Proceedings of the 22nd International Conference on Principles of Knowledge Representation and Reasoning
Main Track

On Strong and Weak Admissibility in Non-Flat Assumption-Based Argumentation

Matti Berthold[1] , **Lydia Blümel**[2] , **Anna Rapberger**[3]
[1]ScaDS.AI, Universität Leipzig
[2]Artificial Intelligence Group, University of Hagen
[3]Imperial College London
berthold@informatik.uni-leipzig.de, lydia.bluemel@fernuni-hagen.de, a.rapberger@imperial.ac.uk

Abstract

In this work, we broaden the investigation of admissibility notions in the context of assumption-based argumentation (ABA). More specifically, we study two prominent alternatives to the standard notion of admissibility from abstract argumentation, namely strong and weak admissibility, and introduce the respective preferred, complete and grounded semantics for general (sometimes called non-flat) ABA. To do so, we use abstract bipolar set-based argumentation frameworks (BSAFs) as formal playground since they concisely capture the relations between assumptions and are expressive enough to represent general non-flat ABA frameworks, as recently shown. While weak admissibility has been recently investigated for a restricted fragment of ABA in which assumptions cannot be derived (flat ABA), strong admissibility has not been investigated for ABA so far. We introduce *strong admissibility* for ABA and investigate desirable properties. We furthermore extend the recent investigations of *weak admissibility* in the flat ABA fragment to the non-flat case. We show that the central modularization property is maintained under classical, strong, and weak admissibility. We also show that strong and weakly admissible semantics in non-flat ABA share some of the shortcomings of standard admissible semantics and discuss ways to address these.

1 Introduction

Computational argumentation is a dynamic and widely studied area within knowledge representation and reasoning (Gabbay et al. 2021). It provides foundational models for human reasoning processes when challenged with conflicting information with the goal of identifying sets of jointly acceptable assumptions, premises, or arguments, representing coherent and defensible viewpoints. A well-established formalism in this domain is assumption-based argumentation (ABA) (Bondarenko et al. 1997; Cyras et al. 2018), which has been thoroughly investigated (Heyninck and Arieli 2024; Lehtonen, Wallner, and Järvisalo 2021; Lehtonen et al. 2023; Rapberger and Ulbricht 2024) and has applications in fields such as decision making (Fan et al. 2014; Cyras et al. 2021), planning (Fan 2018), and explainable AI (Russo, Rapberger, and Toni 2024; Leofante et al. 2024). In ABA frameworks (ABAFs), the reasoning process revolves around identifying acceptable sets of assumptions, which are the defeasible elements of the framework. The focus on inference rules and assumptions places ABA within the subfield of *structured argumentation*, which explicitly accounts for the internal structure of arguments. In contrast, *abstract argumentation* (Dung 1995) models argumentative processes by directed graphs where nodes are abstract arguments and edges represent attacks between them. Various argumentation semantics, largely shared across all major argumentation formalisms, are used to formalize when assumption sets can be considered acceptable, we refer to, e.g., (Baroni, Caminada, and Giacomin 2018; Cyras et al. 2018; Besnard et al. 2014) for an overview.

At the heart of most of these semantics lies the concept of *admissibility* (Dung 1995). This notion, which can be seen as the argumentative counterpart to self-defense, is crucial for defining acceptance and defeat in computational argumentation. Dung (1995) introduced admissibility for abstract argumentation frameworks (AFs), encapsulated with the informal slogan: *the one who has the last word laughs best*. A set of arguments is said to be admissible if it ensures both internal coherence (no self-conflict) and self-defense (ability to counter attacks). In the context of ABA, an assumption a can be challenged (*attacked*) by a set of assumptions S if S derives the negation (the so-called *contrary*) of a. A set of assumptions is then said to be *admissible* if it is conflict-free and counters all of its attackers.

Beyond the classical notion, several refinements of admissibility have been proposed. *Strong admissibility* strengthens defense: a member of a strongly admissible set cannot defend itself. This property has been introduced by Baroni and Giacomin (2007) and first studied as an independent semantics by Caminada (2014) in the context of abstract argumentation; using the following recursive definition: a set S is strongly admissible if each $a \in S$ is strongly defended by a subset of $S \setminus \{a\}$. It addresses concerns about circular justifications which are tolerated under standard admissibility. Strong admissiblity is well understood for AFs (Caminada and Dunne 2020; Caminada and Harikrishnan 2024a) and has been studied in generalizations of AFs (Keshavarzi Zafarghandi, Verbrugge, and Verheij 2022; Bistarelli and Taticchi 2022). In contrast, the realm of structured argumentation lacks similar investigations so far.

On the other end of the spectrum, *weak admissibility* has been proposed as a relaxation of the classical approach. Originally introduced for AFs by Baumann, Brewka, and Ulbricht (2020b), weak admissiblity addresses situations involving paradoxical arguments, such as self-attacking cycles, in which the standard notion of admissibility can be overly restrictive or unintuitive (Dung 1995; Dondio and Longo 2019). The weaker version of the semantics allows for a more relaxed approach to defense that can ignore such paradoxes, focusing instead on maintaining the acceptability of reasonable arguments. Crucial for the notion is the *reduct* which removes arguments whose acceptance status is already decided, i.e., either accepted or defeated. The reduct is the technical prerequisite for the *modularization property* which intuitively captures the idea that the evaluation of a semantics can be broken down into the evaluation on subframeworks, based on the reduct computation. In the context of AFs, weak admissibility is well understood (Baumann, Brewka, and Ulbricht 2022; Dauphin, Rienstra, and van der Torre 2021; Blümel and Ulbricht 2022). In recent work, Blümel, König, and Ulbricht (2024) investigate weak admissibility in the context of ABA. They focus, however, exclusively on the *flat* ABA fragment in which assumptions cannot be derived (Bondarenko et al. 1997). The flat setting restricts the framework's expressiveness when applied to complex domains. In many contexts —such as legal reasoning, ethical deliberation, or cognitive modeling— assumptions are often interdependent, something flat ABA cannot represent. Despite its potential, the non-flat setting remains underexplored and offers a promising direction for extending the applicability and depth of assumption-based frameworks, especially in combination with alternative notions of admissibility.

Therefore, unlike prior work, we do not restrict ourselves to the flat case but explore admissiblity under these more general conditions. Naturally, this poses additional challenges as (i) in addition to attacks, assumptions can now also *support* each other—analogously to the case of attacks, a set of assumptions S supports an assumption a if S derives a—and (ii) assumption sets can carry implicit conflicts due to the assumptions they additionally derive. To address (ii), acceptability in non-flat ABAFs requires in addition to conflict-freeness and defense that the assumption set is *closed*, that is, it contains all assumptions it derives. This ensures that implicit conflicts are exposed and that the acceptability of a set can be meaningfully evaluated only when all its logical consequences are taken into account.

We base our investigations on *bipolar set-based AFs (BSAFs)* (Berthold, Rapberger, and Ulbricht 2024) as they concisely capture the attack and support relations between assumptions and abstract away from all elements such as rules, ordinary atoms, argument trees that are only indirectly used in the evaluation of an ABAF.

Overall, our main contributions are as follows.

- We introduce *strong admissibility* for general (non-flat) ABA. We investigate fundamental properties of strong admissibility such as $(\mathbf{UM})$ which states that the $\subseteq$-maximal strongly admissible set is unique. While some of these properties may be violated in the general case, the semantics behave as expected in the flat ABA fragment, as we show. Section 5
- We generalize *weak admissibility* to non-flat ABA and study its properties. We show that some of the fundamental principles of weakly admissible semantics do not hold for non-flat ABA, e.g., removing self-attacking assumptions $(\mathbf{PA})$ is not always possible. Section 6
- Towards generalizing weak admissiblity to the non-flat case, we define and investigate the BSAF reduct and the closely related modularization property, and show that modularization holds for both weak and strong admissiblity, as well as the classical semantics in non-flat ABA. Section 4
- Inspired by recent attempts to alleviate the undesired behavior of ABA semantics in the non-flat case, we investigate the notion of Γ-closure in the context of weak and strong admissiblity. Intuitively, the Γ-closure of a set S allows to ignore assumptions that are not defended by S. While the adjustment fails to fix weak semantics, we show that the resulting *strongly* Γ*-admissible* semantics retains all desired properties. Section 7

A full version of the paper including all proofs is available online.[1]

2 Background

2.1 Assumption-based Argumentation

We recall the technical definitions of (ABA) (Cyras et al. 2018). We assume a *deductive system*, i.e. a tuple $(\mathcal{L}, \mathcal{R})$, where $\mathcal{L}$ is a set of atoms and $\mathcal{R}$ is a set of inference rules over $\mathcal{L}$. A rule $r \in \mathcal{R}$ has the form $a_0 \leftarrow a_1, \ldots, a_n$, s.t. $a_i \in \mathcal{L}$ for all $0 \leq i \leq n$; $head(r) := a_0$ is the *head* and $body(r) := \{a_1, \ldots, a_n\}$ is the (possibly empty) *body* of r.

