Extension Removal in Abstract Argumentation – An Axiomatic Approach

Ringo Baumann and Gerhard Brewka
Computer Science Institute, Leipzig University, Germany
\{baumann,brewka\}@informatik.uni-leipzig.de

Abstract
This paper continues the rather recent line of research on the dynamics of non-monotonic formalisms. In particular, we consider semantic changes in Dung’s abstract argumentation formalism. One of the most studied problems in this context is the so-called enforcing problem which is concerned with manipulating argumentation frameworks (AFs) such that a certain desired set of arguments becomes an extension. Here we study the inverse problem, namely the extension removal problem: is it possible – and if so how – to modify a given argumentation framework in such a way that certain undesired extensions are no longer generated? Analogously to the well known AGM paradigm we develop an axiomatic approach to the removal problem, i.e. a certain set of axioms will determine suitable manipulations. Although contraction (that is, the elimination of a particular belief) is conceptually quite different from extension removal, there are surprisingly deep connections between the two: it turns out that postulates for removal can be directly obtained as reformulations of the AGM contraction postulates. We prove a series of formal results including conditional and unconditional existence and semantical uniqueness of removal operators as well as various impossibility results – and show possible ways out.

Introduction
Computational models of argumentation have received a lot of interest over the recent years. There are two major lines of research in the area. Structured argumentation (Besnard and Hunter 2008), as the name suggests, deals with the formal structure of arguments, their generation from a possibly inconsistent knowledge base and the identification of relationships among arguments. Abstract argumentation (Baroni et al. 2018) is concerned with the evaluation of arguments, viewed as abstract entities. Here the evaluation is most commonly based on the attack relation among arguments. The leading formal tools in abstract argumentation are Dung’s abstract argumentation frameworks (AFs) (Dung 1995). An AF basically is a graph whose nodes represent arguments whereas the links describe the attack relation. A semantics assigns to each AF a collection of argument sets which constitute a coherent view of the world.

In recent years dynamic aspects of abstract argumentation have become an important focus of research in the area, which is far from surprising as argumentation is an inherently dynamic process. Much of this work has been influenced – in one way or another – by the famous AGM theory (Alchourrón, Gärdenfors, and Makinson 1985), the leading formal account of revision and contraction in the context of propositional logic. Whereas revision potentially replaces information with new knowledge, contraction removes information from a given knowledge base. In contrast to revision where several works already exist (Coste-Marquis et al. 2014a; 2014b; Baumann and Brewka 2015; Diller et al. 2018) to mention a few, contraction in AFs as yet has not received much attention. A notable exception is (Bisquert et al. 2011) which investigates the removal of a single argument.

An intensively studied issue in argumentation is the so-called enforcing problem (Baumann and Brewka 2010; Baumann 2012; Dupin de Saint-Cyr et al. 2016). Enforcement is concerned with manipulating argumentation frameworks (in a certain minimal way) such that a certain desired set of arguments becomes an extension. Recently, (Haret, Wallner, and Woltran 2018) presented a first axiomatic treatment of this manipulating operation and showed its close relationship to classical AGM revision. However, from a conceptual point of view there are still differences. For instance, whereas revision operates on the level of single beliefs, enforcement operates on sets of arguments representing complete views of the world.

The topic of this paper is the inverse problem to extension enforcement, namely the problem of extension removal, that is: given an AF \( F \) and a set of extensions \( E \), identify an AF \( H \) that is as close as possible to \( F \) but has none of the extensions in \( E \). In the same way as enforcement shifts revision to the level of extensions, extension removal shifts contraction to the level of extensions.

To the best of our knowledge the extension removal problem has not been studied in the literature before – which is surprising since removal appears to be as important as enforcement. Assume an argumentation framework \( F \) represents what an agent knows/believes about the world. As discussed earlier, the available arguments may give rise to multiple extensions, that is, multiple equally plausible views of the world “sanctioned” by the available arguments and their relations. It may well be that the agent realizes that for some reason some of these extensions do not reflect adequate op-
impossibility results for other semantics and shows that forbidding to remove all extensions can be a reasonable way out of the impossibility for a high number of representative argumentation semantics. Sect. 7 investigates syntactical desiderata and proves first results in this direction. Finally, Sect. 8 discusses related work and summarizes the core ideas of the paper.

### Background

#### Abstract Argumentation

An argumentation framework (AF) is a directed graph $F = (A, R)$ (Dung 1995). A node $a \in A$ is called an argument and in case of $(a, b) \in R$ we say that $a$ attacks $b$ or $a$ is an attacker of $b$. Furthermore, an argument $b$ is defended by a set $A$ if each attacker of $b$ is counter-attacked by some $a \in A$. In this paper we consider finite AFs only (cf. (Baumann and Spaning 2015; 2017) for a consideration of infinite AFs). We fix an infinite set of arguments $U$ and use $F = \{F = (A, R) \mid A \subseteq U, A \text{ finite}\}$ for the set of all considered AFs. We use $\mathcal{F}(X)$ to denote the power set of a set $X$. A set $E \in \mathcal{F}(\mathcal{F}(U))$ is called an extension-set. For a set $E \in \mathcal{F}(U)$ we use $E^\oplus = E \cup \{b \mid (a, b) \in R, a \in E\}$ and as usual, we denote $(A, R) \in (A', R')$ whenever $A \subseteq A'$ and $R \subseteq R'$.

