, by transitivity and since $x_1 = x_n$ we obtain $x_1 \succ x_1$, which is in contradiction to $\succ$ being asymmetric.

  On the other hand, suppose there is an attack $(a, b) \in \mathcal{F}$ with a path from a to b in $wfp(\mathcal{F})$. Let us denote this path as $x_1, \ldots, x_n$ with $x_1 = a$ and $x_n = b$. By the same argument as above, if there were a PCAF $\mathcal{P} = (A, R', cl, \succ)$ such that $\mathcal{R}_1(\mathcal{P}) = \mathcal{F}$, then $x_n \succ x_1$, i.e., $b \succ a$. But then $(a, b) \notin \mathcal{R}_1(\mathcal{P})$. Contradiction. □

In summary, we have shown that the four new CAF-classes that result from applying preferences to wfCAFs lie strictly in between wfCAFs and general CAFs (see Proposition 1) and that they are distinct from each other (see Propositions 2 and 4). Figure 6 summarizes the relationship between the CAF-classes. Furthermore, we characterize the four classes (see Proposition 3), which allows us to take any CAF, and, in polynomial time, decide whether this CAF belongs to one of the four classes.

From a high-level point of view, these 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) construct 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. Semantic Properties

There are key differences between wfCAFs and general CAFs with respect to semantic properties. It has been shown (Dvořák et al., 2023) that inherited and hybrid variants of stable and preferred semantics coincide on wfCAFs but not on general CAFs (cf. Figure 3). This simplifies the choice of semantics when working with wfCAFs. Moreover, wfCAFs, unlike general CAFs, preserve I-maximality under most maximization-based semantics (cf. Figure 3). This leads to more intuitive behavior of these semantics when considering extensions on the claim-level. As we have seen in Section 4, resolving preferences on wfCAFs results in four new CAF-classes that, from a syntactic perspective, lie in between wfCAFs and general CAFs. We now investigate whether these new CAF-classes retain the benefits of wfCAFs when it comes to semantic properties. We summarize and discuss our results at the end of this section (cf. Theorem 18 and Figure 9).

Firstly, we observe that the basic relations between semantics carry over from general CAFs, i.e., if we have $\sigma_\mu(\mathcal{F}) \subseteq \tau_\nu(\mathcal{F})$ for two CAF-semantics σ_μ, τ_ν and all CAFs $\mathcal{F}$, then we also have also $\sigma^i_\mu(\mathcal{P}) \subseteq \tau^i_\nu(\mathcal{P})$ for all PCAFs $\mathcal{P}$. Likewise, if we have $\sigma_\mu(\mathcal{F}) \not\subseteq \tau_\nu(\mathcal{F})$ for a wfCAF, then we also have $\sigma^i_\mu(\mathcal{P}) \not\subseteq \tau^i_\nu(\mathcal{P})$ for a PCAF.

Secondly, we note that Reductions 2, 3, and 4 cannot entirely remove conflicts between arguments, and that therefore the resolution of preferences has no impact on conflict-free extensions (both on the argument- and claim-level) under these preference reductions.

Lemma 6. *Let $\mathcal{P} = (A, R, cl, \succ)$ be a PCAF and let $\mathcal{R}_i(\mathcal{P}) = (A, R', cl)$ with $i \in \{2, 3, 4\}$. Then $cf((A, R)) = cf((A, R'))$ and $cf_{inh}((A, R, cl)) = cf_{inh}((A, R', cl))$.*

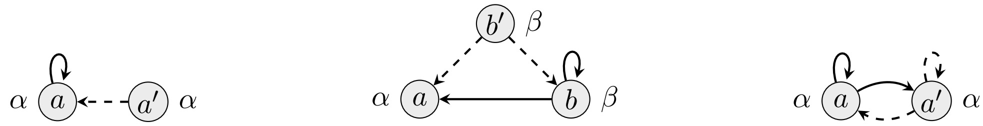
Proof. Let $\mathcal{P} = (A, R, cl, \succ)$ be a PCAF and let $\mathcal{R}_i(\mathcal{P}) = (A, R', cl)$ for $i \in \{2, 3, 4\}$. By definition of Reduction i, if $(a, b) \in R$ then either $(a, b) \in R'$ or $(b, a) \in R'$. Conversely, if $(a, b) \in R'$, then it must be that either $(a, b) \in R$ or $(b, a) \in R$. Thus, for any $S \subseteq A$ we have $S \in cf((A, R))$ iff $S \in cf((A, R'))$. This further implies that for any $C \subseteq cl(A)$ we have $C \in cf_{inh}((A, R, cl))$ iff $C \in cf_{inh}((A, R', cl))$. □

The fact that Reductions 2–4 do not remove conflicts, and the well-formedness of a PCAF's underlying CAF, allow us to show that inherited stable semantics and hybrid admissibility-based stable semantics coincide under Reductions 2–4. Under Reduction 1 the two semantics do not coincide.

Proposition 7. *$stb^i_{inh}(\mathcal{P}) = stb\text{-}adm^i_{hyb}(\mathcal{P})$, where $i \in \{2, 3, 4\}$, holds for every PCAF $\mathcal{P}$.*

Proof. $stb_{inh}(\mathcal{F}) \subseteq stb\text{-}adm_{hyb}(\mathcal{F})$ holds for all CAFs. We must show that $stb\text{-}adm_{hyb}(\mathcal{F}) \subseteq stb_{inh}(\mathcal{F})$ for all $\mathcal{F} \in \mathcal{R}_i$-**CAF**, where $i \in \{2, 3, 4\}$. Let $\mathcal{F} = (A, R, cl) \in \mathcal{R}_i$-**CAF**, and let $\mathcal{P} = (A, R', cl, \succ)$ be a PCAF such that $\mathcal{R}_i(\mathcal{P}) = \mathcal{F}$. Moreover, let $C \in stb\text{-}adm_{hyb}(\mathcal{F})$. Then there is an argument-set $S \subseteq A$ such that $S \in adm(\mathcal{F})$, $cl(S) = C$, and $C \cup S^*_{\mathcal{F}} = cl(A)$. Let $S' = S \cup \{x \in A \setminus S \mid (x, y) \notin R, (y, x) \notin R \text{ for all } y \in S\}$, i.e., S' is obtained by adding all arguments to S that are not in conflict with S. We show that $C \in stb_{inh}(\mathcal{F})$ by showing the following:

1. $cl(S') = C$: clearly, $cl(S) \subseteq cl(S')$. Now consider any $x \in S' \setminus S$. Then S does not defeat $cl(x)$, since there is no conflict between S and x. Since $cl(S) \cup S^*_{\mathcal{F}} = cl(A)$ there must be $x' \in S$ with $cl(x') = cl(x)$. Thus, $cl(S) \supseteq cl(S')$.



(a) $\mathcal{R}_1(\mathcal{P})$ from Proposition 8. (b) $\mathcal{R}_1(\mathcal{P})$ from Proposition 10. (c) $\mathcal{R}_2(\mathcal{P})$ from Proposition 11.

Figure 7: CAFs used to show that some variants of stable semantics do not coincide under Reductions 1 and 2. Dashed arrows are attacks in the wf-problematic part of the CAF.

2. $S' \in cf(A, R)$: since $S \in adm(A, R)$, there is no conflict between any two arguments in S. Moreover, by construction, there is no conflict between arguments in S and arguments in $S' \setminus S$. It remains to show there is no conflict between any two arguments in $S' \setminus S$. Towards a contradiction, assume there are $x', y' \in S' \setminus S$ such that $(x', y') \in R$. Since there is no conflict between S and x' (resp. y'), S does not defeat $cl(x')$ (resp. $cl(y')$). Since $cl(S) \cup S^*_{\mathcal{F}} = cl(A)$, there must be $x, y \in S$ with $cl(x) = cl(x')$ and $cl(y) = cl(y')$. Since Reductions 2,3,4 cannot remove conflicts, we have $(x', y') \in R'$ or $(y', x') \in R'$ in the original PCAF $\mathcal{P}$. By the well-formedness of $\mathcal{P}$, we have $(x, y') \in R'$ or $(y, x') \in R'$. Since Reductions 2,3,4 cannot remove conflicts, if $(x, y') \in R'$ then $(x, y') \in R$ or $(y', x) \in R$, and if $(y, x') \in R'$ then $(y, x') \in R$ or $(x', y) \in R$. But then either $x' \notin S'$ or $y' \notin S'$. Contradiction.

3. for all $z \in A \setminus S'$ there is $x \in S'$ such that $(x, z) \in R$: let $z \in A \setminus S'$. Then there must be $x \in S$ such that either $(x, z) \in R$ or $(z, x) \in R$, otherwise we would have $z \in S'$. If $(z, x) \in R$ but $(x, z) \notin R$, there must be $y \in S$ such that $(y, z) \in R$, otherwise we would have $S \notin adm(\mathcal{F})$. □

Proposition 8. *There is a PCAF $\mathcal{P}$ such that $stb^1_{inh}(\mathcal{P}) \neq stb\text{-}adm^1_{hyb}(\mathcal{P})$.*

Proof. Let $\mathcal{P} = (A, R, cl, \succ)$ with $A = \{a, a'\}$, $R = \{(a, a), (a', a)\}$, $cl(a) = cl(a') = \alpha$, and $a \succ a'$. Figure 7a depicts $\mathcal{R}_1(\mathcal{P}) = (A, R', cl)$, i.e., $R' = \{(a, a)\}$. Note that $stb_{inh}(\mathcal{R}_1(\mathcal{P})) = \emptyset$ while $stb\text{-}adm_{hyb}(\mathcal{R}_1(\mathcal{P})) = stb\text{-}cf_{hyb}(\mathcal{R}_1(\mathcal{P})) = \{\{\alpha\}\}$. □

Similarly, we can show that both variants (conflict-free and admissibility-based) of stable semantics coincide under Reductions 3 and 4, but not under Reductions 1 and 2.

Proposition 9. *$stb\text{-}adm^i_{hyb}(\mathcal{P}) = stb\text{-}cf^i_{hyb}(\mathcal{P})$, where $i \in \{3, 4\}$, holds for every PCAF $\mathcal{P}$.*

Proof. $stb\text{-}adm_{hyb}(\mathcal{F}) \subseteq stb\text{-}cf_{hyb}(\mathcal{F})$ holds for all CAFs. We show that $stb\text{-}cf_{hyb}(\mathcal{F}) \subseteq stb\text{-}adm_{hyb}(\mathcal{F})$ for $\mathcal{F} \in \mathcal{R}_i$-**CAF**, where $i \in \{3, 4\}$. Let $\mathcal{F} = (A, R, cl) \in \mathcal{R}_i$-**CAF**, and let $\mathcal{P} = (A, R', cl, \succ)$ be a PCAF such that $\mathcal{R}_i(\mathcal{P}) = \mathcal{F}$. Moreover, let $C \in stb\text{-}cf_{hyb}(\mathcal{F})$. Then there is an argument-set $S \subseteq A$ such that $S \in cf(A, R)$, $cl(S) = C$, and $C \cup S^*_{\mathcal{F}} = cl(A)$. We show that $C \in stb\text{-}adm_{hyb}(\mathcal{F})$ by showing that $S \in adm(\mathcal{F})$:

Consider any $x \in S$ and $y \in A \setminus S$ such that $(y, x) \in R$ but $(x, y) \notin R$. Under Reductions 3 and 4 a non-symmetric attack (y, x) in $\mathcal{R}_3(\mathcal{P})$ means that (y, x) was also present in the original PCAF $\mathcal{P}$, i.e., $(y, x) \in R'$. Towards a contradiction, assume that S does not defeat $cl(y)$ in $\mathcal{F}$. Since $cl(S) \cup S^*_{\mathcal{F}} = cl(A)$, this means that there is $y' \in S$ with $cl(y') = cl(y)$. By the well-formedness of $\mathcal{P}$ this further implies $(y', x) \in R'$. But

Reductions 3 and 4 cannot remove conflicts, i.e., either $(y', x) \in R$ or $(x, y') \in R$. Thus, $S \notin cf(A, R)$. Contradiction. Therefore, S defeats $cl(y)$ in $\mathcal{F}$, i.e., there is $z \in S$ such that $(z, y) \in R$. We can conclude that $S \in adm(\mathcal{F})$. □

Proposition 10. *There is a PCAF $\mathcal{P}$ such that $stb\text{-}adm^1_{hyb}(\mathcal{P}) \neq stb\text{-}cf^1_{hyb}(\mathcal{P})$.*

Proof. Let $\mathcal{P} = (A, R, cl, \succ)$ with $A = \{a, b, b'\}$, $R = \{(b, a), (b, b), (b', a), (b', b)\}$, $cl(a) = \alpha$, $cl(b) = cl(b') = \beta$, and $a \succ b', b \succ b'$. The attacks (b', a) and (b', b) are deleted in $\mathcal{R}_1(\mathcal{P})$, see Figure 7b. Moreover, $stb\text{-}adm_{hyb}(\mathcal{R}_1(\mathcal{P})) = \emptyset$ but $stb\text{-}cf_{hyb}(\mathcal{R}_1(\mathcal{P})) = \{\{\alpha, \beta\}\}$. □

Proposition 11. *There is a PCAF $\mathcal{P}$ such that $stb\text{-}adm^2_{hyb}(\mathcal{P}) \neq stb\text{-}cf^2_{hyb}(\mathcal{P})$.*

Proof. Consider the PCAF $\mathcal{P} = (A, R, cl, \succ)$ with $A = \{a, a'\}$, $R = \{(a, a), (a', a)\}$, $cl(a) = cl(a') = \alpha$, and $a \succ a'$. Then $\mathcal{R}_2(\mathcal{P}) = (A, R', cl)$ with $R' = \{(a, a), (a, a')\}$, see Figure 7c. Note that $stb_{inh}(\mathcal{R}_2(\mathcal{P})) = stb\text{-}adm_{hyb}(\mathcal{R}_2(\mathcal{P})) = \emptyset$ while $stb\text{-}cf_{hyb}(\mathcal{R}_2(\mathcal{P})) = \{\{\alpha\}\}$. □

Before investigating whether inherited and hybrid preferred semantics coincide, we examine the I-maximality property. The following is analogous to Definition 10.

Definition 17 (I-maximality for PCAFs)**.** *σ^i_μ is I-maximal for PCAFs if, for all PCAFs $\mathcal{P}$ and all $C, D \in \sigma^i_\mu(\mathcal{P})$, $C \subseteq D$ implies $C = D$.*

From known properties of wfCAFs (cf. Figure 3) it follows directly that $naive^i_{inh}$, where $i \in \{1, 2, 3, 4\}$, is not I-maximal for PCAFs. Likewise, from the properties of general CAFs we know that $naive^i_{hyb}$, prf^i_{hyb}, and grd^i_{inh} are I-maximal for all $i \in \{1, 2, 3, 4\}$. It remains to investigate I-maximality of prf^i_{inh} and all inherited and hybrid variants of stable, semi-stable, and stage semantics.

As it turns out, Reduction 3 manages to preserve I-maximality in all cases except for inherited and hybrid stage semantics.

Proposition 12. *prf^3_{inh}, sem^3_{inh}, sem^3_{hyb}, stb^3_{inh}, $stb\text{-}adm^3_{hyb}$, and $stb\text{-}cf^3_{hyb}$ are I-maximal for PCAFs.*

Proof. We show this for prf^3_{inh}. The other results follow from $sem^3_{inh}(\mathcal{P}) \subseteq prf^3_{inh}(\mathcal{P})$ (by properties of general CAFs), $sem^3_{hyb}(\mathcal{P}) \subseteq prf^3_{inh}(\mathcal{P})$ (by properties of general CAFs), and $stb^3_{inh}(\mathcal{P}) = stb\text{-}adm^3_{hyb}(\mathcal{P}) = stb\text{-}cf^3_{hyb}(\mathcal{P}) \subseteq prf^3_{inh}(\mathcal{P})$ (by Propositions 7 and 9 as well as properties of general CAFs). Towards a contradiction, assume there is a PCAF $\mathcal{P} = (A, R, cl, \succ)$ such that $C \subset D$ for some $C, D \in prf^3_{inh}(\mathcal{P})$. Then there must be $S \subseteq A$ such that $S \in prf(\mathcal{R}_3(\mathcal{P}))$ and $cl(S) = C$, as well as $T \subseteq A$ with $T \in prf(\mathcal{R}_3(\mathcal{P}))$ and $cl(T) = D$. Observe that $S \not\subseteq T$, otherwise $S \notin prf(\mathcal{R}_3(\mathcal{P}))$. Thus, there is $x \in S$ (with $cl(x) \in C$) such that $x \notin T$. However, $cl(x) \in D$ since $C \subset D$, i.e., there is some $x' \in T$ such that $cl(x') = cl(x)$. There are two possibilities for why x is not in T:

1. $T \cup \{x\} \notin cf(\mathcal{R}_3(\mathcal{P}))$. By Lemma 6, $T \cup \{x\} \notin cf((A, R, cl))$. Therefore, there is some $y \in T$ such that $y \notin S$ and either $(x, y) \in \mathcal{P}$ or $(y, x) \in \mathcal{P}$. Actually, it cannot be that $(x, y) \in \mathcal{P}$, otherwise, by the well-formedness of (A, R, cl), we would have $(x', y) \in \mathcal{P}$ which, also by Lemma 6, would mean that $T \notin cf(\mathcal{R}_3(\mathcal{P}))$. Thus,

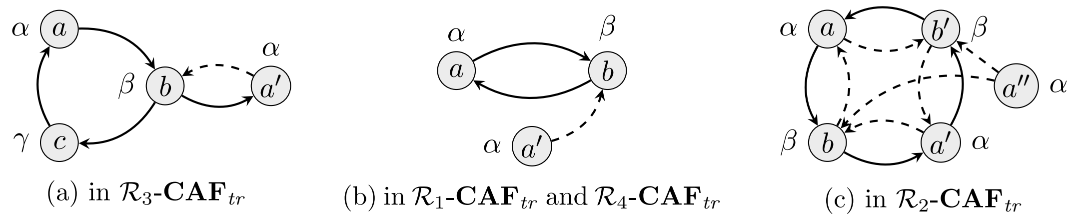


Figure 8: CAFs used as counter examples for I-maximality of some semantics. Dashed arrows are edges in the respective wf-problematic part.

$(y, x) \in \mathcal{P}$. Since $(x, y) \notin \mathcal{P}$, and by the definition of Reduction 3, $(y, x) \in \mathcal{R}_3(\mathcal{P})$. S must defend x from y in $\mathcal{R}_3(\mathcal{P})$, i.e., there is some $z \in S$ such that $(z, y) \in \mathcal{R}_3(\mathcal{P})$. Therefore, also $(z, y) \in \mathcal{P}$. Since we have that $C \subset D$ there is some $z' \in T$ such that $cl(z') = cl(z)$. $(z', y) \in \mathcal{P}$ by the well-formedness of (A, R, cl). But then, by Lemma 6, $T \notin cf(\mathcal{R}_3(\mathcal{P}))$. Contradiction.

2. x is not defended by T. Then there is some $y \in A$ such that $(y, x) \in \mathcal{R}_3(\mathcal{P})$ and such that y is not attacked by any argument in T. But S must defend x against y in $\mathcal{R}_3(\mathcal{P})$, i.e., there is $z \in S$ such that $(z, y) \in \mathcal{R}_3(\mathcal{P})$. Then also $(z, y) \in \mathcal{P}$. Since $C \subset D$ there is some $z' \in T$ such that $cl(z') = cl(z)$. $(z', y) \in \mathcal{P}$ by the well-formedness of (A, R, cl). It cannot be that $(z', y) \in \mathcal{R}_3(\mathcal{P})$, i.e., $y \succ z'$. But then, by the definition of Reduction 3, we must have $(y, z') \in \mathcal{P}$ and also $(y, z') \in \mathcal{R}_3(\mathcal{P})$, which means that T is attacked by y but not defended against it, i.e., $T \notin adm(\mathcal{R}_3(\mathcal{P}))$. Contradiction. □

For negative results, it suffices to show that I-maximality is not preserved for transitive preference orderings to obtain results for the more general case.

Proposition 13. *stg^3_{inh} and stg^3_{hyb} are not I-maximal for PCAFs, even when considering only transitive preferences.*

Proof. Let $\mathcal{F} = (A, R, cl)$ be the CAF shown in Figure 8a, and let $F = (A, R)$ be its underlying AF. Clearly, $\mathcal{F} \in \mathcal{R}_3$-$\mathbf{CAF}_{tr}$.

We can see that $cf(F) = \{\emptyset, \{a\}, \{a'\}, \{b\}, \{c\}, \{a, a'\}, \{a', c\}\}$ and thus $naive(F) = \{\{a, a'\}, \{a', c\}, \{b\}\}$.

Regarding stg^3_{inh}, we have $\{a, a'\}^{\oplus}_F = \{a, a', b\}$, $\{a', c\}^{\oplus}_F = \{a, a', c\}$, and $\{b\}^{\oplus}_F = \{a', b, c\}$. The three ranges are incomparable, i.e., $stg(F) = naive(F)$ and therefore $stg_{inh}(\mathcal{F}) = \{\{\alpha\}, \{\alpha, \gamma\}, \{\beta\}\}$.

Regarding stg^3_{hyb}, $\{a, a'\}$ defeats $\{\beta\}$ while $\{b\}$ defeats $\{\gamma\}$. Thus, $\{a, a'\}^{\circledast}_{\mathcal{F}} = \{\alpha, \beta\}$, $\{a', c\}^{\circledast}_{\mathcal{F}} = \{\alpha, \gamma\}$, and $\{b\}^{\circledast}_{\mathcal{F}} = \{\beta, \gamma\}$. The three claim-ranges are incomparable, and we have $stg_{hyb}(\mathcal{F}) = \{\{\alpha\}, \{\alpha, \gamma\}, \{\beta\}\}$. □

Reductions 1, 2, and 4 lose I-maximality for *all* semantics that are I-maximal on wfCAFs but not on general CAFs.

Proposition 14. *For $i \in \{1, 2, 4\}$, the following semantics are not I-maximal for PCAFs, even if considering only transitive preferences: stb^i_{inh}, $stb\text{-}adm^i_{hyb}$, $stb\text{-}cf^i_{hyb}$, sem^i_{inh}, sem^i_{hyb}, prf^i_{inh}, stg^i_{inh}, stg^i_{hyb}.*

Proof. We show this for stb^i_{inh}. For all other σ^i_μ this follows from $stb^i_{inh}(\mathcal{P}) \subseteq \sigma^i_\mu(\mathcal{P})$ (which holds by the properties of general CAFs).

For $i \in \{1,4\}$, let $\mathcal{F}$ be the CAF shown in Figure 8b. $\mathcal{F} \in \mathcal{R}_1\text{-}\mathbf{CAF}_{tr}$ by Proposition 5. $\mathcal{F} \in \mathcal{R}_4\text{-}\mathbf{CAF}_{tr}$ since $\mathcal{R}_4(\mathcal{P}) = \mathcal{F}$ for $\mathcal{P} = (A, R, cl, \succ)$ with $A = \{a, a', b\}$, $R = \{(b,a)\}$, $cl(a) = cl(a') = \alpha$, $cl(b) = \beta$, and $a \succ b$. As required, the underlying CAF of $\mathcal{P}$ is well-formed. It can be verified that $stb(\mathcal{F}) = \{\{a,a'\},\{a',b\}\}$ and thus $stb_{inh}(\mathcal{F}) = \{\{\alpha\},\{\alpha,\beta\}\}$.

For $i = 2$, let $\mathcal{F}'$ be the CAF of Figure 8c. $\mathcal{F}' \in \mathcal{R}_2\text{-}\mathbf{CAF}_{tr}$ since $\mathcal{R}_2(\mathcal{P}') = \mathcal{F}'$ for the PCAF $\mathcal{P}' = (A', R', cl', \succ)$ with $R' = \{(b,a),(b,a'),(b',a),(b',a')\}$, $a \succ b$, and $a' \succ b'$. As required, the underlying CAF of $\mathcal{P}'$ is well-formed. It can be verified that $stb(\mathcal{F}') = \{\{a,a',a''\},\{a'',b,b'\}\}$ and thus $stb_{inh}(\mathcal{F}') = \{\{\alpha\},\{\alpha,\beta\}\}$. □

We can now use the fact that inherited preferred semantics are I-maximal under Reduction 3 to show that inherited and hybrid preferred semantics coincide under Reduction 3.