Definition 2.1. *An ABA framework (ABAF) is a tuple $(\mathcal{L}, \mathcal{R}, \mathcal{A}, c)$, where $(\mathcal{L}, \mathcal{R})$ is a deductive system, $\mathcal{A} \subseteq \mathcal{L}$ a set of assumptions, and $c : \mathcal{A} \to \mathcal{L}$ a contrary function.*

We fix an arbitrary ABAF $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, c)$ below. The ABAF $\mathcal{D}$ is *flat* iff $head(r) \notin \mathcal{A}$ for all $r \in \mathcal{R}$. In this work we focus on *finite* ABAFs, i.e. $\mathcal{L}$ and $\mathcal{R}$ are finite.

An atom $p \in \mathcal{L}$ is tree-derivable from assumptions $S \subseteq \mathcal{A}$ and rules $R \subseteq \mathcal{R}$, denoted by $S \vdash_R p$, if there is a finite rooted labeled tree t s.t. i) the root of t is labeled with p, ii) the set of labels for the leaves of t is equal to S or $S \cup \{\top\}$, and iii) for each node v that is not a leaf of t there is a rule $r \in R$ such that v is labeled with $head(r)$ and labels of the children correspond to $body(r)$ or $\top$ if $body(r) = \emptyset$. We write $S \vdash p$ iff there exists $R \subseteq \mathcal{R}$ such that $S \vdash_R p$.

Let $S \subseteq \mathcal{A}$. By $c(S) := \{c(a) \mid a \in S\}$ we denote the set of all contraries of S. Set S *attacks* a set $T \subseteq \mathcal{A}$ if there are $S' \subseteq S$ and $a \in T$ s.t. $S' \vdash c(a)$; if S attacks $\{a\}$ we say S attacks a. S is conflict-free ($S \in cf(\mathcal{D})$) if it does not attack itself. The closure $cl(S)$ of S is $cl(S) := Th_{\mathcal{D}}(S) \cap \mathcal{A}$ where $Th_{\mathcal{D}}(S) := \{p \in L \mid \exists S' \subseteq S : S' \vdash p\}$ denotes all derived conclusions. We write $cl(a)$ instead of $cl(\{a\})$ for singletons. We call $S \subseteq \mathcal{A}$ *closed* if $S = cl(S)$.

Now we consider defense (Bondarenko et al. 1997; Cyras et al. 2018). Observe that defense in general ABAFs is only required against closed sets of attackers.

[1] https://zenodo.org/records/16884917

Definition 2.2. *A set S of assumptions defends an assumption a iff for each closed set T which attacks a, we have S attacks T; S defends itself iff S defends each $b \in S$.*

A set E of assumptions is *admissible* ($E \in adm(F)$) iff E is conflict-free, closed and defends itself. We next recall grounded, complete, and preferred ABA semantics.

Definition 2.3. *Let $\mathcal{D}$ be an ABAF and $S \in adm(\mathcal{D})$. Then*
- *$S \in com(\mathcal{D})$ iff it contains every assumption it defends;*
- *$S \in grd(\mathcal{D})$ iff S is $\subseteq$-minimal in $com(\mathcal{D})$;*
- *$S \in pref(\mathcal{D})$ iff S is $\subseteq$-maximal in $adm(\mathcal{D})$.*

We denote by $\Sigma = \{adm, com, grd, pref\}$ the family of (classical, admissible-based) Dung semantics.

To study restrictions of ABAFs, the following notation will be useful. For an ABAF $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, c)$ and $S \subseteq \mathcal{A}$, we write $\mathcal{D}\downarrow_S$ for the framework that arises when restricting $\mathcal{D}$ to S, i.e., $\mathcal{D}\downarrow_S := (\mathcal{L}, \mathcal{R}, S, c|_S)$.

2.2 Bipolar Set-based Abstract Argumentation

Bipolar SETAFs (Berthold, Rapberger, and Ulbricht 2024) combine the ideas underlying argumentation frameworks with collective attacks (SETAFs) (Nielsen and Parsons 2006) and bipolar argumentation frameworks (BAFs) (Cayrol and Lagasquie-Schiex 2005; Amgoud et al. 2008; Ulbricht et al. 2024). Instead of only considering an attack relation, there is also a notion of support. Bipolar SETAFs (BSAFs) can model *collective* attacks and supports.

Definition 2.4. *A bipolar set-argumentation framework (BSAF) is a tuple $F = (A, R, S)$, where A is a finite set of arguments, $R \subseteq 2^A \times A$ is the attack relation and $S \subseteq 2^A \times A$ is the support relation.*

In this work, we consider finite BSAFs only, i.e., A is finite. A SETAF is a BSAF $F = (A, R, S)$ with $S = \emptyset$; an AF is a SETAF with $|T| = 1$ for all $(T, h) \in R$.

Definition 2.5. *Given BSAF $F = (A, R, S)$ and $E \subseteq A$, let*

$$supp_F(E) := E \cup \{h \in A \mid \exists (T, h) \in S : T \subseteq E\}.$$

We define the closure $cl_F(E) := \bigcup_{i \geq 1} supp^i_F(E)$ *of E; E is* closed *if $cl_F(E) = E$.*

Definition 2.6. *Given BSAF $F = (A, R, S)$, a set $E \subseteq A$* defends *$a \in A$ if for each closed attacker $E' \subseteq A$ of a, E attacks E'; E* defends *E' if E defends each $a \in E'$. The* characteristic function *is $\Gamma_F(E) := \{a \in A \mid E \text{ defends } a \text{ in } F\}$.*

We omit the subscript F for Γ and cl if clear from context.

Let us now head to BSAF semantics. A set E is conflict-free ($E \in cf(F)$) if it does not attack itself; E is admissible ($E \in adm(F)$) if it is conflict-free, closed and defends itself.

Definition 2.7. *Let F be an BSAF and let $E \in adm(F)$.*
- *$E \in com(F)$ iff E contains every assumption it defends;*
- *$E \in grd(F)$ iff E is $\subseteq$-minimal in $com(F)$;*
- *$E \in pref(F)$ iff E is $\subseteq$-maximal in $adm(F)$.*

Given a BSAF (A, R, S) and a set of arguments $E \subseteq A$, we denote $E^+_R := \{h \mid \exists T \subseteq A : (T, h) \in R\}$ and the range of E by $E^\oplus_R := E_R \cup E^+_R$. The index R may be omitted, if clear from the context. Graphically, we depict the attack relation of a BSAF by solid edges and the support relation by dashed edges (cf. Example 2.10).

ABA and BSAF BSAFs capture non-flat ABAFs, as shown by Berthold, Rapberger, and Ulbricht (2024).

Definition 2.8. *Let $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, c)$ be an ABAF. Then we set $F_\mathcal{D} := (A, R, S)$, where $A := \mathcal{A}$ and*

$$R := \{(T, h) \mid h \in \mathcal{A},\ T \subseteq \mathcal{A},\ T \vdash c(h)\},$$
$$S := \{(T, h) \mid h \in \mathcal{A},\ T \subseteq \mathcal{A},\ T \vdash h\}\}$$

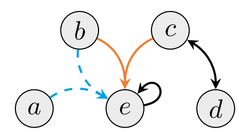
Berthold, Rapberger, and Ulbricht (2024) show that the BSAF abstraction of an ABAF preserves the semantics.

Theorem 2.9. *$\sigma(\mathcal{D}) = \sigma(F_\mathcal{D})$ for any $\sigma \in \Sigma$ and ABAF $\mathcal{D}$.*

Example 2.10. *We consider an ABAF $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, c)$ with literals $\mathcal{L} = \{a, b, c, d, e, a_c, b_c, c_c, d_c, e_c\}$, assumptions $\mathcal{A} = \{a, b, c, d, e\}$, their contraries a_c, b_c, c_c, d_c, e_c, respectively, and rules*

$$c_c \leftarrow d \quad e_c \leftarrow e \quad e \leftarrow a, b \quad d_c \leftarrow c \quad e_c \leftarrow b, c$$

We compute the BSAF $F_\mathcal{D} = (A, R, S)$: the nodes correspond to the assumptions, i.e., $A = \{a, b, c, d, e\}$, the attacks to the tree-derivations that derive contraries, i.e., $R = \{(\{d\}, c), (\{c\}, d), (\{b, c\}, e), (\{e\}, e)\}$, and the supports correspond to the tree-derivations that derive assumptions, in this case $S = \{(\{a, b\}, e)\}$.