An extension-based semantics is a function $\sigma : \mathcal{F} \to \mathcal{F}(\mathcal{F}(U))$ which assigns to any AF $F = (A, R)$ a set of reasonable positions, so-called $\sigma$-extensions, i.e. $\sigma(F) \subseteq \mathcal{F}(A)$. Beside the most basic conflict-free and admissible sets (abbr. $cf$ and $ad$) we consider the following mature semantics, namely stable, stage, semi-stable, complete, preferred, grounded, ideal and eager semantics (abbr. $stb, stg, ss, co, pr, gr, id$ and $eg$ respectively). A very good overview can be found in (Baroni, Caminada, and Giacomin 2011).

**Definition 1.** Let $F = (A, R)$ be an AF and $E \subseteq A$.

1. $E \in cf(F) \iff$ for no $a, b \in E, (a, b) \in R$,
2. $E \in ad(F) \iff E \in cf(F)$ and $E$ defends all its elements,
3. $E \in stb(F) \iff E \in cf(F)$ and $E^\oplus = A$,
4. $E \in stg(F) \iff E \in cf(F)$ and for no $I \in cf(F)$, $E^\oplus \subset I^\oplus$,
5. $E \in ss(F) \iff E \in ad(F)$ and for no $I \in ad(F)$, $E^\oplus \subset I^\oplus$,
6. $E \in co(F) \iff E \in ad(F)$ and for any $a \in A$ defended by $E$, $a \in E$,
7. $E \in pr(F) \iff E \in co(F)$ and for no $I \in co(F)$, $E \subseteq I$,
8. $E \in gr(F) \iff E \in co(F)$ and for any $I \in co(F)$, $E \subseteq I$,
9. $E \in id(F) \iff E \in co(F)$, $E \subseteq \cap pr(F)$ and there is no $I \in co(F)$ satisfying $I \subseteq \cap pr(F)$ s.t. $E \subseteq I$,
10. $E \in eg(F) \iff E \in co(F)$ and $E \subseteq \cap ss(F)$ and there is no $I \in co(F)$ satisfying $I \subseteq \cap ss(F)$ s.t. $E \subseteq I$.

A semantics $\sigma$ is universally defined if $\sigma(F) \neq \emptyset$ for any $F \in \mathcal{F}$. If in addition $|\sigma(F)| = 1$ we say $\sigma$ is uniquely defined, like grounded, ideal and eager semantics. All semantics apart from stable are universally defined. This means, stable semantics may collapse, i.e. there are AFs $F$, s.t. $stb(F) = \emptyset$. This property will play an essential role in this paper. Finally, for a given semantics $\sigma$ we say that two AFs $F$ and $G$ are (ordinarily) $\sigma$-equivalent if $\sigma(F) = \sigma(G)$.
AGM-style Contraction

We now recap the basic AGM postulates for belief contraction (Alchourrón, Gärdenfors, and Makinson 1985). In the AGM paradigm the underlying logic is assumed to be propositional logic and the beliefs are modeled by a deductively closed set of sentences (a so-called belief set). The provided postulates address the problem of how a current belief set \( K \) should be changed in the light of removing a belief \( p \). We use \( \vdash \) for the classical consequence relation, \( K \vdash p \) for the results of contracting a belief \( p \) from \( K \) and \( K + p \) for \( \vdash \)-least deductively closed set of formulas containing both \( K \) and \( p \). The latter operator is called expansion and it simply adds new beliefs without restoring consistency.\(^1\)

\[
\begin{align*}
C1 & \quad K \vdash p \quad \text{is a belief set} & \text{(closure)} \\
C2 & \quad K + p \subseteq K & \text{(inclusion)} \\
C3 & \quad p \not\vdash K \Rightarrow K + p = K & \text{(vacuity)} \\
C4 & \quad \neg p \Rightarrow p \not\vdash K + p & \text{(success)} \\
C5 & \quad K \subseteq (K + p) + p & \text{(recovery)} \\
C6 & \quad (p \leftrightarrow q \Rightarrow K + p = K + q) & \text{(extensionality)}
\end{align*}
\]

Removal Operators: A Semantic Approach

Let us assume that an AF \( F \) reflects the arguments and their interrelationships considered by an agent, and the associated extensions correspond to different views of the world the agent considers plausible (cf. (Coste-Marquis et al. 2014a; Nouioua and Würbel 2014; Diller et al. 2015) for similar approaches). Consequently, removing extensions in AFs can be seen as ruling out certain alternative views a knowledge base admits. Although (w.r.t. sceptical reasoning mode) this amounts to strengthening an agent’s beliefs, it turns out that naturally arising postulates for extension removal can be more or less directly be derived from the AGM contraction postulates. In the following, we will denote the result of removing an extension-set \( E \) from a given AF \( F \) as \( \tau(F, E) \). Clearly, the result will highly depend on the considered semantics. Therefore we parametrize our axiomatization with a certain argumentation semantics \( \sigma \). In order to figure out which axioms should be demanded for suitable removal operator let us start with some reflections:

1. The result of removing alternative views of the world should be in the same format as the input was. This means, in our setup we want to end up with an AF (\( R1^\sigma \)).

2. Moreover, as the name removal problem suggest we definitely do not want to add new extensions, i.e. the resulting views of the world should be a subset of the initial ones (\( R2^\sigma \)).