Proposition 15. *$prf^3_{inh}(\mathcal{P}) = prf^3_{hyb}(\mathcal{P})$ for every PCAF $\mathcal{P}$.*

Proof. $prf_{hyb}(\mathcal{F}) \subseteq prf_{inh}(\mathcal{F})$ holds for all CAFs. We must show $prf_{inh}(\mathcal{F}) \subseteq prf_{hyb}(\mathcal{F})$ for all $\mathcal{F} \in \mathcal{R}_3\text{-}\mathbf{CAF}$. Towards a contradiction, assume there is $\mathcal{F} = (A, R, cl) \in \mathcal{R}_3\text{-}\mathbf{CAF}$ such that $C \in prf_{inh}(\mathcal{F})$ but $C \notin prf_{hyb}(\mathcal{F})$ for some $C \subseteq cl(A)$. Then $C \in adm_{inh}(\mathcal{F})$. Since $C \notin prf_{hyb}(\mathcal{F})$, there must be $D \in prf_{hyb}(\mathcal{F})$ such that $D \supset C$. Since $prf_{hyb}(\mathcal{F}) \subseteq prf_{inh}(\mathcal{F})$ we have $D \in prf_{inh}(\mathcal{F})$. But then we have $C, D \in prf_{inh}(\mathcal{F})$ and $D \supset C$. This means that prf_{inh} is not I-maximal for CAFs in $\mathcal{R}_3\text{-}\mathbf{CAF}$, which contradicts Proposition 12. □

Our results regarding I-maximality also allow us to infer negative results regarding the relationship between semantics: if σ^i_μ is I-maximal while τ^i_ν is not, then there must be a PCAF $\mathcal{P}$ such that $\sigma^i_\mu(\mathcal{P}) \not\subseteq \tau^i_\nu(\mathcal{P})$. Thus, we can conclude:

Proposition 16. *For every $i \in \{1,2,4\}$ there is:*

- *a PCAF $\mathcal{P}$ such that $prf^i_{inh}(\mathcal{P}) \not\subseteq prf^i_{hyb}(\mathcal{P})$;*
- *a PCAF $\mathcal{P}$ such that $sem^i_{inh}(\mathcal{P}) \not\subseteq prf^i_{hyb}(\mathcal{P})$;*
- *a PCAF $\mathcal{P}$ such that $sem_{hyb}(\mathcal{P}) \not\subseteq prf_{hyb}(\mathcal{P})$.*

Proposition 17. *For every $i \in \{1,2,3,4\}$ there is:*

- *a PCAF $\mathcal{P}$ such that $stg^i_{inh}(\mathcal{P}) \not\subseteq naive^i_{hyb}(\mathcal{P})$;*
- *a PCAF $\mathcal{P}$ such that $stg^i_{hyb}(\mathcal{P}) \not\subseteq naive^i_{hyb}(\mathcal{P})$.*

We have now determined the relationship between PCAF-semantics and their properties with respect to I-maximality. In summary:

Theorem 18. *The results depicted in Figure 9 hold, even when considering only PCAFs with transitive preferences.*

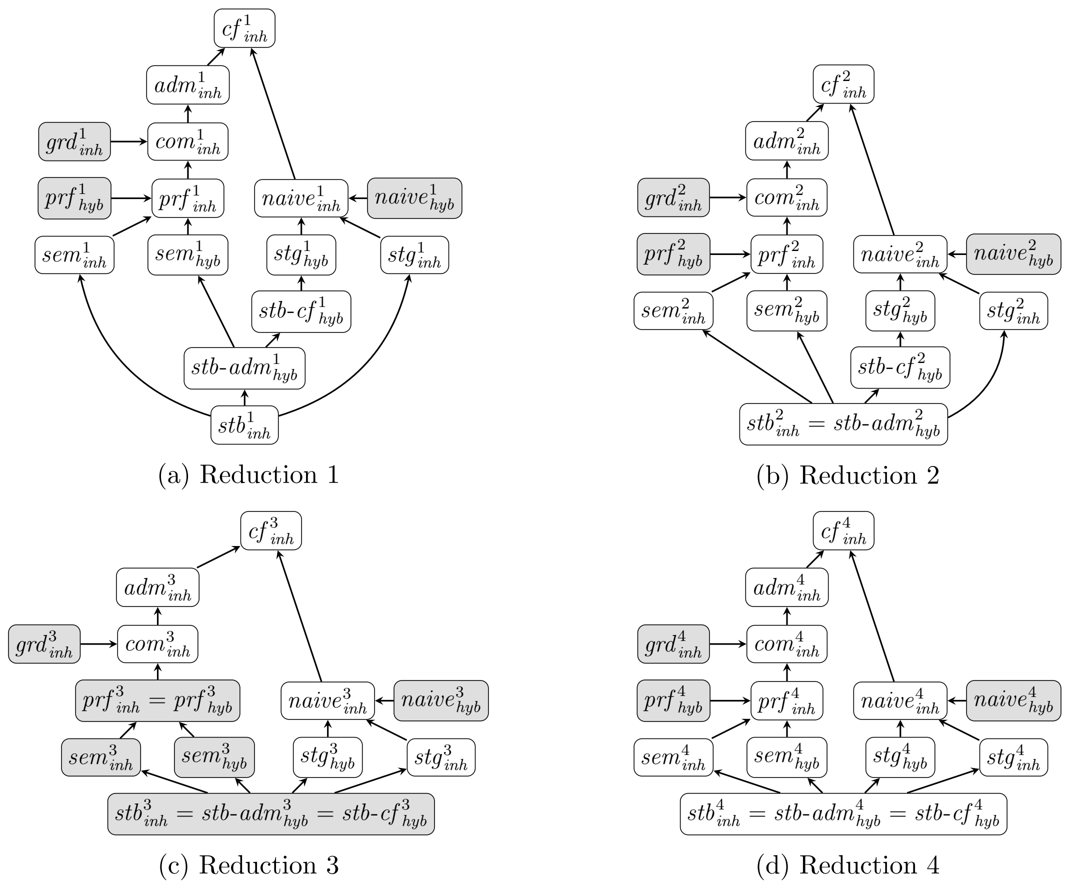


Figure 9: Relations between PCAF-semantics. If there is an arrow from σ to τ, then $\sigma(\mathcal{P}) \subseteq \tau(\mathcal{P})$ for all PCAFs $\mathcal{P}$. Semantics highlighted in gray are I-maximal.

Reduction 3 preserves the properties of wfCAFs for semantics that are based on admissibility (stable, semi-stable, preferred) but not semantics that are based on conflict-freeness (stage, naive). Reductions 1, 2, and 4 on the other hand lose the I-maximality properties of wfCAFs in all cases (except for those semantics that are I-maximal on general CAFs already). Under Reduction 4 all variants of stable semantics coincide, while under Reduction 2 the inherited and admissibility-based hybrid stable semantics coincide. Reduction 1 preserves none of the investigated semantic properties of wfCAFs.

Intuitively, these results can be explained by the fact that Reduction 3 is the most conservative of the reductions, not adding new attacks and preserving conflict-freeness (i.e., given a PCAF $\mathcal{P}$, a set of arguments E is conflict-free in the underlying CAF of $\mathcal{P}$ iff E is conflict-free in $\mathcal{R}_3(\mathcal{P})$). Reductions 2 and 4 preserve conflict-freeness too, but they may introduce new attacks in contrast to Reduction 3. Reduction 1 on the other hand does not preserve conflict-freeness. In fact, it has been deemed problematic for exactly this reason when applied to regular AFs (Amgoud & Vesic, 2014), although it is still discussed and considered in the literature alongside the other reductions (Kaci et al., 2021).

Our results support the decision-making process when choosing how preferences should be resolved (i.e., which preference reduction should be used). For example, if Reduction 3 is chosen then no attention has to be paid to the existence of several variants for preferred or stable semantics, since all the variants coincide. What is more, we know that these semantics are I-maximal and therefore behave 'as expected' on the claim level. If on the other hand Reduction 1 is chosen, then one must be aware that the different variants for stable and preferred semantics may deliver different extensions, and that none of them (except hybrid preferred semantics) are I-maximal.

6. Computational Complexity

In this section, we investigate the impact of preferences on the computational complexity of claim-based reasoning. To this end, we define the three main decision problems for PCAFs analogously to those for CAFs (cf. Definition 11), except that we take a PCAF instead of a CAF as input and appeal to PCAF-semantics σ^i_μ instead of CAF-semantics σ_μ.

Definition 18 (Decision problems for PCAFs)**.** *We consider the following decision problems pertaining to a PCAF-semantics* σ^i_μ*:*

- Credulous Acceptance *(*$Cred^{PCAF}_{\sigma^i_\mu}$*): Given a PCAF* $\mathcal{P}$ *and claim* α*, is* α *contained in some* $C \in \sigma^i_\mu(\mathcal{P})$*?*
- Skeptical Acceptance *(*$Skept^{PCAF}_{\sigma^i_\mu}$*): Given a CAF* $\mathcal{P}$ *and claim* α*, is* α *contained in each* $C \in \sigma^i_\mu(\mathcal{P})$*?*
- Verification *(*$Ver^{PCAF}_{\sigma^i_\mu}$*): Given a CAF* $\mathcal{P}$ *and a set of claims* C*, is* $C \in \sigma^i_\mu(\mathcal{P})$*?*

Membership results for PCAFs can be inferred from results for general CAFs (recall that the preference reductions from PCAFs to CAFs can be done in polynomial time), and hardness results from results for wfCAFs. Thus, except for $naive^i_{hyb}$, the complexity of credulous and skeptical acceptance follows immediately from known results for CAFs and wfCAFs (cf. Table 1):

Observation 19. *Let* $i \in \{1, 2, 3, 4\}$ *and let* σ^i_μ *be any PCAF-semantics considered in this paper.* $Cred^{PCAF}_{\sigma^i_\mu}$ *has the same complexity as* $Cred^{wfCAF}_{\sigma_\mu}$*.* $Skept^{PCAF}_{\sigma^i_\mu}$ *has the same complexity as* $Skept^{wfCAF}_{\sigma_\mu}$*, except for* $\sigma^i_\mu = naive^i_{hyb}$*. Moreover,* $Ver^{PCAF}_{grd^i_{inh}}$ *has the same complexity as* $Ver^{wfCAF}_{grd_{inh}}$*.*

The computational complexity of the verification problem, on the other hand, is one level higher on the polynomial hierarchy for general CAFs compared to wfCAFs (cf. Table 1), i.e., the bounds that existing results yield for PCAFs are not tight. We address this open problem and comprehensively analyze $Ver^{PCAF}_{\sigma^i_\mu}$ for each of the considered reductions and semantics. Moreover, we investigate the complexity of $Skept^{PCAF}_{naive^i_{hyb}}$.

Regarding conflict-free and naive semantics, the fact that Reductions 2–4 do not remove conflicts straightforwardly implies that the properties of wfCAFs are preserved.

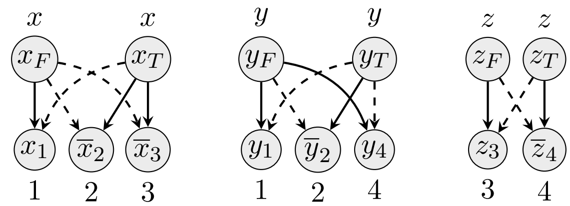


Figure 10: Reduction of 3-SAT-instance $\omega_1 = \{x, y\}$, $\omega_2 = \{\neg x, \neg y\}$, $\omega_3 = \{\neg x, z\}$, $\omega_4 = \{y, \neg z\}$, to an instance $(\mathcal{P}, C)$ of $Ver^{PCAF}_{cf^1_{inh}}$. Dashed arrows are attacks deleted in $\mathcal{R}_1(\mathcal{P})$, i.e., they are edges in $wfp(\mathcal{R}_1(\mathcal{P}))$.

Proposition 20. *$Ver^{PCAF}_{\sigma^i_\mu}$ is in* P *for $\sigma_\mu \in \{cf_{inh}, naive_{inh}, naive_{hyb}\}$ and $i \in \{2, 3, 4\}$.*

Proof. Let $\mathcal{P} = (A, R, cl, \succ)$ be a PCAF, C a set of claims, and $i \in \{2, 3, 4\}$. To check whether $C \in cf^i_{inh}(\mathcal{P})$, by Lemma 6, it suffices to check whether $C \in cf_{inh}((A, R, cl))$. This can be done in polynomial time on wfCAFs (cf. Table 1). Analogously for $naive^i_{inh}(\mathcal{P})$ and $naive^i_{hyb}(\mathcal{P})$. □

Proposition 21. *$Skept^{PCAF}_{naive^i_{hyb}}$ is* coNP*-complete for $\sigma^i_\mu = naive^i_{hyb}$ and $i \in \{2, 3, 4\}$.*

Proof. coNP-hardness follows from known results for wfCAFs (see Table 1). Regarding coNP-membership, let $\mathcal{P} = (A, R, cl, \succ)$ be a PCAF, $\alpha \in cl(A)$, and $i \in \{2, 3, 4\}$. To decide whether α is skeptically accepted in $\mathcal{P}$ under $naive^i_{hyb}$-semantics, by Lemma 6, it suffices to decide whether α is skeptically accepted in the underlying CAF (A, R, cl) of $\mathcal{P}$. This can be done in coNP-time on wfCAFs (cf. Table 1). □

6.1 Hardness under Reduction 1

Since Reduction 1 *does* remove conflicts between arguments, we cannot apply the same reasoning as above when analyzing the complexity of conflict-free and naive semantics under Reduction 1. Indeed, it turns out that we lose the benefits of wfCAFs for these semantics (as well as $stb\text{-}cf_{hyb}$). In the proof of Proposition 22 we make use of Reduction 1's ability to remove conflicts in order to show hardness.

Proposition 22. *$Ver^{PCAF}_{\sigma^i_\mu}$ is* NP*-hard for $\sigma^i_\mu \in \{cf^1_{inh}, naive^1_{inh}, stb\text{-}cf^1_{hyb}\}$, even if we restrict ourselves to PCAFs with transitive preference relations.*

Proof. Let φ be an arbitrary instance of 3-SAT given as a set $\Omega = \{\omega_1, \ldots, \omega_m\}$ of clauses over variables X. Without loss of generality, we can assume that every variable appears both positively and negatively in φ. We construct a PCAF $\mathcal{P} = (A, R, cl, \succ)$ as well as a set of claims C:

- $A = V \cup \overline{V} \cup H$ where $V = \{x_i \mid x \in \omega_i, 1 \leq i \leq m\}$, $\overline{V} = \{\overline{x}_i \mid \neg x \in \omega_i, 1 \leq i \leq m\}$, and $H = \{x_T, x_F \mid x \in X\}$;
- $R = \{(x_T, x_i), (x_F, x_i) \mid x_i \in V\} \cup \{(x_T, \overline{x}_i), (x_F, \overline{x}_i) \mid \overline{x}_i \in \overline{V}\}$;

- $cl(x_i) = cl(\overline{x}_i) = i$ for all $x_i, \overline{x}_i \in V \cup \overline{V}$, $cl(x_T) = cl(x_F) = x$ for all $x \in X$;
- $x_i \succ x_T$ for all $x_i \in V$ and $\overline{x}_i \succ x_F$ for all $\overline{x}_i \in \overline{V}$;
- $C = \{1, \dots, m\} \cup X$.

Figure 10 illustrates the above construction. Note that the preferences $x_i \succ x_T$ remove all conflicts between the 'true' variable arguments x_T and their unnegated occurrences x_i. Likewise for preferences of the form $\overline{x}_i \succ x_F$. Now let $\mathcal{F} = \mathcal{R}_1(\mathcal{P}) = (A, R', cl)$. We must show that φ is satisfiable iff $C \in \sigma_\mu(\mathcal{F})$ for $\sigma_\mu \in \{cf_{inh}, naive_{inh}, stb\text{-}cf_{hyb}\}$.

Assume φ is satisfiable. Then there is an interpretation I such that $I \models \varphi$. Let $S = \{x_T \in H \mid x \in I\} \cup \{x_F \in H \mid x \notin I\} \cup \{x_i \in V \mid x \in I\} \cup \{\overline{x}_i \in \overline{V} \mid x \notin I\}$. It can be easily verified that S is conflict free in (A, R') and that $cl(S) = C$. Note that C contains all claims in $\mathcal{F}$, i.e., $C = cl(A)$. Thus, $C \in stb\text{-}cf_{hyb}(\mathcal{F})$. Moreover, $C \in naive_{inh}(\mathcal{F})$ and $C \in cf_{inh}(\mathcal{F})$ since $stb\text{-}cf_{hyb}(\mathcal{F}) \subseteq naive_{inh}(\mathcal{F}) \subseteq cf_{inh}(\mathcal{F})$.

Assume $C \in cf_{inh}(\mathcal{F})$. Then there is some $S \subseteq A$ such that $S \in cf((A, R'))$ and $cl(S) = C$. Let x be any variable in X. Since $x \in cl(S)$ it must be that either $x_T \in S$ or $x_F \in S$. Thus, for all i, j, we have $x_i \in S \implies \overline{x}_j \notin S$ and $\overline{x}_i \in S \implies x_j \notin S$ (otherwise, we would need both $x_T \notin S$ and $x_F \notin S$ for S to be conflict-free). Furthermore, for any $i \in \{1, \dots, m\}$, there must be some x such that $x_i \in S$ or $\overline{x}_i \in S$. Let $I = \{x \mid x_i \in S \text{ for some } i\}$. Then for every i there is some x such that either $x \in \omega_i$ and $x \in I$ or $\neg x \in \omega_i$ and $x \notin I$. Thus, I satisfies all clauses $\omega_1, \dots, \omega_m$ which means that φ is satisfiable. The proof works likewise if we assume $C \in naive_{inh}(\mathcal{F})$ or $C \in stb\text{-}cf_{hyb}(\mathcal{F})$ since $stb\text{-}cf_{hyb}(\mathcal{F}) \subseteq naive_{inh}(\mathcal{F}) \subseteq cf_{inh}(\mathcal{F})$. □

Note that the above construction does not work for admissible-based semantics, since the variable-arguments x_i resp. $\overline{x}_i$ in the extension S would remain undefended. The existing hardness proof for general CAFs (Dvořák & Woltran, 2020, Proposition 2) cannot be used either, as the constructed CAFs are not in $\mathcal{R}_1$-**CAF**. Specifically, there are symmetric attacks between arguments whose claims occur multiple times, which leads to cycles in the wf-problematic part of the constructed CAF. Instead, we show hardness via a more involved construction in which symmetric attacks are avoided.

Proposition 23. *$Ver^{PCAF}_{\sigma^i_\mu}$ is* NP*-hard for $\sigma^i_\mu \in \{stb^1_{inh}, stb\text{-}adm^1_{hyb}, com^1_{inh}, adm^1_{inh}\}$, even if we restrict ourselves to PCAFs with transitive preference relations.*

Proof. Let φ be an arbitrary 3-SAT-instance given as a set $\Omega = \{\omega_1, \dots, \omega_m\}$ of clauses over variables X. For convenience, we directly construct a CAF $\mathcal{F} = (A, R, cl)$ with $\mathcal{F} \in \mathcal{R}_1\text{-}\mathbf{CAF}_{tr}$ instead of providing a PCAF $\mathcal{P}$ such that $\mathcal{R}_1(\mathcal{P}) = \mathcal{F}$. This is legitimate, as, by our characterization of $\mathcal{R}_1\text{-}\mathbf{CAF}_{tr}$ (see Proposition 5), we can obtain $\mathcal{P}$ by simply adding all edges in $wfp(\mathcal{F})$ to R and defining $\succ$ accordingly. We also construct a set of claims C.

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

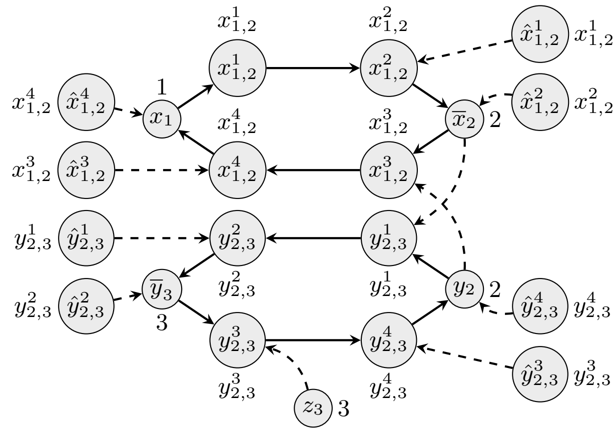


Figure 11: Reduction of 3-SAT-instance $\omega_1 = \{x\}$, $\omega_2 = \{\neg x, y\}$, $\omega_3 = \{\neg y, z\}$, to an instance $(\mathcal{P}, C)$ of $Ver^{PCAF}_{stb^1_{inh}}$. Dashed arrows are attacks deleted in $\mathcal{R}_1(F')$, i.e., they are edges in $wfp(\mathcal{R}_1(\mathcal{P}))$.

- $C = \{1, \ldots, m\} \cup \{cl(a) \mid a \in H\}$.

Figure 11 illustrates the above construction. In general, every $\hat{x}^k_{i,j}$ only has outgoing edges in the wf-problematic part, and no incoming or outgoing attacks in R. Every $x^k_{i,j}$ only has incoming edges in the wf-problematic part. Finally, there can be no edges in the wf-problematic part between any x_i (or $\overline{x}_i$) and any other x_j (or $\overline{x}_j$). From this, and by the above construction, we can infer that (A, R, cl) fulfills all of the conditions to be in $\mathcal{R}_1$-$\mathbf{CAF}_{tr}$ (cf. Proposition 5). It remains to show the correctness of the above construction.

Assume φ is satisfiable. Then there is an interpretation I such that $I \models \varphi$. Let $S = \{x_i \in V \mid x \in I\} \cup \{\overline{x}_i \in \overline{V} \mid x \notin I\} \cup \{x^2_{i,j}, x^3_{i,j} \mid x_i, \overline{x}_j \in A, x \in I\} \cup \{x^1_{i,j}, x^4_{i,j} \mid x_i, \overline{x}_j \in A, x \notin I\} \cup \{\hat{x}^k_{i,j} \mid \hat{x}^k_{i,j} \in A\}$. It can be verified that $S \in stb((A, R))$ and that $cl(S) = C$. Thus, $C \in stb_{inh}(\mathcal{F})$. Moreover, $C \in stb\text{-}adm_{hyb}(\mathcal{F})$, $C \in com_{inh}(\mathcal{F})$, and $C \in adm_{inh}(\mathcal{F})$, since $stb_{inh}(\mathcal{F}) \subseteq stb\text{-}adm_{hyb}(\mathcal{F}) \subseteq com_{inh}(\mathcal{F}) \subseteq adm_{inh}(\mathcal{F})$.

Assume $C \in adm_{inh}(\mathcal{F})$. Then there is some $S \subseteq A$ such that $S \in adm((A, R))$ and $cl(S) = C$. Thus, for any $i \in \{1, \ldots, m\}$, there must be some x such that $x_i \in S$ or $\overline{x}_i \in S$. Consider the case that $x_i \in S$. Since S is admissible, $x^1_{i,j} \notin S$ for any j such that $\overline{x}_j \in A$. This further means that $\overline{x}_j \notin S$ for any $\overline{x}_j \in A$, since we would need $x^1_{i,j} \in S$ to defend $\overline{x}_j$ from the attack by $x^2_{i,j}$. Likewise, if $\overline{x}_i \in S$, then $x_j \notin S$ for all $x_j \in A$. Let $I = \{x \mid x_i \in S \text{ for some } i\}$. Then for every i there is some x such that either $x \in \omega_i$ and $x \in I$ or $\neg x \in \omega_i$ and $x \notin I$. Thus, I satisfies all clauses $\omega_1, \ldots, \omega_m$ which means that φ is satisfiable. The proof works likewise if we assume $C \in stb_{inh}(\mathcal{F})$, $C \in stb\text{-}adm_{hyb}(\mathcal{F})$, or $C \in com_{inh}(\mathcal{F})$, since $stb_{inh}(\mathcal{F}) \subseteq stb\text{-}adm_{hyb}(\mathcal{F}) \subseteq com_{inh}(\mathcal{F}) \subseteq adm_{inh}(\mathcal{F})$. □

Regarding semi-stable and inherited preferred semantics, we can build upon the standard translation for skeptical acceptance of preferred-semantics (Dvořák & Dunne, 2018, Reduction 3.7). We introduce helper arguments and avoid symmetric attacks between arguments of the same claim.

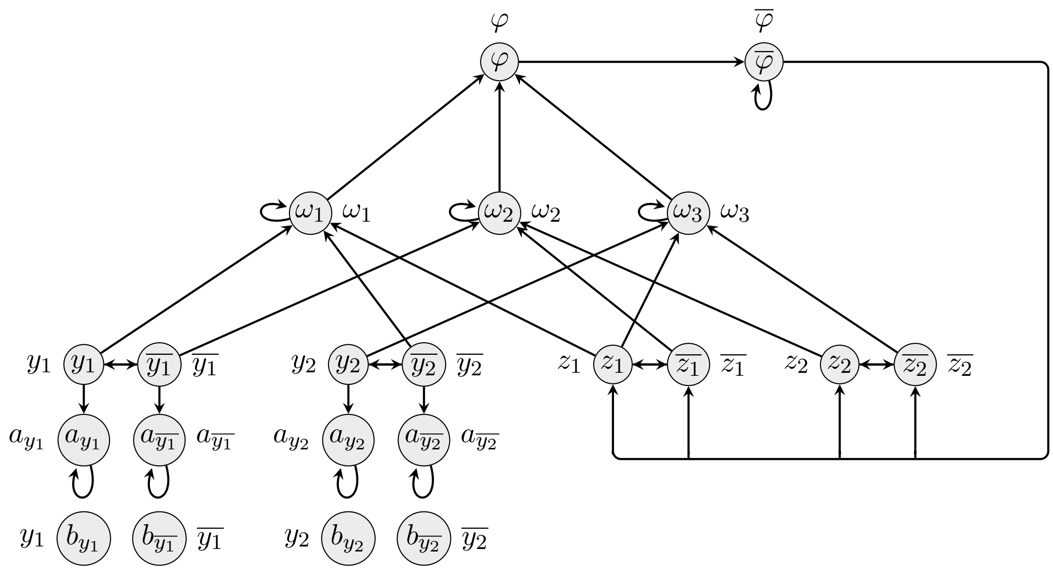


Figure 12: Reduction of the $\text{QBF}^2_\forall$ instance $\Phi = \forall y_1, y_2 \exists z_1, z_2 \varphi$ with φ given by clauses $\omega_1 = \{y_1, \neg y_2, z_1\}$, $\omega_2 = \{\neg y_1, \neg z_1, z_2\}$, $\omega_3 = \{y_2, z_1, \neg z_2\}$ to an instance of $\mathit{Ver}^{PCAF}_{sem^1_{inh}}$.

