From Acceptance to Rejection in Abstract Argumentation

Anne-Marie Heine[1,2] , **Markus Ulbricht**[1,2]
[1]Universität Leipzig
[2]ScaDS.AI Leipzig/Dresden
{aheine, mulbricht}@informatik.uni-leipzig.de

Abstract

Dynamic reasoning environments are among the key aspects in formal argumentation research. Presumably the best understood problem is the so-called *enforcement* problem which asks, generally speaking, whether a given argumentation framework can be modified in a way that a certain desired outcome is ensured. However, enforcement research primarily focuses on the acceptance of arguments or sets thereof. This paper aims to explore the dual problem and investigates means to reject certain unreasonable viewpoints. To achieve this, we use labelling semantics on abstract argumentation frameworks (AFs), since they provide a clearly defined notion of rejection. We consider different kinds of updates for our given AF and provide results on existence as well as minimality of syntactic and semantic changes. For the latter, we define the new concept of *consensus preservation*, formalizing the intuition that formerly acceptable opinions should remain acceptable in the adapted framework. Lastly we discuss how these two notions of minimizing change interact.

1 Introduction

Computational models of argumentation in Artificial Intelligence (AI) (Baroni et al. 2018; Bench-Capon and Dunne 2007) provide formal approaches to reason argumentatively, with a wide variety of application avenues, such as legal reasoning, medical sciences, and e-governmental issues (Atkinson et al. 2017). Reasoning in this way is carried out by instantiation of argument structures from a knowledge base (Bondarenko et al. 1997; Modgil and Prakken 2013; García and Simari 2004; Besnard and Hunter 2008), which represent all that can be argued for, from the standpoint of the knowledge base. Inconsistencies within knowledge bases are then represented by conflicts among arguments, which are modeled via (directed) attacks between arguments, reflecting a counter argument relation. In many such argumentative workflows, the underlying abstract formalism are Dung's abstract argumentation frameworks (AFs) (Dung 1995). By viewing arguments as atomic entities and attacks among them as directed edges, Dung obtains a representation of the given debate as a directed graph $F = (A, R)$. Since then, Dung's AFs have been studied extensively (Baroni et al. 2018).

A highly relevant research direction in KR is concerned with the investigation of dynamical environments, i.e. situations where a given knowledge base changes over time (Gabbay et al. 2021). Since argumentation constitutes inherently dynamic procedure, it is not surprising that researchers investigated dynamic argumentation scenarios extensively. In the context of AFs and extensions thereof, various problems have been investigated like equivalence (Oikarinen and Woltran 2011; Baumann, Rapberger, and Ulbricht 2023; Dvořák, Fandinno, and Woltran 2018), forgetting (Baumann and Berthold 2022; Berthold, Rapberger, and Ulbricht 2023), or repairing a semantical collapse (Baumann and Ulbricht 2019). Dynamics have also been studied in the context of multi-agent systems (Dupuis de Tarlé, Bonzon, and Maudet 2022) and argumentative explanations (Rago, Li, and Toni 2023). Perhaps the most classical and best understood problem is, however, *enforcement* (Baumann and Brewka 2010). Enforcement has been studied extensively, in terms of theoretical issues (Baumann 2012; Dauphin and Satoh 2018), computational aspects (Wallner, Niskanen, and Järvisalo 2017), and the connection to structured argumentation formalisms (Rapberger and Ulbricht 2023; Prakken 2023).

Most research on enforcement in AFs is concerned with the question as to how a given AF F can be modified s.t. a target set of arguments can be rendered acceptable. However, especially in the presence of paradoxical arguments and untrustworthy information –an issue that becomes more and more apparent nowadays– it is arguably an equally important aspect to ensure the *rejection* of certain information.

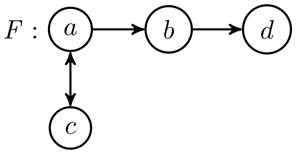
There are different strategies aiming to deal with this kind of information already. A simple approach is simply rendering these arguments as self-attackers. Another method in the realm of preference-based argumentation (see e.g. (Alfano et al. 2023)) assigns a lower value to unwanted arguments thus reducing their influence. While these can be seen as "meta-views" on the discussion, our rejection enforcement aims for a multi-agent view: we want to study how agents involved in the discussion can overcome paradoxical situations instead of adjusting how the debate is modeled. In this case unwanted arguments can simply be skeptical towards the opinion held by the agent instead of downright nonsensical. The goal is to enable an agent to identify weaknesses in their argumentation strategy with regard to their pursued outcome during the progress of the discussion. The so-called updates and enforcements show where to best apply revisions in their argumentation to cover these weak points by adding arguments or attacks and thus facilitate the agent to react efficiently to the changing environment of a discussion. Whether or not this is attainable however depends on the agent's ability to provide logical foundations for the suggested changes.

Let us illustrate the importance of this by means of the following debate about climate change.

Example 1.1. *Suppose our four agents, Anna (a), Bob (b), Carol (c), and Dagobert (d). Suppose our protagonists bring forward the following arguments.*

- *Dagobert: The climate change caused by human beings needs to be stopped as soon as possible.*
- *Bob: No, climate change is something totally natural. There were always hotter and colder periods.*
- *Anna: Sure we are changing the climate, but instead of changing our way of life, it would be much easier to just prepare for the changes.*
- *Carol: I am certain that preventing climate change is much better than handling its dramatic consequences.*

We can model this debate as the following AF:

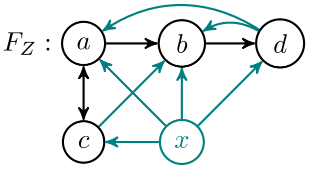


Let us assume agent d judges the opinions of a and b as undesirable and thus wants to reject them. However, under almost all commonly agreed argumentation semantics, this AF F possesses two non-empty sets of accepted arguments, namely $E_1 = \{a, d\}$ and $E_2 = \{c, b\}$; both of them contain one of the undesired arguments a or b.

The goal of this paper is to study how to ensure the rejection of arguments like a and b in the previous example. A common technique to evaluate AFs are extension-based semantics which specify sets E of jointly acceptable arguments. However, they only distinguish between "in" ($a \in E$) or "not in" ($a \notin E$). For our purpose, however, we require a more fine-graded distinction which is why we make use of labellings in order to evaluate AFs which will enable us to label arguments as "in", "undecided", or "out" (see e.g. (Baroni, Caminada, and Giacomin 2011) for an overview).

In a simple AF like the aforementioned one, it is clear that a and b can be rendered rejected. For instance consider the following update:

Example 1.2. *Let F_Z be an expansion of F where the following arguments and attacks have been added:*



While in F_Z, arguments a and b are certainly rejected, this has been achieved by adding numerous attacks and rendering all originally present arguments unacceptable (under almost all argumentation semantics, x is the only acceptable argument here).

This example illustrates various aspects we need to consider to develop reasonable rejection enforcement notions:

- Supposing not every conceivable update of F is attainable in our given argumentation scenario, what can be enforced under different types of expansions?
- How can we reject a given target set of arguments, while preserving the accepted argument sets to the best extent possible?
- What are minimal modifications to the considered AF F in order to obtain the desired changes in the debate?

In this paper, we tackle this issues and thereby lay a thorough theoretical foundation for rejection enforcement in formal argumentation. More specifically, our main contributions can be summarized as follows.