3. Removing a view of the world which is not considered acceptable anyway should not change anything (\( R3^\sigma \)).

4. Now: what about views of the world which have been considered as acceptable before? Clearly, in general we want removal to be successful, but depending on the underlying formalisms some views might not be removable. In consideration of the first item we have to check whether there are certain unremovable views, i.e. extensions which belong to any AF (\( R4^\sigma \)).

5. If we add the undesired extensions to the result of removing them we should recover at least all initial views of the world. This means, a removal operator should not delete an unnecessarily large amount of information (\( R5^\sigma \)).

6. Finally, the more extensions to be removed, the fewer extensions should be acceptable in the resulting framework (\( R6^\sigma \)).

We proceed with a precise definition of unremovable views. In analogy to classical logic we will call them tautologies.

Definition 2. For a semantics \( \sigma \) we define the set of \( \sigma \)-tautologies as

\[
\tau_\sigma = \bigcap_{H \in \mathcal{F}} \sigma(H).
\]

Note that \( \tau_\sigma \subseteq \sigma(F) \) for any semantics \( \sigma \) and any \( F \in \mathcal{F} \). For instance, in case of admissible and conflict-free sets we have \( \tau_{ad} = \tau_{cf} = \{ \emptyset \} \). Whereas for all other considered semantics \( \sigma \) we obtain \( \tau_\sigma = \emptyset \), i.e. there are no tautologies at all.\(^2\)

\[
\begin{align*}
R1^\sigma & \quad \tau(F, E) \text{ is an AF}. \\
R2^\sigma & \quad \sigma(\tau(F, E)) \subseteq \sigma(F). \\
R3^\sigma & \quad \text{For } E' = E \setminus \sigma(F) \text{ we have, } \sigma(F) = \sigma(\tau(F, E')). \\
R4^\sigma & \quad \text{For } E' = E \setminus \tau_\sigma \text{ we have } E' \cap \sigma(\tau(F, E')) = \emptyset. \\
R5^\sigma & \quad \sigma(F) \subseteq \sigma(\tau(F, E')) \cup E. \\
R6^\sigma & \quad E \subseteq E' \Rightarrow \sigma(\tau(F, E')) \subseteq \sigma(\tau(F, E)).
\end{align*}
\]

Definition 3. An operator \( \tau : \mathcal{F} \times \mathcal{P}(\mathcal{P}(U)) \rightarrow \mathcal{F} \) where \( (F, E) \mapsto \tau(F, E) \) is called a \( \sigma \)-removal operator iff axioms \( R1^\sigma-R6^\sigma \) are satisfied.

The attentive reader may have already recognized that apart from the axiom \( R6^\sigma \) all stipulated removal postulates might be almost directly reconstructed from the contraction postulates by equating \( K + p \) and \( K + p \) with the extension-set \( E \). This is exactly what we mean by stating that extension removal shifts contraction to the level of extensions.

Conditional Existence and Uniqueness

Conditional Existence

The very idea underlying extension removal is that certain extensions should no longer exist. Consequently, a natural candidate for the result of removal are those AFs which possess all those initial extensions which do not appear in \( E \). However, as propositional tautologies are exempt from possible contraction in the AGM theory, we also have to accept that some sets of arguments cannot be removed. In order to maintain the possibility of realizing such a set we thus have

\(^1\)The two additional “supplementary” postulates proposed by Gärdenfors do not play a role here and will not be discussed.

\(^2\)The claimed sets of tautologies can be verified by computing the semantics of \( F = \{(a), \emptyset\} \) and \( G = \{(b), \emptyset\} \) as well as accepting that the empty set is always conflict-free and admissible.
to add all tautological extensions since they are accepted by any framework. This leads to the following generic definition of $\sigma$-candidates which plays a central role throughout this section.

**Definition 4.** Given a semantics $\sigma$, an AF $F$ and an extension-set $E$. We define the set of $\sigma$-candidates (for $F$ and $E$) as

$$C^\sigma_{FE} = \{ H \in \mathcal{F} | \sigma(H) = (\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma \}.$$

The following theorem proves the conditional existence of removal operators for any semantics $\sigma$. More precisely, if the set of $\sigma$-candidates is non-empty for any possible pair of an AF and an extension-set, then selecting one AF out of each set yields a $\sigma$-removal operator.

**Theorem 1.** Given semantics $\sigma$, s.t. $C^\sigma_{FE} \neq \emptyset$ for each AF $\mathcal{F}$ and extension-set $E$. Any function $f_\sigma: \mathcal{F} \times \mathcal{P}(\mathcal{U}) \rightarrow \mathcal{F}$ with $f_\sigma(F,E) \in C^\sigma_{FE}$ yields a $\sigma$-removal operator.

**Proof.** Given $f_\sigma$ as described above. We have to prove that all removal postulates are satisfied. In the following we consider $F$ and $E$ as arbitrary but fixed AF or extension-set, respectively.

$R1^\sigma$ Since for any AF $F$ and each extension-set $E$ the non-emptiness of $C^\sigma_{FE}$ is guaranteed by assumption we have that $f_\sigma$ is well-defined, i.e. it is indeed a function. Consequently, $f_\sigma(F,E) \in \mathcal{F}$.