Proposition 24. *$\mathit{Ver}^{PCAF}_{\sigma^i_\mu}$ is Σ^{P}_2-hard for $\sigma^i_\mu \in \{\mathit{prf}^1_{inh}, \mathit{sem}^1_{inh}, \mathit{sem}^1_{hyb}, \mathit{stg}^1_{inh}, \mathit{stg}^1_{hyb}\}$, even if we restrict ourselves to PCAFs with transitive preference relations.*

Proof. We show hardness for $\sigma_\mu \in \{\mathit{prf}^1_{inh}, \mathit{sem}^1_{inh}, \mathit{sem}^1_{hyb}\}$. The remaining cases can be found in Appendix B (Lemma 43). Let $\Phi = \forall Y \exists Z \varphi$ be an instance of $\text{QBF}^2_\forall$, where φ is given by a set Ω of clauses over atoms $X = Y \cup Z$. We provide a reduction to the complementary problem of $\mathit{Ver}^{PCAF}_{\sigma^1_\mu}$. In particular, we construct the CAF $\mathcal{F} = (A, R, cl)$ with underlying AF $F = (A, R)$ and a set of claims C:

- $A = \{\varphi, \overline{\varphi}\} \cup \Omega \cup X \cup \overline{X} \cup Y_a \cup \overline{Y}_a \cup Y_b \cup \overline{Y}_b$, where $\overline{X} = \{\overline{x} \mid x \in X\}$, $Y_a = \{a_y \mid y \in Y\}$, $\overline{Y}_a = \{a_{\overline{y}} \mid y \in Y\}$, $Y_b = \{b_y \mid y \in Y\}$, $\overline{Y}_b = \{b_{\overline{y}} \mid y \in Y\}$;
- $R = \{(x, \overline{x}), (\overline{x}, x) \mid x \in X\} \cup \{(\omega, \omega), (\omega, \varphi) \mid \omega \in \Omega\} \cup \{(\varphi, \overline{\varphi}), (\overline{\varphi}, \overline{\varphi})\} \cup$
  $\{(x, \omega) \mid x \in \omega, \omega \in \Omega\} \cup \{(\overline{x}, \omega) \mid \neg x \in \omega, \omega \in \Omega\} \cup$
  $\{(a_v, a_v), (v, a_v) \mid v \in Y \cup \overline{Y}\} \cup \{(\overline{\varphi}, z), (\overline{\varphi}, \overline{z}) \mid z \in Z\}$;
- $cl(b_v) = v$ for $b_v \in Y_b \cup \overline{Y}_b$ and $cl(v) = v$ else;
- $C = Y \cup \overline{Y}$.

Figure 12 illustrates the above construction. Note that $\mathcal{F} \in \mathcal{R}_1$-$\mathbf{CAF}_{tr}$ (cf. Proposition 5) since all paths in $\mathit{wfp}(\mathcal{F}) = \{(b_a, v) \mid v \in Y \cup \overline{Y}\}$ are of length 1 (only arguments in $Y_b \cup \overline{Y}_b$ have outgoing edges in $\mathit{wfp}(\mathcal{F})$). It remains to verify the correctness of the reduction, i.e., we will show that Φ is valid iff $C \notin \sigma_\mu(\mathcal{F})$.

"$\Longrightarrow$": Assume Φ is valid. Consider any $S \subseteq A$ such that $S \in \mathit{adm}(F)$ and $cl(S) = C$. Then $S \subseteq Y \cup \overline{Y} \cup Y_b \cup \overline{Y}_b$. Let $Y' = S \cap Y$. Since Φ is valid, there is $Z' \subseteq Z$ such that $M = Y' \cup Z'$ is a model of φ. Let $T = M \cup \{\overline{x} \mid x \in X \setminus M\} \cup Y_b \cup \overline{Y}_b \cup \{\varphi\}$.

Note that $S \subset T$ and $T \in cf(F)$ by construction. Moreover, $T \in adm(F)$ since φ defends $v \in Z' \cup \{\overline{z} \mid z \in Z \setminus Z'\}$ against $\overline{\varphi}$; moreover, each argument $v \in X$ defends itself against $\overline{v}$ and vice verca; also, $M \cup \{\overline{x} \mid x \in X \setminus M\}$ defends φ against each attack from clause-arguments $\omega \in \Omega$ since $M \models \varphi$: for each clause $\omega \in \Omega$, there is either $v \in M$ with $v \in \omega$ or $\neg v \in \omega$ for some $v \notin M$. In the first case, $(v, \omega) \in R$ and $v \in S$, in the latter, $(\overline{v}, \omega) \in R$ and $\overline{v} \in S$. Thus, $S \notin prf(F)$. Since S was chosen as an arbitrary admissible set such that $cl(S) = C$, we can conclude that $C \notin prf_{inh}(\mathcal{F})$. Moreover, $C \notin sem_{inh}(\mathcal{F})$ and $C \notin sem_{hyb}(\mathcal{F})$, since $sem_{inh}(\mathcal{F}) \subseteq prf_{inh}(\mathcal{F})$ and $sem_{hyb}(\mathcal{F}) \subseteq prf_{inh}(\mathcal{F})$.

"$\Longleftarrow$": Assume $C \notin sem_{inh}(\mathcal{F})$ (resp. $C \notin sem_{hyb}(\mathcal{F})$). Consider any $Y' \subseteq Y$. We will show that there is some $Z' \subseteq Z$ such that $Y' \cup Z' \models \varphi$. Let $S = Y' \cup \{\overline{y} \mid y \in Y \setminus Y'\} \cup Y_b \cup \overline{Y}_b$. Observe that $cl(S) = C$ and $S \in adm(F)$. Since $C \notin sem_{inh}(\mathcal{F})$ (resp. $C \notin sem_{hyb}(\mathcal{F})$), there is $T \in adm(F)$ with $T \cup T_F^+ \supset S \cup S_F^+$ (resp. $cl(T) \cup T^*_{\mathcal{F}} \supset cl(S) \cup S^*_{\mathcal{F}}$).

In particular, we have $Y' \cup \{\overline{y} \mid y \in Y \setminus Y'\} \subseteq T$ since each $a_v \in S_F^+$ (resp. $a_v \in S^*_{\mathcal{F}}$) with $v \in Y \cup \overline{Y}$ has precisely one non-self-attacking attacker (namely the argument v). Moreover, we can assume that T contains each argument $v \in Y_b \cup \overline{Y}_b$ since each such v is unattacked and does not attack any other argument. We can conclude that $T \supset S$.

It follows that $\varphi \in T$: since $S \subset T$, there is some $v \in A \setminus S$ such that $v \in T$. Clearly, $v \in \{\varphi\} \cup Z \cup \overline{Z}$ since each remaining argument is either self-attacking or attacked by S (and thus also by T). In case $v = \varphi$, we are done; in case $v \in Z \cup \overline{Z}$, we have $\varphi \in T$ by admissibility of T (observe that φ is the only attacker of $\overline{\varphi}$). Consequently, T defends φ against each attack from each clause-argument $\omega \in \Omega$.

Now, let $Z' = Z \cap T$. We show that $M = Y' \cup Z'$ is a model of φ. Consider some arbitrary clause $\omega \in \Omega$. Since $\varphi \in T$, there is some $v \in T$ such that $(v, \omega) \in R$ by admissibility of T. In case $v \in X$, we have $v \in M$ and $v \in \omega$ by construction of $\mathcal{F}$; similarly, in case $v \in \overline{X}$ we have $v \notin M$ and $\neg v \in \omega$. Thus, ω is satisfied by M. Since ω was chosen arbitrarily it follows that $M \models \varphi$. We can conclude that Φ is valid. The proof works likewise if we assume $C \notin prf_{inh}(\mathcal{F})$ since $sem_{inh}(\mathcal{F}) \subseteq prf_{inh}(\mathcal{F})$ (resp. $sem_{hyb}(\mathcal{F}) \subseteq prf_{inh}(\mathcal{F})$). □

Just like inherited preferred/naive semantics, hybrid preferred/naive semantics are DP-complete and thus preserve the high complexity of general CAFs. To show this for prf^1_{hyb}, we adapt an existing reduction from SAT-UNSAT to general CAFs (Dvořák et al., 2023).

Proposition 25. *$Ver^{PCAF}_{\sigma^i_\mu}$ is* DP*-hard for $\sigma^i_\mu \in \{prf^1_{hyb}, naive^1_{hyb}\}$, even if we restrict ourselves to PCAFs with transitive preference relations.*

Proof. We show hardness for $\sigma_\mu = prf^1_{hyb}$. The proof for $\sigma_\mu = naive^1_{hyb}$ can be found in Appendix B (Lemma 44). Let (φ_1, φ_2) be an arbitrary instance of SAT-UNSAT, where φ_i is given over a set of clauses Ω_i and a set of variables X_i such that $X_1 \cap X_2 = \emptyset$. Given X_i, we define $\overline{X_i} = \{\overline{x} \mid x \in X_i\}$. Instead of constructing a PCAF, we directly construct a CAF $\mathcal{F} = (A, R, cl) \in \mathcal{R}_1\text{-}\mathbf{CAF}_{tr}$:

- $A = A_1 \cup A_2$, where $A_i = X_i \cup \overline{X_i} \cup \Omega_i \cup \{\varphi_i\} \cup \{d_x \mid x \in X_i \cup \overline{X_i}\}$;
- $R = R_1 \cup R_2$, where $R_i = \{(x, \omega) \mid \omega \in \Omega_i, x \in \Omega_i\} \cup \{(\overline{x}, \omega) \mid \omega \in \Omega_i, \neg x \in \Omega_i\} \cup \{(x, \overline{x}), (\overline{x}, x) \mid x \in X_i\} \cup \{(\omega, \varphi_i), (\omega, \omega) \mid \omega \in \Omega_i\}$;
- $cl(d_x) = x$ for $x \in X_i \cup \overline{X_i}$, $cl(x) = x$ for all other arguments in A.

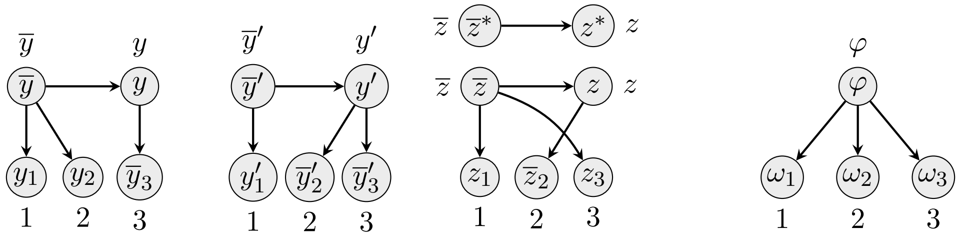


Figure 13: Reduction of the $\text{QBF}^2_\forall$ instance $\Phi = \forall y, y' \exists z \varphi$ with φ given by $\omega_1 = \{y, y', z\}$, $\omega_2 = \{y, \neg y', \neg z\}$, $\omega_3 = \{y, \neg y', z\}$, to an instance of $\mathit{Skept}^{PCAF}_{\mathit{naive}^1_{hyb}}$.

The claim-set to be verified is $C = X_1 \cup \overline{X_1} \cup X_2 \cup \overline{X_2} \cup \{\varphi_1\}$. Observe that $C = \mathit{cl}(A) \setminus (\Omega_1 \cup \Omega_2 \cup \{\varphi_2\})$, i.e., all claims except those of clauses and φ_2 are contained in C. We show that (φ_1, φ_2) is a yes-instance of SAT-UNSAT if and only if $C \in \mathit{prf}_{hyb}(\mathcal{F})$:

Assume (φ_1, φ_2) is a yes-instance of SAT-UNSAT. Then there is an interpretation I such that $I \models \varphi_1$, but there is no interpretation that satisfies φ_2. Thus, φ_2 cannot be part of any admissible extension, since it must be defended against all clause arguments from Ω_2. Let $S = I \cup \{\overline{x} \mid x \in (X_1 \cup X_2) \setminus I\} \cup \{d_x \mid x \in X_1 \cup X_2\} \cup \{\varphi_1\}$. It can be verified that $\mathit{cl}(S) = C$, $S \in \mathit{adm}((A, R))$, and that there is no $S' \in \mathit{adm}((A, R))$ with $\mathit{cl}(S') \supset \mathit{cl}(S)$.

Assume (φ_1, φ_2) is a no-instance of SAT-UNSAT. There are two cases: (1) φ_1 is unsatisfiable. Then φ_1 cannot be part of any admissible extension, i.e., $C \notin \mathit{adm}_{inh}(\mathcal{F})$. (2) Both φ_1 and φ_2 are satisfiable. Since φ_1 and φ_2 share no variables, there is an interpretation I such that $I \models \varphi_1$ and $I \models \varphi_2$. Let $S = I \cup \{\overline{x} \mid x \in (X_1 \cup X_2) \setminus I\} \cup \{d_x \mid x \in X_1 \cup X_2\} \cup \{\varphi_1, \varphi_2\}$. Note that $\mathit{cl}(S) \supset C$ and $S \in \mathit{adm}((A, R))$. Thus, $C \notin \mathit{prf}_{hyb}(\mathcal{F})$. □

It only remains to investigate skeptical acceptance for naive^1_{hyb}, which, as we show, also preserves the higher complexity of general CAFs. This means that Reduction 1 loses the computational benefits of wfCAFs for all semantics considered in this paper.

Proposition 26. *$\mathit{Skept}^{PCAF}_{\sigma^i_\mu}$ is Π^{P}_2-hard for $\sigma^i_\mu = \mathit{naive}^1_{hyb}$, even if we restrict ourselves to PCAFs with transitive preference relations.*

Proof. Let $\Phi = \forall Y \exists Z \varphi$ be an instance of $\text{QBF}^2_\forall$, where φ is given by a set $\Omega = \{\omega_1, \ldots, \omega_m\}$ of clauses over atoms $X = Y \cup Z$. We construct $\mathcal{F} = (A, R, \mathit{cl})$ with

- $A = \{\varphi\} \cup \Omega \cup \{x, \overline{x} \mid x \in X\} \cup \{z^*, \overline{z}^* \mid z \in Z\} \cup$
  $\{x_i \mid x \in X, x \in \omega_i, 1 \leq i \leq m\} \cup \{\overline{x}_i \mid x \in X, \neg x \in \omega_i, 1 \leq i \leq m\}$;
- $R = \{(\varphi, \omega) \mid \omega \in \Omega\} \cup \{(\overline{x}, x) \mid x \in X\} \cup \{(\overline{z}^*, z^*) \mid z \in Z\} \cup$
  $\{(x, \overline{x_i}) \mid x \in X, \neg x \in \omega_i, 1 \leq i \leq m\} \cup \{(\overline{x}, x_i) \mid x \in X, x \in \omega_i, 1 \leq i \leq m\}$;
- $\mathit{cl}(x_i) = \mathit{cl}(\overline{x}_i) = \mathit{cl}(\omega_i) = i$ for $1 \leq i \leq m$,
  $\mathit{cl}(z^*) = z$, $\mathit{cl}(\overline{z}^*) = \mathit{cl}(\overline{z})$, and
  $\mathit{cl}(v) = \mathit{cl}(v)$ for all other $v \in A$.

Figure 13 illustrates the above construction. Note that $\mathcal{F} \in \mathcal{R}_1$-$\mathbf{CAF}_{tr}$ (cf. Proposition 5): the only arguments with both incoming and outgoing edges in $\mathit{wfp}(\mathcal{F})$ are the z^* arguments, with $z \in Z$. The edge leading to z^* comes from $\overline{z}$, and the edge going out of

z^* leads to some $\overline{z}_i$. Thus, there is no cycle in $\mathit{wfp}(\mathcal{F})$. Moreover, the only path in $\mathit{wfp}(\mathcal{F})$ with more than one edge is from $\overline{z}$ to some $\overline{z}_i$, while $(\overline{z}, \overline{z}_i) \notin R$. It remains to verify the correctness of the reduction: we show that Φ is valid iff $\varphi \in C$ for all $C \in \mathit{naive}_{hyb}(\mathcal{F})$.

"$\implies$": Assume Φ is valid. Let $C \in \mathit{naive}_{hyb}(\mathcal{F})$. Note that, for each $y \in Y$, we cannot have $y \in C$ and $\overline{y} \in C$ at the same time. Consider the argument set $S_Y = (Y \cap C) \cup \{\overline{y} \mid y \in Y \setminus C\}$. Note that $\mathit{cl}(S_Y) \supseteq (C \cap \{y, \overline{y} \mid y \in Y\})$. Let $Y' = S_Y \cap Y$. Since Φ is valid there is $Z' \subseteq Z$ such that $I \models \varphi$ for $I = Y' \cup Z'$. Let $S_Z = \{z, \overline{z}^* \mid z \in Z'\} \cup \{\overline{z}, z^* \mid z \in Z \setminus Z'\}$. Note that $\mathit{cl}(S_Z) = \{z, \overline{z} \mid z \in Z\} \supseteq (C \cap \{z, \overline{z} \mid z \in Z\})$. Now let $S_X = S_Y \cup S_Z$ and finally $S = S_X \cup \{x_i \mid x \in S_X, x_i \in A\} \cup \{\overline{x}_i \mid \overline{x} \in S_X, \overline{x}_i \in A\} \cup \{\varphi\}$. Note that $S \in \mathit{cf}((A, R))$ by construction. Moreover, since I satisfies all clauses in Ω we have $\mathit{cl}(S) \supseteq \{1, \ldots, m\} \supseteq (C \cap \{1, \ldots, m\})$. Since also $\varphi \in \mathit{cl}(S)$ we can conclude that $\mathit{cl}(S) \supseteq C$. But $C \in \mathit{naive}_{hyb}(\mathcal{F})$, i.e., it cannot be that $\mathit{cl}(S) \supset C$. Thus, $\mathit{cl}(S) = C$ and therefore $\varphi \in C$.

"$\impliedby$": Assume Φ is not valid. Then there is $Y' \subseteq Y$ such that for all $Z' \subseteq Z$ we have $Y' \cup Z' \not\models \varphi$. Let $S = Y' \cup \{\overline{y} \mid y \in Y \setminus Y'\} \cup \{z, \overline{z}^* \mid z \in Z\} \cup \{\omega_1, \ldots, \omega_m\}$. Towards a contradiction, assume there is $T \in \mathit{cf}((A, R))$ such that $\mathit{cl}(T) \supset \mathit{cl}(S)$. Since for every $y \in Y$ we already have $y \in S$ or $\overline{y} \in S$, and since y and $\overline{y}$ are in conflict, we have $y \in T$ iff $y \in S$ and $\overline{y} \in T$ iff $\overline{y} \in S$. Moreover, since $(\{z, \overline{z}^* \mid z \in Z\} \cup \{1, \ldots, m\}) \subseteq \mathit{cl}(S) \subset \mathit{cl}(T)$ it must be that $\varphi \in T$. This further implies that $\omega_i \notin T$ for all $\omega_i \in \Omega$. Note that for every $z \in Z$ we must have $z, \overline{z}^* \in T$ or $\overline{z}, z^* \in T$ since $\{z, \overline{z}^* \mid z \in Z\} \subset \mathit{cl}(T)$. Let $Z' = T \cap Z$. Since $\{1, \ldots, m\} \subset \mathit{cl}(T)$, we can infer that every clause $\omega_i \in \Omega$ is satisfied by $Y' \cup Z'$, i.e., $Y' \cup Z' \models \varphi$. Contradiction. □

6.2 Efficient Algorithms for Reductions 2–4

We have already seen that the computational benefits of wfCAFs are preserved when using Reductions 2–4 and considering conflict-free/naive semantics (cf. Propositions 20 and 21). In this subsection we show that the benefits of wfCAFs are in fact retained under Reductions 2–4 for the vast majority of admissible-based semantics, with the only exception being complete semantics under Reductions 2 and 4. To do so, we require a more involved algorithm than in the case of conflict-free/naive semantics, since Reductions 2–4 may very well cause certain arguments to be undefended. Consider for example the PCAF $\mathcal{P} = (A, R, \mathit{cl}, \succ)$ with two arguments $A = \{x, y\}$, attacks $R = \{(x, y), (y, x)\}$, claims $\mathit{cl}(x) = x$ and $\mathit{cl}(y) = y$, and the preference $x \succ y$. The preferred claim-extensions before resolving preferences are $\mathit{prf}_{inh}((A, R, \mathit{cl})) = \{\{x\}, \{y\}\}$ while the only preferred claim-extension after resolving preferences is $\mathit{prf}_{inh}(\mathcal{R}_i(\mathcal{P})) = \{\{x\}\}$.

Given a wfCAF $\mathcal{F}$ 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 $\mathcal{F}$ with claim $\mathit{cl}(S) = C$ (Dvořák & Woltran, 2020). Making use of the fact that Reductions 2–4 do not alter conflicts between arguments, we can construct such a maximal set of arguments also for PCAFs: given a PCAF $\mathcal{P}$ and set C of claims, we define the set $E_0(C)$ containing all arguments of $\mathcal{P}$ 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 $\mathcal{P}$; finally, the set $E_*^i(C)$ is obtained by repeatedly removing all arguments not defended by $E_1^i(C)$ in $\mathcal{R}_i(\mathcal{P})$ until a fixed point is reached.

Definition 19. *Given a PCAF $\mathcal{P} = (A, R, cl, \succ)$, a set of claims C, and $i \in \{2, 3, 4\}$, let*

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

The above definition is based on (Dvořák & Woltran, 2020, Definition 5), but with the crucial differences that undefended arguments are (i) computed w.r.t. $\mathcal{R}_i(\mathcal{P})$ and (ii) are iteratively removed until a fixed point is reached.

Lemma 27. *Let $\mathcal{P}$ be a PCAF, C a set of claims, and $i \in \{2, 3, 4\}$. The following holds:*

- *$C \in cf^i_{inh}(\mathcal{P})$ iff $cl(E_1^i(C)) = C$. Moreover, if $C \in cf^i_{inh}(\mathcal{P})$ then $E_1^i(C)$ is the unique maximal conflict-free set S in $\mathcal{R}_i(\mathcal{P})$ such that $cl(S) = C$;*
- *$C \in adm^i_{inh}(\mathcal{P})$ iff $cl(E_*^i(C)) = C$. If $C \in adm^i_{inh}(\mathcal{P})$ then $E_*^i(C)$ is the unique maximal admissible set S in $\mathcal{R}_i(\mathcal{P})$ such that $cl(S) = C$.*

Proof. We consider the two statements separately:

- Conflict-freeness: let $\mathcal{P} = (A, R, cl, \succ)$ be a PCAF. From (Dvořák & Woltran, 2020, Lemma 1) we know that $C \in cf_{inh}((A, R, cl))$ iff $cl(E_1^i(C)) = C$, as well as that, if $C \in cf_{inh}((A, R, cl))$ then $E_1^i(C)$ is the unique maximal conflict-free set S in (A, R, cl) with $cl(S) = C$. From this and our Lemma 6, our result follows immediately.
- Admissibility: let $\mathcal{P} = (A, R, cl, \succ)$ be a PCAF, C a set of claims, and $i \in \{2, 3, 4\}$.

  Assume $cl(E_*^i(C)) = C$. By construction, $E_*^i(C) \in adm(\mathcal{R}_i(\mathcal{P}))$, and thus $C \in adm^i_{inh}(\mathcal{P})$.

  Now assume $C \in adm^i_{inh}(\mathcal{P})$. Then there exists $S \subseteq A$ such that $cl(S) = C$ and $S \in adm(\mathcal{R}_i(\mathcal{P}))$. Furthermore, $C \in cf^i_{inh}(\mathcal{P})$ and therefore $S \subseteq E_1^i(C)$. By construction, $E_*^i(C) \subseteq E_1^i(C)$. Moreover, any $x \in S$ is defended by S in $\mathcal{R}_i(\mathcal{P})$ and therefore also by $E_1^i(C)$. Thus, by definition, $x \in E_2^i(C)$. By the same argument, if $x \in S$ and $x \in E_k^i(C)$ then $x \in E_{k+1}^i(C)$. We can conclude that $S \subseteq E_*^i(C) \subseteq E_1^i(C)$ and thus $cl(E_*^i(C)) = C$. By the above we have that $E_*^i(C)$ is admissible and each $S \subseteq A$ such that $cl(S) = C$ is a subset of $E_*^i(C)$. In other words $E_*^i(C)$ is the unique maximal admissible set S in $\mathcal{R}_i(\mathcal{P})$ such that $cl(S) = C$. □

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

Proposition 28. *$Ver^{PCAF}_{\sigma^i_\mu}$ is in P for $\sigma_\mu \in \{adm_{inh}, stb_{inh}, stb\text{-}adm_{hyb}, stb\text{-}cf_{hyb}\}$ and $i \in \{2, 3, 4\}$, as well as for $\sigma^i_\mu = com^3_{inh}$.*

Proof. Let $\mathcal{P} = (A, R, cl, \succ)$ be a PCAF, let C be a set of claims, and let $i \in \{2, 3, 4\}$. Our goal is to verify that $C \in \sigma^i_\mu(\mathcal{P})$. Note that we can compute $\mathcal{R}_i(\mathcal{P}) = (A, R', cl)$, $E_1^i(C)$, and $E_*^i(C)$ in polynomial time.