- We stipulate a natural notion of *rejection* enforcement under the most common types of updates, i.e. *normal*, *strong*, and *local* expansions. Section 3
- We show that in almost all cases, under mild conditions some rejection exists. Section 4
- We introduce so-called *consensus preserving* rejection enforcement in order to capture the intuition that as many of the acceptable viewpoints as possible should persist. Again, we show under which conditions rejection can still be guaranteed Section 5
- We discuss how to minimize the syntactic changes to the given AF, both for usual as well as consensus preserving rejection. Sections 6 and 7

2 Preliminaries

A *labelling-based semantics* $\mathcal{L}_\sigma : \mathcal{F} \to 2^{\left(2^{\mathcal{U}}\right)^3}$ is a function which assigns to any AF $F = (A, R)$ a set of triples of sets of arguments denoted by $\mathcal{L}_\sigma(F) \subseteq \left(2^A\right)^3$, where A, the set of arguments, is a finite subset of a fixed infinite background set $\mathcal{U}$, and $R \subseteq A \times A$. Each one of them, a so-called *σ-labelling* of F, is a triple $L = (I, O, U)$ indicating that arguments in I, O or U are considered to be *accepted (in)*, *rejected (out)* or *undecided* with respect to F. We assume I, O and U to be disjoint and covering A. We use L^{I} (or $L^{\mathrm{I}}(a)$) to refer to (a is an element of) the first component of the labelling L. Analogously for L^{O} and L^{U}. Additionally we use $\mathcal{L}_\sigma(F)^I = \{E \subseteq A \mid \exists L \in \mathcal{L}_\sigma(F) : E = L^I\}$ for the set of all in-sets.

Definition 2.1. *A labelling L of $F = (A, R)$ is called conflict-free if we have:*

1. *If $a, b \in L^{\mathrm{I}}$, then $(a, b) \notin R$, and*
2. *If $a \in L^{O}$, then there is an $b \in L^{\mathrm{I}}$ with $(b, a) \in R$.*

Definition 2.2. *A labelling L of $F = (A, R)$ is called admissible if we have:*

1. *If $a \in L^{\mathrm{I}}$, then $(b, a) \in R$ implies $b \in L^{O}$.*
2. *$a \in L^{O}$ iff there is some $b \in L^{\mathrm{I}}$ with $(b, a) \in R$.*

Such L is a complete labelling if for each $a \in A$ it holds that

3. *If $b \in L^{O}$ for each $(b, a) \in R$, then $a \in L^{\mathrm{I}}$.*

Note that in the literature, admissibility is usually defined in a way that in the second item, only the left to right direction is required. However since we are primarily interested in the out-sets, adapting the definition this way greatly improves clarity of the notation and results.

Definition 2.3. *Let $F = (A, R)$ be an AF and $L \in \left(2^A\right)^3$ a labelling of F.*

1. *$L \in \mathcal{L}_{cf}(F)$ iff L is a conflict-free labelling of F,*
2. *$L \in \mathcal{L}_{ad}(F)$ iff L is an admissible labelling of F,*
3. *$L \in \mathcal{L}_{co}(F)$ iff L is an complete labelling of F,*
4. *$L \in \mathcal{L}_{pr}(F)$ is a preferred labelling of F iff $L \in \mathcal{L}_{co}(F)$ and there is no $M \in \mathcal{L}_{co}(F)$ s.t. $L^I \subsetneq M^I$,*
5. *$L \in \mathcal{L}_{gr}(F)$ is a grounded labelling of F iff $L \in \mathcal{L}_{co}(F)$ and there is no $M \in \mathcal{L}_{co}(F)$ s.t. $M^I \subsetneq L^I$,*
6. *$L \in \mathcal{L}_{stb}(F)$ s a stable labelling iff $L \in \mathcal{L}_{cf}(F)$ and $L^U = \emptyset$, and*

We sometimes also make use of notations that are typical for *extension-based* semantics: For $\sigma \in \{cf, ad, co, gr, pr, stb\}$ we write $E \in \sigma(F)$ iff $E = L^I$ for some $L \in \mathcal{L}_{cf}(F)$ and call E a σ-extension. For a set E of arguments we define $E^+ = \{a \in A \mid \exists e \in E : (e, a) \in R\}$ and similarly, $E^- = \{a \in A \mid \exists e \in E : (a, e) \in R\}$. We often write e^+ and e^- instead of $\{e\}^+$ and $\{e\}^-$ for singletons. The mapping $\Gamma_F(E) = \{a \in A \mid a^- \subseteq E^+\}$ is the *characteristic function*. If $a \in \Gamma_F(E)$, we say E *defends* a.

Example 2.4. *In our running Example 1.1 we have*

- *$L_1 = (\{a, d\}, \{b, c\}, \{\emptyset\})$, $L_2 = (\{b, c\}, \{a, d\}, \emptyset)$, and $L_3 = (\emptyset, \emptyset, \emptyset)$ are the complete labellings;*
- *L_3 is the unique grounded labelling;*
- *L_1 and L_2 are the preferred labellings;*
- *L_1 and L_2 are also the stable labellings.*

Moreover, throughout this work we assume the reader to be familiar with the complexity classes P, NP, and coNP.

3 Rejection-Enforcement

In this section we introduce the basic rejection enforcement notions we will investigate throughout our study. First of all, let us define what it means for a set Z to be rejected. There are multiple conceivable variations, i.e. Z is rejected in at least one labelling or in each one; moreover, we could ask exactly for Z to be rejected or for some superset. Let us start with the most basic case here, that is, we want Z to correspond to the out labelled arguments of some labelling in F. A "skeptical" counterpart to this notion is discussed in Section 8.

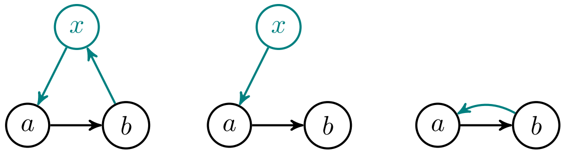
Definition 3.1. *Given a labelling-semantic σ, an AF $F = (A, F)$ and a set $Z \subseteq A$. Then Z is rejected w.r.t. σ iff there exists an $L \in \mathcal{L}_\sigma(F)$ s.t. $Z = L^O$.*

In order to formalize an enforcement notion, we also need to consider which modifications to the given AF F are allowed. Here we utilize the usual expansion notions.

Definition 3.2. *An AF G is an expansion of AF $F = (A, R)$ (for short, $F \preceq_E G$) iff $G = (A \cup B, R \cup S)$ for some sets B and S, s.t. $A \cap B = R \cap S = \emptyset$. An expansion is called*

1. *normal ($F \preceq_N G$) iff $(a, b) \in S$ implies $a \in B$ or $b \in B$;*
   no novel attacks among existing arguments
2. *strong ($F \preceq_S G$) iff $F \preceq_N G$ and $(a, b) \in S$ implies $a \notin A$ or $b \notin B$;*
   no novel attacks from existing to novel arguments
3. *local ($F \preceq_L G$) iff $B = \emptyset$.*
   no novel arguments

Example 3.3. *Consider an AF F consisting of just an attack from some argument a to another one b. Examples for the different types of expansions are depicted below.*



(a) normal (b) strong (c) local

Having settled the notion of rejection and the expansion types, let us formalize expansions achieving our goal.