$R2^\sigma$ According to Definition 4 we have $\sigma(f_\sigma(F,E)) = \sigma(H)$, s.t. $H \in C^\sigma_{FE}$. This means $\sigma(H) = (\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma$. Moreover, obviously, $\sigma(F) \setminus \mathcal{E} \subseteq \sigma(F)$ and due to Definition 2, $\tau_\sigma \subseteq \sigma(F)$. Altogether, $\sigma(f_\sigma(F,E)) \subseteq \sigma(F)$ as required.

$R3^\sigma$ Given $E' = E \setminus \sigma(F)$. Then, $\sigma(f_\sigma(F,E')) = \sigma(f_\sigma(F,E \setminus \sigma(F))) = \sigma(H)$ with $H \in C^\sigma_{FE \setminus \sigma(F)}$. Hence, $\sigma(H) = (\sigma(F) \setminus (E \setminus \sigma(F))) \cup \tau_\sigma$. Applying the following set theoretical identity $C \setminus (B \setminus A) = (A \cap C) \cup (C \setminus B)$ yields $\sigma(H) = \sigma(F) \setminus \tau_\sigma$. Finally, since $\tau_\sigma \subseteq \sigma(F)$ for any semantics $\sigma$ we deduce $\sigma(f_\sigma(F,E')) = \sigma(F)$.

$R4^\sigma$ Assume $E' = E \setminus \tau_\sigma$. We have to show that $E' \cap \sigma(f_\sigma(F,E')) = \emptyset$. In consideration of Definition 4 we have to verify $(E' \setminus \tau_\sigma) \cap ((\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma) = \emptyset$. Striving for a contradiction, let $(E' \setminus \tau_\sigma) \cap ((\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma) \neq \emptyset$. Hence there is an extension-set $E$, s.t. $E \in \mathcal{E} \setminus \tau_\sigma$ and $E \in (\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma$. We deduce $E \in \mathcal{E}$ and $E \notin \tau_\sigma$. Thus, $E \in \sigma(F) \setminus \mathcal{E}$ has to hold which implies $E \notin \mathcal{E}$. Contradiction!

$R5^\sigma$ Removal postulate $R5^\sigma$ can be shown in the following direct manner. $\sigma(F) \subseteq \text{set}h. \ (\sigma(F) \setminus \mathcal{E}) \cup \mathcal{E} = \text{Def}2 \ ((\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma \cup \mathcal{E} = \text{Def}4 \ \sigma(f_\sigma(F,E)) \cup \mathcal{E}).$

$R6^\sigma$ According to Definition 4 we have $\sigma(f_\sigma(F,E')) = (\sigma(F) \setminus E') \cup \tau_\sigma$. Due to the assumption $E' \subseteq \mathcal{E}$ we deduce $(\sigma(F) \setminus E') \cup \tau_\sigma \subseteq (\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma$ which proves the anti-monotonicity in the second component, i.e. $\sigma(f_\sigma(F,E')) \subseteq \sigma(f_\sigma(F,E))$.

**Semantical Uniqueness of Removal Operators**

In the following we show that all removal operators produce the same semantical output. This means, in contrast to classical AGM-style contraction there is no choice for the resulting belief set. Any two removal operators produce ordinarly equivalent AFs for the same input. In order to prove this bold claim we firstly show the main lemma stating that $\sigma$-candidates are the only possible results for any removal operator.

**Lemma 2.** Given a semantics $\sigma$ and a removal operator $r: \mathcal{F} \times \mathcal{P}(\mathcal{U}) \rightarrow \mathcal{F}$. For each AF $F$ and any extension-set $E$ we have: $r(F,E) \in C^\sigma_{FE}$.

**Proof.** Striving for a contradiction, suppose that there are a removal operator $r$, an AF $F$ and an extension-set $E$, s.t. $r(F,E) \notin C^\sigma_{FE}$. Consequently, according to Definition 4 we deduce $\sigma(r(F,E)) \neq (\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma$. This assertion is equivalent to $\sigma(r(F,E)) \notin (\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma$ or $(\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma \notin \sigma(r(F,E))$. We show that both cases lead to a contradiction.

1. Assume $\sigma(r(F,E)) \notin (\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma$. Hence, there is a set $E$ with $E \notin \sigma(r(F,E))$ and $E \notin (\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma$. Consequently, $E \notin \sigma(F) \setminus \mathcal{E}$ and $E \notin \tau_\sigma$. Since $\sigma(r(F,E)) \subseteq \sigma(F)$ by removal postulate $R6^\sigma$ we have $E \in \sigma(F)$. Together with $E \notin \sigma(F) \setminus \mathcal{E}$ we obtain $E \in \mathcal{E}$. Applying $E \notin \tau_\sigma$ yields $E \in \mathcal{E} \setminus \tau_\sigma$. Now, we can employ the removal postulate $R4^\sigma$ which results in $E \notin \sigma(r(F,E))$. Contradiction.

2. Assume $(\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma \notin \sigma(r(F,E))$. In consideration of removal postulate $R5^\sigma$ we obtain $\sigma(F) \subseteq \sigma(r(F,E)) \cup \mathcal{E}$. If $E$ is omitted on both sides we get $\sigma(F) \subseteq \sigma(r(F,E)) \setminus \mathcal{E}$. Thus, even $\sigma(F) \setminus \mathcal{E} \subseteq \sigma(r(F,E))$ is guaranteed. Now, since due to Definition 2 $\tau_\sigma \subseteq \sigma(H)$ for any AF $H$ and moreover, since $r(F,E)$ yields indeed an AF because it is a removal operator we finally obtain $(\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma \subseteq \sigma(r(F,E))$. Contradiction.