- $\sigma^i_\mu = adm^i_{inh}$: by Lemma 27, it suffices to test whether $cl(E^i_*(C)) = C$.

- $\sigma^i_\mu \in \{stb^i_{inh}, stb\text{-}adm^i_{hyb}\}$: note that $stb^i_{inh}(\mathcal{P}) = stb\text{-}adm^i_{hyb}(\mathcal{P})$ for Reductions 2,3,4 (cf. Proposition 7), i.e., we must only verify that $C \in stb^i_{inh}(\mathcal{P})$. We first check whether $C \in adm^i_{inh}(\mathcal{P})$. If not, $C \notin stb^i_{inh}(\mathcal{P})$. If yes, then $cl(E^i_*(C)) = C$ by Lemma 27. We can check in polynomial time if $E^i_*(C) \in stb((A, R'))$. 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 $\mathcal{R}_i(\mathcal{P})$. Moreover, there can be no other $S \in stb((A, R'))$ with $cl(S) = C$ since for any such S we would have $S \subseteq E^i_*(C)$ by Lemma 27, which would imply that S does not attack x and that $x \notin S$.

- $\sigma^i_\mu = stb\text{-}cf^i_{hyb}$: we first check whether $C \in cf^i_{inh}(\mathcal{P})$. If not, $C \notin stb\text{-}cf^i_{hyb}(\mathcal{P})$. If yes, then, by Lemma 27, $cl(E^i_1(C)) = C$. We can check in polynomial time if $E^i_1(C)^{\circledast}_{(A,R')} = cl(A)$. If yes, then $C \in stb\text{-}cf_{hyb}(\mathcal{R}_i(\mathcal{P}))$ and we are done. If no, then there is an argument x such that $x \notin E^i_1(C)$, $cl(x) \notin C$, and x is not attacked by $E^i_1(C)$ in $\mathcal{R}_i(\mathcal{P})$. Moreover, there can be no other $S \in cf((A, R'))$ with $cl(S) = C$ and $S^{\circledast}_{(A,R')} = cl(A)$ since for any such S we would have $S \subseteq E^i_1(C)$ by Lemma 27, which would imply that S does not attack x.

- $\sigma^i_\mu = com^3_{inh}$: we first check if $C \in adm^3_{inh}(\mathcal{P})$. If not, $C \notin com^3_{inh}(\mathcal{P})$. If yes, then $cl(E^3_*(C)) = C$. We can check in polynomial time if $E^3_*(C) \in com(\mathcal{R}_3(\mathcal{P}))$. If no, then $E^3_*(C)$ defends some $x \notin E^3_*(C)$ in $\mathcal{R}_3(\mathcal{P})$. Towards a contradiction, assume there is some $S \subseteq A$ such that $S \in com(\mathcal{R}_3(\mathcal{P}))$ and $cl(S) = C$. By Lemma 27, $S \subseteq E^3_*(C)$, which implies $x \notin S$. Then S cannot defend x in $\mathcal{R}_3(\mathcal{P})$, i.e., there must be y and z such that $y \in E^3_*(C)$, $y \notin S$, $(z, x) \in \mathcal{R}_3(\mathcal{P})$, and $(y, z) \in \mathcal{R}_3(\mathcal{P})$. Then also $(y, z) \in \mathcal{P}$ by the definition of Reduction 3. But there must also be some $y' \in S$ with $cl(y') = cl(y)$, and since the underlying CAF of $\mathcal{P}$ is well-formed there must be $(y', z) \in \mathcal{P}$. Since there cannot be $(y', z) \in \mathcal{R}_3(\mathcal{P})$, otherwise S would defend x, it has to be that $z \succ y'$. For Reduction 3 this further requires $(z, y') \in \mathcal{P}$. Crucially, $(z, y') \in \mathcal{R}_3(\mathcal{P})$. But then S must be defended from z, i.e., there must be some $w \in S$ such that $(w, z) \in \mathcal{R}_3(\mathcal{P})$. But this means that S defends x, i.e., S is not complete. Contradiction. □

Proposition 29. *$Ver^{PCAF}_{\sigma^i_\mu}$ is in* coNP *for $\sigma_\mu \in \{prf_{inh}, prf_{hyb}, sem_{inh}, sem_{hyb}, stg_{inh}, stg_{hyb}\}$ and $i \in \{2, 3, 4\}$.*

Proof. We show that the complementary problem is in NP. Let $\mathcal{P} = (A, R, cl, \succ)$ be a PCAF with $\mathcal{R}_i(\mathcal{P}) = (A, R', cl)$ for $i \in \{2, 3, 4\}$. Let $C \subseteq cl(A)$ be a set of claims. Our algorithm must verify that $C \notin \sigma^i_\mu(\mathcal{P})$ in NP-time. Note that the argument-sets $E^i_1(C)$ and $E^i_*(C)$ can be computed in polynomial time with respect to $\mathcal{P}$ (cf. Definition 19).

- $\sigma^i_\mu \in \{prf^i_{inh}, prf^i_{hyb}, sem^i_{inh}, sem^i_{hyb}\}$: first, guess a set of claims $D \subseteq cl(A)$. Then, check whether $cl(E^i_*(C)) = C$. If no, then, by Lemma 27, $C \notin adm_{inh}(\mathcal{R}_i(\mathcal{P}))$ and we are done. If yes, we proceed differently depending on which semantics we consider:

  - $\sigma^i_\mu = prf^i_{inh}$: verify that $D \in adm^i_{inh}(\mathcal{P})$ and $E^i_*(C) \subset E^i_*(D)$. Since $E^i_*(C)$ is the unique maximal admissible set in $\mathcal{R}_i(\mathcal{P})$ with claim C (cf. Lemma 27), we have $S \subseteq E^i_*(C) \subset E^i_*(D)$ for every $S \in adm((A, R'))$ with $cl(S) = C$. Hence, $C \notin prf^i_{inh}(\mathcal{P})$.

- $\sigma^i_\mu = prf^i_{hyb}$: verify that $D \in adm^i_{inh}(\mathcal{P})$ and $C \subset D$. Then $C \notin prf^i_{hyb}(\mathcal{P})$.
- $\sigma^i_\mu = sem^i_{inh}$: verify that $D \in adm^i_{inh}(\mathcal{P})$ and $E^i_*(C)^{\oplus}_{(A,R')} \subset E^i_*(D)^{\oplus}_{(A,R')}$. As above, we have $S \subseteq E^i_*(C)$ and therefore also $S^{\oplus}_{(A,R')} \subseteq E^i_*(C)^{\oplus}_{(A,R')} \subset E^i_*(D)^{\oplus}_{(A,R')}$ for every $S \in adm((A, R'))$ with $cl(S) = C$. Hence, $C \notin sem^i_{inh}(\mathcal{P})$.
- $\sigma^i_\mu = sem^i_{hyb}$: verify that $D \in adm^i_{inh}(\mathcal{P})$ and $E^i_*(C)^{\circledast}_{\mathcal{R}_i(P)} \subset E^i_*(D)^{\circledast}_{\mathcal{R}_i(P)}$. As above, we have $S \subseteq E^i_*(C)$ and therefore also $S^{\circledast}_{\mathcal{R}_i(P)} \subseteq E^i_*(C)^{\circledast}_{\mathcal{R}_i(P)} \subset E^i_*(D)^{\circledast}_{\mathcal{R}_i(P)}$ for every $S \in adm((A, R'))$ with $cl(S) = C$. Hence, $C \notin sem^i_{hyb}(\mathcal{P})$.

• $\sigma^i_\mu \in \{stg^i_{inh}, stg^i_{hyb}\}$: first, guess a set of claims $D \subseteq cl(A)$. Then, check whether $cl(E^i_1(C)) = C$. If no, then, by Lemma 27, $C \notin cf_{inh}(\mathcal{R}_i(\mathcal{P}))$ and we are done. If yes, we proceed differently depending on which semantics we consider:

- $\sigma^i_\mu = stg^i_{inh}$: verify that $D \in cf^i_{inh}(\mathcal{P})$ and $E^i_1(C)^{\oplus}_{(A,R')} \subset E^i_1(D)^{\oplus}_{(A,R')}$. Since $E^i_1(C)$ is the unique maximal conflict-free argument set in $\mathcal{R}_i(\mathcal{P})$ with claim C (cf. Lemma 27), we have $S \subseteq E^i_1(C)$ and therefore also $S^{\oplus}_{(A,R')} \subseteq E^i_1(C)^{\oplus}_{(A,R')} \subset E^i_1(D)^{\oplus}_{(A,R')}$ for every $S \in cf((A, R'))$ with $cl(S) = C$. Hence, $C \notin stg^i_{inh}(\mathcal{P})$.
- $\sigma^i_\mu = stg^i_{hyb}$: verify that $D \in cf^i_{inh}(\mathcal{P})$ and $E^i_1(C)^{\circledast}_{\mathcal{R}_i(P)} \subset E^i_1(D)^{\circledast}_{\mathcal{R}_i(P)}$. As above, we have $S \subseteq E^i_1(C)$ and therefore also $S^{\circledast}_{\mathcal{R}_i(P)} \subseteq E^i_1(C)^{\circledast}_{\mathcal{R}_i(P)} \subset E^i_1(D)^{\circledast}_{\mathcal{R}_i(P)}$ for every $S \in cf((A, R'))$ with $cl(S) = C$. Hence, $C \notin stg^i_{hyb}(\mathcal{P})$. □

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 30. *$Ver^{PCAF}_{\sigma^i_\mu}$ is* NP*-hard for $\sigma_\mu = com_{inh}$ and $i \in \{2, 4\}$, even if we restrict ourselves to PCAFs with transitive preference relations.*

Proof. We show NP-hardness for $\sigma^i_\mu = com^4_{inh}$. The proof for $\sigma^i_\mu = com^2_{inh}$ is similar and can be found in Appendix B (Lemma 45).

Let φ be an arbitrary instance of 3-SAT given as a set Ω of clauses over variables X and let $\overline{X} = \{\overline{x} \mid x \in X\}$. We construct a PCAF $\mathcal{P} = (A, R, cl, \succ)$ as well as a set of claims C:

• $A = \{\varphi\} \cup \Omega \cup X \cup \overline{X} \cup \{a_x \mid x \in X \cup \overline{X}\} \cup \{b_x \mid x \in X\}$;

• $R = \{(\omega, \varphi) \mid \omega \in \Omega\} \cup \{(\omega, \omega) \mid \omega \in \Omega\} \cup$
$\{(\omega, x) \mid x \in \omega, \omega \in \Omega\} \cup \{(\omega, \overline{x}) \mid \neg x \in \omega, \omega \in \Omega\} \cup$
$\{(a_x, x) \mid x \in X \cup \overline{X}\} \cup \{(a_x, b_x), (a_{\overline{x}}, b_x) \mid x \in X\}$;

• $cl(x) = cl(\overline{x}) = x$ for $x \in X$, $cl(v) = v$ otherwise;

• $x \succ \omega$, $x \succ a_x$ for all $x \in X \cup \overline{X}$ and all $\omega \in \Omega$;

• $C = X \cup \{\varphi\}$.

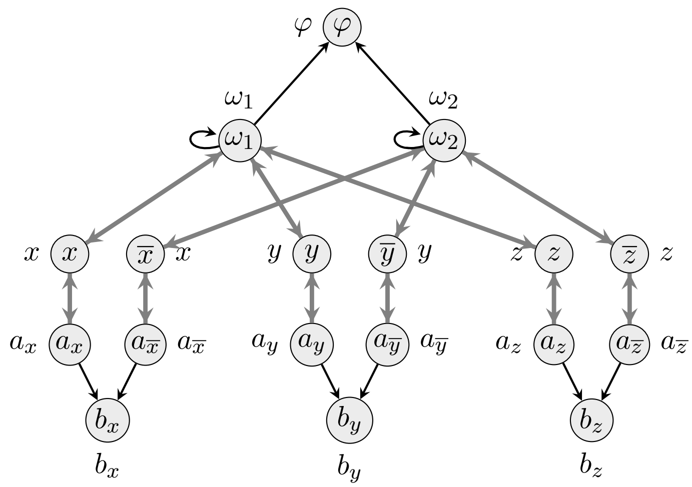


Figure 14: $\mathcal{R}_4(\mathcal{P})$ from the proof of Proposition 30, with φ given by clauses $\omega_1 = \{x, y, z\}$, $\omega_2 = \{\neg x, \neg y, \neg z\}$. Symmetric attacks (gray/thick) have been introduced by Reduction 4.

Figure 14 illustrates the above construction. It remains to show that φ is satisfiable if and only if $C \in com_{inh}(\mathcal{R}_4(\mathcal{P}))$.

Assume φ is satisfiable. Then there is an interpretation I such that $I \models \varphi$. Let $S = \{x \mid x \in X, x \in I\} \cup \{\overline{x} \mid x \in X, x \notin I\} \cup \{\varphi\}$. Clearly, $cl(S) = C$. Furthermore, S defends φ in $\mathcal{R}_4(\mathcal{P})$ since each clause is satisfied by I, and thus each clause argument ω_j is attacked by some x (or $\overline{x}$) in S. Each variable $x \in X$ clearly defends itself. Moreover, if $x \in S$, then $\overline{x} \notin S$ and none of b_x, $\overline{x}$, or $a_{\overline{x}}$ is defended by S. Analogously for the case that $\overline{x} \in S$. Thus, S is admissible, and contains all arguments it defends, i.e., $S \in com(\mathcal{R}_4(\mathcal{P}))$.

Assume $C \in com_{inh}(\mathcal{R}_4(\mathcal{P}))$. Then there is $S \subseteq A$ such that $cl(S) = C$ and $S \in com(\mathcal{R}_4(\mathcal{P}))$. For each $x \in X$, at least one of $x, \overline{x}$ must be contained in S. In fact, it cannot be that $x \in S$ and $\overline{x} \in S$, otherwise b_x would be defended by S and we would have $cl(S) \neq C$. Thus, for each $x \in X$, there is either $x \in S$ or $\overline{x} \in S$, but not both. Furthermore, S defends φ, i.e., S attacks all clause arguments ω_j. Thus, $I \models \varphi$ for $I = X \cap S$. □

6.3 Summary and Impact of Complexity Results

When using Reduction 1 we obtain the same complexity as for general CAFs, i.e., the benefits of wfCAFs are lost. On the other hand, Reductions 2–4 preserve the lower complexity of wfCAFs for almost all semantics. Intuitively, this can be explained by the fact that these reductions do not remove conflicts between arguments. This in turn means that Reductions 2–4 retain enough of the structure of wfCAFs in order to, given a claim, efficiently compute a subset-maximal admissible argument set with that claim. The only outlier are complete semantics, for which verification remains hard under Reductions 2 and 4 but not Reduction 3. Here, the fact that Reductions 2 and 4 can introduce new attacks leads to an increase in complexity. We conclude:

Theorem 31. *The complexity results in Table 2 hold, even if we restrict ourselves to PCAFs with transitive preference relations.*

Table 2: Computational Complexity of PCAFs. Results in boldface had to be proven explicitly. All other results follow directly from known properties (cf. Observation 19).

σ_μ^i	$Cred_{\sigma_\mu^i}^{PCAF}$	$Skept_{\sigma_\mu^i}^{PCAF}$		$Ver_{\sigma_\mu^i}^{PCAF}$		
	$i \in \{1,2,3,4\}$	$i=1$	$i \in \{2,3,4\}$	$i=1$	$i \in \{2,4\}$	$i=3$
cf_{inh}	in P	trivial		**NP-c**	**in P**	
adm_{inh}	NP-c	trivial		**NP-c**	**in P**	
com_{inh}	NP-c	P-c		**NP-c**		**in P**
grd_{inh}	in P	in P		in P		
stb_{inh}, $stb\text{-}adm_{hyb}$, $stb\text{-}cf_{hyb}$	NP-c	coNP-c		**NP-c**	**in P**	
$naive_{inh}$	in P	coNP-c		**NP-c**	**in P**	
$naive_{hyb}$		**Π_2^P-c**	**coNP-c**	**DP-c**		
prf_{inh}	NP-c	Π_2^P-c		**Σ_2^P-c**	**coNP-c**	
prf_{hyb}				**DP-c**		
sem_{inh}, sem_{hyb}	Σ_2^P-c	Π_2^P-c		**Σ_2^P-c**	**coNP-c**	
stg_{inh}, stg_{hyb}	Σ_2^P-c	Π_2^P-c		**Σ_2^P-c**	**coNP-c**	

The lower complexity of the verification problem is crucial for enumerating claim-extensions in wfCAFs (Dvořák & Woltran, 2020). Indeed, this is also true for PCAFs using Reductions 2–4. If claim sets can be verified in polynomial time we can simply iterate through all claim sets. For preferred, semi-stable, and stage semantics the algorithm builds heavily on the existence and polynomial-time computability of unique maximal realizations for conflict-free and admissible claim-sets, i.e., $E_1^i(C)$ and $E_*^i(C)$ (cf. Definition 19).

Proposition 32. *Consider PCAFs $\mathcal{P} = (A, R, cl, \succ)$ with $|A| \leq n$ and $|cl(A)| \leq k$.*

- *If $Ver_{\sigma_\mu^i}^{PCAF}$ is in* P *for a PCAF-semantics σ_μ^i, then there is a polynomial $poly(\cdot)$ such that $\sigma_\mu^i(\mathcal{P})$ can be enumerated in $O(2^k \cdot poly(n))$ time.*
- *For σ_μ^i with $\sigma_\mu \in \{prf_{inh}, prf_{hyb}, sem_{inh}, sem_{hyb}, stg_{inh}, stg_{hyb}\}$ and $i \in \{2,3,4\}$ there is a polynomial $poly(\cdot)$ such that $\sigma_\mu^i(\mathcal{P})$ can be enumerated in $O(4^k \cdot poly(n))$ time.*

Proof. If $Ver_{\sigma_\mu^i}^{PCAF}$ is in P we can iterate through all 2^k claim-sets $C \subseteq cl(A)$ and check whether $C \in \sigma^i(\mathcal{F})$ in polynomial time. This procedure runs in $O(2^k \cdot poly(n))$ time.

For σ_μ^i with $\sigma_\mu \in \{prf_{inh}, prf_{hyb}, sem_{inh}, sem_{hyb}, stg_{inh}, stg_{hyb}\}$ and $i \in \{2,3,4\}$, recall the proof of Proposition 29. There, to decide that $C \notin \sigma_\mu^i(\mathcal{P})$, we guessed a claim-set $D \subseteq cl(A)$ and performed some checks in polynomial time. Instead of guessing D, we can iterate through all 2^k claim-sets $D \subseteq cl(A)$. If $C \in adm_{inh}(\mathcal{P})$ (resp. $C \in cf_{inh}(\mathcal{P})$ in case $\sigma^i_\mu \in \{stg^i_{inh}, stg^i_{hyb}\}$), and if no D that witnesses $C \notin \sigma^i_\mu(\mathcal{P})$ is found, we have $C \in \sigma^i_\mu(\mathcal{P})$. Therefore, to enumerate $\sigma^i_\mu(\mathcal{P})$ we can iterate through all $(2^k)^2 = 4^k$ pairs (C, D) of claim-sets. This procedure runs in $O(4^k \cdot poly(n))$ time. □

Proposition 32 directly implies that deciding the main decision problems is tractable if the number of claims is bounded by a constant k. In particular, these problems are fixed parameter tractable (FPT) with respect to the number of claims.

Corollary 33. *For all PCAF-semantics σ^i_μ considered in this paper, except for those such that $i = 1$ and except for com^2_{inh} and com^4_{inh}, there is a polynomial $poly(\cdot)$ such that $Cred^{PCAF}_{\sigma^i_\mu}$, $Skept^{PCAF}_{\sigma^i_\mu}$, and $Ver^{PCAF}_{\sigma^i_\mu}$ can be solved in time $O(4^k \cdot poly(n))$ for PCAFs $(A, R, cl, \succ)$ with $|cl(A)| \leq k$.*

7. Excursion: Instantiating ABA$^+$

PCAFs and CAFs are a natural target formalism for many structured argumentation formalisms in which claims/conclusions play a central role. As we have seen, the semantical and computational properties of a given framework depend on how preferences are dealt with, i.e., which preference reduction is used. Thus, given a CAF instantiated from some structured formalism, we can infer some of its properties by checking which CAF-class it belongs to, i.e., which preference reduction has been used to obtain the CAF.

Preferences play an important role in structured argumentation formalisms (Modgil & Prakken, 2018; Cyras, Fan, Schulz, & Toni, 2018). Structured formalisms allow for qualitative comparisons between defeasible elements of the knowledge base, e.g., between assumptions (Cyras et al., 2018) or defeasible rules (Modgil & Prakken, 2018). Hence, in contrast to abstract argumentation, preferences are usually not directly given between arguments. Similar to abstract formalisms, popular preference incorporation techniques alter the attack relation before the outcome of the framework is computed, given a specific semantics. In ASPIC$^+$, for instance, preferences result in the deletion of attacks, similar to Reduction 1. In this work, we focus on preference incorporation techniques in *assumption-based argumentation (ABA)* (Bondarenko, Dung, Kowalski, & Toni, 1997) which is one of the most prominent structured argumentation formalism. Central concepts of ABA are assumptions, their contraries and inference rules; notions of attacks, defense, and semantics in ABA are defined on sets of assumptions. ABA and abstract formalisms such as AFs and CAFs are closely related; instantiating an ABA framework as (C)AF preserves the assumption-based semantics of the framework (Cyras et al., 2018; König, Rapberger, & Ulbricht, 2022). In ABA, arguments are built from rules and assumptions; each argument has a claim that determines outgoing attacks of the arguments: if the claim corresponds to the contrary of an assumption a, then each argument that contains a is attacked. Hence, arguments with the same claim attack the same arguments. Indeed, each ABA framework satisfies the well-formedness condition, as already observed by Kö nig, Rapberger and Ulbricht (2022). This fundamental property is however violated if preferences are taken into account. ABA with preferences (ABA$^+$) (Cyras et al., 2018) incorporates preferences between assumptions by reversing attacks between assumption sets, which can lead to a violation of well-formedness.

In this section, we examine the close relation between ABA$^+$ and PCAFs. Specifically, we show that, for a prominent fragment of ABA$^+$, the application of Reduction 2 in the PCAF captures the way how preferences are handled on the assumption level in the corresponding ABA$^+$ framework. It is well-known that generalized ABA formalisms, including ABA$^+$, often cannot be instantiated via the classical AF model. Bao, Cyras, and Toni (2017) show that, under certain restrictions, it is possible to instantiate ABA$^+$ frameworks as AFs by using an instantiation method that associates sets of assumptions with arguments. The central condition that is necessary to accomplish this is called *weak contraposition*. In their work, the authors show that the presented AF instantiation preserves the classical complete-based Dung semantics (grounded, complete, preferred, stable semantics) if the ABA$^+$ framework satisfies weak contraposition. In this work, we show that PCAFs admit an even closer correspondence to the ABA$^+$ fragment. We show that each ABA$^+$ framework satisfying weak contraposition can be instantiated as a PCAF which preserves inherited grounded, complete, preferred, and stable semantics under Reduction 2. We furthermore show that hybrid semi-stable semantics preserve the native ABA semi-stable semantics, i.e., we provide an abstract counterpart for assumption-based semi-stable semantics.

Below, we recall the necessary background before presenting our instantiation. We need to pay special attention to the particular way in which the preferences on the assumptions are lifted to the arguments, because the preference relation is no longer strict. In Section 7.2, we present our main result and show that this instantiation preserves the classical complete-based inherited Dung semantics (grounded, complete, preferred, stable semantics) and hybrid semi-stable semantics under Reduction 2.

7.1 Background

In this section, we recall *assumption-based argumentation with preferences* (Cyras et al., 2018). We assume a deductive system $(\mathcal{L}, \mathcal{R})$, where $\mathcal{L}$ is a formal language, i.e., a set of sentences, and $\mathcal{R}$ is a set of inference rules over $\mathcal{L}$. A rule $r \in \mathcal{R}$ has the form

$$a_0 \leftarrow a_1, \ldots, a_n$$

where $a_i \in \mathcal{L}$, $head(r) = a_0$ is the head, and $body(r) = \{a_1, \ldots, a_n\}$ is the (possibly empty) body of r.