Definition 3.4. *Given an AF F, a semantics σ, and expansion type T, and a set of argument Z. An AF F_Z is called a σ* rejection *for Z under T expansions iff*

- *$F \preceq_T F_Z$ and*
- *Z is σ-rejected in F_Z.*

We call F_Z the σ-rejecting AF. *Moreover, we call Z σ-*rejectable *under T expansions if such F_Z exists.*

Example 3.5. *Recall F as in Example 1.1 with expansion F_Z given in Example 1.2. The AF F_Z is a strong expansion because x does not receive incoming attacks. Moreover, F_Z is a stb-rejection for $Z = \{a, b\}$. Thus, Z is rejectable under strong expansions in F.*

Before heading to our enforcement results, let us make a computational remark. The foundation for our study is the question "Under which circumstances can Z be rejected?" Leaving expansions out of the equation for a moment, let us answer this question within a given AF F. This problem is somewhat dual to the *verification problem*, i.e. given a set E, "Is $E \in \sigma(F)$ some extension?". This problem is well-studied and known to be tractable for all semantics considered in this paper except pr. For rejection, we need to take into consideration that several sets E might reject the same $Z \subseteq A$. This induces some search space for rejecting Z.

So suppose we are given $F = (A, R)$ with Z to be rejected. An important observation is that in case of stable semantics, if Z is supposed to be rejected, the only candidate labelling is $L = (A \setminus Z, Z, \{\})$ since stb is two-valued. This leads to the following simple observation.

Fact 3.6. *Let an AF $F = (A, R)$ and a set $Z \subseteq A$ be given. Then Z is stable rejected iff $(A \setminus Z)^+ = Z$.*

In case of admissible semantics, we note that if E rejects Z, then it must hold that $E \subseteq \{a \in A \setminus Z \mid a^+ \cup a^- \subseteq Z\}$. The reason is that i) E cannot contain arguments in Z, ii) E cannot attack arguments outside Z, and iii) E cannot be attacked by arguments outside Z because then it would have to counter this attack in order to be admissible. Due to the meaning of this set for the admissible case, let us set $E_{ad}(Z) = \{a \in A \setminus Z \mid a^+ \cup a^- \subseteq Z\}$.

Consequently we could reduce the candidate sets to $A \setminus Z$ and $E_{ad}(Z)$. For co, we argue analogously and gr can be computed easily. Thus we note:

Proposition 3.7. *On input F and Z, deciding whether Z is rejected in F w.r.t. σ can be done in polynomial time for $\sigma \in \{ad, gr, co, stb\}$.*

As can be shown by a simple modification to the *standard construction* (Dvorák and Dunne 2018), pr is hard.

Proposition 3.8. *On input F and Z, deciding whether Z is rejected in F w.r.t. pr is* coNP*-complete.*

For the reasons we explained, the sets $A \setminus Z$ and $E_{ad}(Z)$ will be the main protagonists for many results in our study.

4 Existence of an Enforcement

Let us start with the most basic question, namely whether or not some rejection of a given set Z exists. We proceed by the various types of expansions.

4.1 Strong and Normal Expansions

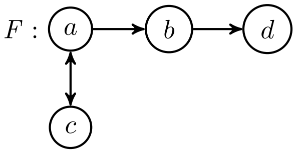
Strong expansions are the most powerful types of expansions since we are allowed to introduce novel arguments attacking F as well as attacks between the novel arguments. However the existing ones are not allowed to defend themselves.

As expected, we obtain the strongest results within this context. Interestingly, however, for stb and ad semantics it does not make any difference whether we consider strong or normal expansions. Consequently, let us discuss these cases simultaneously.

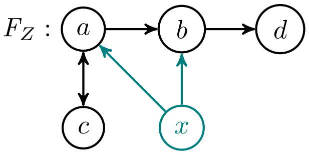
We start with stable semantics. Speaking in terms of labellings, stable semantics are simple in the sense that no undecided arguments exist, i.e. $L = (L^I, L^O, \emptyset)$ for each labelling. Consequently, if we strive for $Z = L^O$ in a fixed AF F, the set of in-labelled arguments is already determined (as already noted in Fact 3.6) This observation yields the foundation for rejection enforcement: We require $A \setminus Z$ to be conflict-free; the potentially missing attacks can be added in our expansion.

Proposition 4.1. *Let an AF $F = (A, R)$ and a set $Z \subseteq A$ be given. Then Z can be stable rejected under both strong and normal expansion iff $A \setminus Z \in cf(F)$.*

Example 4.2. *We consider the AF F:*



with the stable labellings $L_1 = (\{a, d\}, \{b, c\}, \{\emptyset\})$, and $L_2 = (\{b, c\}, \{a, d\}, \emptyset)$ as discussed in Example 2.4. Suppose our goal is to reject $Z = \{a, b\}$ under strong expansion. As we can see in the table above, Z is not already rejected in F. Thus we have to adjust F. We use a strong expansion and construct a rejection F_Z as follows:



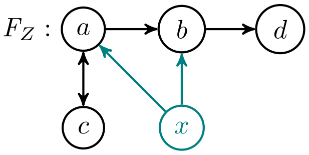
Looking at the rejection F_Z, we see that Z is indeed the out-set of the stable labelling $(\{c, d, x\}, \{a, b\}, \emptyset)$.

Let us now head to admissible-based semantics. Conceptually, admissible semantics differ from stable since they are 3-valued, i.e. undec-labelled arguments might exist in some labellings. Due to this additional flexibility, we get an even stronger result w.r.t. our rejection notion: any set Z can be σ-rejected for $\sigma \in \{ad, co, gr, pr\}$.

Proposition 4.3. *Let an AF $F = (A, R)$ and a set $Z \subseteq A$ be given. Let $\sigma \in \{ad, co, pr, gr\}$. Then Z can always be σ-rejected under both strong and normal expansion.*

Heading back to our running example, we see that the same construction does the carries out the function for admissible semantics.

Example 4.4. *Recall our running example AF $F = (A, R)$. Suppose we let $Z = \{a, b\}$ again. Then our previous rejection F_Z also works for admissible semantics:*



Indeed, we have $L = (\{x\}, \{a, b\}, \{c, d\})$ as an admissible labelling. For the other semantics, we would additionally have to ensure that c and d do not occur in the in-labelled arguments. This can be done by e.g. introducing a self-attacking y attacking these arguments.

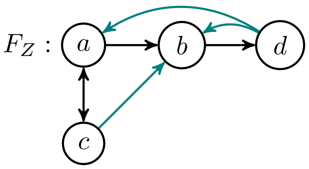
4.2 Local Expansion

Let us now head to local expansions, that is, we do not allow the introduction of novel arguments. Interestingly, for stable semantics the situation does not change: if $A \setminus Z$ is conflict-free, then we can add out-going attacks ensuring that $A \setminus Z$ is stable.

Proposition 4.5. *Let an AF $F = (A, R)$ and a set $Z \subseteq A$ be given. Then Z can be stable rejected under local expansions iff $A \setminus Z \in cf(F)$.*

We illustrate the underlying idea in our running example: We can simply add the necessary out-going attacks to $A \setminus Z$.

