Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning
Main Track

Defining Defense and Defeat in Abstract Argumentation From Scratch – A Generalizing Approach

Lydia Blümel and **Markus Ulbricht**
Department of Computer Science, Leipzig University
mulbricht@informatik.uni-leipzig.de

Abstract

We propose a general framework to investigate semantics of Dung-style argumentation frameworks (AFs) by means of a generic defeat notion formalized by refute operators. After establishing the technical foundations, we propose natural generic versions of Dung's classical semantics. We demonstrate how classical as well as recent proposals can be captured by our approach when utilizing suitable notions of refutal. We perform an investigation of basic properties which semantics inherit from the underlying refute operator. In particular, we show under which conditions a counterpart to Dung's fundamental lemma can be inferred and how it ensures the existence of the generalized version of complete extensions. We contribute to a principle-based study of AF semantics by discussing properties tailored to compare different refute operators. Finally, we report computational complexity results for basic reasoning tasks which hold in our general framework.

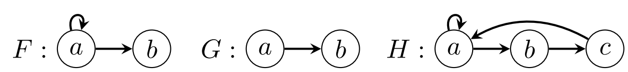
1 Introduction

Since Dung's seminal 1995 paper (Dung 1995) introducing abstract argumentation frameworks (AFs), the field of formal argumentation has become a vibrant research area in Artificial Intelligence. Research in AFs is driven by investigating the behavior of semantics formalizing jointly acceptable sets of arguments (cf. (Baroni, Caminada, and Giacomin 2018) for an overview). Over the years, many fundamental properties typically considered for non-monotonic formalisms have been studied for abstract argumentation semantics. Among others, central issues like existence and uniqueness (Weydert 2011; Baumann and Spanring 2015), replaceability (Oikarinen and Woltran 2011; Baumann 2016), expressibility (Dunne et al. 2015), as well as general modularity and locality properties were studied (Baroni and Giacomin 2007; Baumann 2011; Baroni, Giacomin, and Liao 2018).

Due to the considerable amount of semantics that have been proposed in the literature, researchers have developed proposals in order to objectively assess their quality for different scenarios. This resulted in a comprehensive collection of so-called *principles* (Baroni and Giacomin 2007; van der Torre and Vesic 2018), i.e. properties that formalize (un-)desirable characteristics of semantics. The principles range from basic concepts which hold by definition for most semantics to rather technical requirements which formalize different concepts of step-wise computability of extensions.

The principle-based analysis reveals that most commonly accepted semantics are either based on the notion of admissibility or naivity. In a nutshell, admissibility requires that a set of arguments is capable of defending itself (that is, counter-attacking any argument challenging it) in the given AF, whereas naivity is not concerned with defense at all.

However, these two concepts do not always yield intuitive results. Let us illustrate the sometimes undesired behavior by considering the following example AFs. The formal theoretical background for AFs can be found in Section 2.



Suppose we are in a setting where we wish to neglect the self-attacker a in F. Then b should be acceptable which is however not the case for any admissible-based semantics. As a possible solution we could consider naivity-based semantics. But since they completely disregard the notion of defense we then would also accept b in G. This is clearly undesired though since in G, b is attacked by the undisputed argument a. Naivity-based semantics which also make use of the SCC-recursive scheme (Baroni, Giacomin, and Guida 2005) would handle F and G quite well, but due to their reliance on the structure of the given graph, the AF H would be treated unsatisfactory: It consists of a single SCC only and therefore this approach collapses to the underlying base function (we refer the reader to (Baroni, Giacomin, and Guida 2005) for more details on SCC-recursive semantics).

These observations motivated the proposal of some recent approaches which are tailored to find "good" ways to disregard some of the attacking arguments in certain situations (Bodanza and Tohmé 2009; Dondio and Longo 2021; Baumann, Brewka, and Ulbricht 2020b; Dauphin, Rienstra, and van der Torre 2021). Comparing the characteristics of such non-classical semantics in a formal way is a challenging endeavor, since all of them make (more or less implicit) use of different defense notions by disregarding certain attackers of the given extension. Thus, a comprehensive understanding of these semantics requires delving into and comparing the varying underlying concepts. Moreover, the usual principles considered in the literature (Baroni and Giacomin 2007; van der Torre and Vesic 2018) are not concerned with different versions of defense making it hard to use them when assessing what a "good" weak version of admissibility is.

In this paper, we contribute to this line of research and propose a unifying approach for investigating defense notions, their implications for the induced semantics, and their suitability for certain scenarios. The key observation in our proposal is that Dung's defense is based on the conception that each attacker of a given argument shall be *refuted* by the defending set. Therefore, by altering the definition of "refute" we obtain versions of Dung's theory, which are familiar in their spirit, but suitable for our purpose.

Our main contributions can be summarized as follows:

- After recalling the necessary AF background in Section 2 we introduce our general framework in Section 3. Based on a generic refute operator δ we develop a defense notion and a generalization of the characteristic function, yielding counterparts to Dung's semantics.
- We demonstrate how recent proposals from the literature can be captured by our approach and the underlying defense notion of the semantics is revealed (Section 4).
- We discuss formal properties of our general semantics based on the characteristics of the specified refute operator. Most notably, we obtain a natural generalization of Dung's fundamental lemma (Section 5).
- We propose abstract principles, tailored to investigate refute operators and evaluate the ones we introduced throughout the paper (Section 6).
- We investigate the computational complexity of basic reasoning tasks for semantics based on polynomial-time computable refute operators (Section 7).