Definition 20 (ABA$^+$ framework)**.** *An ABA$^+$ framework is a tuple $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$, where*

- *$(\mathcal{L}, \mathcal{R})$ is a deductive system,*
- *$\mathcal{A} \subseteq \mathcal{L}$ a non-empty set of assumptions,*
- *$^-$ is a function mapping assumptions $a \in \mathcal{A}$ to sentences $\mathcal{L}$ (contrary function),*
- *and $\leq$ is a transitive binary relation on $\mathcal{A}$.*

As usual, we write $a < b$ if $a \leq b$ and $b \not\leq a$. $\mathcal{D}$ is an ABA framework (without preferences) if $\leq$ is empty.

In this work, we focus on frameworks which are *flat*, i.e., $head(r) \notin \mathcal{A}$ for each rule $r \in \mathcal{R}$, and *finite*, i.e., $\mathcal{L}$, $\mathcal{R}$, $\mathcal{A}$ are finite.

Definition 21 (Tree-derivation). *A sentence $p \in \mathcal{L}$ is* tree-derivable *from assumptions $\mathcal{U} \subseteq \mathcal{A}$ and rules $\mathcal{R}' \subseteq \mathcal{R}$, denoted by $\mathcal{U} \vdash_{\mathcal{R}'} p$, if there is a finite rooted labeled tree $\mathcal{V}$ such that the root is labeled with p, the set of labels for the leaves of $\mathcal{V}$ is equal to $\mathcal{U}$ or $\mathcal{U} \cup \{\top\}$, and there is a surjective mapping from the set of internal nodes to $\mathcal{R}'$ satisfying for each internal node q there is a rule $r \in \mathcal{R}'$ such that q is labeled with $head(r)$ and the set of all successor nodes corresponds to $body(r)$ or $\top$ if $body(r) = \emptyset$.*

Definition 22 (ABA arguments). *We call $\mathcal{U} \vdash p$ an* (ABA) argument *iff there is a tree-derivation $\mathcal{U} \vdash_{\mathcal{R}'} p$ for some set of rules $\mathcal{R}' \subseteq \mathcal{R}$.*

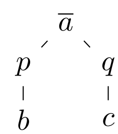
Each assumption $a \in \mathcal{A}$ gives rise to an argument $\{a\} \vdash a$. Since we consider only flat arguments, $\{a\} \vdash a$ is the unique argument with claim a in a given ABA framework.

Example 7. *Let $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$ be an ABA$^+$ framework with $\mathcal{L} = \{a, b, c, p, q, \overline{a}, \overline{b}, \overline{c}\}$, assumptions $\mathcal{A} = \{a, b, c\}$, their contraries $\overline{a}$, $\overline{b}$ and $\overline{c}$, preference $b < a$, and rules*

$$\overline{a} \leftarrow p, q \qquad p \leftarrow b \qquad q \leftarrow c \qquad \overline{b} \leftarrow a \qquad \overline{c} \leftarrow b.$$

We can derive the contrary $\overline{a}$ from assumptions b, c via the following tree derivation

$\{b, c\} \vdash_{\{\overline{a} \leftarrow p,q; p \leftarrow b; q \leftarrow c\}} \overline{a}$:



The leaves are labeled with b and c, and the root of the tree is $\overline{a}$. The tree derivation corresponds to the argument $\{b, c\} \vdash \overline{a}$. Further arguments of our $\mathcal{D}$ are $\{a\} \vdash a$, $\{b\} \vdash b$, $\{c\} \vdash c$ (corresponding to the assumptions a, b, and c, respectively), $\{a\} \vdash \overline{b}$, $\{b\} \vdash \overline{c}$, $\{b\} \vdash o$, and $\{c\} \vdash q$.

Definition 23 (ABA$^+$ attacks). *Given an ABA$^+$ framework $(\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$, a set of assumptions $\mathcal{U} \subseteq \mathcal{A}$* attacks *a set of assumptions $\mathcal{V} \subseteq \mathcal{A}$ iff*

1. *there is some $\mathcal{U}' \subseteq \mathcal{U}$ such that $\mathcal{U}' \vdash \overline{q}$ for some $q \in \mathcal{V}$, and there is no $p \in \mathcal{U}'$ with $p < q$ (we say that $\mathcal{U}$ normally attacks $\mathcal{V}$); or*
2. *there is some $\mathcal{V}' \subseteq \mathcal{V}$ such that $\mathcal{V}' \vdash \overline{p}$ for some $p \in \mathcal{U}$, and there is some $q \in \mathcal{V}'$ with $q < p$ (we say that $\mathcal{U}$ reversely attacks $\mathcal{V}$).*

For ABA without preferences, only normal attacks can take place: a set of assumptions $\mathcal{U}$ attacks another set of assumptions $\mathcal{V}$ iff (a subset of) $\mathcal{U}$ derives a contrary of some assumption in $\mathcal{V}$. Taking preferences into account might result into an attack reversal, as formalized in item two.

Example 8. *Consider again our ABA$^+$ framework $\mathcal{D}$ from Example 7. The set $\mathcal{U} = \{a\}$ reversely attacks the set $\mathcal{V} = \{b, c\}$ because $\{b, c\} \vdash \overline{a}$ and $b < a$, i.e., $\mathcal{V}$ contains an element which is weaker than a. As a result, the attack is reversed.*

Let $\mathcal{U}^+_{\mathcal{D}} = \{x \in \mathcal{A} \mid \mathcal{U} \text{ attacks } \{x\}\}$ denote the set of all attacked assumptions. We drop subscript $\mathcal{D}$ if clear from context.

Definition 24 (ABA$^+$ conflict-freeness and defense)**.** *For a set of assumptions $\mathcal{U} \subseteq \mathcal{A}$,*

- *$\mathcal{U}$ is conflict-free iff it does not attack itself;*
- *$\mathcal{U}$ defends itself iff $\mathcal{U}$ attacks all attackers $\mathcal{V} \subseteq \mathcal{A}$ of $\mathcal{U}$.*

Next we recall admissible, complete, grounded, preferred, stable, and semi-stable semantics for ABA$^+$ (abbreviated by *adm*, *com*, *grd*, *prf*, *stb*, *sem*).

Definition 25 (ABA$^+$ semantics)**.** *Let $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$ be an ABA$^+$ framework. Further, let $\mathcal{U} \subseteq \mathcal{A}$ be conflict-free. Then*

- *$\mathcal{U} \in adm(\mathcal{D})$ iff $\mathcal{U}$ defends itself;*
- *$\mathcal{U} \in com(\mathcal{D})$ iff $\mathcal{U}$ is admissible and contains every assumption set it defends;*
- *$\mathcal{U} \in grd(\mathcal{D})$ iff $\mathcal{U}$ is $\subseteq$-minimal in $com(\mathcal{D})$;*
- *$\mathcal{U} \in prf(\mathcal{D})$ iff $\mathcal{U}$ is $\subseteq$-maximal in $com(\mathcal{D})$;*
- *$\mathcal{U} \in stb(\mathcal{D})$ iff $\mathcal{U}$ attacks each $\{x\}$ for every $x \in \mathcal{A} \setminus \mathcal{U}$;*
- *$\mathcal{U} \in sem(\mathcal{D})$ iff $\mathcal{U} \in adm(\mathcal{D})$ and $\mathcal{U} \cup \mathcal{U}^+_{\mathcal{D}}$ is $\subseteq$-maximal among admissible sets.*

Notation 34. *Let $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$ be an ABA$^+$ framework. For an argument $x = \mathcal{U} \vdash p$, we use $asms(x) = \mathcal{U}$ and $cl(x) = p$ to denote the assumptions and the claim of x, respectively. We generalize the notion to sets of arguments as expected, i.e., for a set of arguments X, we let $asms(X) = \bigcup_{x \in X} asms(x)$ and $cl(X) = \bigcup_{x \in X} cl(x)$. For a set $\mathcal{U} \subseteq \mathcal{A}$, we let $\overline{\mathcal{U}} = \{\overline{a} \mid a \in \mathcal{U}\}$.*

Example 9. *Let $\mathcal{D}$ be the ABA$^+$ framework from Example 7. The admissible extensions of $\mathcal{D}$ are $\emptyset$, $\{a\}$, and $\{a, c\}$: first note that the set $\{a\}$ is unattacked in $\mathcal{D}$. Since $b < a$ the attack from $\{b, c\}$ to $\{a\}$ is reversed; hence, $\{a\}$ attacks $\{b, c\}$. The set $\{a, c\}$ is admissible because $\{a\}$ defends $\{c\}$ against $\{b\}$. The remaining sets are not admissible: $\{b\}$ cannot defend itself against the attack from $\{a\}$; likewise, $\{c\}$ cannot defend itself against $\{b\}$; $\{a, b\}$ is conflicting and $\{b, c\}$ does not defend itself. It can be checked that $\{a, c\}$ is furthermore the unique complete, grounded, preferred, semi-stable and stable extension.*

It is well known that each ABA framework without preferences can be captured by an AF (Cyras et al., 2018). In this work, we instantiate only arguments that either derive the contrary of an assumption or that correspond to a single assumption (following Bao, Cyras, and Toni (2017)).

Notation 35. *We write $Args_{\mathcal{U}} = \{\mathcal{U}' \vdash p, \mid \mathcal{U}' \subseteq \mathcal{U}, p \in \mathcal{A} \cup \overline{\mathcal{A}}\}$ to denote the set of all arguments having either a contrary or an assumption as claim that can be obtained from $\mathcal{U}$.*

Definition 26 (AF instantiation of ABA$^+$)**.** *The associated AF $F_{\mathcal{D}} = (A, R)$ of an ABA$^+$ framework $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$ is given by $A = Args_{\mathcal{A}}$ and attack relation $(\mathcal{U} \vdash p, \mathcal{U}' \vdash p') \in R$ iff $p \in \overline{\mathcal{U}'}$.*

Since each assumption $a \in \mathcal{A}$ gives rise to argument $\{a\} \vdash a$, each assumption is represented on claim-level by a unique argument (recall that we consider flat frameworks).

We recall the correspondence between ABA without preferences and abstract argumentation (Cyras et al., 2018, Theorem 4.3).

Theorem 36. *Given an ABA framework without preferences $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$, its corresponding AF $F_\mathcal{D}$ and a semantics $\sigma \in \{grd, com, prf, stb\}$, the following holds:*

- *if $S \in \sigma(F_\mathcal{D})$ then $asms(S) \in \sigma(\mathcal{D})$;*
- *if $\mathcal{U} \in \sigma(\mathcal{D})$ then $Args_\mathcal{U} \in \sigma(F_\mathcal{D})$.*

The result has been extended to well-formed CAFs for conclusion extensions of ABA without preferences (König et al., 2022). However, when considering preferences, the correspondence is no longer preserved. The underlying issue can be traced back to the violation of the *fundamental lemma*, stating that if an admissible set $\mathcal{U}$ defends an assumption a, then $\mathcal{U} \cup \{a\}$ is admissible as well. A sufficient criteria for the fundamental lemma to hold is the *Axiom of Weak Contraposition* (Bao et al., 2017).

Definition 27 (Weak contraposition)**.** *An ABA$^+$ framework $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$ satisfies the* Axiom of Weak Contraposition (WCP) *iff for all $\mathcal{U} \subseteq \mathcal{A}$ and $b \in \mathcal{A}$, it holds that, if $\mathcal{U} \vdash \overline{b}$ and there is $a' \in \mathcal{U}$ with $a' < b$, then, for some $\leq$-minimal $a \in \mathcal{U}$ with $a < b$, there is $\mathcal{U}' \subseteq (\mathcal{U} \setminus \{a\}) \cup \{b\}$ such that $\mathcal{U}' \vdash \overline{a}$.*

Example 10. *The ABA$^+$ framework in Example 7 satisfies WCP: it holds that $\{b, c\} \vdash \overline{a}$ and $b < a$, hence, WCP is satisfied if there is some argument $\mathcal{U} \vdash \overline{b}$ with $\mathcal{U} \subseteq (\{b, c\} \setminus \{b\}) \cup \{a\}$. Since $\{a\} \vdash \overline{b}$, we indeed obtain that WCP is satisfied in $\mathcal{D}$.*

Bao, Cyras, and Toni (2017) show that the semantic correspondence between ABA$^+$ instances and an abstract representation can be preserved if WCP is satisfied. Instead of the standard instantiation (Definition 26), they focus on the so-called *assumption graph* which is obtained by computing all ABA arguments and extracting the corresponding set of assumptions.

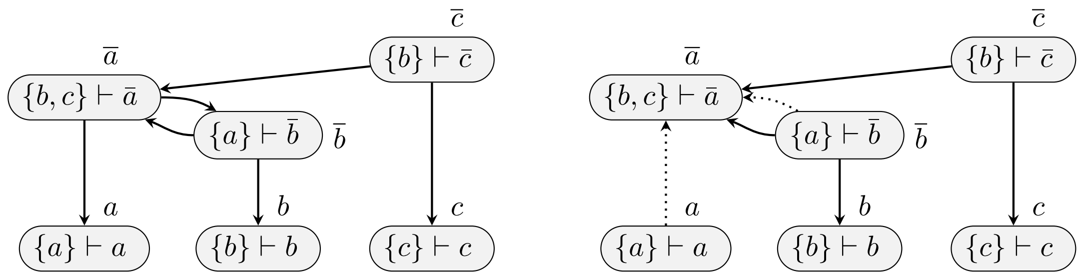
7.2 Capture ABA$^+$ as PCAF

In this subsection, we show that PCAFs can be used to more directly capture ABA$^+$ when WCP is satisfied. In contrast to Bao, Cyras, and Toni (2017), we use the standard ABA$^+$-instantiation, which demonstrates that PCAFs constitute a natural formalism to directly capture ABA$^+$. Moreover, we obtain a semantical correspondence also for semi-stable semantics.

As a first step, we will lift the preference relation of ABA$^+$ from sets of assumptions to sets of arguments.

Definition 28 (Preferences over ABA$^+$ arguments)**.** *Let $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$ be an ABA$^+$ framework. We define the preference relation $\preceq$ over the set of all arguments of $\mathcal{D}$ as follows:*

$$\mathcal{U} \vdash p \preceq \mathcal{U}' \vdash p' \quad \Leftrightarrow \quad \overline{a} = p \text{ for some } a \in \mathcal{U}' \text{ and there is some } b \in \mathcal{U} \text{ s.t. } b < a.$$



(a) PCAF $\mathcal{P}_\mathcal{D}$, $\{b,c\} \vdash \overline{a} \prec \{a\} \vdash a$, $\{b,c\} \vdash \overline{a} \prec \{a\} \vdash \overline{b}$. (b) CAF $\mathcal{F}_\mathcal{D}$. Reversed attacks are dotted.

Figure 15: PCAF and CAF instantiation of the ABA$^+$ framework $\mathcal{D}$ from Example 7.

We write $\mathcal{U} \vdash p \prec \mathcal{U}' \vdash p'$ iff $\mathcal{U} \vdash p \preceq \mathcal{U}' \vdash p'$ and $\mathcal{U}' \vdash p' \not\preceq \mathcal{U} \vdash p$. Moreover, we write $\mathcal{U} \vdash p \approx \mathcal{U}' \vdash p'$ iff $\mathcal{U} \vdash p \preceq \mathcal{U}' \vdash p'$ and $\mathcal{U}' \vdash p' \preceq \mathcal{U} \vdash p$. We observe that the resulting ordering is not strict, that is, it can indeed be the case that $\mathcal{U} \vdash p \approx \mathcal{U}' \vdash p'$ for two ABA arguments.

Example 11. *Consider the arguments $\{a,b\} \vdash \overline{c}$, $\{c,d\} \vdash \overline{b}$, and the preference relation $a < c$ and $d < b$. The first inequality yields $\{a,b\} \vdash \overline{c} \preceq \{c,d\} \vdash \overline{b}$, and the second inequality yields $\{a,b\} \vdash \overline{c} \succeq \{c,d\} \vdash \overline{b}$.*

Since these equalities between arguments can only appear in case the arguments symmetrically attack each other, it is safe to simply ignore these cases. For our PCAF instantiation, we consider only the strict preference relations between ABA arguments.

Definition 29 (PCAF instantiation of ABA$^+$)**.** *Let $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$ be an ABA$^+$ framework. Let $F_\mathcal{D} = (A, R)$ be the associated AF of $\mathcal{D}$ (cf. Definition 26). The associated PCAF $\mathcal{P}_\mathcal{D}$ of $\mathcal{D}$ is given by $\mathcal{P}_\mathcal{D} = (A, R, cl, \succ)$ with claim function $cl(\mathcal{U} \vdash p) = p$ and preference relation*

$$\mathcal{U} \vdash p \prec \mathcal{U}' \vdash p' \quad \Leftrightarrow \quad \mathcal{U} \vdash p \preceq \mathcal{U}' \vdash p' \text{ and } \mathcal{U}' \vdash p' \not\preceq \mathcal{U} \vdash p.$$

In terms of the standard construction, the CAF associated to an ABA framework with preferences looks as follows:

Definition 30 (CAF instantiation of ABA$^+$)**.** *Let $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$ be an ABA$^+$ framework. We define the associated CAF $\mathcal{F}_\mathcal{D} = (A, R, cl)$ with $A = Args_\mathcal{A}$, $R = R_{nor} \cup R_{rev}$ where*

$$R_{nor} = \{(\mathcal{U} \vdash p, \mathcal{U}' \vdash p') \mid p = \overline{a} \in \overline{\mathcal{U}'} \text{ and there is no } b \in \mathcal{U} \text{ with } b < a\}$$
$$R_{rev} = \{(\mathcal{U} \vdash p, \mathcal{U}' \vdash p') \mid p' = \overline{a}' \in \overline{\mathcal{U}} \text{ and there is } b \in \mathcal{U}' \text{ such that } b < a'\}$$

and claim function $cl(\mathcal{U} \vdash p) = p$. We call attacks in R_{nor} normal attacks *and R_{rev}* reversed attacks *between arguments.*

Applying Reduction 2 to a PCAF $\mathcal{P}_\mathcal{D}$ yields $\mathcal{F}_\mathcal{D}$, as expressed by the following proposition (see Appendix C for the proof).

Proposition 37. *Let $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$ be an ABA$^+$ framework, let $\mathcal{P}_\mathcal{D} = (A, R, cl, \prec)$ be the corresponding PCAF (cf. Definition 29), and let $\mathcal{F}_\mathcal{D} = (A', R', cl')$ be the corresponding CAF (cf. Definition 30). It holds that $\mathcal{R}_2(\mathcal{P}_\mathcal{D}) = \mathcal{F}_\mathcal{D}$.*

Figure 15 shows the PCAF $\mathcal{P}_\mathcal{D}$ and CAF $\mathcal{F}_\mathcal{D}$ instantiated from the ABA$^+$ framework $\mathcal{D}$ of Example 7.

To prove that PCAFs capture ABA$^+$ instances if WCP is satisfied, we will first show that crucial properties of the AF standard instantiation can be preserved when moving from ABA to ABA with preferences under WCP. We show that each admissible set of arguments induces an admissible set of assumptions; moreover, for each admissible set of assumptions $\mathcal{U}$, it holds that the set of all arguments $Args_\mathcal{U}$ that can be constructed is admissible in the corresponding AF. Most crucially, as formalized in the following proposition, we show that each of complete, grounded, preferred, and stable ABA semantics are in one-to-one correspondence to the respective AF semantics (see Appendix C for the proof).

Proposition 38. *Let $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$ be an ABA$^+$ framework that satisfies weak contraposition, $\mathcal{F}_\mathcal{D} = (A, R, cl)$ its associated CAF, and $F_\mathcal{D} = (A, R)$ the underlying AF, and let $\sigma \in \{grd, com, prf, stb\}$ be a semantics.*

- *If $S \in \sigma(F_\mathcal{D})$ then $asms(S) \in \sigma(\mathcal{D})$;*
- *if $\mathcal{U} \in \sigma(\mathcal{D})$ then $Args_\mathcal{U} \in \sigma(F_\mathcal{D})$;*
- *the correspondence is one-to-one; i.e., $S = Args_{asms(S)}$ for each $S \in \sigma(F_\mathcal{D})$ and $\mathcal{U} = asms(Args_\mathcal{U})$ for all $\mathcal{U} \in \sigma(\mathcal{D})$.*

In contrast to complete, grounded, preferred, and stable semantics, semi-stable semantics for ABA and AF cannot be translated into each other. As discussed by Caminada et al. (2015), semi-stable semantics are not preserved, even for ABA without preferences. However, as recently shown (Rapberger, 2023), hybrid semi-stable CAF semantics provide an abstract counterpart for semi-stable ABA semantics for ABA without preferences. We show that this correspondence can be preserved when moving to ABA$^+$ frameworks that satisfy WCP.

We are ready to give the proof for the considered semantics.

Theorem 39. *Let $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$ be an ABA$^+$ framework that satisfies weak contraposition, and let $\mathcal{P}_\mathcal{D} = (A, R, cl, \succ)$ be the associated PCAF of $\mathcal{D}$. It holds that*

$$\sigma(\mathcal{D}) = \{C \cap \mathcal{A} \mid C \in \sigma^2_{inh}(\mathcal{P}_\mathcal{D})\} \text{ for } \sigma \in \{grd, com, prf, stb\},$$
$$sem(\mathcal{D}) = \{C \cap \mathcal{A} \mid C \in sem^2_{hyb}(\mathcal{P}_\mathcal{D})\}.$$

Proof. First, we show the statement for $\sigma \in \{grd, com, prf, stb\}$.

($\subseteq$) Let $\mathcal{U} \in \sigma(\mathcal{D})$. From Proposition 38, we obtain that $Args_\mathcal{U} \in \sigma(F_\mathcal{D})$, where $F_\mathcal{D}$ denotes the corresponding AF of $\mathcal{D}$. By Proposition 37, it holds that $\mathcal{R}_2(\mathcal{P}_\mathcal{D}) = \mathcal{F}_\mathcal{D}$. By definition of PCAF semantics, $cl(Args_\mathcal{U}) \in \sigma^2_{inh}(\mathcal{P}_\mathcal{D})$. For each assumption $a \in \mathcal{U}$, the argument $\{a\} \vdash a$ is contained in $Args_\mathcal{U}$. Therefore, we obtain $\mathcal{U} = cl(Args_\mathcal{U}) \cap \mathcal{A}$.

($\supseteq$) For the other direction, we consider $C \cap \mathcal{A}$ for some $C \in \sigma^2_{inh}(\mathcal{P}_{\mathcal{D}})$. By Proposition 37, it holds that $\mathcal{R}_2(\mathcal{P}_{\mathcal{D}}) = \mathcal{F}_{\mathcal{D}}$. Let $F_{\mathcal{D}}$ denote the corresponding AF, and let S denote the σ_{inh}-realisation of $C \cap \mathcal{A}$ in $F_{\mathcal{D}}$. From Proposition 38, we obtain $\mathcal{U} = asms(S) \in \sigma(\mathcal{D})$.

It remains to prove the statement for semi-stable semantics. By Proposition 37 we have $\mathcal{R}_2(\mathcal{P}_{\mathcal{D}}) = \mathcal{F}_{\mathcal{D}}$. We prove $sem(\mathcal{D}) = \{C \cap \mathcal{A} \mid C \in sem^2_{hyb}(\mathcal{P}_{\mathcal{D}})\}$. Let $F_{\mathcal{D}}$ denote the corresponding AF. Again, we can exploit the one-to-one correspondence between complete sets. Let $\mathcal{U} \in sem(\mathcal{D})$ and let $S = Args_{\mathcal{U}}$ denote the corresponding complete set of arguments in $F_{\mathcal{D}}$ corresponding to $\mathcal{U}$. It holds that S and $\mathcal{U}$ attack the same assumptions ($*$): $\mathcal{U}$ attacks assumption $a \in \mathcal{A}$ iff $\mathcal{U}$ derives the contrary of a, which implies that the unique argument with claim a is defeated in $\mathcal{F}_{\mathcal{D}}$ by S. Moreover ($**$), we observe that S defeats a conclusion $\overline{a}$ iff S attacks all arguments that derive $\overline{a}$ iff the assumption a (the corresponding argument $\{a\} \vdash a$) is contained in S.

($\subseteq$) First, let $\mathcal{U} \in sem(\mathcal{D})$ and let $S = Args_{\mathcal{U}}$ denote the corresponding complete set of arguments in $F_{\mathcal{D}}$. We show that $cl(S)$ is hybrid semi-stable in $\mathcal{F}_{\mathcal{D}}$. Towards a contradiction, assume there is $S' \in com(F_{\mathcal{D}})$ with $S^{\circledast}_{\mathcal{F}_{\mathcal{D}}} \subset S'^{\circledast}_{\mathcal{F}_{\mathcal{D}}}$. By Proposition 38, $\mathcal{U}' = asms(S')$ is complete in $\mathcal{D}$. Moreover, S' and $\mathcal{U}'$ attack the same assumptions by ($*$). It holds that $\mathcal{U} \cup \mathcal{U}^+_{\mathcal{D}} \subseteq S^{\circledast}_{\mathcal{F}_{\mathcal{D}}}$ and $\mathcal{U}' \cup (\mathcal{U}')^+_{\mathcal{D}} \subseteq S'^{\circledast}_{\mathcal{F}_{\mathcal{D}}}$. We obtain $\mathcal{U} \cup \mathcal{U}^+_{\mathcal{D}} \subset S'^{\circledast}_{\mathcal{F}_{\mathcal{D}}}$. Since $\mathcal{U}$ is semi-stable in $\mathcal{F}_{\mathcal{D}}$, it follows that $\mathcal{U} \cup \mathcal{U}^+_{\mathcal{D}} = \mathcal{U}' \cup (\mathcal{U}')^+_{\mathcal{D}}$, and $S = S'$. This proves that $sem(\mathcal{D}) \subseteq \{C \cap \mathcal{A} \mid C \in sem^2_{hyb}(\mathcal{P}_{\mathcal{D}})\}$.