Having Lemma 2 and Definition 4 at hand we immediately obtain the claimed semantical uniqueness result. For ease of notation we introduce equivalence among operators first. That is, two operators are equivalent iff they produce ordinarly equivalent AFs for any initial framework and extension-set. We want to stress that the theorem below holds for any semantics.

**Definition 5.** Given a semantics $\sigma$ and two $\sigma$-removal operators $r, r': \mathcal{F} \times \mathcal{P}(\mathcal{U}) \rightarrow \mathcal{F}$. We say $r$ and $r'$ are $\sigma$-equivalent (denoted $r \equiv^\sigma r'$) if for any AF $F$ and any extension-set $E$ we have: $r(F,E) \equiv^\sigma r'(F,E)$.

**Theorem 3.** Given a semantics $\sigma$ and two $\sigma$-removal operators $r$ and $r'$. Then $r \equiv^\sigma r'$. 
Removal Operators for Stable Semantics

So far we did not tackle the question whether there are removal operators at all. In consideration of the previous results (Theorem 1, Lemma 2) we deduce that the existence of removal operators depends on the non-emptiness of the set of \( \sigma \)-candidates. In consideration of Definition 4 the existence problem of removal operators shifts to a realizability problem of argumentation semantics. More precisely, if a semantics \( \sigma \) allows us to realize the relative complement of a given set of \( \sigma \)-extensions and an arbitrary extension-set, then the existence of \( \sigma \)-removal operators is guaranteed. We mention that expressibility issues highly depend on the logical formalisms. For instance, in case of propositional logic any finite set of two-valued interpretations is realizable. This means, given such a finite set \( I \), we always find a set of formulas \( T \), s.t. \( \text{Mod}(T) = I \). A first systematic study for abstract argumentation semantics was given in (Dunne et al. 2015). One of the main insights was that representational limits highly depend on the chosen semantics (cf. (Baumann 2017, Table 2) for a comprehensive overview).

Let us consider stable semantics first. It was shown that tightness and incomparability are the decisive properties for realizability under stable semantics. Consider the following definition and theorem.

**Definition 6** ((Dunne et al. 2015)). Given an extension-set \( E \in \mathcal{E}(\mathcal{P}(U)) \). Then \( E \) is called

1. tight if for all \( E \in \mathcal{E} \) and \( a \in \bigcup E \) we have: if \( S \cup \{a\} \notin \mathcal{E} \), then there is an \( s \in S \), s.t. \( (a,s) \notin \{(a,b) \mid \exists E \in \mathcal{E} : \{a,b\} \subseteq E \} \) and
2. incomparable if for each \( E, E' \in \mathcal{E} \), \( E \notin \mathcal{E} \).

**Theorem 4** ((Dunne et al. 2015)). Given an extension-set \( E \in \mathcal{E}(\mathcal{P}(U)) \). Then,

\[ F \in \mathcal{F} \text{ s.t. } \text{stb}(F) = \mathcal{E} \text{ exists } \iff \mathcal{E} \text{ is incomparable and tight.} \]

Roughly speaking, tightness encodes that if an argument \( a \) does not occur in some extension \( E \) there must be a reason for that, e.g. an attack between \( a \) and \( E \). Incomparability is just another name for forming a \( \preceq \)-antichain. We encourage the reader to verify the mentioned properties for the running example \( F \) depicted in Example 1.

The following theorem proves that the set of \( \text{stb} \)-candidates is guaranteed to be non-empty. The main reason for this is that the characterizing properties transfer to any subset of a given set of stable extensions.

**Theorem 5.** For each AF \( F \) and extension-set \( E \), \( \mathcal{C}^{\text{stb}}_{F,E} \neq \emptyset \).

**Proof.** Given an AF \( F \) and an extension-set \( E \). In accordance with Theorem 4 we have that \( \text{stb}(F) \) is tight and incomparable. Since incomparability implies that tightness of a set transfers to any subset of it (Baumann 2017, Lemma 3.10, item 2) we have tightness of \( \text{stb}(F) \setminus E \) (\( = \text{stb}(F) \setminus \mathcal{E} \) \( \cup \mathcal{T}^{\text{stb}} \) since \( \mathcal{T}^{\text{stb}} = \emptyset \)). Moreover, being a \( \preceq \)-antichain obviously transfers to any subset. This means, \( \text{stb}(F) \setminus \mathcal{E} \) is tight and incomparable which implies its realizability according to Theorem 4. Hence, \( \mathcal{C}^{\text{stb}}_{F,E} \neq \emptyset \) concluding the proof. \( \square \)

Now, we are ready to present the main theorem for stable semantics stating the existence as well as the range of removal operators.

**Theorem 6.** Let \( R = \{r \mid r \text{ is a } \text{stb}-\text{removal operator} \} \). Then, \( R = \{f \mid f: \mathcal{F} \times \mathcal{P}(\mathcal{U}) \rightarrow \mathcal{F} \text{ with } f(F,E) \in \mathcal{C}^{\text{stb}}_{F,E} \} \) and moreover, \( R \neq \emptyset \).