2 Background

We fix a non-finite background set $\mathcal{U}$. An *argumentation framework* (AF) (Dung 1995) is a directed graph $F = (A, R)$ with a set of arguments $A \subseteq \mathcal{U}$ and the attack relation $R \subseteq A \times A$. If not stated otherwise we assume any AF to be finite and we use $\mathcal{F}$ for the set of all these graphs. If $(a, b) \in R$ we say that a *attacks* b as well as a *attacks* (the set) E given that $b \in E \subseteq A$. This situation is denoted as $a \to b$ or $a \to E$, respectively; we define $E \to a$ analogously.

We frequently use the so-called *range* of a set E defined as $E^\oplus = E \cup E^+$ where $E^+ = \{a \in A \mid E \text{ attacks } a\}$. We also let $E^- = \{a \in A \mid a \text{ attacks } E\}$. If $E = \{a\}$ is a singleton, we write a^+ and a^- instead of $\{a\}^+$ and $\{a\}^-$. By $F{\downarrow_E}$ we denote F projected to the arguments in E, i.e. $F{\downarrow_E} = (E, R \cap (E \times E))$. The *E-reduct of F* is the AF $F^E = (E^*, R \cap (E^* \times E^*))$ where $E^* = A \setminus E^\oplus$. This means, F^E is the subframework of F obtained by removing the range of E, i.e. $F^E = F{\downarrow_{A \setminus E^\oplus}}$. For an AF $F = (B, S)$ we let $A(F) = B$ and $R(F) = S$.

A *semantics* is a function $\sigma : \mathcal{F} \to 2^{2^{\mathcal{U}}}$ with $F = (A, R) \mapsto \sigma(F) \subseteq 2^A$. This means, a semantics returns a set of subsets of A, so-called *σ-extensions*. We say that an argument $a \in A$ is *credulously accepted* if $a \in \bigcup \sigma(F)$. Similarly, a is considered as *skeptically accepted* if $|\sigma(F)| \geq 1$ and $a \in \bigcap \sigma(F)$. In case of uniquely defined semantics, i.e. $|\sigma(F)| = 1$ for any F we may simply speak of *accepted* arguments as both notions coincide.

In this paper we consider several semantics which are in spirit based on Dung's classical *admissible*, *complete*, *preferred*, *grounded* and *stable* semantics (abbr. ad, co, pr, gr, stb). All mentioned semantics satisfy conflict-freeness. A set $E \subseteq A$ is conflict-free in F (for short, $E \in cf(F)$) iff for no $a, b \in E$, $a \to b$. For the present paper it will be convenient to utilize the so-called *characteristic function* Γ_F to define the semantics. For $E \subseteq A$ we have $\Gamma_F(E) = \{a \in A \mid E \text{ defends } a\}$. A set E *defends* a if $b \to a$ implies $E \to b$.

Definition 2.1. *Let $F = (A, R)$ be an AF and $E \in cf(F)$.*

1. *$E \in ad(F)$ iff $E \subseteq \Gamma_F(E)$,*
2. *$E \in co(F)$ iff $E = \Gamma_F(E)$,*
3. *$E \in pr(F)$ iff E is $\subseteq$-maximal in $co(F)$,*
4. *$E \in gr(F)$ iff E is $\subseteq$-minimal in $co(F)$,*
5. *$E \in stb(F)$ iff E attacks any $a \in A \setminus E$.*

3 Defeat and Defense

Defense and defeat are fundamental cornerstones in abstract argumentation and before generalizing these notions, we shall reflect about which basic principles we want to modify, and which we want to preserve.

The first observation we are going to make is that the notion of "defense" is actually based on the attack relation, i.e. a set E of arguments defends some $a \in A$ if $E \to b$ for each b s.t. $b \to a$. Since Dung AFs represent attacks between arguments, it is natural to stipulate that an argument a is acceptable by default, and for each attacker b there needs to be some reason to reject b, i.e. arguments attacking a represent threats for the acceptance of it and hence need to be taken care of. We therefore want to preserve the basic idea that an argument is *defended* if each attacker is countered.

So we observe that if "defense" means each attacking argument is defeated, it suffices to adjust the notion of defeat. For Dung's classical semantics, the notion of "E defeats a" is defined as $\exists b \in E : b \to a$. So defeat is merely syntactical (and existential quantified); a choice which is natural from an intuitive point of view, but technically speaking rather arbitrary. The aforementioned choice also ties together the notion of "conflict" (there is an attack between two arguments) and "defeat" (an extension defeats an argument).

Let us separate these notions by introducing a general concept of refutal. Given an AF $F = (A, R)$, we expect a refute operator to assign to a set E of arguments the subset of A which is refuted by E. We do not impose any further restriction and therefore end up with the following.