The BSAF provides an easy-to-understand graphical representation to evaluate the semantics in $\mathcal{D}$. The sets $\{a\}$ and $\{b\}$ are admissible since they are unattacked; however, they cannot be accepted together because they jointly support the self-attacker e. Thus, the BSAF (and therefore the ABAF) has neither complete nor grounded extensions. It has three preferred extensions, $\{a, c\}$, $\{b, c\}$ and $\{b, d\}$.

3 Strong and Weak Admissibility in ABA

Our goal is to study semantics based on strong and weak admissibility in ABA in a principled way. We will thus identify desirable properties that we expect our new families of semantics adhere to.

To get a better understanding of the oftentimes involved technicalities of the recursive definitions of strong and weak admissibility, we make use of the close relationship between BSAFs and ABAFs. BSAFs concisely capture the attack and support relations between assumptions while abstracting away from all components of an ABAF that are only implicitly needed to compute the extensions, as discussed in Example 2.10. We will therefore utilize BSAFs as our formal playground to rigorously define and investigate our novel semantics.

In the remainder of this section, we consider an arbitrary but fixed ABAF $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, c)$.

Global Desiderata We discuss desired properties for both strong and weak admissiblility. A central property of argumentation semantics is the fundamental lemma, introduced by Dung (1995). It states that each assumption defended by an admissible set E can be accepted together with E.

($\mathbf{F}$) Fundamental Lemma: If $S \in adm(\mathcal{D})$ defends a, then $S \cup \{a\} \in adm(\mathcal{D})$.

We recall basic semantics relations; also, all semantics are expected to return a result.

($\mathbf{SR}$) Semantics Relations: $pref(\mathcal{D}) \subseteq com(\mathcal{D}) \subseteq adm(\mathcal{D})$ and $grd(\mathcal{D}) \subseteq com(\mathcal{D})$

($\mathbf{NE}$) Non-empty: $adm(\mathcal{D}), com(\mathcal{D}), pref(\mathcal{D}), grd(\mathcal{D}) \neq \emptyset$

We note that ($\mathbf{SR}$) implies the *maximal complete principle* ($\mathbf{MC}$) (Blümel, König, and Ulbricht 2024) which states that the preferred extensions correspond to the maximal complete extensions of a given framework.

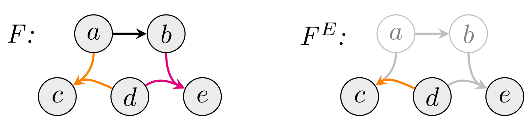
Proposition 3.1. ($\mathbf{SR}$) $\Rightarrow$ ($\mathbf{MC}$) *where* ($\mathbf{MC}$) *denotes* $pref(\mathcal{D}) = \{E \subseteq A \mid E \text{ is maximal in } com(\mathcal{D})\}$

Central to both strong and weak admissibility is the modularization property. A semantics satisfies modularization if its extensions can be computed iteratively by projecting away all elements that are already known to be either accepted or defeated. Crucial for this property is the reduct of a framework, originally introduced in (Baumann, Brewka, and Ulbricht 2020b). Below, we recall the definition for SETAFs as they correspond to flat ABAFs (Dvořák et al. 2021).

Definition 3.2. *Given a SETAF* $F = (A, R)$ *and* $E \subseteq A$, *the* E-reduct *of* F *is the SETAF* $F^E := (A^E, R^E)$, *with*

$$\begin{aligned} A^E &:= A \setminus E^{\oplus}_R \\ R^E &:= \{(T \setminus E, h) \mid (T, h) \in R,\, T \cap E^{+}_R = \emptyset, \\ &\qquad T \not\subseteq E,\, h \in A^E\} \end{aligned}$$

Example 3.3. *Consider the SETAF* F *and its reduct* F^E *wrt.* $E = \{a\}$ *(with removed arguments and attacks in lightgray), depicted below.*



The reduct F^E *of* F *wrt.* E *assumes* a *to be true. Thus,* b *is assumed to be false, since it is attacked by* E. *Both arguments* a *and* b *are therefore removed. The attack* $(\{b, d\}, e)$ *is deactivated because* b *is out, thus it is removed entirely; the attack* $(\{a, d\}, c)$ *can still fire, since* a *is in, thus, we adjust it accordingly and keep* $(\{d\}, c)$ *in* F^E.

We are ready to define *modularization* which allows to compute extensions in a modular fashion. The property holds for (SET)AFs (Baumann, Brewka, and Ulbricht 2020a; Dvořák et al. 2024), and we expect this property to hold for BSAFs (that resemble non-flat ABAFs) as well.

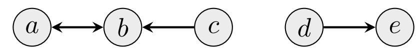
($\mathbf{M}$) Modularization: A semantics σ satisfies modularization iff for each $E \subseteq A, E' \subseteq A^E$ we have $E \in \sigma(F)$ and $E' \in \sigma(F^E)$ implies $E \cup E' \in \sigma(F)$.

Strong Admissibility Desiderata Strong admissibility strengthens the standard notion by requiring that each member of a strongly admissible set is defended by a strongly admissible subset that does not contain it. The notion has first been introduced in the scope of a principle-based analysis by Baroni and Giacomin (2007) for AFs and has first been studied as an independent semantics concept by Caminada (2014). We recall the AF definition.

Definition 3.4. *Let* $F = (A, R)$ *be an AF. A conflict-free set* $E \in cf(F)$ *is strongly admissible in* F *iff each* $a \in E$ *is defended by a strongly admissible set* $E' \subseteq E \setminus \{a\}$.

Similar to grounded semantics, strongly admissible extensions can be computed starting from a set of undefeated arguments. We discuss the idea in the following example.

Example 3.5. *Consider the following AF.*



A strongly admissible extension may, or may not contain c *or* d, *since they are unattacked. Accordingly, an extension may contain* a, *if it contains* c. *Note that* a *is not contained in a strongly admissible extension if* c *is not present. We have* $adm^s(F) = \{\emptyset, \{c\}, \{c, a\}, \{d\}, \{c, d\}, \{c, d, a\}\}$.

Analogous to the AF case, we expect that strong admissibility for ABA strengthens admissibility.

($\mathbf{S}$) Strengthening: It holds that $adm^s(\mathcal{D}) \subseteq adm(\mathcal{D})$.

We consider the following three desiderata, inspired by Caminada (2014). Unique maximum states that the set of strongly admissible extensions has a unique maximal element; Unique relative maximum states that this property also holds relative to a given admissible set; and complete containment states that each strongly admissible set is contained in each complete extension.

($\mathbf{UM}$) Unique Maximum: The $\subseteq$-maximal strongly admissible extension is unique.

($\mathbf{URM}$) Unique Relative Maximum: Each admissible set has a unique $\subseteq$-maximal strongly admissible subset.

($\mathbf{CC}$) Complete Containment: $E \in adm^s(\mathcal{D})$ and $E' \in com(\mathcal{D})$ imply $E \subseteq E'$.

Note that for AFs, ($\mathbf{CC}$) implies that all strongly admissible sets are a subset of the (unique) grounded extension.

Weak Admissibility Desiderata Blümel, König, and Ulbricht (2024) introduced weak admissiblity for the flat ABA fragment, basing their definition on SETAFs.

Definition 3.6. *Let* $F = (A, R)$ *be a SETAF, let* $E \subseteq A$ *be a set of arguments, and* $F^E = (A^E, R^E)$ *its* E*-reduct. Then* E *is called* weakly admissible *in* F *(*$E \in adm^w(F)$*) iff*

1. $E \in cf(F)$ *and*
2. *for each* $(T, h) \in R$ *with* $h \in E$, *and* $T \cap E^+_R = \emptyset$ *it holds* $\nexists E' \in adm^w(F^E)$ *s.t.* $T \cap A^E \subseteq E'$.

Let $\mathcal{D} = (\mathcal{L}, \mathcal{A}, \mathcal{R}, c)$ *be an ABAF and* $F_{\mathcal{D}}$ *the corresponding SETAF (cf. Definition 2.8). A set* $E \subseteq \mathcal{A}$ *of assumptions is* weakly admissible *(*$E \in adm^w(\mathcal{D})$*) iff* $E \in adm^w(F_{\mathcal{D}})$.