**Proof.** Apply Theorems 1, 5 as well as Lemma 2. \( \square \)

As a matter of fact, knowing that a certain set is realizable does not provide one automatically with a witnessing AF. Consequently, we have to ask: how to obtain a framework which possesses the required sets of extensions? Fortunately, in case of stable semantics we are equipped with a two-step procedure showing realizability in a constructive fashion (Baumann 2017, Definition 3.18, Proposition 3.19). This means, we are able to build a \( \text{stb} \)-candidate from scratch. The main idea of the canonical construction is as follows: For a given extension-set \( E \) we start with a framework \( G \) possessing all arguments in \( U \) and add an attack between any two arguments iff they do not occur jointly in any set \( E \in \mathcal{E} \). In this way it is guaranteed that any set in \( E \) becomes a stable extension of \( G \) (generating step). In a second step, we augment the initial framework \( G \) to \( I \), s.t. only elements in \( E \) become stable. For any undesired stable extension \( S \in G \) we add a self-attacking argument \( s \) which is attacked by any argument in \( U \setminus S \) (eliminating step).

**Example 2** (Standard Construction). In Example 1 we presented two ad hoc “solutions” (AFs \( H \) and \( H' \)) to the problem of removing the undesired set \( E \) from the AF \( F \). However, both frameworks can not stem from a removal operator \( r \) in the sense of Definition 3 since several postulates are not satisfied, e.g. \( R5^{\text{stb}} \) is violated by \( H \) and \( R2^{\text{stb}} \) is not fulfilled by \( H' \).

The following framework \( I \) is the above mentioned standard construction for the required extension-set \( E = \{\{a,e\}, \{c,d,f\} \} \). This means, \( I \in \mathcal{C}^{\text{stb}}_{F,E} \). We mention that the framework \( I \) is already obtained after the generating step since no additional stable extensions are constructed.

\[ \begin{align*}
I : f & \rightarrow e \\
\backslash & \in \text{ stb}
\end{align*} \]
one. More precisely, any mature universally defined semantics \( \sigma \) is not able to realize no extensions which is enforced if we want to remove all initial extensions (cf. Lemma 2). However, on the positive side, if we forbid to give up all extensions, i.e. if requesting to start with a tabula rasa is disallowed, then one may define a reasonable form of removal in the spirit of stable semantics.

**Theorem 7.** There is no \( \sigma \)-removal operator for each semantics \( \sigma \in \{ cf, ad, stg, stb, ss, co, pr, gr, id, eg \} \).

**Proof.** Consider \( \sigma \in \{ stg, stb, ss, co, pr, gr, id, eg \} \). We have, \( \tau_\sigma = \emptyset \). Striving for a contradiction assume that there is a \( \sigma \)-removal operator \( r : \mathcal{F} \times \mathcal{P}(\mathcal{U}) \rightarrow \mathcal{F} \) and let \( F = \{ \{ a \}, \emptyset \} \) and \( \mathcal{E} = \{ \{ a \} \} \). Hence, by Lemma 2 and Definition 4 we obtain \( \sigma( r(F, \mathcal{E}) ) = (\sigma(F) \setminus \mathcal{E}) \cup \tau_\sigma = \emptyset \cup \emptyset = \emptyset \). Since \( \sigma \) is universally defined we obtain \( \sigma(H) \neq \emptyset \) for any \( H \in \mathcal{F} \). This means, \( r(F, \mathcal{E}) \) fails to be an AF.

Consider now \( \sigma \in \{ cf, ad \} \). We have \( \tau_\sigma = \{ \emptyset \} \).

Suppose for a contradiction that there is a \( \sigma \)-removal operator \( r \). Consider \( F'' = \{ \{ a, b, a \}, \emptyset \} \) and \( \mathcal{E}'' = \{ \emptyset, \{ a, b, a \} \} \). Hence, according to Lemma 2 and Definition 4 we obtain \( \sigma( r(F'', \mathcal{E}'') ) = (\sigma(F'') \setminus \mathcal{E}'') \cup \tau_{cf} = (\emptyset, \{ a, b, a \}) \setminus \{ \emptyset, \{ a, b, a \} \} \cup \{ \emptyset, \{ a, b, a \} \} = \{ \emptyset, \{ a, b, a \} \} \).

Clearly, if \( \{ a, b \} \in cf(H) \), then \( \{ a \} \in cf(H) \). Thus, \( cf( r(F'', \mathcal{E}'') ) \) is not realizable by any AF \( H \in \mathcal{F} \).

Finally, let us turn to admissible sets. Consider \( F''' = \{ \{ a, b, a, a \}, \emptyset \} \) and \( \mathcal{E}''' = \{ \emptyset, \{ a, b, a, a \} \} \). Hence, \( ad(F''') = \mathcal{P}(\{ a, b, a, a \}) \) and thus, according to Lemma 2 and Definition 4 we deduce \( \sigma( r(F''', \mathcal{E}''') ) = \mathcal{P}(\{ a, b, a, a \}) \setminus \{ \{ a, b, a, a \} \} \).

It can be checked that this set does not fulfill so-called conflict-sensitivity which implies that \( r(F''', \mathcal{E}''') \) fails to be an AF (Baumann 2017, Definition 3.21, Theorem 3.24).

We close this section with a positive result for several mature semantics. In analogy to Theorem 5 we may show an existence result as long as removing all extensions is forbidden. The proof uses non-trivial realizability results first shown in (Dunne et al. 2015).