Definition 3.1. *Let $F = (A, R)$ be an AF. A* refute operator *is a mapping $\delta : 2^A \to 2^A$; E δ-refutes a if $a \in \delta(E)$.*

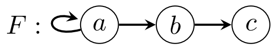
If the underlying AF is not clear from context, we denote the refute operator by $\delta(F, E)$ to emphasize F. If δ and δ' are two refute operators we write $\delta \subseteq \delta'$ if $\delta(E) \subseteq \delta'(E)$ holds for each set E of arguments in each AF $F = (A, R)$.

Definition 3.2 (Classical Refutal)**.** *Let $F = (A, R)$ be an AF. The classical refute operator is $\delta_c(E) = E^+$.*

Let us introduce a novel refute operator as well. For this, suppose we want it to behave as close as possible to the classical δ_c, but disregard self-attacking arguments.

Definition 3.3. *Let $F = (A, R)$ be an AF. For $E \subseteq A$ we define δ_{self} as $\delta_{self}(E) = E^+ \cup \{a \in A \mid a \to a\}$.*

Example 3.4. *Consider the following AF F and $E = \{b\}$.*



Then $\delta_c(E) = \{c\}$, whereas $\delta_{self}(E) = \{a, c\}$. For $E' = \{a, b\}$ we get $\delta_c(E') = \delta_{self}(E') = \{a, b, c\}$.

Having established a general notion of refutal, let us now define conflict-freeness. As mentioned before, the intuitive idea of our proposal is to separate the notions of "conflict' and "defeat". We interpret $a \to E$ resp. $E \to a$ as "conflict between E and a" and $a \in \delta(E)$ as "E refutes a". For a set E to be conflict-free, we insist both notions are respected.

Definition 3.5. *Let $F = (A, R)$ be an AF and δ a refute operator. A set E of arguments is δ-conflict free ($E \in cf_\delta(F)$) iff i) $E \in cf(F)$ and ii) $E \cap \delta(E) = \emptyset$.*

Example 3.6. *Consider again our AF from Example 3.4. The only δ_c-conflict free sets are $cf_{\delta_c}(F) = \{\emptyset, \{b\}, \{c\}\}$. Since δ_c and δ_{self} only disagree regarding self-attacking arguments, we infer $cf_{\delta_{self}}(F) = cf_{\delta_c}(F)$.*

Let us collect some basic properties of cf_δ.

Proposition 3.7. *For any AF F and refute operator δ, i) $\emptyset \in cf_\delta(F)$, ii) $cf_{\delta_c}(F) = cf(F)$, and iii) if δ is monotonic, then $E' \notin cf_\delta(F)$ if there is $E \subseteq E'$ with $E \notin cf_\delta(F)$.*

For our definition of δ-defense we view each argument a with $a \to E$ as a potential threat to our extension. In order to take care of this threat, E must refute a, i.e. $a \in \delta(E)$. As for usual defense, this condition must hold for each threat.

Definition 3.8. *For an AF $F = (A, R)$ and refute operator δ, $E \subseteq A$ δ-defends $a \in A$ iff $b \in \delta(E)$ for each $b \in a^-$.*

Example 3.9. *Returning to our Example 3.4, the set $E = \{b\}$ δ_{self}-defends itself because $a \in \delta_{self}(E)$, while it does not δ_c-defend itself since $a \notin \delta_c(E) = E^+$. Again, both notions agree that $E' = \{a, b\}$ defends itself.*

Based on our generalized version of defense, let us define the natural induced characteristic function as well.

Definition 3.10. *Let $F = (A, R)$ be an AF and δ a refute operator. The δ-characteristic function Γ_δ is the mapping $\Gamma_\delta : 2^A \to 2^A$ with $\Gamma_\delta(E) = \{a \in A \mid E\ \delta\text{-defends } a\}$.*

Note $\Gamma(E) = \Gamma_{\delta_c}(E)$ for the classical refute operator δ_c. The characteristic function is a fundamental concept in AFs and therefore requires a suitable generalization for our setting. However, as the following example illustrates we need to carefully assess the behavior we expect.

Example 3.11. *Recall our AF F from Example 3.4. The set $E = \{b\}$ δ_{self}-defends the self-attacker a: We already saw $\delta_{self}(E) = \{a, c\}$. Since $a^- = \{a\}$ we get $a^- \subseteq \delta_{self}(E)$ and thus $a \in \Gamma_{\delta_{self}}(E)$. In contrast, $a \notin \Gamma(E)$.*

We therefore anticipate that a generalized version of complete semantics would insist that E cannot be an extension without a if δ_{self} is the considered refute operator. That is, we would be forced to include a self-attacking argument in our extension. The problematic behavior stems from the fact that a conflict-free set E δ_{self}-defends an argument which is also refuted by E. For Dung-semantics this cannot happen.

Proposition 3.12. *Let $F = (A, R)$ be an AF. If $E \in cf(F)$, then $\Gamma(E) = \Gamma(E) \setminus E^+$. That is, for the classical refute operator δ_c we have $\Gamma_{\delta_c}(E) = \Gamma_{\delta_c}(E) \setminus \delta_c(E)$.*

This property is a cornerstone in Dung's theory as it is the main ingredient for the proof of his fundamental lemma. However, as we already saw for the quite simple refute operator δ_{self}, we cannot guarantee this behavior in general. We will thus refine our δ-characteristic function before defining the semantics. Later in Section 5.3 we will revisit this requirement and investigate in which cases it is necessary and under which conditions we can drop it.