Example 4.6. *Again set $Z = \{a, b\}$ in our running example. We construct the following rejection-AF $F_Z = (A, R_Z)$.*



Indeed, we see that $L = (\{d, c\}, \{a, b, \}, \emptyset) \in \mathcal{L}_{stb}(F_Z)$.

Now let us head to the 3-valued semantics. This time, it becomes slightly more involved to check for the existence of some rejection F_Z. The intuitive reason is as follows: suppose we want to reject Z (and only Z), but without introducing novel arguments. Then, we need to find some set E of arguments that does not attack any argument outside Z, because then, acceptance of E would reject too many arguments. In the same vein, E cannot be attacked by arguments outside Z, because then defense would again require E to counter-attack arguments outside Z (and consequently, reject too much). We end up with the following characterization.

Proposition 4.7. *Let an AF $F = (A, R)$ and a set $Z \subsetneq A$ be given. Let $\sigma \in \{ad, co, pr\}$. Then Z can be σ-rejected under local expansion iff there is an $E \subseteq A \setminus Z$ s.t $E^+ \cup E^- \subseteq Z$.*

Example 4.8. *Recall our running example AF $F = (A, R)$. Suppose this time we choose $Z = \{a, c\}$. Then there are 3 possibilities for the set $E \subseteq A \setminus Z$:*

- *$E = \{b, d\}$. Then $E^+ = \{d\} \nsubseteq Z$ i.e. has an inner conflict.*
- *$E = \{b\}$. Then $E^+ = \{d\} \nsubseteq Z$ i.e. the out-set would include $d \notin Z$.*
- *$E = \{d\}$. Then $E^- = \{b\} \nsubseteq Z$ i.e. E would have to defend against the attack from b including $b \notin Z$ in the out-set.*

Hence none of the possible choices of E satisfy our demanded property and Z is indeed not admissible rejectable in F under local expansion.

For gr semantics we get an analogous result, but one of the elements in $E_{ad}(Z)$ needs to be unattacked, because otherwise, the grounded extension will be empty.

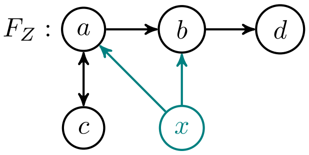
Proposition 4.9. *Let an AF $F = (A, R)$ and a set $Z \subseteq A$ be given. Then Z can be grounded rejected under local expansion iff*

- *there is an $E \subseteq A \setminus Z$ s.t $E^+ \cup E^- \subseteq Z$ and*
- *there exists $e \in E_{ad}(F)$ s.t. $e^- = \emptyset$.*

5 Consensus Preservation

So far, we have settled basic existence results for our semantics under the different expansion types. However, our constructions did not follow any specific additional goal like minimizing the considered modifications in F_Z or preserving the existing labellings in F. Indeed, already in our small running example, we did not preserve existing labellings at all. The underlying intuition behind this notion is interesting for agents who want to preserve certain, desirable viewpoints (e.g. their own ones) and seek for ways to justify them, while attacking arguments they want to ensure are rejected. Let us recall the following scenario.

Example 5.1. *Consider again our running example F with $Z = \{a, b\}$ and rejection F_Z:*



We observe that in the original AF F, we have two stable labellings $L_1 = (\{a, d\}, \{b, c\}, \emptyset)$, $L_2 = (\{b, c\}, \{a, d\}, \emptyset)$, while in F_Z the unique stable labelling is $L = (\{c, d, x\}, \{a, b\}, \emptyset)$. Neither of the in-labelled sets L_1^I or L_2^I is preserved.

The next natural step is thus to incorporate such aspects in our enforcement notions. Striving to formalize this, let us develop a suitable notion of *consensus preserving* enforcement. The intuitive idea is that all accepted sets in F are also accepted in F_Z, i.e. we require $\mathcal{L}_\sigma^I(F) \subseteq \mathcal{L}_\sigma^I(F_Z)$.

Definition 5.2. *Let $F = (A, R)$ be an AF and $F_Z = (A_Z, R_Z)$ a rejection of $Z \subseteq A$. We call F_Z consensus preserving iff $\mathcal{L}_\sigma^I(F) \subseteq \mathcal{L}_\sigma^I(F_Z)$.*

That is, F_Z is a rejection where all σ-accepted sets $\mathcal{L}_\sigma^I(F)$ remain σ-accepted in the rejection AF.

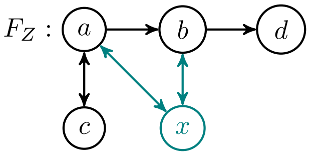
Example 5.3. *Heading back to our previous example, F_Z is not consensus preserving since $L^I = \{c, d, x\}$ coincides neither with L_1^I nor L_2^I.*

Now we again proceed by the underlying expansion notions and explore what is possible and what is not.

5.1 Normal Expansions

As before, we will start with normal expansions. Since we obtain different results compared to strong expansions, we cannot discuss both cases simultaneously this time.

Example 5.4. *We recall the rejection F_Z used in Example 5.1. We discuss stable semantics here, but the same reasoning applies to admissible as well. As we already argued, quite a few of the former labellings are lost. This can be traced back to the fact that a set E attacking each non-included argument in F loses this characteristic due to the newly added argument x. This problem can be fixed by allowing x to be attacked back i.e. using a normal instead of a strong expansion. We obtain the following rejection F_Z:*



Indeed, the previous labellings $L_1 = (\{a, d\}, \{b, c\}, \emptyset)$ and $L_2 = (\{b, c\}, \{a, d\}, \emptyset)$, are preserved, while at the same time, we still have the novel $L = (\{c, d, x\}, \{a, b\}, \emptyset)$ ensuring rejection of a and b in at least one labelling.

Indeed, as the following proposition shows this idea generalizes to any given AF F. Beforehand we state a short but interesting auxiliary result concerning the stable case.

Lemma 5.5. *Let an AF $F = (A, R)$ and a set $Z \subseteq A$ be given. If F itself does not stable reject Z, then each $E \in \mathcal{L}_{stb}(F)^I$ contains at least one $z \in Z$.*

Now we are ready to state our enforcement result, for any considered semantics.

Proposition 5.6. *Let an AF $F = (A, R)$ and a set $Z \subseteq A$ be given. Let $\sigma \in \{ad, stb, co, pr\}$. If Z is σ-rejectable under normal expansion, then there exists a consensus preserving rejection.*

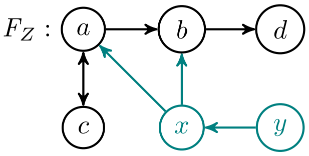
For gr semantics, note that $L \in \mathcal{L}_{gr}$ is uniquely determined. Thus a consensus preserving rejection must work in a way that the exact same unique grounded labelling persists. This is of course only possible if F already accomplishes the objective.

Proposition 5.7. *Let an AF $F = (A, R)$ and a set $Z \subseteq A$ be given. There is a consensus preserving gr-rejection under normal expansions for Z iff F itself gr-rejects Z.*

5.2 Strong Expansions

When trying to extend the previous Proposition 5.6 to strong expansions, we face our first negative results. To this end we note that in a strong expansion, the novel argument x can never be attacked by the original arguments in F. However, when attacking x with a second novel argument y, we still add y to the in-sets, thus again not achieving consensus preservation.