($\supseteq$) Next, consider a set $C \in sem_{hyb}(\mathcal{F}_{\mathcal{D}})$, and let S denote a sem_{hyb}-realization of C in $F_{\mathcal{D}}$. We can assume that S is complete. Let $\mathcal{U} = asms(S)$. We show that $\mathcal{U}$ is semi-stable in $\mathcal{D}$. Towards a contradiction, assume there is another set of assumptions $\mathcal{U}'$ in $\mathcal{D}$ with $\mathcal{U} \cup \mathcal{U}^+_{\mathcal{D}} \subset \mathcal{U}' \cup (\mathcal{U}')^+_{\mathcal{D}}$. Let $S' = Args_{\mathcal{U}'}$ denote the corresponding set of arguments in $F_{\mathcal{D}}$.

From observations ($*$) and ($**$), we obtain $S^{\circledast}_{\mathcal{F}_{\mathcal{D}}} = \mathcal{U} \cup \mathcal{U}^+_{\mathcal{D}} \cup \overline{\mathcal{U}}$.

From ($*$), we obtain $\mathcal{U}' \cup (\mathcal{U}')^+_{\mathcal{D}} \subseteq S'^{\circledast}_{\mathcal{F}_{\mathcal{D}}}$, hence, also $\mathcal{U} \cup \mathcal{U}^+_{\mathcal{D}} \subseteq S'^{\circledast}_{\mathcal{F}_{\mathcal{D}}}$. We show that $\overline{\mathcal{U}} \subseteq S'^{\circledast}_{\mathcal{F}_{\mathcal{D}}}$. Let $\overline{a} \in \overline{\mathcal{U}}$. Then $a \in \mathcal{U}$, hence, $a \in \mathcal{U}' \cup (\mathcal{U}')^+_{\mathcal{D}}$. In case $a \in \mathcal{U}'$ we have that a is defended by $\mathcal{U}'$, and hence, $\overline{a}$ is defeated by S'. In case $a \in (\mathcal{U}')^+_{\mathcal{D}}$ we have that $\{a\} \vdash a$ is attacked, and hence, there is an argument with claim $\overline{a} \in S'$. In both cases, we obtain $\overline{\mathcal{U}} \subseteq S'^{\circledast}_{\mathcal{F}_{\mathcal{D}}}$. Therefore, $S^{\circledast}_{\mathcal{F}_{\mathcal{D}}} = \mathcal{U} \cup \mathcal{U}^+_{\mathcal{D}} \cup \overline{\mathcal{U}} \subseteq S'^{\circledast}_{\mathcal{F}_{\mathcal{D}}}$, contradiction to the assumption.

We have shown that $sem(\mathcal{D}) \supseteq \{C \cap \mathcal{A} \mid C \in sem^2_{hyb}(\mathcal{P}_{\mathcal{D}})\}$. □

In summary, we have shown that PCAFs are a natural target formalism to capture ABA with preferences if WCP is satisfied. Instead of reversing the attacks and instantiating the resulting CAF, it is equally possible to instantiate the ABA without preferences and perform attack reversal on the instantiated framework. This can be beneficial when considering dynamic updates (Kaci et al., 2018; Rapberger & Ulbricht, 2023; Cayrol, de Saint-Cyr, & Lagasquie-Schiex, 2010), particularly in situations in which the preference relation evolves over time. Instead of instantiating the framework from scratch after each update, we can keep the CAF instantiation of the ABA knowledge base and adapt the preferences, as needed. Moreover, we extend a result by Rapberger (2023) who shows that hybrid semi-stable semantics capture ABA semi-stable semantics, even when considering preferences, which is not possible under traditional AF semantics.

The instantiation furthermore offers alternative ways to handle preferences in ABA by considering other reductions for preference handling. Instead of applying Reduction 2, for instance, it could be beneficial to adapt a more cautious approach and revert attacks only between symmetric attacks (that is, apply Reduction 3). Further studies and comparisons of these different approaches would be an interesting avenue for future research.

8. Conclusion

Many approaches to argumentation assume that arguments with the same claims attack the same arguments. This gives rise to the natural class of wfCAFs, which enjoy several desired semantic and computational properties (Dvořák et al., 2023, 2023). However, in formalisms in which preferences are used, well-formedness cannot be assumed in general. In this paper, we analyzed whether the desired properties of wfCAFs still hold when preferences are taken into account. To this end, we introduced Preference-based CAFs (PCAFs) and investigated the impact of the four commonly used preference reductions on PCAFs.

We examined and characterized resulting CAF-classes, yielding insights into the expressiveness of argumentation formalisms that can be instantiated as CAFs and allow for preference incorporation. Furthermore, we investigated PCAFs with respect to semantic properties, computational complexity, and their relationship to structured formalisms. Preserving semantic properties such as I-maximality can be 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., (Baumeister, Järvisalo, Neugebauer, Niskanen, & Rothe, 2021; Fazzinga, Flesca, & Furfaro, 2020), for a similar research endeavor in terms of incomplete AFs).

Our results show that (i) Reduction 3 exhibits the same properties as wfCAFs regarding computational complexity, and mostly preserves semantic properties such as I-maximality; (ii) Reductions 2 and 4 retain the advantages of wfCAFs regarding complexity for all but complete semantics, but do not preserve I-maximality; (iii) under Reduction 1, neither complexity properties nor semantic properties 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. (2021).

Regarding future work, one could consider different methods for handling preferences, and examine the effect of these methods in (well-formed) CAFs. In this paper, we dealt with preferences via preference reductions that modify the attack relation, which means that hard- and soft-constraints are closely interlinked (Bernreiter, 2024). Another approach is to lift orderings over arguments to sets of arguments and select extensions in this way (Brewka, Truszczynski, & Woltran, 2010; Amgoud & Vesic, 2014; Kaci et al., 2018; Alfano, Greco, Parisi, & Trubitsyna, 2022, 2023). These two paradigms interpret the meaning of preferences between arguments differently: using reductions, $x \succ y$ expresses that x is stronger than

y, while in the second approach $x \succ y$ expresses that it is preferred to have outcomes with x rather than with y. Interestingly, under Reduction 3, the admissible/complete/stable extensions of a preference-based AF are also extensions in the underlying AF (Kaci et al., 2021). Thus, Reduction 3 selects the 'best' extensions from the underlying AF in these cases. A similar dichotomy concerning preference handling can be observed in related areas such as logic programs, where preferences are incorporated either on the syntactic level (Delgrande, Schaub, & Tompits, 2003) or by ranking the outcome (Sakama & Inoue, 2000).

Another possibility for future work is to investigate whether the resolution of preferences in wfCAFs affects semantical properties (van der Torre & Vesic, 2017; Dvořák et al., 2023) other than I-maximality. Moreover, one could lower the level of abstraction used, e.g., by incorporating more structure into arguments, by allowing arguments to act in support of other arguments as is done in bipolar AFs (Amgoud, Cayrol, Lagasquie, & Livet, 2008), or by preserving more information about the claims of arguments. Regarding the latter point, recent research (Wakaki, 2020) has shown that formalisms that permit strong negation require careful examination with regards to consistency.

Acknowledgments

We thank the anonymous reviewers for their valuable and constructive feedback. This work was funded by the Austrian Science Fund (FWF) under grants Y698, W1255-N23, and 10.557766/COE12, and by the Vienna Science and Technology Fund (WWTF) under grant ICT19-065.

Appendix A. Remaining Proofs for Section 4

Lemma 40. *Let $\mathcal{F} = (A, R, cl)$ be a CAF. $\mathcal{F} \in \mathcal{R}_2$-**CAF** iff there are no arguments a, a', b, b' in $\mathcal{F}$ with $cl(a) = cl(a')$ and $cl(b) = cl(b')$ such that $(a, b) \in wfp(\mathcal{F})$, $(b, a) \notin R$, $(a', b) \in R$, and either $(b, a') \in R$ or $((a', b') \notin R$ and $(b', a') \notin R)$.*

Proof. "$\implies$": By contrapositive. Suppose that there are $a, a', b, b' \in A$ with $cl(a') = cl(a)$ and $cl(b') = cl(b)$ such that $(a, b) \in wfp(\mathcal{F})$, $(b, a) \notin R$, $(a', b) \in R$, and either $(b, a') \in R$ or $((a', b') \notin R$ and $(b', a') \notin R)$. Towards a contradiction, assume that $\mathcal{F} \in \mathcal{R}_2$-**CAF**. Then there must be a PCAF $\mathcal{P} = (A, R', cl, \succ)$ such that $\mathcal{R}_2(\mathcal{P}) = \mathcal{F}$. Reduction 2 cannot completely remove conflicts between arguments. Since there is no conflict between a and b in $\mathcal{F}$ there can be no conflict in $\mathcal{P}$ either, i.e., $(a, b) \notin R'$ and $(b, a) \notin R'$. Therefore, since the underlying CAF (A, R', cl) of $\mathcal{P}$ must be well-formed, $(a', b) \notin R'$. Since $(a', b) \in R$ it must be that $(b, a') \in R'$ and $a' \succ b$. Then $(b, a') \notin \mathcal{R}_2(\mathcal{P})$. Furthermore, by the well-formedness of (A, R', cl), we have that $(b', a') \in R'$ and therefore either $(a', b') \in \mathcal{R}_2(\mathcal{P})$ or $(b', a') \in \mathcal{R}_2(\mathcal{P})$. Contradiction to $\mathcal{R}_2(\mathcal{P}) = \mathcal{F}$.

"$\impliedby$": Our underlying assumption is that there are no arguments a, a', b, b' in $\mathcal{F}$ with $cl(a) = cl(a')$ and $cl(b) = cl(b')$ such that $(a, b) \in wfp(\mathcal{F})$, $(b, a) \notin R$, $(a', b) \in R$, and either $(b, a') \in R$ or $((a', b') \notin R$ and $(b', a') \notin R)$. We will construct a PCAF $\mathcal{P} = (A, R'', cl, \succ)$ such that $\mathcal{R}_2(\mathcal{P}) = \mathcal{F}$.

But first, as an intermediate step, we construct the CAF $\mathcal{F}' = (A, R', cl)$. We say that (b, a) is forced in $\mathcal{F}$ if $(a, b) \in R$ and if there is an argument a' with $cl(a') = cl(a)$

such that $(a',b) \notin R$ and $(b,a') \notin R$. Observe that if (b,a) is forced in $\mathcal{F}$, then (a,b) cannot be forced in $\mathcal{F}$ by our underlying assumption. Furthermore, if (b,a) is forced in $\mathcal{F}$, then $(b,a) \notin R$, again by our underlying assumption. We construct $R' = (R \cup \{(b,a) \mid (b,a) \text{ is forced in } \mathcal{F}\}) \setminus \{(a,b) \mid (b,a) \text{ is forced in } \mathcal{F}\}$. Note that $(a,b) \in \mathit{wfp}(\mathcal{F}')$ implies $(b,a) \in R'$ for all arguments a,b: towards a contradiction, assume otherwise. Then there is some $(a,b) \in \mathit{wfp}(\mathcal{F}')$ such that $(b,a) \notin R'$. Then $(a,b) \notin R$ and $(b,a) \notin R$ by construction of R'. Furthermore, since $(a,b) \in \mathit{wfp}(\mathcal{F}')$, there must be some a' with $cl(a') = cl(a)$ and $(a',b) \in R'$. It cannot be that $(a',b) \in R$, otherwise (b,a') would be forced in $\mathcal{F}$ and $(a',b) \notin R'$. Thus, $(b,a') \in R$ and (a',b) was added to R' because it is forced in $\mathcal{F}$. But this is only possible if there is some b' with $cl(b') = cl(b)$ and $(a',b') \notin R$ and $(b',a') \notin R$. This contradicts our underlying assumption: $(b',a') \in \mathit{wfp}(\mathcal{F})$, $(a',b') \notin R$, $(b,a') \in R$, $(a,b) \notin R$, and $(b,a) \notin R$.

Now we construct $R'' = R' \cup \{(a,b) \mid (a,b) \in \mathit{wfp}(\mathcal{F}')\}$. Furthermore, $b \succ a \iff (a,b) \in R'' \setminus R$. This gives us $\mathcal{P} = (A, R'', cl, \succ)$. The underlying CAF of $\mathcal{P}$ is well-formed since $\mathit{wfp}((A, R'', cl)) = \emptyset$ by construction. Moreover, $\succ$ is asymmetric since if $(a,b) \in R''$ and $(b,a) \in R''$ then, by construction of R' and R'', either $(a,b) \in R$ or $(b,a) \in R$. Lastly, we show that $\mathcal{R}_2(\mathcal{P}) = \mathcal{F}$: if $(a,b) \in R'' \setminus R$, then we defined $b \succ a$ and thus $(a,b) \notin \mathcal{R}_2(\mathcal{P})$. If $(a,b) \in R \setminus R''$, then (b,a) was forced in $\mathcal{F}$, i.e., $(b,a) \notin R$ but $(b,a) \in R'$ and therefore also $(b,a) \in R''$. Thus, we define $a \succ b$ which means that $(a,b) \in \mathcal{R}_2(\mathcal{P})$. □

Lemma 41. *Let $\mathcal{F} = (A, R, cl)$ be a CAF. $\mathcal{F} \in \mathcal{R}_3$-**CAF** iff $(a,b) \in \mathit{wfp}(\mathcal{F})$ implies $(b,a) \in R$.*

Proof. "$\implies$": By contrapositive. Suppose there is $(a,b) \in \mathit{wfp}(\mathcal{F})$ such that $(b,a) \notin R$. Towards a contradiction, assume $\mathcal{F} \in \mathcal{R}_3$-**CAF**. Then there is a PCAF $\mathcal{P} = (A, R', cl, \succ)$ such that $\mathcal{R}_3(\mathcal{P}) = \mathcal{F}$. Since Reduction 3 can only delete but not introduce attacks, and since (A, R', cl) must be well-formed, $(a,b) \in R'$. However, Reduction 3 cannot completely remove conflicts between arguments, i.e., either $(a,b) \in \mathcal{R}_3(\mathcal{P})$ or $(b,a) \in \mathcal{R}_3(\mathcal{P})$. Contradiction.

"$\impliedby$": Suppose $(a,b) \in \mathit{wfp}(\mathcal{F})$ implies $(b,a) \in R$. Then $\mathcal{R}_3(\mathcal{P}) = \mathcal{F}$ for the PCAF $\mathcal{P} = (A, R', cl, \succ)$ with $R' = R \cup \{(a,b) \mid (a,b) \in \mathit{wfp}(\mathcal{F})\}$ as well as $a \succ b \iff (b,a) \in R' \setminus R$. (A, R', cl) is well-formed since $\mathit{wfp}((A, R', cl)) = \emptyset$. Furthermore, $\succ$ is asymmetric by construction. □

Lemma 42. *Let $\mathcal{F} = (A, R, cl)$ be a CAF. $\mathcal{F} \in \mathcal{R}_4$-**CAF** iff there are no arguments a, a', b, b' in $\mathcal{F}$ with $cl(a) = cl(a')$ and $cl(b) = cl(b')$ such that $(a,b) \in \mathit{wfp}(\mathcal{F})$, $(b,a) \notin R$, $(a',b) \in R$, and either $(b,a') \notin R$ or $((a',b') \notin R$ and $(b',a') \notin R)$.*

Proof. Similar to the proof of Lemma 40:

"$\implies$": By contrapositive. Suppose that there are $a, a', b, b' \in A$ with $cl(a') = cl(a)$ and $cl(b') = cl(b)$ such that $(a,b) \in \mathit{wfp}(\mathcal{F})$, $(b,a) \notin R$, $(a',b) \in R$, and either $(b,a') \notin R$ or $((a',b') \notin R$ and $(b',a') \notin R)$. Towards a contradiction, assume that $\mathcal{F} \in \mathcal{R}_4$-**CAF**. Then there must be a PCAF $\mathcal{P} = (A, R', cl, \succ)$ such that $\mathcal{R}_4(\mathcal{P}) = \mathcal{F}$. Reduction 4 cannot completely remove conflicts between arguments. Since there is no conflict between a and b in $\mathcal{F}$ there can be no conflict in $\mathcal{P}$ either, i.e., $(a,b) \notin R'$ and $(b,a) \notin R'$. Therefore, since the underlying CAF of $\mathcal{P}$ must be well-formed, $(a', b) \notin R'$. The only way to obtain $(a', b) \in R$ from $(a', b) \notin R'$ via Reduction 4 is to have $(b, a') \in R'$ and $a' \succ b$. Then $(b, a') \in \mathcal{R}_4(\mathcal{P})$. Furthermore, by the well-formedness of (A, R', cl), we have that $(b', a') \in R'$ and therefore either $(a', b') \in \mathcal{R}_4(\mathcal{P})$ or $(b', a') \in \mathcal{R}_4(\mathcal{P})$. Contradiction to $\mathcal{R}_4(\mathcal{P}) = \mathcal{F}$.

"$\Longleftarrow$": Our underlying assumption is that there are no arguments a, a', b, b' in $\mathcal{F}$ with $cl(a) = cl(a')$ and $cl(b) = cl(b')$ such that $(a, b) \in wfp(\mathcal{F})$, $(b, a) \notin R$, $(a', b) \in R$, and either $(b, a') \notin R$ or $((a', b') \notin R$ and $(b', a') \notin R)$. We will construct a PCAF $\mathcal{P} = (A, R'', cl, \succ)$ such that $\mathcal{R}_4(\mathcal{P}) = \mathcal{F}$.

But first, as an intermediate step, we construct the CAF $\mathcal{F}' = (A, R', cl)$. We say that (b, a) is forced in $\mathcal{F}$ if $(a, b) \in R$, $(b, a) \in R$, and if there is an argument a' with $cl(a') = cl(a)$ such that $(a', b) \notin R$ and $(b, a') \notin R$. Observe that if (b, a) is forced in $\mathcal{F}$, then (a, b) cannot be forced in $\mathcal{F}$ by our underlying assumption. We construct $R' = R \setminus \{(a, b) \mid (b, a) \text{ is forced in } \mathcal{F}\}$. Note that $(a, b) \in wfp(\mathcal{F}')$ implies $(b, a) \in R'$ for all arguments a, b: towards a contradiction, assume otherwise. Then there is some $(a, b) \in wfp(\mathcal{F}')$ such that $(b, a) \notin R'$. Then $(a, b) \notin R$ and $(b, a) \notin R$ by construction of R'. Furthermore, there must be some a' with $cl(a') = cl(a)$ and $(a', b) \in R'$. It cannot be that $(a', b) \in R$ and $(b, a') \in R$, otherwise (b, a') would be forced in $\mathcal{F}$ and $(a', b) \notin R'$. Thus, $(a', b) \in R$ and $(b, a') \notin R$ by construction of $\mathcal{F}'$. But this contradicts our underlying assumption: $(a, b) \in wfp(\mathcal{F})$, $(b, a) \notin R$, $(a', b) \in R$, and $(b, a') \notin R$.

Now we construct $R'' = R' \cup \{(a, b) \mid (a, b) \in wfp(\mathcal{F}')\}$. Furthermore, $b \succ a \iff (a, b) \in R'' \setminus R$ or $(b, a) \in R \setminus R''$. This gives us, $\mathcal{P} = (A, R'', cl, \succ)$. The underlying CAF of $\mathcal{P}$ is well-formed since $wfp((A, R'', cl)) = \emptyset$ by construction. Moreover, $\succ$ is asymmetric: if $b \succ a$, there are two cases.

1. $(a, b) \in R'' \setminus R$. Clearly, $(a, b) \notin R \setminus R''$. Moreover, $(a, b) \in R'' \setminus R$ implies $(b, a) \in R$ since we did not add attacks to R'' if there was no conflict between these attacks in R. Thus, $(b, a) \notin R'' \setminus R$. We can conclude $a \not\succ b$.

2. $(b, a) \in R \setminus R''$. Clearly, $(b, a) \notin R'' \setminus R$. Moreover, $(b, a) \in R \setminus R''$ implies $(a, b) \in R''$, since we never completely removed conflicts when constructing R'' from R. Thus, $(a, b) \notin R \setminus R''$. We can conclude $a \not\succ b$.

Lastly, we show that $\mathcal{R}_4(\mathcal{P}) = \mathcal{F}$: if $(a, b) \in R'' \setminus R$, then we defined $b \succ a$. As above, $(a, b) \in R'' \setminus R$ implies $(b, a) \in R$. The only possible reason for why we added (a, b) to R'' is because $(a, b) \in wfp(\mathcal{F}')$. As previously discussed, this means that $(b, a) \in R'$ and therefore also $(b, a) \in R''$. Thus, $(a, b) \notin \mathcal{R}_4(\mathcal{P})$. If $(a, b) \in R \setminus R''$, then $a \succ b$. As above, this implies $(b, a) \in R''$, and therefore $(a, b) \in \mathcal{R}_4(\mathcal{P})$. □

Appendix B. Remaining Proofs for Section 6

Lemma 43. *$Ver^{PCAF}_{\sigma^i_\mu}$ is Σ^{P}_2-hard for $\sigma^i_\mu \in \{stg^1_{inh}, stg^1_{hyb}\}$, even if we restrict ourselves to PCAFs with transitive preference relations.*

Proof. We provide a reduction from $\text{QBF}^2_\forall$ to the complementary problem. Let $\Phi = \forall Y \exists Z \varphi$ be an instance of $\text{QBF}^2_\forall$, where φ is given by a set Ω of clauses over atoms $X = Y \cup Z$. We construct the CAF $\mathcal{F} = (A, R, cl)$ with underlying AF $F = (A, R)$ and a set of claims C:

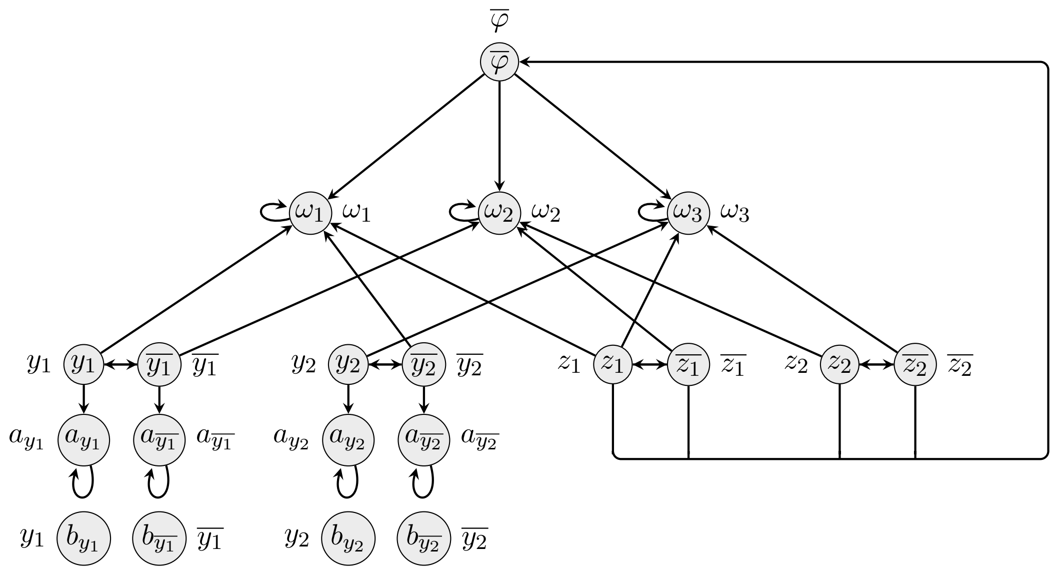


Figure 16: Reduction of the $\mathrm{QBF}_\forall^2$ instance $\Phi = \forall y_1, y_2 \exists z_1, z_2 \varphi$ with φ given by clauses $\omega_1 = \{y_1, \neg y_2, z_1\}$, $\omega_2 = \{\neg y_1, \neg z_1, z_2\}$, $\omega_3 = \{y_2, z_1, \neg z_2\}$ to an instance of $Ver^{PCAF}_{stg^1_{inh}}$.