Definition 3.13. *Let $F = (A, R)$ be an AF and δ a refute operator. The polished δ-characteristic function χ_δ is the mapping $\chi_\delta : 2^A \to 2^A$ with $\chi_\delta(E) = \Gamma_\delta(E) \setminus \delta(E)$.*

Example 3.14. *In Example 3.4 we have $\chi_{\delta_{self}}(\{b\}) = \{b\}$.*

We are ready to define the generalized versions of Dung's semantics from the paper (Dung 1995).

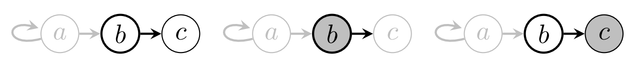
Definition 3.15. *Let F be an AF and δ a refute operator.*

1. *$E \in ad_\delta(F)$ iff $E \subseteq \chi_\delta(E)$,*
2. *$E \in co_\delta(F)$ iff $E = \chi_\delta(E)$,*
3. *$E \in pr_\delta(F)$ iff E is $\subseteq$-maximal in $co_\delta(F)$,*
4. *$E \in gr_\delta(F)$ iff E is $\subseteq$-minimal in $co_\delta(F)$,*
5. *$E \in stb_\delta(F)$ iff $E \in cf_\delta(F)$ and $E \cup \delta(E) = A(F)$.*

The attentive reader may have observed that we did not impose $E \in cf_\delta(F)$ as a precondition. Indeed, using the polished characteristic function this property is implicit.

Proposition 3.16. *For any AF F and refute operator E we have that $E \in ad_\delta(F)$ implies $E \in cf_\delta(F)$.*

Example 3.17. *For the AF F from before we already found $cf_{\delta_{self}}(F) = \{\emptyset, \{b\}, \{c\}\}$. The three sets and their refuted arguments are as follows.*



$\chi_{\delta_{self}}(\emptyset) = \{b\}$ $\quad$ $\chi_{\delta_{self}}(\{b\}) = \{b\}$ $\quad$ $\chi_{\delta_{self}}(\{c\}) = \{b\}$

We have $b \notin \delta_{self}(\{c\})$ and hence $\{c\} \notin ad_{\delta_{self}}(F)$. Moreover, $\emptyset \in ad_{\delta_{self}}(F)$ is easy to see, but $\chi_{\delta_{self}}(\emptyset) = \{b\}$ yields $\emptyset \notin co_{\delta_{self}}(F)$. From the previous example we recall $\chi_{\delta_{self}}(\{b\}) = \{b\}$ and hence $\{b\} \in co_{\delta_{self}}(F)$. This also implies $\{b\} \in pr_{\delta_{self}}(F) = gr_{\delta_{self}}(F) = co_{\delta_{self}}(F)$. From $\delta_{self}(\{b\}) = \{a, c\} = A \setminus \{b\}$ we also infer that $\{b\} \in stb_{\delta_{self}}(F)$; there is no other δ-stable extension.

Using χ_δ is not necessary to define ad_δ as we show next. For the general version of admissibility, we can always use Γ_δ instead. The issue described in Example 3.18 which gave rise to define the polished version χ_δ only applies to complete-based semantics.

Proposition 3.18. *Given an AF F and a refute operator δ, we have that $E \in ad_\delta(F)$ iff $E \in cf_\delta(F)$ and $E \subseteq \Gamma_\delta(E)$.*

Let us finish this section by formalizing that the semantics from Definition 3.15 faithfully generalize Dung's semantics.

Proposition 3.19. *For $\delta = \delta_c$ we have $\sigma_\delta(F) = \sigma(F)$ for each AF F and $\sigma \in \{ad, co, gr, pr, stb\}$.*

4 Defense Notions of Non-classical Semantics

Before delving into various formal properties of our new notions, let us demonstrate how our approach captures several non-classical semantics in a quite natural way. In this section, we consider three different families of semantics based on proposals from the literature. Since we do not intend to discriminate based on any specific order, we will simply proceed chronologically.

The aim of this section is to provide the reader with a selection of examples of possible refute operators capturing recently proposed non-classical semantics from the literature. The subsequent sections are designed in a way that a full understanding of these examples is not necessary in order to follow the remainder of the paper.

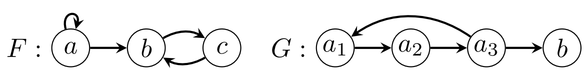
4.1 Cogent Semantics

Bodanza and Tohmé presented alternative semantics with the goal to limit the effect of self-defeating arguments (Bodanza and Tohmé 2009). In a nutshell, the idea is that defense is not required against each argument in the given AF, but only against ones which are serious enough in a certain sense. The latter concept is implemented as follows.