Example 5.8. *Let us extend our running example as described with two additional arguments x and y:*



Then each labelling L must contain either x or y in L^I, making the requirement from consensus preserving rejection impossible.

Indeed, the following result states that under stable semantics, we can only reject Z under trivial conditions.

Proposition 5.9. *Let an AF $F = (A, R)$ and a set $Z \subseteq A$ be given. There is a consensus preserving stb-rejection under strong expansions for Z iff F itself stb-rejects Z.*

We note that this is the case iff $A \setminus Z$ is stable in F.

Corollary 5.10. *Let an AF $F = (A, R)$ and a set $Z \subseteq A$ be given. There is a consensus preserving stb-rejection under strong expansions for Z iff $(A \setminus Z, Z, \emptyset) \in \mathcal{L}_{stb}(F)$.*

For admissible semantics, we have slightly more freedom as we do not necessarily need to attack the novel arguments X. However, for ensuring consensus preservation, acceptable arguments would need to defend themselves against X. Since, as we argued, this is impossible in the context of strong expansions, this implies that X is not allowed to attack any credulously accepted argument.

Proposition 5.11. *Let an AF $F = (A, R)$ and a set $Z \subseteq A$ be given. Then Z is consensus preserving ad-rejectable under strong expansions iff there is some $E \subseteq A \setminus Z$ s.t.*

- *$E^+ \cup E^- \subseteq Z$,*
- *$Z \setminus E^+$ does not contain any credulously accepted argument.*

For the other semantics co and pr we conjecture that no such simple characterization exists. The reason is that under the above conditions, we can add some x attacking each $z \in Z \setminus E^*$. However, for complete and preferred extensions, it might be the case that we then need to include further arguments into our extension; how to handle this in a consensus preserving way is a challenging question for future work. However, also for ad, this characterization lays the foundation for our first intractability result. The underlying issue is that we need to check for acceptance of certain arguments. Thus checking whether such a rejection F_Z exists is a computationally hard problem.

Proposition 5.12. *On input an AF F and a set Z of arguments, deciding whether Z is consensus preserving σ-rejectable under strong expansions is intractable for $\sigma \in \{ad, co, pr\}$.*

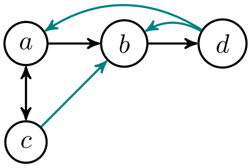
5.3 Local Expansion

Deciding whether a consensus preserving rejection exists under local expansions, is even harder: this time, for any semantics (except gr) we find intractability of the underlying decision problem.

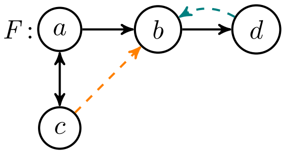
Proposition 5.13. *On input of an AF F and a set Z of arguments, deciding whether Z is consensus preserving σ-rejectable under local expansions is intractable for $\sigma \in \{ad, co, pr, stb\}$.*

However, for many cases we can at least give an easily checked sufficient property for stb semantics. To see this let us recall Example 4.2.

Example 5.14. *In the expansion from Example 4.6 we lost labellings since now, only $(\{c, d\}, \{a, b\}, \emptyset)$ is left.*



One can observe that we lose the accepted $\{a, d\}$ due to the added attack (d, a). Now the question is: do we actually need this attack? The goal of the construction in this example is that the set $\{a, b\}$ is attacked by the set $\{c, d\}$; a is already attacked by c, so no further attacks on a are necessary. Thus (d, a) is superfluous. However b still needs to be attacked. To achieve this, there are two possibilities, $c \rightarrow b$ or $d \rightarrow b$. (marked orange and blue resp. in the following AF).



In the case of adding $d \rightarrow b$ (blue), all former labellings are preserved and only one labelling, realizing the rejection we desire, is added. The result is a consensus preserving rejection AF.

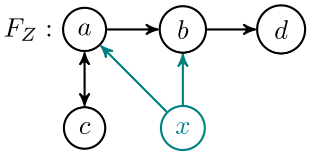
The observations we made in the previous example generalize to arbitrary AFs as follows.

Proposition 5.15. *Let an AF $F = (A, R)$ and a set $Z \subseteq A$ be given and Z is stable rejectable under local expansion. Then there exists a consensus preserving rejection under local expansion, if for each $z \in Z$ exists an $a \in A \setminus Z$ s.t $(z, a) \in R$.*

6 Minimizing Syntactic Change

So far, our intuition of preserving the original AF was based on a semantical point of view: we tried to preserve existing extensions as good as possible. Let us now have a look at the necessary modifications instead. That is, in this section, we discuss how to preserve the structure of the argumentation graph to the best extent possible. To illustrate this, let us head back to our running example with the usual expansion.

Example 6.1. *Consider again our running example F with rejection F_Z:*



While $\{a, b\}$ is successfully rejected, we observe that the attack towards a was actually not necessary; c would already have done the job for us. In this sense, the modification imposed by F_Z is not minimal.

We base our definition of syntactic minimal change on (Baumann 2012). While this work only deals with changing attacks, we add an additional term to also consider arguments. Since in our scenarios attacks as well as arguments are only ever added and never deleted the definition can be simplified. We end up with the following notion.

Definition 6.2 (syntactic minimal change)**.** *The syntactic distance between an AFs F and associated expansion F_Z is a natural number defined via*

$$d_{st}(F, F_Z) = |R(F_Z) \setminus R(F)| + |A(F_Z) \setminus A(F)|.$$

Let $F_Z = (A_Z, R_Z)$ be a rejection of $Z \subseteq A$ over F. Then F_Z is a rejection with minimal syntactic change *iff*

$$d_{st}(F, F_Z) = \min\{d_{st}(F, H) \mid H \textit{ enforces } Z \textit{ over } F\}.$$

Thus a rejection with syntactic minimal change is a rejection which is among the "optimal" ones judged by the number of added arguments and attacks. Within the context of minimizing change, we will focus on stb and ad semantics since under complete-based semantics, we often need to introduce self-attackers to block the iteration of Γ_F; this is contrary to the intuition of minimizing changes to F.

6.1 Strong and Normal Expansion

Following our usual structure, let us start with strong expansions. Let us familiarize with the setting by means of the following example.

Example 6.3. *Take Example 6.1. Minimizing the number of attacks is quite straightforward in this example: We need the set $A_Z \setminus Z$ to reject a as well as b. We achieved this by letting the new argument x attack both. However the attack to a is redundant, since $A \setminus Z$ already attacks a. Thus adding just the attack (x, b) is sufficient.*

Indeed, under stable semantics, this idea generalizes to arbitrary AFs as follows.

Definition 6.4. *If $F = (A, R)$ and a set $Z \subseteq A$ are given, the AF $F_Z^{sm} = (A_Z^{sm}, R_Z^{sm})$ is contructed as follows:*

- *$A_Z^{sm} = A \cup \{x\}$ where $x \notin A$*
- *$R_Z^{sm} = R \cup \{(x, b) \mid b \in Z, \nexists a \in (A \setminus Z) : (a, b) \in R\}$*

Proposition 6.5. *Let $F = (A, R)$ and a set $Z \subseteq A$ be given. If Z is stable rejectable under normal resp. strong expansion, then the minimal syntactic distance of a stable rejection for Z under normal resp. strong is*

$$d(F, F_Z) = |Z| - |(A \setminus Z)^+| + 1$$

which is attained by the expansion F_Z^{sm}.