- $A = \{\overline{\varphi}\} \cup \Omega \cup X \cup \overline{X} \cup Y_a \cup \overline{Y}_a \cup Y_b \cup \overline{Y}_b$, where $\overline{X} = \{\overline{x} \mid x \in X\}$, $Y_a = \{a_y \mid y \in Y\}$, $\overline{Y}_a = \{a_{\overline{y}} \mid y \in Y\}$, $Y_b = \{b_y \mid y \in Y\}$, $\overline{Y}_b = \{b_{\overline{y}} \mid y \in Y\}$;
- $R = \{(x, \overline{x}), (\overline{x}, x) \mid x \in X\} \cup \{(\omega, \omega), (\overline{\varphi}, \omega) \mid \omega \in \Omega\} \cup$
  $\{(x, \omega) \mid x \in \omega, \omega \in \Omega\} \cup \{(\overline{x}, \omega) \mid \neg x \in \omega, \omega \in \Omega\} \cup$
  $\{(a_v, a_v), (v, a_v) \mid v \in Y \cup \overline{Y}\} \cup \{(z, \overline{\varphi}), (\overline{z}, \overline{\varphi}) \mid z \in Z\}$;
- $cl(b_v) = v$ for $b_v \in Y_b \cup \overline{Y}_b$ and $cl(v) = v$ else;
- $C = Y \cup \overline{Y} \cup \{\overline{\varphi}\}$.

Figure 16 illustrates the above construction. Note that $\mathcal{F} \in \mathcal{R}_1$-$\mathbf{CAF}_{tr}$ since all paths in $wfp(\mathcal{F}) = \{(b_v, v) \mid v \in Y \cup \overline{Y}\}$ are of length 1 (only arguments in $Y_b \cup \overline{Y}_b$ have outgoing edges in $wfp(\mathcal{F})$). It remains to verify the correctness of the reduction, i.e., we will show that Φ is valid iff $C \notin \sigma_\mu(\mathcal{F})$. The proof proceeds similar as the proof of Proposition 24.

"$\Longrightarrow$": Assume Φ is valid. Consider any $S \subseteq A$ such that $S \in cf(F)$ and $cl(S) = C$. Then $S \subseteq Y \cup \overline{Y} \cup Y_b \cup \overline{Y}_b \cup \{\overline{\varphi}\}$. Let $Y' = S \cap Y$. Since Φ is valid, there is $Z' \subseteq Z$ such that $M = Y' \cup Z'$ is a model of φ. Let $T = M \cup \{\overline{x} \mid x \in X \setminus M\} \cup Y_b \cup \overline{Y}_b$. Note that $T \in cf(F)$ by construction. Moreover, $S \setminus \{\overline{\varphi}\} \subseteq T$. Since for each $z \in Z$ we have either $z \in T$ or $\overline{z} \in T$, and since $(z, \overline{\varphi}), (\overline{z}, \overline{\varphi}) \in R$, we have $\overline{\varphi} \in T^+_F$ (resp. $\overline{\varphi} \in T^*_F$). Since $M \models \varphi$, all clause-arguments $\omega \in \Omega$ are attacked by T and we have $\{\overline{\varphi}\}^+_F = \Omega \subseteq T^+_F$ (resp. $\{\overline{\varphi}\}^*_{\mathcal{F}} = \Omega \subseteq T^*_F$). We can conclude that $S \cup S^+_F \subset T \cup T^+_F$ (resp. $cl(S) \cup S^*_{\mathcal{F}} \subset cl(T) \cup T^*_{\mathcal{F}}$), i.e., $C \notin stg_{inh}(\mathcal{F})$ (resp. $C \notin stg_{hyb}(\mathcal{F})$).

"$\Longleftarrow$": Assume $C \notin stg_{inh}(\mathcal{F})$ (resp. $C \notin stg_{hyb}(\mathcal{F})$) and consider an arbitrary subset $Y' \subseteq Y$. We must show that there is $Z' \subseteq Z$ such that $Y' \cup Z' \models \varphi$. Let $S = Y' \cup \{\overline{y} \mid y \in Y \setminus Y'\} \cup Y^* \cup \overline{Y}^* \cup \{\overline{\varphi}\}$. Observe that $cl(S) = C$ and that $S \in cf(F)$. By $C \notin stg_{inh}(\mathcal{F})$ (resp. $C \notin stg_{hyb}(\mathcal{F})$) there is some $T \in cf(F)$ with $S \cup S^+_F \subset T \cup T^+_F$ (resp. $cl(S) \cup S^*_{\mathcal{F}} \subset cl(T) \cup T^*_{\mathcal{F}}$).

In particular, we have $Y' \cup \{\overline{y} \mid y \in Y \setminus Y'\} \subseteq T$ since each $a_v \in S_F^+$ (resp. $a_v \in S_{\mathcal{F}}^*$) with $v \in Y \cup \overline{Y}$ has precisely one non-self-attacking attacker (namely the argument v). Moreover, we can assume that T contains each argument $v \in Y_b \cup \overline{Y}_b$ since each such v is unattacked and does not attack any other argument. Thus, $T \supseteq S \setminus \{\overline{\varphi}\}$.

Furthermore, $\overline{\varphi} \notin T$ since $S \in naive(F)$ (note that $\overline{\varphi}$ attacks each $z, \overline{z}$ with $z \in Z$ as well as every clause-argument $\omega \in \Omega$ and thus cannot be extended any further). Therefore, it must be that $\overline{\varphi} \in T_F^+$ (resp. $\overline{\varphi} \in T_{\mathcal{F}}^*$). Also, we have that T attacks each clause-argument $\omega \in \Omega$ since $\Omega \subseteq S_F^+$ (resp. $\Omega \subseteq S_{\mathcal{F}}^*$), and since each clause-argument $\omega \in \Omega$ is self-attacking.

Now, let $Z' = Z \cap T$. We show that $M = Y' \cup Z'$ is a model of φ. Consider some arbitrary clause $\omega \in \Omega$. Then there is some argument $v \in T$ such that $(v, \omega) \in R$. As outlined above, $v \neq \overline{\varphi}$ since $\overline{\varphi}$ is not contained in T. Consequently, we have $v \in X \cup \overline{X}$. In case $v \in X$ we have $v \in M \cap \omega$, in case $v \in \overline{X}$ we have $\neg v \in \omega$ and $v \notin M$ by definition of R. In every case, the clause ω is satisfied by M. As ω was chosen arbitrary it follows that $M \models \varphi$. We can conclude that Φ is valid. □

Lemma 44. *$Ver_{\sigma_\mu^i}^{PCAF}$ is* DP*-hard for $\sigma_\mu^i = naive_{hyb}^1$, even if we restrict ourselves to PCAFs with transitive preference relations.*

Proof. Before showing DP-hardness, we show NP- and coNP-hardness separately:

Let $(\mathcal{P}, C)$ be an instance of $Ver_{cf_{inh}^1}^{PCAF}$, i.e., $\mathcal{P} = (A, R, cl, \succ)$ is a PCAF and $C \subseteq cl(A)$ is the claim-set to be verified for conflict-freeness. Recall that $Ver_{cf_{inh}^1}^{PCAF}$ is NP-complete, even when restricted to transitive preferences (see Proposition 22).

- First, we construct a PCAF $\mathcal{P}' = (A', R', cl', \succ')$ with $A' = \{x \mid x \in A, cl(x) \in C\}$ as well as $R' = \{(x, y) \mid x, y \in A', (x, y) \in R\}$, $cl'(x) = cl(x)$ for all $x \in A'$, and $x \succ' y$ iff $x \succ y$ and $x, y \in A'$. Observe that (A', R', cl') is still well-formed. Furthermore, if $\succ$ is transitive, then so is $\succ'$. It is easy to see that $C \in cf_{inh}(\mathcal{R}_1(\mathcal{P}))$ iff $C \in cf_{inh}(\mathcal{R}_1(\mathcal{P}'))$. Since $C = cl(A')$, $C \in cf_{inh}(\mathcal{R}_1(\mathcal{P}))$ iff $C \in naive_{hyb}(\mathcal{R}_1(\mathcal{P}'))$.

- Second, we construct another PCAF $\mathcal{P}'' = (A'', R'', cl'', \succ'')$. Without loss of generality, we can assume $C \neq \emptyset$. We fix an arbitrary claim $c \in C$ and for each claim $d \in C \setminus \{c\}$ introduce a fresh argument z_d. Let Z be the set of those fresh arguments. Then $A'' = A' \cup Z$, $R'' = R' \cup \{(x, z_d) \mid cl(x) = c, z_d \in Z\} \cup \{(z_d, y) \mid z_d \in Z, y \in A'$, there exists $x \in A'$ with $cl(x) = d$ such that $(x, y) \in R'\}$, $cl''(x) = cl'(x)$ for all $x \in A'$, $cl''(z_d) = d$ for all $z_d \in Z$, and $x \succ'' y$ iff $x \succ' y$. (A'', R'', cl'') is well-formed by construction, and $\succ''$ can still be assumed to be transitive. Now we show that $C \notin cf_{inh}(\mathcal{R}_1(\mathcal{P}))$ iff $C \setminus \{c\} \in naive_{hyb}(\mathcal{R}_1(\mathcal{P}''))$: (1) assume $C \in cf_{inh}(\mathcal{R}_1(\mathcal{P}))$. Then also $C \in cf_{inh}(\mathcal{R}_1(\mathcal{P}''))$ and thus $C \setminus \{c\} \notin naive_{hyb}(\mathcal{R}_1(\mathcal{P}''))$. (2) assume $C \notin cf_{inh}(\mathcal{R}_1(\mathcal{P}))$. Then also $C \notin cf_{inh}(\mathcal{R}_1(\mathcal{P}''))$ since all arguments x with $cl(x) = c$ are in conflict with the fresh arguments z_d. But because the fresh arguments z_d do not attack each other, $C \setminus \{c\} \in cf_{inh}(\mathcal{R}_1(\mathcal{P}''))$. Since $C = cl(A'')$, $C \setminus \{c\} \in naive_{hyb}(\mathcal{R}_1(\mathcal{P}''))$.

The construction of $\mathcal{P}'$ shows NP-hardness, while the construction of $\mathcal{P}''$ shows coNP-hardness. Now we show DP-hardness: let (φ_1, φ_2) be an arbitrary instance of SAT-UNSAT, with φ_1 and φ_2 sharing no variables. We can construct instances $(\mathcal{P}_1 = (A_1, R_1, cl_1, \succ_1$

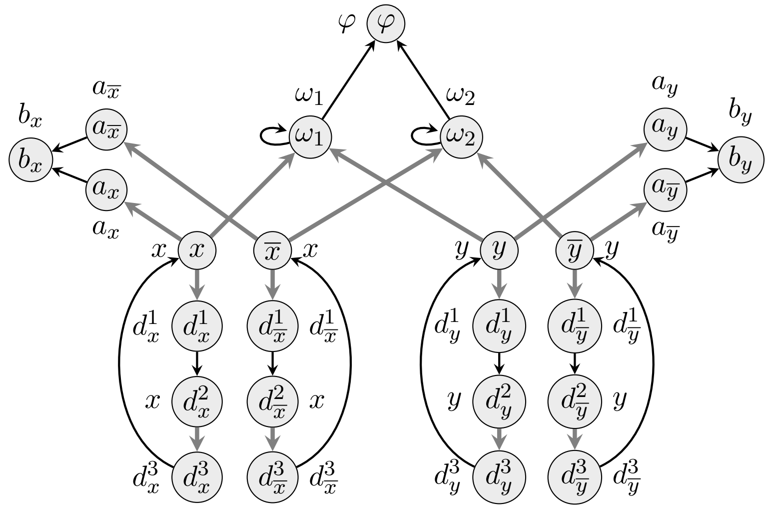


Figure 17: $\mathcal{R}_2(\mathcal{P})$ from the proof of Lemma 45, with φ given by clauses $\omega_1 = \{x, y\}$, $\omega_2 = \{\neg x, \neg y\}$. Gray/thick attacks have been reversed by Reduction 2.

$), C_1)$ and $(\mathcal{P}_2 = (A_2, R_2, cl_2, \succ_2), C_2)$ of $Ver^{PCAF}_{naive^1_{hyb}}$ such that φ_1 is satisfiable iff $C_1 \in naive_{hyb}(\mathcal{R}_1(\mathcal{P}_1))$ and φ_2 is unsatisfiable iff $C_2 \in naive_{hyb}(\mathcal{R}_1(\mathcal{P}_2))$. Note that we can assume $\mathcal{P}_1$ and $\mathcal{P}_2$ to be disjoint, i.e., they share no arguments and claims. Let $\mathcal{P} = (A_1 \cup A_2, R_1 \cup R_2, cl_1 \cup cl_2, \succ_1 \cup \succ_2)$ be the combination of $\mathcal{P}_1$ and $\mathcal{P}_2$. Observe that $C_1 \cup C_2 \in naive_{hyb}(\mathcal{R}_1(\mathcal{P}))$ iff $C_1 \in naive_{hyb}(\mathcal{R}_1(\mathcal{P}_1))$ and $C_2 \in naive_{hyb}(\mathcal{R}_1(\mathcal{P}_2))$. Thus, (φ_1, φ_2) is a yes-instance of SAT-UNSAT iff $C_1 \cup C_2 \in naive_{hyb}(\mathcal{R}_1(\mathcal{P}))$. □

Lemma 45. *$Ver^{PCAF}_{\sigma^i_\mu}$ is* NP*-hard for $\sigma^i_\mu = com^2_{inh}$, even if we restrict ourselves to PCAFs with transitive preference relations.*

Proof. Let φ be an arbitrary instance of 3-SAT given as a set Ω of clauses over variables X and let $\overline{X} = \{\overline{x} \mid x \in X\}$. We construct a PCAF $\mathcal{P} = (A, R, cl, \succ)$ as well as a set of claims C:

- $A = \{\varphi\} \cup \Omega \cup X \cup \overline{X} \cup \{a_x \mid x \in X \cup \overline{X}\} \cup \{b_x \mid x \in X\} \cup \{d^j_x \mid x \in X \cup \overline{X}, 1 \leq j \leq 3\}$;
- $R = \{(\omega, \varphi) \mid \omega \in \Omega\} \cup \{(\omega, \omega) \mid \omega \in \Omega\} \cup$
  $\{(\omega, x) \mid x \in \omega, \omega \in \Omega\} \cup \{(\omega, \overline{x}) \mid \neg x \in \omega, \omega \in \Omega\} \cup$
  $\{(d^1_x, x), (d^1_x, d^2_x), (d^3_x, d^2_x), (d^3_x, x), (a_x, x) \mid x \in X \cup \overline{X}\} \cup$
  $\{(a_x, b_x), (a_{\overline{x}}, b_x) \mid x \in X\}$;
- $cl(x) = cl(\overline{x}) = cl(d^2_x) = cl(d^2_{\overline{x}}) = x$ for $x \in X$, $cl(v) = v$ else;
- $x \succ \omega$, $x \succ d^1_x$, $x \succ a_x$, $d^2_x \succ d^3_x$ for all $x \in X \cup \overline{X}$ and all $\omega \in \Omega$;
- $C = X \cup \{\varphi\}$.

Figure 17 illustrates the above construction. It remains to show that φ is satisfiable if and only if $C \in com_{inh}(\mathcal{R}_2(\mathcal{P}))$.

Assume φ is satisfiable. Then there is an interpretation I such that $I \models \varphi$. Let $S = \{x, d_x^2 \mid x \in X, x \in I\} \cup \{\overline{x}, d_{\overline{x}}^2 \mid x \in X, x \notin I\} \cup \{\varphi\}$. Clearly, $cl(S) = C$. Furthermore, S defends φ in $\mathcal{R}_2(\mathcal{P})$ since each clause is satisfied by I, and thus each clause argument ω_j is attacked by some x (or $\overline{x}$) in S. For each variable x, if $x \in S$, then x defends d_x^2 and d_x^2 defends x. Moreover, if $x \in S$, then $\overline{x} \notin S$ and none of $\overline{x}$, $a_{\overline{x}}$, b_x, or $d_{\overline{x}}^j$ with $1 \leq j \leq 3$ is defended by S. Analogously for the case that $\overline{x} \in S$. Thus, S is admissible, and contains all arguments it defends, i.e., $S \in com(\mathcal{R}_2(\mathcal{P}))$.

Assume $C \in com_{inh}(\mathcal{R}_2(\mathcal{P}))$. Then there is $S \subseteq A$ such that $cl(S) = C$ and $S \in com(\mathcal{R}_2(\mathcal{P}))$. For each $x \in X$, at least one of $x, \overline{x}, d_x^2, d_{\overline{x}}^2$ must be contained in S. In fact, if $x \in S$, then also $d_x^2 \in S$ and vice versa. Analogous for $\overline{x}$ and $d_{\overline{x}}^2$. However, it cannot be that $x \in S$ and $\overline{x} \in S$, otherwise b_x would be defended by S and we would have $cl(S) \neq C$. Thus, for each $x \in X$, there is either $x \in S$ or $\overline{x} \in S$, but not both. Furthermore, S defends φ, i.e., S attacks all clause arguments ω_j. Therefore, $I \models \varphi$ for $I = X \cap S$. □

Appendix C. Remaining Proofs for Section 7

Proposition 37. *Let $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$ be an ABA$^+$ framework, let $\mathcal{P}_\mathcal{D} = (A, R, cl, \prec)$ be the corresponding PCAF (cf. Definition 29), and let $\mathcal{F}_\mathcal{D} = (A', R', cl')$ be the corresponding CAF (cf. Definition 30). It holds that $\mathcal{R}_2(\mathcal{P}_\mathcal{D}) = \mathcal{F}_\mathcal{D}$.*

Proof. Let $\mathcal{R}_2(\mathcal{P}_\mathcal{D}) = (A'', R'', cl)$. By Definition 4, $(a, b) \in R''$ if and only if (i) $((a, b) \in R$, $b \not\succ a)$ or (ii) $((b, a) \in R, (a, b) \notin R, a \succ b)$. We show that $R'' = R'$.

($\subseteq$) Let $(\mathcal{U} \vdash p, \mathcal{U}' \vdash p') \in R''$. In case (i) we have $(\mathcal{U} \vdash p, \mathcal{U}' \vdash p') \in R$ and $\mathcal{U}' \vdash p' \not\succ \mathcal{U} \vdash p$. Hence there is some $a \in \mathcal{A}$ such that $\overline{a} = p$ and either there is no $b \in \mathcal{U}$ that is strictly weaker than a or $\mathcal{U} \vdash p \approx \mathcal{U}' \vdash p'$. In the latter case, $\mathcal{U} \vdash p$ and $\mathcal{U}' \vdash p'$ symmetrically attack each other, hence $p' = \overline{a}' \in \overline{\mathcal{U}}$ and there is $b \in \mathcal{U}'$ such that $b < a'$. In both cases, we obtain $(\mathcal{U} \vdash p, \mathcal{U}' \vdash p') \in R'$. In case (ii) we have $(\mathcal{U}' \vdash p', \mathcal{U} \vdash p) \in R$, $(\mathcal{U} \vdash p, \mathcal{U}' \vdash p') \notin R$, and $\mathcal{U} \vdash p \succ \mathcal{U}' \vdash p'$, i.e., $p' = \overline{a}' \in \overline{\mathcal{U}}$ and there is $b \in \mathcal{U}'$ such that $b < a'$ (and either p is not the contrary of an assumption or $p = \overline{c}$ for some $c \in \mathcal{U}'$ there is no assumption $d \in \mathcal{U}$ which is strictly weaker than c). We obtain $(\mathcal{U} \vdash p, \mathcal{U}' \vdash p') \in R'$. Hence, we have shown that $R'' \subseteq R'$.

($\supseteq$) Let $(\mathcal{U} \vdash p, \mathcal{U}' \vdash p') \in R'$. Then either (i) ($p = \overline{a} \in \overline{\mathcal{U}}'$ and there is no $b \in \mathcal{U}$ with $b < a$) or (ii) ($p' = \overline{a}' \in \overline{\mathcal{U}}$ and there is $b \in \mathcal{U}'$ such that $b < a'$). In case (i), we obtain $\mathcal{U}' \vdash p' \not\prec \mathcal{U} \vdash p$ by Definition. We obtain $(\mathcal{U} \vdash p, \mathcal{U}' \vdash p') \in R''$. In case (ii) we have $(\mathcal{U}' \vdash p', \mathcal{U} \vdash p) \in R$ and $\mathcal{U} \vdash p \succeq \mathcal{U}' \vdash p'$. In case $\mathcal{U} \vdash p \not\approx \mathcal{U}' \vdash p'$ it holds that $\mathcal{U} \vdash p \succ \mathcal{U}' \vdash p'$ and $(\mathcal{U} \vdash p, \mathcal{U}' \vdash p') \notin R$, hence $(\mathcal{U} \vdash p, \mathcal{U}' \vdash p') \in R''$ by Definition 4 (cf. case (ii) above). In case $\mathcal{U} \vdash p \approx \mathcal{U}' \vdash p'$ we have $(\mathcal{U} \vdash p, \mathcal{U}' \vdash p') \in R$ and $\mathcal{U}' \vdash p' \succeq \mathcal{U} \vdash p$, i.e., both attacks from $\mathcal{U}' \vdash p'$ to $\mathcal{U} \vdash p$ and vice versa are reversed in the corresponding CAF. Consequently, we obtain $(\mathcal{U} \vdash p, \mathcal{U}' \vdash p') \in R''$. Hence, $R'' \supseteq R'$. □

Lemma 46. *Let $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$ be an ABA$^+$ framework that satisfies weak contraposition, $\mathcal{F}_\mathcal{D} = (A, R, cl)$ its associated CAF, and $F_\mathcal{D} = (A, R)$ the underlying AF of $\mathcal{F}_\mathcal{D}$. Then $S \in adm(F_\mathcal{D})$ implies $asms(S) \in adm(\mathcal{D})$.*

Proof. Let $S \in adm(F_\mathcal{D})$, and let $\mathcal{U} = asms(S)$.

We show that $\mathcal{U}$ is conflict-free. Towards a contradiction, assume $\mathcal{U}$ is conflicting. That is, $\mathcal{U}$ either normally or reversely attacks itself.

In the first case, there is $\mathcal{U}' \vdash \overline{a}$ with $\mathcal{U}' \subseteq \mathcal{U}$ and $a \in \mathcal{U}$, and no assumption $b \in \mathcal{U}'$ is weaker than a. Hence, some argument $x \in S$ is attacked by $\mathcal{U}' \vdash \overline{a}$ in $F_{\mathcal{D}}$. Since S is admissible in $F_{\mathcal{D}}$, there is some counter-attack on $\mathcal{U}' \vdash \overline{a}$. Since no assumption $b \in \mathcal{U}'$ is weaker than a, it cannot be the case that the attack on the argument $\mathcal{U}' \vdash \overline{a}$ stems from a normal attack in the ABA$^+$ framework $\mathcal{D}$: Towards a contradiction, assume there is an argument $\mathcal{W} \vdash \overline{c}$ which reversely attacks the argument $\mathcal{U}' \vdash \overline{a}$ in the underlying ABA. By Definition 30, it holds that $a \in \mathcal{W}$ and there is $b \in \mathcal{U}'$ such that $b < a$. This is a contradiction to our assumption. Hence, there is $y \in S$ which normally attacks the argument $\mathcal{U}' \vdash \overline{a}$ on some claim $u \in \mathcal{U}'$. However, this implies that there is an argument $z \in S$ with $u \in asms(z)$ which is attacked by y, contradiction to conflict-freeness of S.

Now assume $\mathcal{U}$ reversely attacks itself. Then there is $\mathcal{U}' \vdash \overline{a}$, $\mathcal{U}' \subseteq \mathcal{U}$, $a \in \mathcal{U}$, and there is $b \in \mathcal{U}'$ with $b < a$. Wlog, let b be $<$-minimal in $\mathcal{U}'$. By WCP, there is an argument $\mathcal{V} \vdash \overline{b}$ with $\mathcal{V} \subseteq (\mathcal{U}' \setminus \{b\}) \cup \{a\}$. By $<$-minimality of b, no originating attacks from the argument are reversed. Hence, $\mathcal{U}$ normally attacks itself ($\mathcal{V} \subseteq \mathcal{U}$ and $b \in \mathcal{U}$). We proceed as in the previous case.

Next, we show that $\mathcal{U}$ is admissible. Let $\mathcal{V} \subseteq \mathcal{A}$ attack $\mathcal{U}$. Again, we distinguish the cases where $\mathcal{V}$ normally resp reversely attacks $\mathcal{U}$.

In the first case, there is $\mathcal{V}' \subseteq \mathcal{V}$, $a \in \mathcal{U}$ such that $\mathcal{V}' \vdash \overline{a}$. Hence, there is an argument $x \in S$ which is attacked by $\mathcal{V}' \vdash \overline{a}$ in $F_{\mathcal{D}}$. Since S is admissible in $F_{\mathcal{D}}$, there is some counter-attack on $\mathcal{V}' \vdash \overline{a}$. Since no assumption $b \in \mathcal{V}'$ is weaker than a, the only way to attack $\mathcal{V}' \vdash \overline{a}$ is via a normal attack. Hence, there is $y \in S$ which normally attacks the argument $\mathcal{V}' \vdash \overline{a}$, i.e., $y = \mathcal{U}' \vdash \overline{v}$ with $\mathcal{U}' \subseteq \mathcal{U}$ and $v \in \mathcal{V}$. We have shown that $\mathcal{U}$ defends itself against the attack from $\mathcal{V}$.