**Theorem 8.** Given a semantics \( \sigma \in \{ stg, ss, pr, gr, id, eg \} \). For any AF \( F \) and any extension-set \( \mathcal{E} \) we have \( C_{F, \mathcal{E}}^\sigma \neq \emptyset \), whenever \( \sigma(F) \setminus \mathcal{E} \neq \emptyset \).

**Proof.** For any considered semantics we have \( \tau_\sigma = \emptyset \). This means, \( C_{F, \mathcal{E}}^\sigma \neq \emptyset \) if and only if there is an AF \( H \), s.t. \( \sigma(H) = \sigma(F) \setminus \mathcal{E} \). Consider first the uniquely defined semantics \( \sigma \in \{ gr, id, eg \} \). Since \( \sigma(F) \setminus \mathcal{E} \neq \emptyset \) is assumed we deduce \( \sigma(F) \setminus \mathcal{E} = \{ a \} \). Hence, in accordance with (Baumann 2017, Theorem 3.33) we obtain \( C_{F, \mathcal{E}}^\sigma \neq \emptyset \).

Consider now \( \sigma \in \{ stg, ss, pr \} \). The characterization theorems require non-empty extension sets, incomparability, and conflict-sensitivity or tightness (Baumann 2017, Theorems 3.12, 3.24). The latter both transfer to subsets given incomparability (Baumann 2017, Lemmata 3.10, 3.23). Incomparability itself transfers to any subset by definition. Finally, non-emptiness is implied by the assumption \( \sigma(F) \setminus \mathcal{E} \neq \emptyset \).

Hence, \( C_{F, \mathcal{E}}^\sigma \neq \emptyset \) concluding the proof.

We emphasize that conflict-free and admisible sets as well as complete semantics are not included in the theorem above since the assertion would be false. Moreover, just like in case of stable semantics witnessing frameworks for the considered semantics can be constructed from scratch. In contrast to stage semantics where the same standard construction as for stable semantics is used we have to rely on a more involved construction for preferred semantics (Baumann 2017, Propositions 3.19, 3.31). Due to translation results we may use the latter even for semi-stable semantics (Dvorák and Woltran 2011).

**Syntactical Desiderata**

One of the main assumptions of belief revision is that of minimal change. The removal postulates \( R1^p - R6^p \) encode this core idea with respect to the semantical side of AFs. In this section we will briefly discuss which kind of syntactical desiderata can be fulfilled on top of the semantical ones. Consider the following example.

**Example 3 (Syntactically Desirable Results).** In Example 2 we presented the standard construction which realizes the required set of extensions. The following two AFs \( G \) and \( G' \) present further possible results for removal operators \( r \) since \( stb(G) = stb(G') = \{ \{ a, c \}, \{ c, d, f \} \} \), i.e. \( G, G' \in C_{F,E}^{stb} \).

![Diagram](image.png)

Which result should be preferred for \( r(F, \{ E \}) \)? One may argue that the resulting framework should be as similar as possible (w.r.t. some distance measure) to the initial AF. In this case \( AF \ H \) should be preferred over \( H' \) as well as \( I \). Another plausible option is to require as few syntactical material as possible. Hence, in this setup \( AF \ H' \) should be preferred over \( H \) and \( I \).

As a first result in this direction we show that it is impossible to guarantee that there is a removal result which is a subgraph of the initial framework (as satisfied by \( H \) and \( H' \) depicted above). We consider stable semantics only since for the other semantics the general impossibility is already shown (Theorem 7).

**Theorem 9.** There is no \( stb \)-removal operator \( r \), s.t. for any AF \( F \) and any extension-set \( \mathcal{E} \) we have, \( r(F, \mathcal{E}) \in F \).

**Proof.** Striving for a contradiction assume the existence of a \( stb \)-removal operator \( r \). Consider the AF \( J = \{ \{ a, b, c, d \}, \{ \{ a, b \}, \{ b, a \}, \{ c, d \}, \{ d, c \} \} \} \) and \( \mathcal{E} = \{ \{ a, c \} \} \). We have \( stb(J) = \{ \{ a, c \}, \{ a, d \}, \{ b, c \}, \{ b, d \} \} \).

According to Theorem 6 we deduce \( stb(r(J, \mathcal{E})) = \)
\{\{a, d\}, \{b, c\}, \{b, d\}\}. Since any argument of \(J\) occurs in at least one extension we deduce that \(J\) and \(r(J, \mathcal{E})\) possess the same arguments. Moreover, even deleting one attack would lead to one unattacked argument \(x\). Hence, \(x\) is necessarily contained in any extension in contrast to the required set of stable extension \(stb(r(J, \mathcal{E}))\). Contradiction!

On the positive side, we may show that a removal result can always be found as a supergraph of the initial framework. Moreover, the distance (in terms of added arguments) is guaranteed to be low given that the number of extensions which has to be removed is low too. A further detailed consideration of this issue will be part of future work.