Let us now proceed to admissible semantics. Again we have a bit more freedom here since not all arguments in $A \setminus Z$ need to be accepted in order to reject Z; we can consider a suitable subset of $A \setminus Z$ as well. Hence, we want to choose a set $E \subseteq A \setminus Z$, that is optimal in a way that it already attacks as many arguments of Z as possible. As was already common in previous observations regarding ad, we see that $E^+ \cup E^- \subseteq Z$ must hold. If we use a set E satisfying these two requirements, then we only need to add attacks from a newly added argument to each $z \in Z$ that is not attacked by E. The maximal number of arguments we can already attack is attained by $E_{ad}(Z) = \{a \in A \setminus Z \mid a^+ \cup a^- \subseteq Z\}$.

Example 6.6. *We use the example AF $F = (A, F)$ and want to admissible reject $Z = \{a, b\}$ in the AF F (under both normal and strong expansion). First we find our set a as described above: $A \setminus Z = \{c, d\}$, so we only have to check the remaining arguments c and d.*

- *$\{c\}^+ = \{a\} \subseteq Z$ and $\{c\}^- = \{a\} \subseteq Z$*
- *$\{d\}^+ = \{\} \subseteq Z$ and $\{d\}^- = \{b\} \subseteq Z$*

Thus $E_{ad}(Z) = \{c, d\}$. Our set already rejects a, meaning the newly added argument only needs to attack b. The syntactic distance thus was reduced by one attack.

Definition 6.7. *If $F = (A, R)$ and a set $Z \subseteq A$ are given, the AF $F_Z^{sm} = (A_Z^{sm}, R_Z^{sm})$ is contructed as follows:*

- *$A_Z^{sm} = A \cup \{x\}$ where $x \notin A$*
- *$R_Z^{sm} = R \cup \{(x, z) \mid z \in Z, z \notin (E_{ad}(Z))^+\}$*

The following proposition establishes that this AF F_Z minimizes the syntactic change.

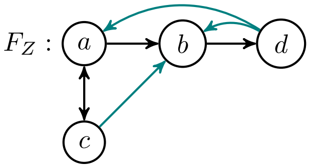
Proposition 6.8. *Let $F = (A, R)$ and a set $Z \subseteq A$ be given. If Z is admissible rejectable under both normal and strong expansion, then a admissible rejection for Z is F_Z^{sm} with the minimal syntactic distance*

$$d(F, F_Z) = |Z| - |E_{ad}(Z)^+| + 1$$

6.2 Local Expansion

Minimizing the syntactic change in local expansions for under stb semantics can be achieved based on the previously considered construction.

Example 6.9. *We start by considering the rejection from Example 4.6.*



Starting from $A \setminus Z = \{c, d\}$ we see that a is already attacked and one attack to b would suffice. Thus we can reduce the number of additional attacks to just one.

It is clear that any argument in $A \setminus Z$ can be used to ensure rejection of Z. Consequently, in the following we just pick one argument $a \in A \setminus Z$ arbitrarily. Needless to say, the following construction is by no means unique.

Definition 6.10. *Given $F = (A, R)$ and a set $Z \subseteq A$. Let $a \in A \setminus Z$. The AF $F^{am}_{Z,a} = (A, R^{am}_{Z,a})$ is constructed via*

- *$R^{sm}_{Z,a} = R \cup Q$ where $Q = \{(a, z) \mid z \notin (A \setminus Z)^+\}$.*

It is important to observe that in

$$Q = \{(a, z) \mid z \notin (A \setminus Z)^+\},$$

for each $z \notin (A \setminus Z)^+$, exactly one attack is added.

Proposition 6.11. *Let $F = (A, R)$ and a set $Z \subseteq A$ be given. If Z is stable rejectable under local expansion, then a minimal rejection for Z is $F^{sm}_{Z,a}$ for any $a \in A \setminus Z$, with the syntactic distance*

$$d(F, F^{sm}_{Z,a}) = |Z| - |(A \setminus Z)^+|.$$

The last case we want to examine is the syntactic minimization of admissible rejection under local expansion. The attentive reader might expect that this involves a similar technique, but moves from $A \setminus Z$ to the usually considered $E_{ad}(Z) = \{a \in A \setminus Z \mid a^+ \cup a^- \subseteq Z\}$. As this is true, let us proceed directly to the rejecting AF $F^{sm}_{Z,a}$.

Definition 6.12. *Given $F = (A, R)$ and a set $Z \subseteq A$. Let $a \in E_{ad}(Z)$. The AF $F^{sm}_{Z,a} = (A, R^{sm}_{Z,a})$ is constructed via*

- *$R^{sm}_{Z,a} = R \cup Q$ where $Q = \{(a, z) \mid z \notin (E_{ad}(Z))^+\}$.*

Again, we emphasize that each $z \notin (E_{ad}(Z))^+$ receives exactly one incoming attack from Q.

Proposition 6.13. *Let $F = (A, R)$ and a set $Z \subseteq A$ be given. If Z is rejectable, then for any $a \in E_{ad}(Z)$, the AF $F^{sm}_{Z,a}$ is a minimal admissible rejection for Z under local expansion with the syntactic distance*

$$d(F, F^{sm}_{Z,a}) = |Z| - |E_{ad}(Z)^+|$$

Example 6.14. *Let us again consider Example 4.6. First we determine $E_{ad}(Z) = \{c, d\}$. We end up in the same situation as in stable example above (in this case our newly constructed labelling happens to be stable) and the number of added attack can be reduced to one, either (c, b) or (d, b).*

7 Minimization and Consensus Preservation

In this section we want to study what happens if we strive for syntactic and semantic minimal change at the same time. As usual, we proceed by the types of expansions, focusing on ad and stb semantics.

7.1 Normal and Strong Expansion

As usual we consider the stable semantics first. Generally it is not possible to optimize both aspects under normal resp. strong expansion. The syntactic minimal distance is achieved under strong expansion. To achieve consensus preservation the newly added arguments X have to be rejected, otherwise every formerly stable extension is lost. Under normal expansion we sustain these extension by adding a reverse attack for each from X outgoing attack. However adding these attacks contradicts the desire for syntactical minimization. These counterattacks do not contribute to the rejection of Z. That is the cause why we need a strong expansion to realize a stable rejection with the minimal syntactic distance. Under strong expansion consensus preservation is only possible if the AF itself was already a rejection. Thus we can reuse the characterization from Proposition 5.9:

Proposition 7.1. *Input an AF $F = (A, R)$ and a set $Z \subseteq A$. A stable rejection F_Z that is consensus preserving as well as possesses the minimal syntactic distance is not achievable under normal resp. strong expansion, except $F = F_Z$.*

While we are not able to achieve the minimal syntactic distance, it is possible to undergo some syntactical optimization. For one, after constructing a consensus preserving stable rejection under normal expansion, unnecessary outgoing attack from X can be omitted.

Example 7.2. *Recall the consensus preserving stable rejection F_Z modified under normal expansion given in Example 5.4). This rejection has the syntactical distance*

$$d(F, F_Z) = 2|Z| + 1 = 5.$$

We can minimize the added attacks in the same way as in Example 6.3. Every argument in Z needs to rejected by $A \setminus Z \cup \{x\}$. However a is already attacked by c. The new attack (x, a) is redundant. Thus one attacked can be omitted.