In the second case, there is $\mathcal{U}' \vdash \overline{a}$, $a \in \mathcal{V}$, $\mathcal{U}' \subseteq \mathcal{U}$, and there is $b \in \mathcal{U}'$ with $b < a$. Wlog, let b be $<$-minimal in $\mathcal{U}'$. By WCP, there is an argument $\mathcal{W} \vdash \overline{b}$ with $\mathcal{W} \subseteq (\mathcal{U}' \setminus \{b\}) \cup \{a\}$. By $<$-minimality of b, no originating attacks from the argument are reversed. Hence, to defend itself from the attack, S normally attacks $\mathcal{W} \vdash \overline{b}$. Let $y \in S$ normally attack the argument $\mathcal{W} \vdash \overline{b}$, i.e., $y = \mathcal{U}' \vdash \overline{w}$ with $\mathcal{U}' \subseteq \mathcal{U}$ and $w \in \mathcal{W}$. We have $w = a$ (otherwise, $w \in \mathcal{U}'$ and $\mathcal{U}$ is not conflict-free). Hence, $\mathcal{U}$ attacks $\mathcal{V}$ on a. We have shown that $\mathcal{U}$ defends itself against the attack. □

Lemma 47. *Let $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$ be an ABA$^+$ framework that satisfies weak contraposition, $\mathcal{F}_{\mathcal{D}} = (A, R, cl)$ its associated CAF, and $F_{\mathcal{D}} = (A, R)$ the underlying AF. Then $\mathcal{U} \in adm(\mathcal{D})$ implies $Args_{\mathcal{U}} \in adm(F_{\mathcal{D}})$.*

Proof. First, we show that $Args_{\mathcal{U}}$ is conflict-free. Let $x = \mathcal{V} \vdash q$ and $y = \mathcal{U}' \vdash p$ denote two arguments in $Args_{\mathcal{U}}$. It holds that $\mathcal{V} \subseteq \mathcal{U}$ and $\mathcal{U}' \subseteq \mathcal{U}$. Hence, in case x attacks y we have $\mathcal{V}$ (normally or reversely) attacks $\mathcal{U}$. Hence we obtain that $Args_{\mathcal{U}}$ is conflict-free.

Next, we show that $Args_{\mathcal{U}}$ is admissible. Let $x = \mathcal{V} \vdash q$ denote an argument in A which attacks an argument $y = \mathcal{U}' \vdash p$ in $Args_{\mathcal{U}}$ with $\mathcal{U}' \subseteq \mathcal{U}$. Since $\mathcal{U}$ defends itself, there is either a normal or a reversed counter-attack. In the first case, there is $\mathcal{U}'' \subseteq \mathcal{U}$ such that $\mathcal{U}'' \vdash \overline{b}$ for some $b \in \mathcal{V}$. Then $\mathcal{U}'' \vdash \overline{b} \in Args_{\mathcal{U}}$ and $Args_{\mathcal{U}}$ defends itself against the attack from x, as desired. In the second case, there is some assumption $d \in \mathcal{A}$ with $q = \overline{d}$ and some $c \in \mathcal{V}$ with $c < d$. The derivation $\mathcal{V} \vdash \overline{d}$ can be seen as failed attempt to attack the assumption d

(and every set containing d). Hence, the argument $z = \{d\} \vdash d$ which denotes the argument corresponding to the assumption d defends y against the attack from x. □

Proposition 38. *Let $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-, \leq)$ be an ABA$^+$ framework that satisfies weak contraposition, $\mathcal{F}_\mathcal{D} = (A, R, cl)$ its associated CAF, and $F_\mathcal{D} = (A, R)$ the underlying AF, and let $\sigma \in \{grd, com, prf, stb\}$ be a semantics.*

- *If $S \in \sigma(F_\mathcal{D})$ then $asms(S) \in \sigma(\mathcal{D})$;*
- *if $\mathcal{U} \in \sigma(\mathcal{D})$ then $Args_\mathcal{U} \in \sigma(F_\mathcal{D})$;*
- *the correspondence is one-to-one; i.e., $S = Args_{asms(S)}$ for each $S \in \sigma(F_\mathcal{D})$ and $\mathcal{U} = asms(Args_\mathcal{U})$ for all $\mathcal{U} \in \sigma(\mathcal{D})$.*

Proof. We first prove the statement for complete semantics. Since the fundamental lemma is satisfied, each preferred, grounded, and stable assumption set is complete, hence the proof for the remaining semantics will be based on complete semantics.

We proceed in four steps. First, we show that each complete set S in $F_\mathcal{D}$ is of the form $Args_{asms(S)}$ (first part of the 1-1 correspondence). Second, we show that for each $S \in com(F_\mathcal{D})$, $asms(S)$ is complete. Third, we prove that for each complete set $\mathcal{U}$, the set of arguments $Args_\mathcal{U}$ is complete. Lastly, we show that $\mathcal{U} = asms(Args_\mathcal{U})$ for each complete assumption set $\mathcal{U}$, to finish the proof of the 1-1 correspondence for complete sets.

- We show that for each $S \in com(F_\mathcal{D})$, it holds that $S = Args_{asms(S)}$. Let $\mathcal{U} = asms(S)$. By Lemma 46, $\mathcal{U}$ is admissible. We show that S defends all arguments $Args_\mathcal{U}$.

  Towards a contradiction, assume there is $\mathcal{U}' \vdash p$, $\mathcal{U}' \subseteq \mathcal{U}$, which is not contained in S. Let x denote an attack on $\mathcal{U}' \vdash p$. First assume x normally attacks $\mathcal{U}' \vdash p$. Then $cl(x) \in \mathcal{U}$ and hence x attacks S. Therefore, there is $y \in S$ which counter-attacks the attack. Next, assume the attack is reversed. Then $p = \overline{a}$ for some $a \in asms(x)$, and there is $b \in \mathcal{U}'$ with $b < a$. Wlog, let b be $<$-minimal in $\mathcal{U}'$. By WCP, there is an argument $\mathcal{W} \vdash \overline{b}$ with $\mathcal{W} \subseteq (\mathcal{U}' \setminus \{b\}) \cup \{a\}$. By $<$-minimality of b, no originating attacks from the argument are reversed. Hence, to defend itself from the attack, S normally attacks $\mathcal{W} \vdash \overline{b}$. Since S is admissible, $a \in \mathcal{W}$ and there is $y \in S$ with $cl(\overline{a})$ which normally attacks $\mathcal{W} \vdash \overline{b}$. Therefore, y attacks x.

  We have shown that each argument in $Args_\mathcal{U}$ is defended by S. Hence, we obtain $S = Args_\mathcal{U}$.

- Let $S \in com(F_\mathcal{D})$. We show that $asms(S) \in com(\mathcal{D})$. Let $\mathcal{U} = asms(S)$. By Lemma 46, $\mathcal{U} \in adm(\mathcal{D})$. It remains to show that $\mathcal{U}$ contains all assumption sets it defends. As shown in the above item, it holds that $S = Args_\mathcal{U}$. Let $\mathcal{V}$ denote an assumption set that is defended by $\mathcal{U}$. We show that S defends all arguments constructible from $\mathcal{V}$.

  Let $\mathcal{V}' \vdash q$ with $\mathcal{V}' \subseteq \mathcal{V}$ and let x denote an attacker of $\mathcal{V}' \vdash q$.

  First assume x normally attacks $\mathcal{V}' \vdash q$. Then $x = \mathcal{W} \vdash \overline{v}$ for some $v \in \mathcal{V}'$. Moreover, there is no $w \in \mathcal{W}$ with $w < v$ (otherwise, the attack would not have been successful). Since $\mathcal{U}$ defends $\mathcal{V}$ there is $\mathcal{U}' \vdash \overline{w}$ for some $w \in \mathcal{W}$. Hence S defends $\mathcal{V}' \vdash q$.

Next assume x reversely attacks $\mathcal{V}' \vdash q$. That is, $q = \overline{a}$, $b < a$ for some $b \in \mathcal{V}'$, and $a \in \mathit{asms}(x)$. Wlog, let b be $<$-minimal in $\mathit{asms}(x)$. By WCP, there is an argument $\mathcal{W} \vdash \overline{b}$ with $\mathcal{W} \subseteq (\mathcal{V}' \setminus \{b\}) \cup \{a\}$. By $<$-minimality of b, no originating attacks from the argument are reversed. Since $\mathcal{U}$ defends $\mathcal{V}$ there is $\mathcal{U}' \vdash \overline{w}$ for some $w \in \mathcal{W}, \mathcal{U}' \subseteq \mathcal{U}$. We have $w = a$, otherwise $\mathcal{U}$ must attack itself in order to defend $\mathcal{V}$, contradiction. It follows that x is attacked on a, hence, S defends $\mathcal{V}' \vdash q$.

We have shown that S defends all arguments constructible from $\mathcal{V}$.

- Let $\mathcal{U} \in \mathit{com}(\mathcal{D})$. We show that $\mathit{Args}_{\mathcal{U}} \in \mathit{com}(F_{\mathcal{D}})$. By Lemma 47, we have that $\mathit{Args}_{\mathcal{U}}$ is admissible. It remains to show that $\mathit{Args}_{\mathcal{U}}$ contains each argument it defends. Let $x = \mathcal{V} \vdash q$ denote an argument in A which is defended by $\mathit{Args}_{\mathcal{U}}$. We show that $\mathcal{U}$ defends $\mathcal{V}$ in $\mathcal{D}$.

  Let $\mathcal{W}$ be a set of assumptions attacking $\mathcal{V}$ in $\mathcal{D}$.

  First assume $\mathcal{W}$ normally attacks $\mathcal{V}$. Then there is $\mathcal{W}' \vdash \overline{a}$ with $\mathcal{W}' \subseteq \mathcal{W}$, $a \in \mathcal{V}$, and a is not weaker than any assumption in $\mathcal{W}'$. Then $\mathcal{W}' \vdash \overline{a}$ attacks the argument $\mathcal{V} \vdash q$ in $F_{\mathcal{D}}$. By assumption, there is an argument $y \in \mathit{Args}_{\mathcal{U}}$ that attacks $\mathcal{W}' \vdash \overline{a}$. Since a is not weaker than any assumption in $\mathcal{W}'$, the argument $\mathcal{W}' \vdash \overline{a}$ has no incoming reversed attack (otherwise, there would be some $b \in \mathcal{W}'$ such that $b < a$, contradiction). Hence, there is $\mathcal{U}' \subseteq \mathcal{U}$ such that $\mathcal{U}' \vdash \overline{b}$ for some $b \in \mathcal{W}'$. We obtain that $\mathcal{U}$ defends $\mathcal{V}$ against the attack from $\mathcal{W}$.

  Now assume $\mathcal{W}$ reversely attacks $\mathcal{V}$. By definition, there is $\mathcal{V}' \subseteq \mathcal{V}$ such that $\mathcal{V}' \vdash \overline{w}$ for some $w \in \mathcal{W}$ and there is $v \in \mathcal{V}'$ with $v < w$. Wlog, let v be the $<$-minimal element in $\mathcal{V}'$ with $v < w$. By WCP, there is $\mathcal{W}'' \vdash \overline{v}$ with $\mathcal{W}'' \subseteq (\mathcal{V}' \setminus \{v\}) \cup \{w\}$. Hence, the argument $\mathcal{W}'' \vdash \overline{v}$ attacks $\mathcal{V} \vdash q$ in $F_{\mathcal{D}}$. By assumption, there is an argument $y \in \mathit{Args}_{\mathcal{U}}$ which attacks $\mathcal{W}'' \vdash \overline{v}$. Since v is the $<$-minimal element among $\mathcal{W}'' \cup \{v\}$, the only way to attack the argument is via a normal attack. Hence, $y = \mathcal{W}^* \vdash \overline{w''}$ for some $w'' \in \mathcal{W}''$, $\mathcal{W}^* \subseteq \mathcal{U}$, and there is no element in $\mathcal{W}^*$ that is weaker than w''. Now, from $\mathcal{W}'' \subseteq (\mathcal{V}' \setminus \{v\}) \cup \{w\}$ we get that either $w'' \in \mathcal{V}'$ or $w'' = w$. In the first case, we derive a contradiction: if $w'' \in \mathcal{V}'$ then y attacks $\mathcal{V} \vdash q$. Hence, in order to defend the argument, $\mathit{Args}_{\mathcal{U}}$ must attack $y \in \mathit{Args}_{\mathcal{U}}$, and hence $\mathit{Args}_{\mathcal{U}}$ is not conflict-free, contradiction. In the other case, i.e., if $w'' = w$, we have that y attacks the argument $\mathcal{W}'' \vdash \overline{v}$ on the assumption $w \in \mathcal{W}$. Hence we obtain that $\mathcal{U}$ defends $\mathcal{V}$ against the attack from $\mathcal{W}$.

  We obtain that $\mathcal{U}$ defends $\mathcal{V}$ in $\mathcal{D}$. Since $\mathcal{U}$ is complete, we have $\mathcal{V} \subseteq \mathcal{U}$. Consequently, $x = \mathcal{V} \vdash q \in \mathit{Args}_{\mathcal{U}}$ and we obtain that $\mathit{Args}_{\mathcal{U}}$ defends each argument, as desired.

- We show that $\mathcal{U} = \mathit{asms}(\mathit{Args}_{\mathcal{U}})$ for all $\mathcal{U} \in \mathit{com}(\mathcal{D})$: Let $\mathcal{U} \in \mathit{com}(\mathcal{D})$. By the second item, $\mathit{Args}_{\mathcal{U}}$ is complete in $F_{\mathcal{D}}$. By the first item, $\mathit{asms}(\mathit{Args}_{\mathcal{U}})$ is complete in $\mathcal{D}$. We obtain $\mathcal{U} = \mathit{asms}(\mathit{Args}_{\mathcal{U}})$ because each argument of the form $\{a\} \vdash a$ is contained in $\mathit{Args}_{\mathcal{U}}$.

We have shown that the instantiation preserves complete semantics when $\mathcal{D}$ satisfies WCP. Moreover, the correspondence between the extensions of $F_{\mathcal{D}}$ and $\mathcal{D}$ is one-to-one. Hence, for $\sigma \in \{\mathit{grd}, \mathit{prf}\}$, the statement follows immediately by taking the $\subseteq$-minimal resp. $\subseteq$-maximal extensions.

It remains to prove the statement for stable semantics.

- First, let $S \in stb(F_{\mathcal{D}})$, and let $\mathcal{U} = asms(S)$. Since each stable extension is complete, we can assume $S = Args_{\mathcal{U}}$. To show that $\mathcal{U} \in stb(\mathcal{D})$, it remains to prove that each assumption $a \in \mathcal{A} \setminus \mathcal{U}$ is attacked. Let $a \in \mathcal{A} \setminus \mathcal{U}$, and tet $x = \{a\} \vdash a$ denote the argument corresponding to a. Since S is stable, S attacks x. An assumption argument receives only normal attacks (since the support cannot be weaker than the claim), hence, we can construct an argument $\mathcal{U}' \vdash \overline{a}$ with $\mathcal{U}' \vdash \mathcal{U}$ and obtain that $\mathcal{U}$ attacks a. Since a was arbitrary, it follows that $\mathcal{U}$ is stable in $\mathcal{D}$.

- Now, let $\mathcal{U} \in stb(\mathcal{D})$, and let $S = Args_{\mathcal{U}}$. We show that $S \in stb(F_{\mathcal{D}})$. Let x be an argument in $F_{\mathcal{D}}$. We show that x is either attacked by or contained in S. Let $x = \mathcal{V} \vdash q$. In case there is $a \in \mathcal{V}$ with $a \in \mathcal{A} \setminus \mathcal{U}$, it holds that x is attacked by S since $\mathcal{U}$ attacks each assumption in $\mathcal{A} \setminus \mathcal{U}$. In case $\mathcal{V} \subseteq \mathcal{U}$ we have $x \in Args_{\mathcal{U}} = S$. This proves the statement. □

References

Alfano, G., Greco, S., Parisi, F., & Trubitsyna, I. (2022). On preferences and priority rules in abstract argumentation. In *Proc. IJCAI'22*, pp. 2517–2524. IJCAI Organization.

Alfano, G., Greco, S., Parisi, F., & Trubitsyna, I. (2023). Abstract argumentation framework with conditional preferences. In *Proc. AAAI'23*, pp. 6218–6227. AAAI Press.

Amgoud, L., & Cayrol, C. (1998). On the acceptability of arguments in preference-based argumentation. In *Proc. UAI'98*, pp. 1–7. Morgan Kaufmann.

Amgoud, L., Cayrol, C., Lagasquie, M.-C., & Livet, P. (2008). On bipolarity in argumentation frameworks. *Int. J. Intelligent Systems*, *23*, 1–32.

Amgoud, L., & Vesic, S. (2014). Rich preference-based argumentation frameworks. *Int. J. Approx. Reason.*, *55*(2), 585–606.

Arora, S., & Barak, B. (2009). *Computational Complexity: A Modern Approach* (1st edition). Cambridge University Press, New York, NY, USA.

Atkinson, K., Baroni, P., Giacomin, M., Hunter, A., Prakken, H., Reed, C., Simari, G. R., Thimm, M., & Villata, S. (2017). Towards artificial argumentation. *AI Magazine*, *38*(3), 25–36.

Atkinson, K., & Bench-Capon, T. J. M. (2021). Value-based argumentation. In *Handbook of Formal Argumentation, Volume 2*, pp. 299–354. College Publications. also appears in IfCoLog Journal of Logics and their Applications 8(6):1543–1588.

Bao, Z., Cyras, K., & Toni, F. (2017). Abaplus: Attack reversal in abstract and structured argumentation with preferences. In *Proc. PRIMA'17*, Vol. 10621 of *Lecture Notes in Computer Science*, pp. 420–437. Springer.

Baroni, P., Caminada, M., & Giacomin, M. (2018). Abstract argumentation frameworks and their semantics. In *Handbook of Formal Argumentation*, chap. 4, pp. 159–236. College Publications.

Baroni, P., & Giacomin, M. (2007). On principle-based evaluation of extension-based argumentation semantics. *Artif. Intell.*, *171*(10-15), 675–700.

Baroni, P., Governatori, G., & Riveret, R. (2016). On labelling statements in multi-labelling argumentation. In *Proc. ECAI'16*, Vol. 285 of *FAIA*, pp. 489–497. IOS Press.

Baroni, P., & Riveret, R. (2019). Enhancing statement evaluation in argumentation via multi-labelling systems. *J. Artif. Intell. Res.*, *66*, 793–860.

Baumeister, D., Järvisalo, M., Neugebauer, D., Niskanen, A., & Rothe, J. (2021). Acceptance in incomplete argumentation frameworks. *Artif. Intell.*, *295*, 103470.

Bernreiter, M. (2024). *Integrated Preferences in Logic and Abstract Argumentation*. Ph.D. thesis, TU Wien (to appear).

Bernreiter, M., Dvořák, W., Rapberger, A., & Woltran, S. (2023). The effect of preferences in abstract argumentation under a claim-centric view. In *Proc. AAAI'23*, pp. 6253–6261. AAAI Press.

Bernreiter, M., Dvořák, W., & Woltran, S. (2024). Abstract argumentation with conditional preferences. *Argument & Computation*, *15*(2), 161–189.

Bistarelli, S., & Santini, F. (2021). Weighted argumentation. In *Handbook of Formal Argumentation, Volume 2*, pp. 355–395. College Publications. also appears in IfCoLog Journal of Logics and their Applications 8(6):1589–1622.

Bondarenko, A., Dung, P. M., Kowalski, R. A., & Toni, F. (1997). An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, *93*, 63–101.

Brewka, G., Truszczynski, M., & Woltran, S. (2010). Representing preferences among sets. In *Proc. AAAI'10*, pp. 273–278. AAAI Press.

Caminada, M., Sá, S., Alcântara, J., & Dvořák, W. (2015). On the equivalence between logic programming semantics and argumentation semantics. *Int. J. Approx. Reasoning*, *58*, 87–111.

Cayrol, C., de Saint-Cyr, F. D., & Lagasquie-Schiex, M. (2010). Change in abstract argumentation frameworks: Adding an argument. *J. Artif. Intell. Res.*, *38*, 49–84.

Cyras, K., Fan, X., Schulz, C., & Toni, F. (2018). Assumption-based argumentation: Disputes, explanations, preferences. In *Handbook of Formal Argumentation*, chap. 7, pp. 365–408. College Publications. also appears in IfCoLog Journal of Logics and their Applications 4(8):2407–2456.

Cyras, K., & Toni, F. (2016). ABA+: assumption-based argumentation with preferences. In *Proc.KR'16*, pp. 553–556. AAAI Press.

Delgrande, J. P., Schaub, T., & Tompits, H. (2003). A framework for compiling preferences in logic programs. *Theory and Practice of Logic Programming*, *3*(2), 129–187.

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.*, *77*(2), 321–358.

Dung, P. M., Kowalski, R. A., & Toni, F. (2009). Assumption-based argumentation. In *Argumentation in Artificial Intelligence*, pp. 199–218. Springer.

Dunne, P. E., Dvořák, W., Linsbichler, T., & Woltran, S. (2015). Characteristics of multiple viewpoints in abstract argumentation. *Artif. Intell.*, *228*, 153–178.

Dvořák, W., & Dunne, P. E. (2018). Computational problems in formal argumentation and their complexity. In *Handbook of Formal Argumentation*, chap. 14, pp. 631–687. College Publications. also appears in IfCoLog Journal of Logics and their Applications 4(8):2557–2622.

Dvořák, W., Greßler, A., Rapberger, A., & Woltran, S. (2023). The complexity landscape of claim-augmented argumentation frameworks. *Artif. Intell.*, *317*, 103873.

Dvořák, W., Rapberger, A., & Woltran, S. (2020). Argumentation semantics under a claim-centric view: Properties, expressiveness and relation to SETAFs. In *Proc. KR'20*, pp. 341–350. IJCAI.org.

Dvořák, W., Rapberger, A., & Woltran, S. (2023). A claim-centric perspective on abstract argumentation semantics: Claim-defeat, principles, and expressiveness. *Artif. Intell.*, *324*, 104011.

Dvořák, W., & Woltran, S. (2020). Complexity of abstract argumentation under a claim-centric view. *Artif. Intell.*, *285*, 103290.

Fazzinga, B., Flesca, S., & Furfaro, F. (2020). Revisiting the notion of extension over incomplete abstract argumentation frameworks.. In *Proc. IJCAI'20*, pp. 1712–1718.

Horty, J. F. (2002). Skepticism and floating conclusions. *Artif. Intell.*, *135*(1-2), 55–72.

Kaci, S., van der Torre, L. W. N., Vesic, S., & Villata, S. (2021). Preference in abstract argumentation. In *Handbook of Formal Argumentation, Volume 2*, pp. 211–248. College Publications.

Kaci, S., van der Torre, L. W. N., & Villata, S. (2018). Preference in abstract argumentation. In *Proc. COMMA'18*, Vol. 305 of *FAIA*, pp. 405–412. IOS Press.

König, M., Rapberger, A., & Ulbricht, M. (2022). Just a matter of perspective. In *Proc.COMMA'22*, Vol. 353 of *FAIA*, pp. 212–223. IOS Press.

Li, H., Oren, N., & Norman, T. J. (2011). Probabilistic argumentation frameworks. In *Proc. TAFA'11*, Vol. 7132 of *LNCS*, pp. 1–16. Springer.

Modgil, S. (2009). Reasoning about preferences in argumentation frameworks. *Artif. Intell.*, *173*(9-10), 901–934.

Modgil, S., & Prakken, H. (2013). A general account of argumentation with preferences. *Artif. Intell.*, *195*, 361–397.

Modgil, S., & Prakken, H. (2018). Abstract rule-based argumentation. In *Handbook of Formal Argumentation*, chap. 6, pp. 287–364. College Publications.

Prakken, H., & Vreeswijk, G. A. (2002). Logics for defeasible argumentation. In *Handbook of Philosophical Logic*, Vol. 4. Springer.

Rapberger, A. (2020). Defining argumentation semantics under a claim-centric view. In *Proc. STAIRS'20*, Vol. 2655 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Rapberger, A. (2023). *Unpacking The Argument – A Claim-centric View On Abstract Argumentation*. Ph.D. thesis, TU Wien.

Rapberger, A., & Ulbricht, M. (2023). On dynamics in structured argumentation formalisms. *J. Artif. Intell. Res.*, *77*, 563–643.

Rocha, V. H. N., & Cozman, F. G. (2022a). Bipolar argumentation frameworks with explicit conclusions: Connecting argumentation and logic programming. In *Proc.NMR'22*, Vol. 3197 of *CEUR Workshop Proceedings*, pp. 49–60. CEUR-WS.org.

Rocha, V. H. N., & Cozman, F. G. (2022b). A credal least undefined stable semantics for probabilistic logic programs and probabilistic argumentation. In *Proc. KR'22*, pp. 309–319. IJCAI Organization.

Sakama, C., & Inoue, K. (2000). Prioritized logic programming and its application to commonsense reasoning. *Artif. Intell.*, *123*(1-2), 185–222.

van der Torre, L., & Vesic, S. (2017). The principle-based approach to abstract argumentation semantics. *IfCoLog Journal of Logic and its Applications*, *4*(8), 2735–2778.

Wakaki, T. (2020). Consistency in assumption-based argumentation. In *Proc.COMMA'20*, Vol. 326 of *FAIA*, pp. 371–382. IOS Press.