**Theorem 10.** There exists a \(stb\)-removal operator \(r\), s.t. for any \(AF\) \(F = (A, R)\) and any extension-set \(\mathcal{E}\) we have:

\[ F \in r(F, \mathcal{E}) = (A', R') \text{ and } |A'| = |A| + |stb(F) \cap \mathcal{E}|. \]

**Proof.** Given an \(AF\) \(F = (A, R)\) and an extension-set \(\mathcal{E}\). Moreover, let \(\mathcal{E}' = stb(F) \cap \mathcal{E}\). Now, define \(A' = A \cup \{E | E \in \mathcal{E}'\}\) and \(R' = R \cup \{ (E, E') | E \in \mathcal{E}' \cup \{ (a, \overline{E}) | E \in \mathcal{E}, a \in A \setminus E \}.\) By construction we have \((A, R') \in (A', R')\) and \(|A'| = |A| + |stb(F) \cap \mathcal{E}|\) as required.

It remains to show that \(F' = (A', R') \in C^t_{\mathcal{E}\mathcal{E}}\). This means, according to Definition 4 we have to prove \(stb(F') = \overline{stb(F)} \cdot \mathcal{E}\). We prove \(stb(F') \supseteq \overline{stb(F)} \cdot \mathcal{E}\) only and state that the \(\varepsilon\)-direction can be shown in a similar fashion.

Let \(E' \in stb(F) \cdot \mathcal{E}\). Consequently, \(E' \in stb(F)\) and \(E' \notin \mathcal{E}\) justifying \(E' \notin stb(F) \cdot \mathcal{E}\). \(E'\) attacks all arguments in \(A\) since \(E' \in stb(F)\). Moreover it remains conflict-free in \(F'\) since new attacks involve at least one new argument in \(\{E | E \in \mathcal{E}'\}\). Since \(stb(F)\) forms a \(\varepsilon\)-antichain we have that \(E'\) possesses at least one distinct element \(a_E\) w.r.t. each \(E \in \mathcal{E}' \subseteq stb(F).\) This means, \(a_E \in E' \setminus E\) and therefore \(a_E \in A \setminus E\). Thus, any new argument in \(\{E | E \in \mathcal{E}'\}\) is attacked by \(E\) according to the definition of \(R'\). Eventually, \(E \in stb(F')\).

**Summary and Related Work**

In this paper we presented an axiomatic approach to extension removal in abstract argumentation. An extension removal operator is a function that is given a Dung argumentation framework and a collection of “undesired” extensions as input. The outcome is an argumentation framework that does not possess any of the undesired extensions.

We introduced 6 removal axioms which formalize reasonable properties a removal operator should possess. Although removal and contraction are conceptually quite different, it turned out that the postulates could be obtained as reformulations of the AGM postulates for contraction. We formally studied extension removal operators and obtained various – as we believe quite interesting – results for a representative number of argumentation semantics. This includes existence, uniqueness and impossibility of various special cases. Moreover, in case of impossibility results we studied how to weaken the initial requirements in order to obtain reasonable removal operators. Stable semantics played an outstanding role in this study since it guarantees the unconditional existence of removal operators. Moreover, we showed how to essentially construct such removal results and proved first results regarding the distances of such solutions. As indicated in that section the topic of syntactical desiderata is one central issue for future work.

As mentioned in the Introduction, we are not aware of any alternative approaches to extension removal for AFs. The closest work to ours is (Boella, Kaci, and van der Torre 2009) where the authors studied removing arguments and/or attacks, s.t. the initial semantics remain unchanged. This problem might be called extension conserving problem and they studied it for the uniquely defined grounded semantics.

Among the existing work related to ours is work on contracting defeasible logic programs (García et al. 2011), a specific approach to structured argumentation, and contracting logic programs in general (Binnewies, Zhuang, and Wang 2015). However, the techniques used there are specific for the formalisms they are intended for and it is difficult to see how they could be adapted to AFs. Moreover, they are concerned with the removal of single beliefs rather than whole world views, which are what extensions represent.

In (Coste-Marquis et al. 2014a) the emphasis is on finding assignments of values to arguments which reflect intended revisions - expressed in a flexible language - and are as close as possible to the original assignments. A drawback is that the result of revision may not be representable using a single AF. Diller and colleagues (Diller et al. 2018) take up this issue and present an extension-based approach where the revision of an AF is actually guaranteed to be representable by an AF. To prove a representation theorem, they make use of recent advances in both areas of argumentation and belief change. In particular, they use the concept of realizability in argumentation and the concept of compliance as introduced in Horn revision. Finally, (Coste-Marquis et al. 2014b) present an approach which is based on translating an AF into a propositional formula which is then revised using a standard propositional revision operator. This latter translation-based approach could without much effort be adapted for contraction. However, the outcome of such an approach heavily depends on the chosen representation of AFs in propositional logic, and there is certainly more than one way of choosing such a representation.

There are several obvious directions for future work, in particular the development and implementation of algorithms for extension removal. Furthermore, we will also extend our analysis to further semantics, e.g. ranking-based semantics (Amgoud and Ben-Naim 2013; Amgoud et al. 2016) and abstract dialectical frameworks (ADFs) (Brewka et al. 2013) which generalize AFs significantly. We are confident that the techniques developed in this paper will also be applicable to this more general formalism. We will also investigate the suitability of our approach to other nonmonotonic formalisms, e.g. to extension removal in default logic (Reiter 1980) or answer set removal in answer set programming (Brewka, Eiter, and Truszczyński 2011).

**Acknowledgments**

This work was partially funded by DFG (1817/7-2) and DAAD (57407370). We want to thank Felix Linker for fruitful discussions on an earlier version of this paper.
References