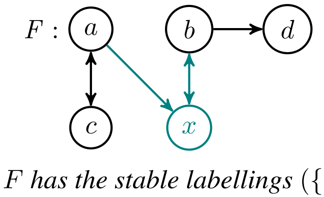
In general, we can use this heuristic to approximate the minimal syntactic change to the best extent possible in polynomial time.

Proposition 7.3. *Let $F = (A, R)$ and a set $Z \subseteq A$ be given. If Z is stable rejectable under normal expansion, then there is is a consensus preserving stable rejection F_Z computable in polynomial time for Z with syntactic distance of*

$$d(F, F_Z) = |Z| + (|Z| - |(A \setminus Z)^+|) + 1.$$

The attentive reader might ask at this point why the ingoing attacks on X were not considered in this syntactic optimization. The following example illustrates, however, that trying to refine this even further is hard in general.

Example 7.4. *We change up our running example AF.*



The initial AF F has the stable labellings $(\{a, b\}, \{c, d\}, \emptyset)$ and $(\{c, b\}, \{a, d\}, \emptyset)$. We construct the consensus preserving stable rejection and modify the number of added attacks using the just introduced heuristic. While the rejection is consensus preserving as intended, we can see that the attack (a, x) is still redundant, since a is only stable in combination with b and b already attacks x. This distinction relies on the knowledge of the initial stable labelling. Attaining this knowledge, however, is intractable.

		stb	ad	co	pr	gr
str	$\exists$	$A \setminus Z \in cf(F)$	$\top$	$\top$	$\top$	$\top$
	cp	$F = F_Z$	intrc	intrc	intrc	intrc
	sm	$\downarrow \exists$	$\downarrow \exists$			
	$cpsm$	$F = F_Z$	intrc			
nor	$\exists$	$A \setminus Z \in cf(F)$	$\top$	$\top$	$\top$	$\top$
	cp	$\downarrow \exists$	$\downarrow \exists$	$\downarrow \exists$	$\downarrow \exists$	$F = F_Z$
	sm	$\downarrow \exists$	$\downarrow \exists$			
	$cpsm$	$F = F_Z$	intrc			
loc	$\exists$	$A \setminus Z \in cf(F)$	$\exists E \in A \setminus Z : E^+ \cup E^- \subseteq Z$	$\rightarrow$	$\rightarrow$	$\rightarrow \wedge \exists e \in E_{ad}(Z) : e^- = \emptyset$
	cp	intrc*	intrc	intrc	intrc	$F = F_Z$
	sm	$\downarrow \exists$	$\downarrow \exists$			
	$cpsm$	intrc*	intrc			

Table 1: Summary of Rejection Enforcement: $\top$ - always possible, $\downarrow \exists$ - follows from existence, $\rightarrow$ - as case to the left, intrc(*) - intractable (with pol-time sufficient crit.)

Now let us head to the admissible case. Here we reason the same way as before. The counterattacks added in a normal expansion to keep the defense property of the initial admissible extension intact, does not contribute to the rejection of Z. Consequently we need to asses the problem under strong expansion. We reuse the results from Proposition 5.11:

Proposition 7.5. *Let an AF $F = (A, R)$ and a set $Z \subseteq A$ be given. A admissible rejection F_Z that is consensus preserving as well as possesses the minimal syntactic distance is only achievable under strong expansion and then only if there is some $E \subseteq A \setminus Z$ s.t. $E^+ \cup E^- \subseteq Z$ and $Z \setminus E^+$ does not contain any credulously accepted argument.*

In summary, it is technically possible to construct a consensus preserving admissible rejection under strong expansion. However we have shown in Proposition 5.12 that checking the second property is an intractable problem. Again consensus preservation is inconciliable with minimal syntactic change in polynomial time.

In the same way as in the stable case it is at least possible to approximate the minimal syntactic change by omitting superfluous outgoing attacks from X. Minimizing the incoming attacks is also intractable following the same deliberation from Example 7.4.

Proposition 7.6. *If Z is admissible rejectable under normal expansion, then there is is a consensus preserving stable rejection F_Z computable in polynomial time for Z with syntactic distance of $d(F, F_Z) = |Z| + (|Z| - |E_{ad}(Z)^+|) + 1$.*

7.2 Local Expansion

In the case of local expansion the preservation of consensus as well as the syntactic minimization do not counteract each other. In a local expansion each added attack poses the risk of introducing a new conflict between the initial arguments. Thus in contrast to the normal (strong resp.) expansion keeping the number of added attacks low is also an essential part of attaining consensus preservation. Consensus preservation is an intractable problem (Proposition 5.13) for stable as well as admissible semantics. Even so at least in the stable case we can fall back to the sufficient polynomial condition given Proposition 5.15.

Proposition 7.7. *Let $F = (A, R)$ and an under local expansion stable rejectable $Z \subseteq A$ be given. Then there exists a consensus preserving stable rejection F_Z under local expansions with the minimal syntactic distance if for each $z \in Z$ there is an $a \in A \setminus Z$ s.t. $(z, a) \in R$.*

8 On Skeptical Rejection

At first glance, it may appear more reasonable to reject Z by requiring that Z is out-labelled in *each* acceptable viewpoint, that is, in each labelling $L \in \mathcal{L}_\sigma(F)$. In this section, we briefly want to discuss this notion as well as problems which arise. First of all, it would be too restrictive to stipulate that $L^O = Z$ for each $L \in \mathcal{L}_\sigma(F)$ as the following shows.

Proposition 8.1. *Given $F = (A, R)$ and $Z \subseteq A$. If $L^O = Z$ for each $L \in \mathcal{L}_\sigma(F)$, then $|\sigma(F)| = 1$ for each $\sigma \in \{co, gr, pr, stb\}$.*

That is, by requiring Z to be the rejected set of arguments in each labelling, we force F to have only one model left (which the exception of ad). Consequently, each rejection F_Z w.r.t. this notion would be a somewhat trivial AF.

So we consider a notion of weak skeptical rejection where it suffices for Z to be a subset of the out-labelled arguments.

Definition 8.2. *Given a labelling-semantic σ, an AF $F = (A, F)$ and a set $Z \subseteq A$. Then Z is weakly skeptically rejected w.r.t. σ iff for each $L \in \mathcal{L}_\sigma(F)$, it holds $Z \subseteq L^O$.*

However, also under this notion, we do not get novel theoretical insights: we can attain the existence results from Section 4 with constructions that yield a rejection F_Z with one extension only. Consequently, the requirements from Definition 8.2 are also satisfied.

Finally, if we strive for the more interesting notion of consensus preservation, we get that in F, none of the arguments in Z are allowed to be accepted. Formally:

Proposition 8.3. *Given a labelling-semantic σ, an AF $F = (A, F)$ and a set $Z \subseteq A$. If there is any expansion F_Z s.t.*

- *F_Z weakly skeptically rejects Z, and* rejection
- *$\mathcal{L}^I_\sigma(F) \subseteq \mathcal{L}^I_\sigma(F_Z)$,* consensus

then no argument in Z is credulously σ-accepted in F.