Definition 4.1. *Let $F = (A, R)$ be an AF. A set $E \subseteq A$ is called cogent ($E \in cog(F)$) iff for any X the following holds: If $X \in ad(F{\downarrow}_{X\cup E})$, then $E \in ad(F{\downarrow}_{X\cup E})$.*

Then *sustainable* extensions are defined as maximal cogent, i.e. $E \in sus(F)$ iff E is maximal in $cog(F)$.

Example 4.2. *Consider the following AFs F and G*



We verify that $E = \{b\} \in cog(F)$. Observe that $X \in ad(F{\downarrow}_{X\cup E})$ can never be the case for some set X containing the self-attacker a. Thus, the only conflict-free candidates for X are $\emptyset$, $\{b\}$, and $\{c\}$; in all cases, $E \in ad(F{\downarrow}_{X\cup E})$ holds. Similarly, we find $\{c\} \in cog(F)$. Since the emptyset is also cogent, $cog(F) = \{\emptyset, \{b\}, \{c\}\}$. Consider now G. In this case, $E = \{b\} \notin cog(G)$. The reason is that we can set $X = \{a_3\}$ and in the subframework $F{\downarrow}_{X\cup E}$ the argument b cannot counterattack a_3. A similar reasoning can be applied to any argument in G and in summary we obtain $cog(G) = \{\emptyset\}$.

It turns out that cogency is closely related to classical admissibility. Formally, the only difference is that self-attacking arguments are disregarded since they can never be in any cogent set. Consequently, cogent semantics can be captured by the refute operator δ_{self} we already introduced.

Definition 4.3. *Let $F = (A, R)$ be an AF. The* cogency *refute operator δ_{cog} is given as $\delta_{cog} = \delta_{self}$.*

The following theorem formalizes that δ_{self} is the suitable refute operator to capture the cogent semantics family.

Theorem 4.4. *For any AF F we have $E \in cog(F)$ iff $E \in ad_{\delta_{self}}(F)$, $E \in sus(F)$ iff $E \in pr_{\delta_{self}}(F)$.*

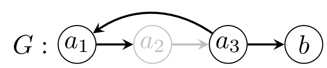
Thus, we have successfully extracted the underlying defense notion of cogent semantics. As a by-product, we obtain natural notions of "complete" and "grounded" semantics by considering $co_{\delta_{self}}$ and $gr_{\delta_{self}}$, respectively.

4.2 Undecidedness Blocking

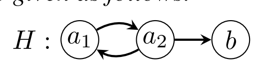
The paper (Dondio and Longo 2021) proposes a weakened version of admissibility which is based on the following idea of "undecidedness blocking": Suppose we are given an argument a which is undecided w.r.t. some admissible extension E in the sense that $a \notin E^{\oplus}$. Then, each argument a' attacked by a cannot be accepted, i.e. $a' \notin E$. If even $a' \notin E^{+}$ holds, then a' is also undecided. This may lead to chains of undecided arguments of an arbitrary length. Dondio and Longo therefore propose an approach where the effect of undecided arguments is limited. The original definition is given in terms of labelings. However, we can find a convenient equivalent characterization via extensions.

Definition 4.5. *A set $E \subseteq A$ is called weakly ub-complete ($E \in co^{wu}(F)$) if $E \in cf(F)$ and $\Gamma(E) \subseteq E$.*

Example 4.6. *Recall our two AFs form Example 4.2. We find $co^{wu}(F) = \{\emptyset, \{b\}, \{c\}\}$ so one might wonder whether the definition of co^{wu} yields some kind of "defense" notion. The answer is affirmative, but the underlying notion is quite subtle which is nicely illustrated in G: Here, $\{b\} \in co^{wu}(G)$ as well since b clearly does not defend either of the a_i. However, consider the situation when choosing e.g. a_1:*



We have $\Gamma(\{a_1\}) = \{a_3\}$ so it does not hold that $\Gamma(\{a_1\}) \subseteq \{a_1\}$. However, simply iterating Γ does not help to fix this issue since $\{a_1, a_3, b\}$ is not conflict-free in G. Indeed, $co^{wu}(G) = \{\emptyset, \{b\}\}$, i.e. the arguments in the odd cycle cannot be chosen. We want to mention though that the acceptability of b does not depend on a_1, a_2, a_3 forming an odd cycle. Consider H given as follows.



Here $\{b\} \in co^{wu}(H)$ as well since $\{b\}$ does neither defend a_1 nor a_2. In summary, $co^{wu}(H) = \{\emptyset, \{b\}, \{a_1\}, \{a_2\}\}$.

The behavior of the previous example generalizes to the following idea: Whenever E defends some set E' which in turn attacks E, then $E \notin co^{wu}(F)$. The same is true if E' defends some E'' attacking E and so on. A concise way to formalize this observation is the reduct $F^{E} = F{\downarrow}_{A\setminus E^{\oplus}}$.

Proposition 4.7. *Let $F = (A, R)$ be an AF and $E \in cf(F)$. There is some $E' \in co^{wu}(F)$ with $E \subseteq E'$ iff $G \in gr(F^{E})$ does not attack E in F.*

We can consequently think of a "defense" notion which is based on the grounded extension of F^{E}. Indeed, we can capture the semantics co^{ub} as follows.