In their work, they identified desirable properties for weakly admissible semantics which we will recall below. First, the semantics is expected to weaken the traditional admissibility notion.

($\mathbf{L}$) Liberalization: It holds that $adm(\mathcal{D}) \subseteq adm^{w}(\mathcal{D})$.

A fundamental principle of weak admissibility is that it allows for deleting so-called paradoxical components. In AFs, these components are, for instance, self-attackers or odd cycles in general. In the case of ABAFs, Blümel, König, and Ulbricht (2024) identified the following counterpart for paradoxical assumptions.

($\mathbf{PA}$) Paradoxical Assumptions: If $\{a\} \vdash c(a)$, then it holds that $adm^{w}(\mathcal{D}) = adm^{w}(\mathcal{D} \downarrow_{A\setminus\{a\}})$.

Finally, we consider a novel principle similar to the paradoxical rule principle in (Blümel, König, and Ulbricht 2024). The principle involves attacks and supports in the ABAF that we deem paradoxical. We introduce the concept (in terms of BSAFs as they resemble ABA attacks and supports) below.

Definition 3.7. *Given a BSAF $F = (A, R, S)$. An attack $r = (T, h) \in R$ is* paradoxical *iff $T \neq \emptyset$ and for every $t \in T$ there is a $T' \subseteq T, T' \neq \emptyset$ s.t.*

- *there exists $(T', t) \in R$ and $h \notin T'$.*

A support $s = (T, h) \in S$ is paradoxical *iff $T \neq \emptyset$ and for every $t \in T$ there is a $T' \subseteq T, T' \neq \emptyset$ s.t.*

- *there exists $(T', t) \in R$.*

($\mathbf{PRS}$) Paradoxical Attacks/Supports: Removing a paradoxical attack r or support s does not alter the models of F, i.e. $adm^{w}(F) = adm^{w}(F')$ where $F' = (A, R \setminus \{r\}, S)$ (resp. $F' = (A, R, S \setminus \{s\})$).

Section Outline In the following sections, we will define strong and weak admissibility and study them in terms of the desiderata that we identified, thereby utilizing BSAFs as a formal playground. First, we will introduce the *BSAF reduct* and study modularization for classical Dung semantics in general (potentially non-flat) ABA (cf. Section 4). Second, we introduce *strong admissibility*, first for BSAFs, and subsequently for ABA in Section 5 where we study its behavior for non-flat ABA and in the flat fragment. Third, we discuss *weak admissibilty* for non-flat ABA in Section 6.

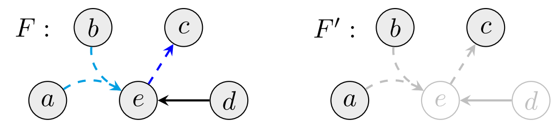
The classical semantics general (potentially non-flat) ABA are known to admit undesired behavior, as, e.g., discussed in (Heyninck and Arieli 2024; Berthold, Rapberger, and Ulbricht 2024). We (correctly) anticipate similar issues with the generalizations of the semantics based on strong respectively weak admissibility. We tackle these issues in Section 7 and consider revised versions of the semantics to reinstate ($\mathbf{F}$),($\mathbf{SR}$) and ($\mathbf{NE}$), as recently proposed in (Berthold, Rapberger, and Ulbricht 2024).

4 The BSAF Reduct

Towards modularization and weak admissibility for BSAFs, we first introduce the *E-reduct for BSAFs*. As for the AF and SETAF reduct (Baumann, Brewka, and Ulbricht 2020b; Dvořák et al. 2024), the BSAF reduct should capture the intuition that the arguments in a given set E are true, and all arguments in E^{+} are false.

It is tempting to proceed analogous to the (SET)AF case and set all arguments that are attacked to false (and remove them). However, when doing so, we encounter some issues.

Example 4.1. *Consider the BSAF F as depicted below (left) and the result F' of the SETAF reduct computation wrt. $E = \{d\}$ (cf. Definition 3.2) (right).*

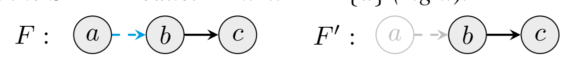


If we proceed as in the SETAF case, the argument e is defeated and thus can be removed. The resulting BSAF thus contains no attacks and supports, thus all arguments can be accepted in the next iteration. With supports present, however, we need to be a bit more careful since accepting the arguments a and b would require to accept the previously defeated argument e as well. Thus, we need to ensure that the set a and b cannot be jointly accepted. To do so, we will add attacks from the set $\{a, b\}$ to each of its members.

As shown in the previous example, the removal of an attack has consequences for all its supporting sets. To prevent that all arguments in a set S that supports an already defeated argument can get accepted, we add self-attacks (S, a) to all $a \in S$; this acts as a constraint. As soon as one of the arguments in the set gets defeated the attacks will be removed and the remaining arguments can be accepted.

We encounter another subtlety.

Example 4.2. *Consider the BSAF F as depicted below (left) and the SETAF reduct F' wrt. $E = \{a\}$ (right).*



However, since a supports b, we already know that b can be deemed accepted as well; and, consequently, c is defeated.

In the BSAF reduct, we will thus remove all true and defeated arguments wrt. the *closure* of E.

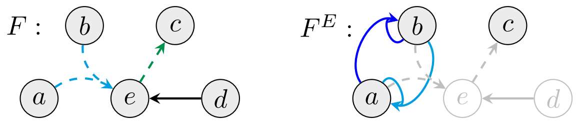
We define the reduct as follows.

- First, we add constraints, for each support (T, h) with $h \in cl(E)^{+}_{R}$, we add attacks (T, t) for all $t \in T$;
- Next, we remove all true and defeated arguments: compute the closure $cl(E)$ of E, remove all arguments that are contained in $cl(E)$ since they are true wrt. E, remove all arguments in $cl(E)^{+}_{R}$ since they are defeated wrt. E;
- Now, we remove all attacks and supports (T, h) whenever $h \in cl(E)^{\oplus}$ or $T \cap cl(E)^{+}_{R} \neq \emptyset$;
- Finally, we adjust the attacks and supports by restricting the remaining attacks and supports to $(T \setminus cl(E), h)$.

Definition 4.3. *Given a BSAF $F = (A, R, S)$ and $E \subseteq A$, the E*-reduct *of F is the BSAF $F^{E} := (A^{E}, R^{E}, S^{E})$, with*

$$
\begin{aligned}
A^{E} &:= A \setminus (cl(E)^{\oplus}_{R}) \\
R^{E} &:= \{(T \setminus cl(E), t) \mid \exists h \in cl(E)^{+}_{R} : (T, h) \in S, \\
&\qquad\qquad t \in T \cap A^{E}\} \cup \\
&\qquad \{(T \setminus cl(E), h) \mid T \cap cl(E)^{+}_{R} = \emptyset, \\
&\qquad\qquad (T, h) \in R^{E}, h \in A^{E}\} \\
S^{E} &:= \{(T \setminus cl(E), h) \mid T \cap cl(E)^{+}_{R} = \emptyset, \\
&\qquad\qquad (T, h) \in S, h \in A^{E}\}
\end{aligned}
$$

Example 4.1 (continued). *We depict the previous BSAF F and its BSAF reduct F^E for $E = \{d\}$ below.*



The reduct generalizes the SETAF-reduct (and the AF-reduct) in that for each BSAF (A, R, S) with $S = \emptyset$ (and $|H| = 1$ for all $(H, t) \in R$) the two reduct notions coincide.

We show that the reduct is compatible with union for conflict-free and closed sets.

Proposition 4.4. *Let $F = (A, R, S)$ be a BSAF, $E \subseteq A$ closed and conflict-free in F,$E' \subseteq A^E$ closed and conflict-free in F^E. Then $F^{E \cup E'} = (F^E)^{E'}$.*

Let us inspect whether our newly defined reduct behaves as expected with respect to modularization $(\mathbf{M})$. Towards proving this property, we first observe that the reduct guarantees to preserve the closedness of a set of arguments.

Proposition 4.5. *Let $F = (A, R, S)$ be a BSAF, then for each $E \subseteq A, E' \subseteq A^E$ it holds that*

- *$E \in cf(F)$ and closed F and $E' \in cf(F^E)$ and closed in $F^E \Rightarrow E \cup E'$ closed in F*
- *$E, E \cup E' \in cf(F)$ and closed in $F \Rightarrow E'$ closed in F^E*

We show that all semantics in Σ satisfy modularization (cf. $(\mathbf{M})$ in Section 3), using our newly defined reduct.