Intuitively, this means that a consensus preserving skeptical (weak) rejection for Z can only exist whenever the goal of rejecting Z is already fulfilled in F; at least to the degree that none of the arguments can be accepted. This, of course, begs the question as to why one should modify F in the first place. In view of these observations, we leave a thorough study of natural skeptical rejection notions for future work.

9 Conclusion and Future Work

In this paper, we studied as to how certain, undesired arguments in a debate can be rejected. To this end we developed a notion of rejection enforcement in the context of AFs as proposed by Dung (Dung 1995). As is common in the context of enforcement research, we studied different types of expansions for the AF under consideration. Our study reveals that a basic notion of rejection can be ensured in many cases and, if it is impossible to reject Z, this can be decided by simply evaluating the syntax of the given AF. We then moved on to more elaborate notions which also strive to preserve existing extensions or minimize the syntactical modifications to the AF F. It turns out that these cases are much more involved and several impossibility results are derived (see Table 1 for a summary of our study).

Acceptance enforcement has been studied extensively, both theoretical (Baumann et al. 2021) as well as computational aspects (Wallner, Niskanen, and Järvisalo 2017; Niskanen, Wallner, and Järvisalo 2018). A natural future work direction would be to also consider algorithms and empirical evaluations for our rejection enforcement notions. Moreover, as our discussion in skeptical rejection demonstrates, finding a suitable notion to skeptically reject undesired arguments is an interesting, yet challenging future work direction.

Another avenue for future work is the connection between abstract and structured argumentation formalisms. Indeed, it is common for dynamic reasoning tasks like enforcement, equivalence (Oikarinen and Woltran 2011), or forgetting (Baumann and Berthold 2022; Berthold, Rapberger, and Ulbricht 2023) that investigating updates of an abstract AF does not correspond well to updating an underlying knowledge base. However, the rich literature on argumentation research provides us with tools to tackle these issues by suitably extending the AF (Baumann, Rapberger, and Ulbricht 2023; Rapberger and Ulbricht 2023; Prakken 2023; Dvorák, Rapberger, and Woltran 2023; Rocha and Cozman 2022; Bernreiter et al. 2023). Such *semi-abstract formalisms* provide a stronger connection between the AF and the underlying knowledge base. It would be worthwhile to study these formalisms also in the context of rejection enforcement.

Acknowledgments

This work was supported by the German Research Foundation (DFG, BA 6170/3-1) and by the German Federal Ministry of Education and Research (BMBF, 01/S18026A-F) by funding the competence center for Big Data and AI “ScaDS.AI” Dresden/Leipzig.

References

Alfano, G.; Greco, S.; Parisi, F.; and Trubitsyna, I. 2023. Preferences and constraints in abstract argumentation. In *Proceedings of (IJCAI-23)*. IJCAI.

Atkinson, K.; Baroni, P.; Giacomin, M.; Hunter, A.; Prakken, H.; Reed, C.; Simari, G. R.; Thimm, M.; and Villata, S. 2017. Towards artificial argumentation. *AI Magazine* 38.

Baroni, P.; Gabbay, D.; Giacomin, M.; and van der Torre, L., eds. 2018. *Handbook of Formal Argumentation*. College Publications.

Baroni, P.; Caminada, M.; and Giacomin, M. 2011. An introduction to argumentation semantics. *The Knowledge Engineering Review*.

Baumann, R., and Berthold, M. 2022. Limits and possibilities of forgetting in abstract argumentation. In *Proceedings of (IJCAI-22)*. IJCAI.

Baumann, R., and Brewka, G. 2010. Expanding argumentation frameworks: Enforcing and monotonicity results. In *Proceedings of (COMMA-10)*. IOS Press.

Baumann, R., and Ulbricht, M. 2019. If nothing is accepted–repairing argumentation frameworks. *Journal of Artificial Intelligence Research*.

Baumann, R.; Doutre, S.; Mailly, J.; and Wallner, J. P. 2021. Enforcement in formal argumentation. *FLAP*.

Baumann, R.; Rapberger, A.; and Ulbricht, M. 2023. Equivalence in argumentation frameworks with a claim-centric view: Classical results with novel ingredients. *Journal of Artificial Intelligence Research* 77.

Baumann, R. 2012. What does it take to enforce an argument? minimal change in abstract argumentation. In *Proceedings of (ECAI-12)*. IOS Press.

Bench-Capon, T. J. M., and Dunne, P. E. 2007. Argumentation in artificial intelligence. *Artificial Intelligence* 171.

Bernreiter, M.; Dvorák, W.; Rapberger, A.; and Woltran, S. 2023. The effect of preferences in abstract argumentation under a claim-centric view. In *Proceedings of (AAAI-23)*. AAAI Press.

Berthold, M.; Rapberger, A.; and Ulbricht, M. 2023. Forgetting aspects in assumption-based argumentation. In *Proceedings of (KR-23)*, 86–96.

Besnard, P., and Hunter, A. 2008. *Elements of Argumentation*. MIT Press.

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

Dauphin, J., and Satoh, K. 2018. Dialogue games for enforcement of argument acceptance and rejection via attack removal. In *Proceedings of (PRIMA-18)*. Springer-Verlag.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 7.

Dupuis de Tarlé, L.; Bonzon, E.; and Maudet, N. 2022. Multiagent Dynamics of Gradual Argumentation Semantics. In *Proceedings of (AAMAS-22)*. IFAAMAS.

Dvorák, W., and Dunne, P. E. 2018. Computational problems in formal argumentation and their complexity. In *Handbook of Formal Argumentation*. College Publications.

Dvorák, W.; Rapberger, A.; and Woltran, S. 2023. A claim-centric perspective on abstract argumentation semantics: Claim-defeat, principles, and expressiveness. *Artificial Intelligence* 324.

Dvořák, W.; Fandinno, J.; and Woltran, S. 2018. On the expressive power of collective attacks. In *Proceedings of (COMMA-18)*. IOS Press.

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

García, A. J., and Simari, G. R. 2004. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming* 4.

Modgil, S., and Prakken, H. 2013. A general account of argumentation with preferences. *Artificial Intelligence* 195:361–397.

Niskanen, A.; Wallner, J. P.; and Järvisalo, M. 2018. Extension enforcement under grounded semantics in abstract argumentation. In *Proceedings of (KR-18)*. AAAI Press.

Oikarinen, E., and Woltran, S. 2011. Characterizing strong equivalence for argumentation frameworks. *Artificial Intelligence* 175.

Prakken, H. 2023. Relating abstract and structured accounts of argumentation dynamics: the case of expansions. In *Proceedings of (KR-23)*. AAAI Press.

Rago, A.; Li, H.; and Toni, F. 2023. Interactive Explanations by Conflict Resolution via Argumentative Exchanges. In *Proceedings of (KR-23)*. AAAI Press.

Rapberger, A., and Ulbricht, M. 2023. On dynamics in structured argumentation formalisms. *Journal of Artificial Intelligence Research* 77.

Rocha, V. H. N., and Cozman, F. G. 2022. Bipolar argumentation frameworks with explicit conclusions: Connecting argumentation and logic programming. In *Proceedings of (NMR-22)*, CEUR Workshop Proceedings. CEUR-WS.org.

Wallner, J. P.; Niskanen, A.; and Järvisalo, M. 2017. Complexity results and algorithms for extension enforcement in abstract argumentation. *Journal of Artificial Intelligence Research* 60.