Definition 4.8. *For an $F = (A, R)$ and $E \subseteq A$ define the* wu *operator as $\delta_{wu}(E) = \{a \in A \mid a \notin (gr(F^{E}) \cup E)\}$.*

Theorem 4.9. *Let $F = (A, R)$ be an AF and $E \subseteq A$. Then $E \in co^{wu}(F)$ iff $E \in co_{\delta_{wu}}(F)$.*

We hence extracted the underlying defense notion of the undecidedness blocking principle. Versions of "admissible", "grounded", and "preferred" semantics can be obtained by considering $ad_{\delta_{wu}}$, $gr_{\delta_{wu}}$, and $pr_{\delta_{wu}}$, respectively.

4.3 Recursive "Weak Admissibility"

Weak admissibility as proposed in (Baumann, Brewka, and Ulbricht 2020b) is based on a recursive definition utilizing the reduct F^E. The intuition is that arguments in E are accepted, those in E^+ are rejected and out of those in $A \setminus E^{\oplus}$, attackers y of E are of concern. In order to assess such a y as "serious" enough to be a threat, it has to be weakly admissible in the reduct F^E, leading to the following definition.

Definition 4.10. *Let $F = (A, R)$ be an AF. $E \subseteq A$ is called* weakly admissible *in F ($E \in ad^w(F)$) iff*

1. *$E \in cf(F)$, and*
2. *for any attacker y of E we have $y \notin \bigcup ad^w(F^E)$.*

Example 4.11. *Let us discuss our examples from the previous subsections. Regarding F, for both $E = \{b\}$ and $E' = \{c\}$ it holds that the corresponding reduct F^E and $F^{E'}$ contains only the self-attacker a. Since self-attacking arguments can never be weakly admissible, $ad^w(F) = \{\emptyset, \{b\}, \{c\}\}$ follows. Observe that in an AF consisting of an odd cycle a_1, a_2, a_3, no argument occurs in a weakly admissible extension. It follows that $ad^w(G) = \{\emptyset, \{b\}\}$. Since $\{a_1\}, \{a_2\} \in ad^w(H)$, b cannot survive without a_1, so we conclude $ad^w(H) = \{\emptyset, \{a_1\}, \{a_2\}, \{a_1, b\}\}$.*

The decisive feature of weak admissibility is that arguments which are not acceptable in the reduct F^E are disregarded. The natural refute operator hence is the following.

Definition 4.12. *Let $F = (A, R)$ be an AF. The* weak refute operator *δ_w is for each set E of arguments given as $\delta_w(E) = E^+ \cup \{a \in F^E \mid a \notin \bigcup ad^w(F^E)\}$.*

Intuitively, an argument a is δ_w-refuted by E iff it is either attacked by E or occurring in the reduct without being credulously accepted there. We formalize that δ_w is suitable for our purpose as it captures weakly admissible semantics.

Theorem 4.13. *Let $F = (A, R)$ be an AF and $E \subseteq A$. Then $E \in ad^w(F)$ iff $E \in ad_{\delta_w}(F)$.*

The paper also presents a notion of defense and the induced weak complete, preferred, and grounded semantics.

Definition 4.14. *Let $F = (A, R)$ be an AF. Given two sets $E, X \subseteq A$. We say E weakly defends (or simply, w-defends) X iff for any attacker y of X we have,*

1. *E attacks y, or*
2. *$y \notin E$, $y \notin \bigcup ad^w(F^E)$ and $X \subseteq X' \in ad^w(F)$.*

Definition 4.15. *Let $F = (A, R)$ be an AF. A set $E \subseteq A$ is called* weakly complete *in F ($E \in co^w(F)$) iff $E \in ad^w(F)$ and if E w-defends $X \supseteq E$, then $X = E$.*

Weakly grounded (gr^w) and preferred (pr^w) are defined as minimal resp. maximal weakly complete sets. In the paper (Dauphin, Rienstra, and van der Torre 2021) the notion of weak defense has been further investigated and interesting alternatives are proposed, but within the scope of this contribution we stick with Definition 4.15. The following result establishes a natural fixed-point characterization for co^w.

Theorem 4.16. *Let $F = (A, R)$ be an AF and $E \subseteq A$. Then $E \in co^w(F)$ iff $E \in co_{\delta_w}(F)$.*

Corollary 4.17. *Let F be an AF. Then i) $E \in pr^w(F)$ iff $E \in pr_{\delta_w}(F)$ and ii) $E \in gr^w(F)$ iff $E \in gr_{\delta_w}(F)$.*

5 Basic Properties

In this section we examine basic properties of semantics based on general refute operators. This investigation is inspired by the behavior of the classical semantics, i.e. we particularly focus on the following questions.

- Is there a suitable version of the Fundamental Lemma?
- Under which conditions do complete extensions exist?
- Can we infer that the grounded extension is unique?

Our investigation requires caution since the classical refute operator $\delta_c(E) = E^+$ is a simple syntactical one with intuitive properties. Moving to a general mapping δ is thus a challenging endeavour at first glance. However, it turns out that under some mild conditions we can answer the above questions affirmatively for a general refute operator δ. First, let us mention that $\emptyset$ is admissible for any operator.