Proposition 4.6. *σ satisfies modularization $(\mathbf{M})$ for $\sigma \in \Sigma$.*

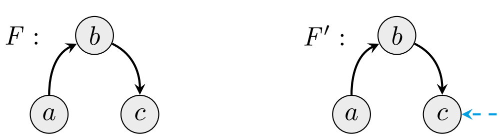
5 Strong Admissibility

Following the investigation of the abstract case in (Caminada 2014; Baumann, Linsbichler, and Woltran 2016), we lift strong admissibility to BSAFs, before we turn our attention to the consequences this has for ABA.

5.1 Strong Admissibility for BSAFs

Towards lifting strong admissibility to BSAF, lets recall that on AF it is defined recursively, where a set E is strongly admissible, if each arguments $a \in E$ is defended by a strongly admissible set $E' \subseteq E \setminus \{a\}$ (cf. Def 3.4). The support relation introduces a particularity that prevents us from using directly a similar recursion for BSAF: Clearly, a strongly admissible set should be closed, yet if strong admissibility is defined recursively on itself, as it is in AF, we would need a closed set on each step of the defense. Let us look at an example to illustrate this counter-intuitive behavior.

Example 5.1. *Consider the following two BSAFs F and F':*



In F we expect $\emptyset$ and $\{a\}$ to be strongly admissible, further $\{a, c\}$ is strongly admissible, since c is defended by the strongly admissible set $\{a\}$. In F', however, the same set $\{a, c\}$ is not strongly admissible, if we require c to be defended by a strongly admissible set, since $\{a\}$ is not closed. This is counterintuitive, since additional support on an extension should not speak against it.

We therefore decouple closedness and strong defense: an extension is strongly defended if it is defended by a smaller strongly defended extension, but does not need to be closed.

Definition 5.2. *Let $F = (A, R, S)$ be a BSAF. A set of arguments $E \subseteq A$ is* strongly defended *($E \in sd(F)$) iff E is conflict-free and for every $a \in E$ there exists a strongly defended subset $E' \subseteq E \setminus \{a\}$, which defends a.*

We are now ready to define strong admissibility on BSAF.

Definition 5.3. *Let $F = (A, R, S)$ be a BSAF. A set of arguments $E \subseteq A$ is* strongly admissible *($E \in adm^s(F)$) iff $E \in sd(F)$ and E is closed.*

Like for AFs, (Baumann, Linsbichler, and Woltran 2016), we can now give a constructive characterization of strongly admissible extensions.

Proposition 5.4. *Let $F = (A, R, S)$ be a BSAF, and $E \subseteq A$. Then $E \in adm^s(F)$ iff E is conflict-free, closed and there exists a finite sequence of pairwise disjoint sets $E_1, ..., E_n$ such that $E_1 = \emptyset$, $E = \bigcup_{i=1}^{n} E_i$ and for each $i \geq 1$ it holds that E_i is defended by $\bigcup_{j<i} E_j$.*

In the same vein we are able to proof the satisfaction of modularization $(\mathbf{M})$ wrt. strong admissible semantics.

Proposition 5.5. *Strongly admissible semantics satisfies modularization $(\mathbf{M})$.*

We will now discuss desirable properties specifically for for strong admissibility. We observe that strongly admissible semantics are a subset of admissible semantics, i.e., strengthening $(\mathbf{S})$ is satisfied. Likewise, the complete containment property $(\mathbf{CC})$ that formalizes that each strongly admissible set is contained in each complete extension holds for strongly admissible semantics for non-flat ABA.

Proposition 5.6. *Let $F = (A, R, S)$ be a BSAF. Then,*

1. *$adm^s(F) \subseteq adm(F)$; and*
2. *$E \in adm^s(F)$ and $E' \in com(F)$ implies $E \subseteq E'$.*

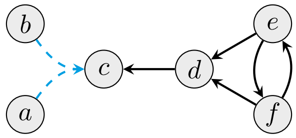
As a consequence, each BSAF has at most one complete strongly admissible extension; and if such a complete extension exists, it is also the unique grounded and the unique $\subseteq$-maximal strongly admissible extension.

Proposition 5.7. *Let $F = (A, R, S)$ be a BSAF. Then,*

1. *$|adm^s(F) \cap com(F)| \leq 1$; and*
2. *if $E \in adm^s(F) \cap com(F)$, then E is the unique grounded extension of F ($E \in grd(F)$) and the unique subset-maximal strongly admissible extension.*

In contrast, unique relative maximum $(\mathbf{URM})$ and unique maximum $(\mathbf{UM})$ are not satisfied, as illustrated by the following example. This example also shows that the complete strongly admissible extension does not always exist. It also demonstrates that the union of two strongly admissible sets may fail to be strongly admissible, thereby violating another property discussed in (Caminada 2014).

Example 5.8. *Consider the following BSAF F*



In F, the sets $\emptyset, \{a\}, \{b\}, \{a,b,c,e\}, \{a,b,c,f\}$ are admissible; the latter two are complete, grounded, and preferred. The strongly admissible sets are $adm^s(F) = \{\emptyset, \{a\}, \{b\}\}$.

- $(\mathbf{URM})$ *Both $\{a,b,c,e\}$ and $\{a,b,c,f\}$ have two $\subseteq$-maximal strongly admissible subsets: $\{a\}$ and $\{b\}$;*
- $(\mathbf{UM})$ *Both $\{a\}$ and $\{b\}$ are $\subseteq$-maximal in $adm^s(F)$;*
- *Their union $\{a,b\}$ is not strongly admissible.*

As the example shows, strongly admissible extensions take over an important function of the grounded semantics, i.e. they identify arguments with a stronger justification. In the abstract setting, the unique grounded extension contains only arguments, whose defense can be traced back to the empty set, and which therefore do not rely on cycles for their defense. In non-flat ABA, this is not true for members of grounded extensions in general, but still holds for arguments accepted under strongly admissible semantics. In general there can be more than one subset-maximal strongly admissible extension, which warrants the definition of *strongly preferred* and *strongly complete* semantics as a BSAF-semantics in its own right.

Definition 5.9. *Let F be a BSAF and let $E \in adm^s(F)$.*

- *$E \in com^s(F)$ iff E contains every assumption it defends;*
- *$E \in pref^s(F)$ iff E is $\subseteq$-maximal in $adm^s(F)$;*

We omit strongly grounded semantics since it coincides with strongly complete semantics (as observed below Proposition 5.6, $com^s(F)$ has at most one member). We write $\Sigma^s = \{adm^s, com^s, pref^s\}$ to denote the family of strongly admissible semantics. Note that the framework F above has two complete extensions $\{a,b,c,e\}$, and $\{a,b,c,f\}$. The strong variant does not have an extension in F, i.e. $com^s(F) = \emptyset$. Example 5.8 shows that the fundamental lemma $(\mathbf{F})$ is not satisfied, and that in general $pref^s(F) \subseteq com^s(F)$ does not hold.

Example 5.8 (continued). *The strongly admissible set $\{a\}$ defends the unattacked argument b. Since $\{a,b\}$ is not closed it is not strongly admissible. Thus the fundamental lemma $(\mathbf{F})$ is violated and we have no grounded extension, instead $\{a\}, \{b\}$ are our two strongly preferred extensions.*

5.2 Concequences for (flat and non-flat) ABA

In the last section we proved the (un)satisfiability of several desiderata that pose requirements of the ABA semantics directly in BSAF. We utilize these findings to define and investigate semantics based on strong admissibility for ABAFs.

Definition 5.10. *Given an ABAF $\mathcal{D} = (\mathcal{L}, \mathcal{R}, \mathcal{A}, c)$ and a semantics $\sigma \in \Sigma^s$, then $E \in \sigma(\mathcal{D})$ iff $E \in \sigma(F_{\mathcal{D}})$.*

Due to the close correspondence of ABAFs and BSAFs, the results of the previous section directly transfer to ABA. Strongly admissible semantics for ABA satisfy modularization $(\mathbf{M})$ since the corresponding BSAF semantics does, as shown in Proposition 5.5; moreover, strengthening $(\mathbf{S})$ and complete containment $(\mathbf{CC})$ is satisfied by Proposition 5.6. For the remaining cases, the counter-examples carry over.

Theorem 5.11. *The strongly admissible semantics for ABA satisfies $(\mathbf{M})$, $(\mathbf{S})$, and $(\mathbf{CC})$, but does not satisfy $(\mathbf{F})$, $(\mathbf{SR})$, $(\mathbf{NE})$, $(\mathbf{MC})$, $(\mathbf{UM})$, nor $(\mathbf{URM})$.*

While central properties are satisfied, our results show that, as anticipated, the family of strongly admissible semantics admits in some aspects undesired behavior.

For the flat ABA fragment, these issues do not occur. Flat ABA, directly corresponds to SETAF, meaning that the instantiation $F_{\mathcal{D}} = (A, R, S)$ of a flat ABAF $\mathcal{D}$ (via Def. 2.8) has no support relations, i.e. $S = \emptyset$. It holds that $(\mathbf{M})$, $(\mathbf{S})$ and $(\mathbf{CC})$ are satisfied since any SETAF is a BSAF. Further, if we take a look at the counter-examples used to show non-satisfaction, we notice all of them use at least one support. It turns out that all $(\mathbf{F})$, $(\mathbf{SR})$, $(\mathbf{NE})$, $(\mathbf{MC})$, $(\mathbf{UM})$, and $(\mathbf{URM})$ are indeed satisfied for flat ABA.

Theorem 5.12. *The strongly admissible semantics for flat ABA and SETAFs satisfy all of $(\mathbf{M})$, $(\mathbf{S})$, $(\mathbf{CC})$, $(\mathbf{F})$, $(\mathbf{SR})$, $(\mathbf{NE})$, $(\mathbf{MC})$, $(\mathbf{UM})$, and $(\mathbf{URM})$.*

6 Weak Admissibility

In contrast to strong admissibility our goal is now to accept as much as we reasonably can. In this section, we generalize the definition of weak admissibility by Blümel, König, and Ulbricht (2024) to non-flat ABA.

6.1 Weak Admissibility for BSAFs

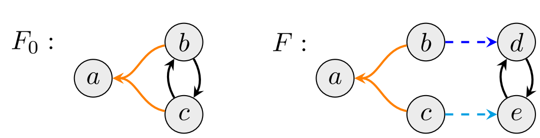
The E-reduct for BSAFs gives us the tools to generalize weak admissibility. Note that the definition is recursive, but well-defined as in each recursion step the reduct contains fewer arguments and we deal only with finite BSAFs.

Definition 6.1. *Let $F = (A, R, S)$ be a BSAF, $E \subseteq A$ a set of arguments, and $F^E = (A^E, R^E, S^E)$ its E-reduct. Then E is called* weakly admissible *in F ($E \in adm^w(F)$) iff*

1. *$E \in cf(F)$, E closed and*
2. *for each $(T, h) \in R$ with $h \in E$, and $T \cap E_R^+ = \emptyset$ it holds $\nexists E' \in adm^w(F^E)$ s.t. $T \cap A^E \subseteq E'$.*

As it is the case for SETAFs, the tail of a joint attack has to be part of a weakly admissible set (of the reduct) as a whole for it to be considered an attack one has to defend against. In BSAFs, we additionally have to take closedness into consideration. On the one hand, we limit the set of weakly admissible extensions to closed sets. On the other hand, this allows us to ignore even more involved types of unreasonable attacks, e.g. indirectly conflicting sets of attackers.

Example 6.2. *Compare the SETAF F_0 and BSAF F:*



In both frameworks $\{a\}$ is attacked by $\{b,c\}$, and weakly admissible. In F_0 the joint attack does not fire, because $\{b,c\}$ is not conflict-free.

At first sight, nothing seems to be wrong with $\{b, c\}$ in F, the set is not only conflict-free, it is even unattacked. However, b and c support conflicting arguments, so even though the closed subsets $\{b, d\}$ and $\{c, e\}$ are weakly admissible in the reduct $F^{\{a\}}$, there is no closed and conflict-free set containing both b and c, so the joint attack on a can be ignored.

Closedness neutralizes attacks from indirectly self-conflicting sets under weak admissibility. Next, we show that closedness of attackers already holds by definition.

Proposition 6.3. *Let $F = (A, R, S)$ be a BSAF, let $E \subseteq A$ be a set of arguments, and $F^E = (A^E, R^E, S^E)$ its E-reduct. Then $E \in adm^w(F)$ iff E is conflict-free, closed, and for every closed set E' which attacks E it holds that either E attacks E' or $(E' \setminus E) \notin adm^w(F^E)$.*

The notion of weak defense, as well as, weakly complete, weakly grounded and weakly preferred semantics generalize to BSAFs in the natural way.

Definition 6.4. *Let $F = (A, R, S)$ be a BSAF and let $E, X \subseteq A$. We say E* weakly defends *X (abbr. E w-defends X) if for each $(T, h) \in R$ with $h \in X$, and for every closed E' with $T \subseteq E'$ one of the following two conditions hold:*

- *E attacks E', or*
- *the following two conditions hold simultaneously:*
  1. *there is no $E^* \in adm^w(F^E)$ with $E' \subseteq E \cup E^*$,*
  2. *there is some X' s.t. $X \subseteq X' \in adm^w(F)$.*

Definition 6.5. *Let F be a BSAF and let $E \in adm^w(F)$:*

- *E is* weakly complete*, $E \in com^w(F)$, iff for each $X \supseteq E$ s.t. E w-defends X, we have $E = X$,*
- *E is* weakly preferred*, $E \in pref^w(F)$, iff E is maximal wrt. $\subseteq$ in $adm^w(F)$,*
- *E is* weakly grounded*, $E \in grd^w(F)$, iff E is minimal wrt. $\subseteq$ in $com^w(F)$.*

We denote by $\Sigma^w = \{adm^w, com^w, grd^w, pref^w\}$ the family of weakly admissible semantics.

Turning now to modularization $(\mathbf{M})$, we find that an even stronger result can be proved. As for AFs (Baumann, Brewka, and Ulbricht 2020a), also the other direction holds.

Proposition 6.6. *Let $\sigma \in \Sigma^w$, and $F = (A, R, S)$, then for each $E \subseteq A, E' \subseteq A^E$ we have*

- *$E \in \sigma(F)$, $E' \in \sigma(F^E) \Rightarrow E \cup E' \in \sigma(F)$*
- *$E, E \cup E' \in \sigma(F) \Rightarrow E' \in \sigma(F^E)$*

6.2 Consequences for ABA

We are ready to introduce weak admissibility for ABA. As in the case of strong admissibility, we define the semantics with respect to the instantiated BSAF.

Definition 6.7. *Let $\mathcal{D}$ be an ABAF and $F_{\mathcal{D}}$ be the instantiated BSAF. For any $\sigma^w \in \Sigma^w$ we let $\sigma^w(\mathcal{D}) = \sigma^w(F_{\mathcal{D}})$.*

Note that the semantics faithfully generalize the weakly admissible semantics of the flat ABA fragment (Blümel, König, and Ulbricht 2024).

We examine our novel semantics with respect to the desiderata stated in Section 3. Observe first, that weak admissibility satisfies liberalization $(\mathbf{L})$, i.e. classic admissibility is a special case of weak admissibility. Modularization $(\mathbf{M})$ is also preserved, which indicates our generalization is well-behaved for the most part. The indifference towards the presence of self-attacking assumptions $(\mathbf{PA})$, however, a key feature of weak admissibility in AFs and SETAFs, is no longer given in BSAFs. Due to the additional requirement of closedness and the sensitivity towards indirect conflicts accompanying it, self-attackers influence the set of weakly admissible extensions indirectly via their incoming supports.

Example 6.8. *Consider $F = (\{a\}, \{(\{a\}, a)\}, \{(\emptyset, a)\})$. This BSAF contains a single self-attacking assumption a which is supported by the empty set, rendering the empty set not acceptable under weak admissibility since it is not closed. Now if a is removed, $\emptyset$ immediately becomes a weakly admissible set.*

It turns out that weak admissibility for BSAFs, and consequently for ABA, takes a more differentiated look at self-attackers than for SETAFs/flat ABA. On the one hand, self-attackers compromise the assumptions supporting them, on the other hand, they do not impact the acceptance of assumptions attacked by them under weak admissibility. The later is validated by the fact that weak admissibility satisfies paradoxical attacks/supports $(\mathbf{PRS})$ in BSAFs.

Observe that, as for classic admissible semantics, the fundamental lemma $(\mathbf{F})$ is not satisfied for weakly admissible semantics and the family of weak semantics admits unwanted behavior. We summarize our findings below.