Lemma 5.1. *Let $F = (A, R)$ be an AF and δ a refute operator. Then $\emptyset \in ad_\delta(F)$. In particular, $ad_\delta(F) \neq \emptyset$.*

5.1 The Fundamental Lemma

The above Lemma 5.1 is simple, yet our first major building block in finding conditions ensuring $co_\delta(F)$ exists. Dung's classical proof for the existence of complete extensions starts with some admissible set and then applies:

Lemma 5.2 (Fundamental lemma)**.** *Let $F = (A, R)$ be an AF and $E \in ad(F)$. If E defends a, then $E \cup \{a\} \in ad(F)$.*

Unfortunately, we cannot expect this property to hold for any conceivable refute operator. Our goal is thus to find conditions ensuring this property. However, before doing so let us establish that a natural counterpart to the fundamental lemma indeed yields the existence of complete extensions.

Definition 5.3. *A refute operator* induces the fundamental lemma *if for any AF $F = (A, R)$ and $E \in ad_\delta(F)$ the following holds: If $a \in \chi_\delta(E)$, then $E \cup \{a\} \in ad_\delta(F)$.*

Theorem 5.4. *Let $F = (A, R)$ be an AF. If δ induces the fundamental lemma, then $co_\delta(F) \neq \emptyset$.*

Proof. We have $\emptyset \subseteq \chi_\delta(\emptyset)$. If even "=" holds, we are done. So let $E_1 = \{a_1\}$ for some $a \in \chi_\delta(\emptyset)$ and for each $i \geq 2$ we set $E_i = E_i \cup \{a_i\}$ for some $a_i \in \chi_\delta(E_{i-1}) \setminus E_{i-1}$. Since δ induces the fundamental lemma, a simple induction shows that $E_i \cup \{a_i\} \in ad_\delta(F)$ for each i and $a_i \notin E_{i-1}$, i.e. $E_{i-1} \subsetneq E_i$. By finiteness of A, this procedure stops after at most $|A|$ steps with some E_i satisfying $E_i = \chi_\delta(E_i)$. ☐

Our goal is thus to find natural conditions ensuring that our given refute operator δ induces the fundamental lemma. We show that this can be ensured by the following two conditions (which are both satisfied by δ_c):

Definition 5.5. *We say χ_δ satisfies*

- cf-monotonicity *if for any AF $F = (A, R)$ and $E, E' \in cf_\delta(F)$ we have that $E \subseteq E'$ implies $\chi_\delta(E) \subseteq \chi_\delta(E')$;*
- conflict-free transfer *if for any AF $F = (A, R)$ and $E \in ad_\delta(F)$ we have $E \cup \{a\} \in cf_\delta(F)$ for each $a \in \chi_\delta(E)$.*

We can show that in general, this two properties indeed suffice in order to infer the fundamental lemma.

Proposition 5.6. *Let F be an AF and δ a refute operator. If χ_δ satisfies both cf-monotonicity and conflict-free transfer, then δ induces the fundamental lemma.*

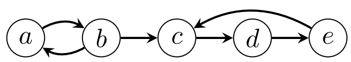
Proof. Let $E \in ad_\delta(F)$ and $a \in \chi_\delta(E)$. By conflict-free transfer, $E \cup \{a\} \in cf_\delta(F)$. Since $E \in ad_\delta(F)$ we can apply cf-monotonicity and find $\chi_\delta(E) \subseteq \chi_\delta(E \cup \{a\})$. By definition of $E \in ad_\delta(F)$, we have $E \subseteq \chi_\delta(E)$. From $a \in \chi_\delta(E)$ we can therefore also infer $E \cup \{a\} \subseteq \chi_\delta(E)$. □

As we show next, Proposition 5.6 is applicable to δ_{cog}. Although this is not the case for δ_{wu}, this refute operator induces the fundamental lemma nonetheless.

Proposition 5.7. *The refute operator δ_{self} satisfies both cf-monotonicity and conflict-free transfer, δ_{wu} induces the fundamental lemma but does not satisfy cf-monotonicity.*

In contrast, the following shows that δ_w does not induce the fundamental lemma and thus by Proposition 5.6 cannot satisfy both cf-monotonicity and conflict-free transfer.

Example 5.8. *In the AF F below $\{d\}$ is δ_w-defended by $\emptyset$ but $\{d\} \notin ad_{\delta_w}(F)$.*

I :


Nonetheless $co_{\delta_w}(F) \neq \emptyset$ for any AF F, so the fundamental lemma is no necessary condition.

5.2 Grounded and Preferred Extensions

Next we want to give a criterion ensuring that gr_δ is always unique. As before our inspiration comes from the behavior of the classical semantics. Notably, the following property is a natural counterpart to conflict-free transfer.