Theorem 6.9. *The weakly admissible semantics for ABA satisfies $(\mathbf{L})$, $(\mathbf{PRS})$, and $(\mathbf{M})$, and does not satisfy $(\mathbf{F})$, $(\mathbf{PA})$, $(\mathbf{NE})$, $(\mathbf{SR})$, $(\mathbf{MC})$.*

7 Fixing Strong Admissibility (and why Fixing Weak Admissibility Fails)

The undesired behavior of ABA semantics has received quite some attention in the literature recently (Heyninck and Arieli 2024; Berthold, Rapberger, and Ulbricht 2024). To overcome some of the issues, Berthold, Rapberger, and Ulbricht (2024) propose alternatives to the classical semantics that reinstate some of the desired properties of the semantics. Inspired by their investigations, we discuss ways how to address the observed shortcomings in our setting. To do so, we focus on the so-called Γ-semantics as they address issues of admissible-based semantics by modifying the closure. In this section, we focus on BSAF semantics; as in the previous sections, the results transfer to ABA as well.

We recall Γ-closure below. Intuitively, an argument a only counts as supported by a set E if E is strong enough to defend a against each attack (Γ is defined in Definition 2.6).

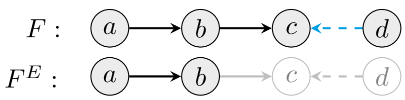
Definition 7.1. *Given a BSAF $F = (A, R, S)$, $E \subseteq A$, and $a \in A$. Then E Γ-supports a iff $a \in cl(E)$ and $a \in \Gamma(E)$; E is Γ-closed iff E contains all arguments it Γ-supports. By $cl_\Gamma(E)$ we denote the Γ-closure of E.*

The Γ-closure induces versions of admissible, preferred, complete, and grounded semantics, which we denote by σ_Γ. We state the definition of Γ-admissible semantics; the remaining semantics are defined analogously.

Definition 7.2. *Let F be a BSAF. $E \in cf(F)$ is* Γ-admissible *($E \in adm_\Gamma(F)$) iff E defends itself and is Γ-closed.*

As discussed in (Berthold, Rapberger, and Ulbricht 2024), Γ-admissible semantics satisfy many of the desired semantics properties, especially for admissible-based semantics. The objective of this section is to introduce Γ-closure as a fix to the strong and weak admissible semantics to address some of the undesired properties for our semantics. As it turns out, however, Γ-semantics lack an important prerequisite, especially for weak semantics: the modularization property is not satisfied. We demonstrate this issue below.

Example 7.3. *Consider the following BSAF $F = (A, R, S)$ and its reduct wrt. $E = \{d\}$ below.*



E is Γ-closed since E does not defend c although it is in the closure of E. Thus, E is Γ-admissible. Let $E' = \{a\}$. Thus, E is weakly Γ-admissible in F, E' is weakly Γ-admissible in F^E. However, $E \cup E'$ is not weakly Γ-admissible in F.

Without modularization the Γ-closure is not well suited for fixing weakly admissible semantics, because the idea of rejecting an attacker that is not acceptable in the reduct has to be justified wrt. the framework as a whole. For strong admissibility, on the other hand, the adaptation of Γ-closedenss offers several desirable results, as we discuss below. First, let us define the Γ-version of strong admissibility, and the semantics based on strong admissibility.

Definition 7.4. *Let F be a BSAF; a set $E \in cf(F)$ is* strongly Γ-admissible *($E \in adm^s_\Gamma(F)$) iff $E \in sd(F)$ and E is Γ-closed. Further, given $E \in adm^s_\Gamma(F)$:*

- *$E \in com^s_\Gamma(F)$ iff E contains every assumption it defends;*
- *$E \in grd^s_\Gamma(F)$ iff E is $\subseteq$-minimal in $com^s_\Gamma(F)$;*
- *$E \in pref^s_\Gamma(F)$ iff E is $\subseteq$-maximal in $adm^s_\Gamma(F)$;*

Note that strong Γ-admissible semantics are guaranteed to return some extension since $cl_\Gamma(\emptyset)$ can be closed.

Proposition 7.5. *$adm^s_\Gamma(F) \neq \emptyset$ for each BSAF F.*

As a result, non-emptiness ($\mathbf{NE}$) is satisfied by strongly Γ-admissible semantics.

Next, we give a constructive definition of strongly Γ-admissible semantics.

Proposition 7.6. *Let $F = (A, R, S)$ be a BSAF. A set of arguments $E \subseteq A$ is strongly Γ-admissible iff it is conflict-free, Γ-closed and there exists a finite sequence of pairwise disjoint sets $E_1, ..., E_n$ such that $E_1 = \emptyset$, $E = \bigcup_{i=1}^{n} E_i$ and for each $i \geq 1$ it holds that E_i is defended by $\bigcup_{j<i} E_j$.*

We also obtain the following weaker version of the fundamental lemma ($\mathbf{F}$) for strongly Γ-admissible semantics.

($\mathbf{WF}$) Weakened Fundamental Lemma: If $S \in adm(F)$ defends a, then there exists $S' \in adm(F)$, s.t. $S \cup \{a\} \subseteq S'$.

Given a strongly Γ-admissible set S and an argument a it defends, we can use the constructive definition of the semantics to compute the Γ-closure of $S \cup \{a\}$ in an iterative way. In each step, we add all arguments that lie in the Γ-closure and are not already contained in S. The resulting sequence satisfies the requirements from Proposition 7.6.

Proposition 7.7. *If $E \in adm^s_\Gamma(F)$ defends a, then there exists $E' \in adm^s_\Gamma(F)$, such that $E \cup \{a\} \subseteq E'$.*

Note that the weakened fundamental lemma ($\mathbf{WF}$) is not satisfied for *adm*^{s}, as Example 5.1 shows: a defends the unattacked argument b but there is no strongly admissible extension that contains both.

As a consequence of Proposition 7.7, we obtain that each BSAF contains a unique maximal Γ-admissible extension. Also, strongly Γ-admissible semantics satisfies strengthening ($\mathbf{S}$) and complete containment ($\mathbf{CC}$). Overall, strongly Γ-admissible semantics satisfy several properties that are not satisfied wrt. strong admissibility, as summarized below.

Theorem 7.8. *The strongly Γ-admissible semantics for BSAF satisfies ($\mathbf{S}$), ($\mathbf{NE}$), ($\mathbf{WF}$), ($\mathbf{URM}$), ($\mathbf{UM}$), ($\mathbf{CC}$), ($\mathbf{SR}$), and ($\mathbf{MC}$), but does not satisfy ($\mathbf{M}$) and ($\mathbf{F}$).*

8 Conclusion

This work introduces strong and weak admissibility for non-flat ABA. We generalize the notion of reduct to BSAF and show that modularity is satisfied by standard, weak, and strong admissibility. We furthermore propose semantics based on strong Γ-admissibility, which satisfy several desirable properties. Overall, our results reveal a fundamental trade-off between general semantics properties and a sufficiently well-behaved notion of closedness, e.g., strongly Γ-admissible semantics are Γ-closed but not modular.

Our findings contribute to ongoing work in ABA and abstract argumentation in a multitude of ways, touching open issues regarding existence and computation of extensions, principle satisfaction and interrepresentability of formalisms wrt. three well-established notions of admissibility. Our investigations pave the way for sequential computation of extensions (Caminada and Harikrishnan 2024b; Bengel and Thimm 2022) for a large variety of semantics in BSAF and general ABA. Moreover, successfully capturing weak admissibility for general ABA provides valuable insights towards a native reduct notion for ABA and an ABA-semantics satisfying long-standing rationality postulates like non-interference (Borg and Straßer 2018) in the future. The novel families of semantics can be beneficial for several ABA applications since they can model real world scenarios where the classical semantics may be too strict or not strict enough,e.g., in the planning approach by Fan (2018) which uses flat ABA or in the causal discovery setting which uses non-flat ABA (Russo, Rapberger, and Toni 2024).

Our results demonstrate that no existing semantics satisfies all weak admissibility desiderata in the general case. As shown in Theorem 6.9, paradoxical assumptions cannot be avoided, and attempts to resolve this using Γ-admissible semantics fall short due to the lack of modularity. Identifying a suitable middle ground that ensures more of the desired properties remains an interesting and challenging direction for future work. One promising candidate is the Δ-semantics proposed in (Berthold, Rapberger, and Ulbricht 2024) to address related issues of complete-based semantics under strong and weak admissibility.

Acknowledgments

The research reported here was partially supported by the Deutsche Forschungsgemeinschaft (grant 550735820). Furthermore, the authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany and by Sächsische Staatsministerium für Wissenschaft, Kultur und Tourismus in the programme Center of Excellence for AI-research "Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig´´, project identification number: ScaDS.AI.

References

Amgoud, L.; Cayrol, C.; Lagasquie, M.-C.; and Livet, P. 2008. On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems* 23:1–32.

Baroni, P., and Giacomin, M. 2007. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence* 171(10-15):675–700.

Baroni, P.; Caminada, M.; and Giacomin, M. 2018. Abstract argumentation frameworks and their semantics. In Baroni, P.; Gabbay, D.; Giacomin, M.; and van der Torre, L., eds., *Handbook of Formal Argumentation*. College Publications. chapter 4, 159–236.

Baumann, R.; Brewka, G.; and Ulbricht, M. 2020a. Comparing weak admissibility semantics to their dung-style counterparts - reduct, modularization, and strong equivalence in abstract argumentation. In Calvanese, D.; Erdem, E.; and Thielscher, M., eds., *Proceedings of (KR-20)*, 79–88.

Baumann, R.; Brewka, G.; and Ulbricht, M. 2020b. Revisiting the foundations of abstract argumentation - semantics based on weak admissibility and weak defense. In *Proceeding of (AAAI-20)*, 2742–2749. AAAI Press.

Baumann, R.; Brewka, G.; and Ulbricht, M. 2022. Shedding new light on the foundations of abstract argumentation: Modularization and weak admissibility. *Artificial Intelligence* 310:103742.

Baumann, R.; Linsbichler, T.; and Woltran, S. 2016. Verifiability of argumentation semantics. In *Computational Models of Argument*. IOS Press. 83–94.

Bengel, L., and Thimm, M. 2022. Serialisable semantics for abstract argumentation. In Toni, F.; Polberg, S.; Booth, R.; Caminada, M.; and Kido, H., eds., *Proceedings of (COMMA-22)*, volume 353 of *Frontiers in Artificial Intelligence and Applications*, 80–91. IOS Press.

Berthold, M.; Rapberger, A.; and Ulbricht, M. 2024. Capturing non-flat assumption-based argumentation with bipolar setafs. In Marquis, P.; Ortiz, M.; and Pagnucco, M., eds., *Proceedings of (KR-24)*.

Besnard, P.; García, A. J.; Hunter, A.; Modgil, S.; Prakken, H.; Simari, G. R.; and Toni, F. 2014. Introduction to structured argumentation. *Argument & Computation* 5(1):1–4.

Bistarelli, S., and Taticchi, C. 2022. A labelling semantics and strong admissibility for weighted argumentation frameworks. *Journal of Logic and Computation* 32(2):281–306.

Blümel, L., and Ulbricht, M. 2022. Defining defense and defeat in abstract argumentation from scratch - A generalizing approach. In *Proceeding of (KR-22)*.

Blümel, L.; König, M.; and Ulbricht, M. 2024. Weak admissibility for ABA via abstract set-attacks. In Marquis, P.; Ortiz, M.; and Pagnucco, M., eds., *Proceedings of (KR-24)*.

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

Borg, A., and Straßer, C. 2018. Relevance in structured argumentation. In Lang, J., ed., *Proceedings of (IJCAI-18)*, 1753–1759. ijcai.org.

Caminada, M., and Dunne, P. E. 2020. Minimal strong admissibility: A complexity analysis. In Prakken, H.; Bistarelli, S.; Santini, F.; and Taticchi, C., eds., *Proceedings of (COMMA-20)*, volume 326 of *Frontiers in Artificial Intelligence and Applications*, 135–146. IOS Press.

Caminada, M., and Harikrishnan, S. 2024a. An evaluation of algorithms for strong admissibility. In Borg, A.; Ellmauthaler, S.; Mailly, J.; and Niskanen, A., eds., *Proceedings of (SAFA-24)*, volume 3757 of *CEUR Workshop Proceedings*, 69–82. CEUR-WS.org.

Caminada, M., and Harikrishnan, S. 2024b. An evaluation of algorithms for strong admissibility. In Borg, A.; Ellmauthaler, S.; Mailly, J.; and Niskanen, A., eds., *Proceedings of (COMMA-24)*, volume 3757 of *CEUR Workshop Proceedings*, 69–82. CEUR-WS.org.

Caminada, M. 2014. Strong admissibility revisited. In *Proceedings of (COMMA-14)*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, 197–208. IOS Press.

Cayrol, C., and Lagasquie-Schiex, M. 2005. On the acceptability of arguments in bipolar argumentation frameworks. In Godo, L., ed., *Proceedings of (ECSQARU-05)*, volume 3571 of *Lecture Notes in Computer Science*, 378–389. Springer.

Cyras, K.; Fan, X.; Schulz, C.; and Toni, F. 2018. Assumption-based argumentation: Disputes, explanations, preferences. In Baroni, P.; Gabbay, D.; Giacomin, M.; and van der Torre, L., eds., *Handbook of Formal Argumentation*. College Publications. chapter 7, 365–408.

Cyras, K.; Oliveira, T.; Karamlou, A.; and Toni, F. 2021. Assumption-based argumentation with preferences and goals for patient-centric reasoning with interacting clinical guidelines. *Argument & Computation* 12(2):149–189.

Dauphin, J.; Rienstra, T.; and van der Torre, L. 2021. New weak admissibility semantics for abstract argumentation. In *Proceeding of (CLAR-21)*, volume 13040 of *Lecture Notes in Computer Science*, 112–126. Springer.

Dondio, P., and Longo, L. 2019. Beyond reasonable doubt: A proposal for undecidedness blocking in abstract argumentation. *Intelligenza Artificiale* 13(2):123–135.

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

Dvořák, W.; König, M.; Ulbricht, M.; and Woltran, S. 2021. A reduct-driven study of argumentation frameworks with collective attacks. In *Proceedings of (NMR-21)*, 285–294.

Dvořák, W.; König, M.; Ulbricht, M.; and Woltran, S. 2024. Principles and their computational consequences for argumentation frameworks with collective attacks. *Journal of Artificial Intelligence Research (JAIR)* 79:69–136.

Fan, X.; Toni, F.; Mocanu, A.; and Williams, M. 2014. Dialogical two-agent decision making with assumption-based argumentation. In Bazzan, A. L. C.; Huhns, M. N.; Lomuscio, A.; and Scerri, P., eds., *Proceedings of (AAMAS-14)*, 533–540. IFAAMAS/ACM.

Fan, X. 2018. A temporal planning example with assumption-based argumentation. In *Proceedings of (PRIMA-18)*, 362–370.

Gabbay, D.; Giacomin, M.; Simari, G. R.; and Thimm, M., eds. 2021. *Handbook of Formal Argumentation*, volume 2. College Publications.

Heyninck, J., and Arieli, O. 2024. Semantics for non-flat assumption-based argumentation, revisited. In *Proceedings of (IJCAI-24)*, 3413–3420. ijcai.org.

Keshavarzi Zafarghandi, A.; Verbrugge, R.; and Verheij, B. 2022. Strong admissibility for abstract dialectical frameworks. *Argument & Computation* 13(3):249–289.

Lehtonen, T.; Rapberger, A.; Ulbricht, M.; and Wallner, J. P. 2023. Argumentation frameworks induced by assumption-based argumentation: Relating size and complexity. In *Proceedings of (KR-23)*, 440–450.

Lehtonen, T.; Wallner, J. P.; and Järvisalo, M. 2021. Harnessing incremental answer set solving for reasoning in assumption-based argumentation. *Theory and Practice of Logic Programming* 21(6):717–734.

Leofante, F.; Ayoobi, H.; Dejl, A.; Freedman, G.; Gorur, D.; Jiang, J.; Paulino-Passos, G.; Rago, A.; Rapberger, A.; Russo, F.; Yin, X.; Zhang, D.; and Toni, F. 2024. Contestable AI needs computational argumentation. In Marquis, P.; Ortiz, M.; and Pagnucco, M., eds., *Proceedings of (KR-24)*.

Nielsen, S. H., and Parsons, S. 2006. A generalization of Dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In *Proceeding of (ArgMAS-06)*, volume 4766 of *Lecture Notes in Computer Science*, 54–73. Springer.

Rapberger, A., and Ulbricht, M. 2024. Repairing assumption-based argumentation frameworks. In Marquis, P.; Ortiz, M.; and Pagnucco, M., eds., *Proceedings of (KR-24)*.

Russo, F.; Rapberger, A.; and Toni, F. 2024. Argumentative causal discovery. In Marquis, P.; Ortiz, M.; and Pagnucco, M., eds., *Proceedings of (KR-24)*.

Ulbricht, M.; Potyka, N.; Rapberger, A.; and Toni, F. 2024. Non-flat ABA is an instance of bipolar argumentation. In Wooldridge, M. J.; Dy, J. G.; and Natarajan, S., eds., *Proceedings of (AAAI-24)*, 10723–10731. AAAI Press.