Definition 5.9. *Let δ a refute operator. The polished characteristic function χ_δ satisfies* admissibility transfer *if for any AF F and $E \in ad_\delta(F)$ we have $\chi_\delta(E) \in ad_\delta(F)$.*

Theorem 5.10. *Let $F = (A, R)$ be an AF and δ a refute operator. If χ_δ satisfies cf-monotonicity and admissibility transfer, then $|gr_\delta(F)| = 1$ and the unique δ-grounded extension $G \in gr_\delta(F)$ satisfies $G = \bigcup_{i\in\mathbb{N}} \chi^i_\delta(\emptyset)$.*

Proof. Let $E \in co_\delta(F)$. By cf-monotonicity we have $\chi_\delta(\emptyset) \subseteq \chi_\delta(E)$. Now by induction and due to admissibility transfer we infer $\chi^i_\delta(\emptyset) \subseteq \chi^i_\delta(E)$ for each integer i. For some i, $\chi^i_\delta(\emptyset) = \chi^{i+1}_\delta(\emptyset)$, i.e. G is complete Since $E \in co_\delta(F)$ we have $E = \chi^i_\delta(E)$ and thus, $G \subseteq E$. □

Theorem 5.10 is applicable to both δ_{cog} and δ_{wu}.

Proposition 5.11. *The refute operators δ_{self} and δ_{wu} satisfy admissibility transfer.*

Since gr_{δ_w} is not unique in general (Baumann, Brewka, and Ulbricht 2020b) we can be certain that δ_w does not satisfy both cf-monotonicity and admissibility transfer. The final property of classical semantics we inspect more closely here is regarding preferred semantics: They are defined as maximal complete extensions, but this is equivalent to maximal admissible sets. Remarkably, the same properties ensuring uniqueness of gr also yield this equivalence. First consider an auxiliary lemma which establishes the technical property we require. Then we infer the desired property.

Lemma 5.12. *Let $F = (A, R)$ be an AF and δ a refute operator. Let $E \in ad_\delta(F)$. If the polished characteristic function χ_δ satisfies both cf-monotonicity and admissibility transfer, then there is some $E' \in co_\delta(F)$ with $E \subseteq E'$.*

Proposition 5.13. *Let $F = (A, R)$ be an AF and δ a refute operator. If the polished characteristic function χ_δ satisfies both cf-monotonicity and admissibility transfer, then $pr_\delta(F) = \{E \subseteq A \mid E$ is maximal in $ad_\delta(F)\}$.*

Proof. For ($\subseteq$) let $E \in co_\delta(F)$. If E is not maximal in $ad_\delta(F)$, then there is some $E' \in ad_\delta(F)$ with $E \subsetneq E'$. By Lemma 5.12, there is $E'' \in co_\delta(F)$ with $E \subsetneq E' \subseteq E''$, contradiction. The other direction ($\supseteq$) is analogous. □

Observe that we can thus infer the following result which, to the best of our knowledge, has not been stated explicitly in the original papers where the semantics have been proposed.

Corollary 5.14. *For any AF F we have $pr_\delta(F) = \{E \subseteq A \mid E$ is maximal in $ad_\delta(F)\}$ for $\delta \in \{\delta_{self}, \delta_{wu}\}$.*

Proposition 5.13 is not applicable to δ_w, but due to (Baumann, Brewka, and Ulbricht 2020b), $pr^w(F) = \{E \subseteq A \mid E$ is maximal in $ad^w(F)\}$ holds nonetheless.

5.3 Polished Refute Operators

We recall that our semantics based on a refute operator δ were defined in terms of the polished characteristic function $\chi_\delta(E) = \Gamma_\delta(E) \setminus \delta(E)$. Since $\chi_\delta(E)$ and $\Gamma_\delta(E)$ coincide for Dung's semantics whenever E is conflict-free (see Proposition 3.12), one could argue that χ_δ is a faithful generalization of the classical characteristic function on its own. Nonetheless, let us investigate situations where $\Gamma_\delta = \chi_\delta$.

Definition 5.15. *We call δ* polished *if for each AF $F = (A, R)$ we have $\chi_\delta(E) = \Gamma_\delta(E)$ for each $E \in cf_\delta(F)$.*

We can give the following general negative result. Let us stipulate the notation $A^\circ = \{a \in A \mid a \to a\}$.

Proposition 5.16. *Let $F = (A, R)$ be an AF. A refute operator δ s.t. $A^\circ \subseteq \delta(E)$ for each $E \subseteq A$ is not polished.*

Proof. Let $F = (\{a\}, \{(a, a)\})$. By assumption $\delta(\emptyset) = \{a\}$. However, this means $\emptyset$ refutes each attacker of a which implies $a \in \Gamma_\delta(\emptyset)$ as well. Thus, $\chi_\delta(\emptyset) = \emptyset \subsetneq \{a\} = \Gamma_\delta(\emptyset)$. Since $\emptyset \in cf_\delta(F)$, δ is not polished. □

As we will later see when considering the *self attack refute* principle, δ_{self} and δ_w satisfy $A^\circ \subseteq \delta(E)$ for each $E \subseteq A$ in any AF F (see Table 1). We hence infer:

Corollary 5.17. *Both δ_{self} and δ_w are not polished.*

Recall that the semantics based on δ_{self}, δ_{wu}, and δ_w are tailored to limit the impact self-defeating arguments have on the reasoning in F. We thus assume a setting where it is desired to disregard self-attackers. Hence, we want to emphasize that the above corollary is not a flaw of the refute operators. Rather, Proposition 5.16 suggests that refute operators serving this purpose will in general not tend to be polished. As the last observation we make in this section we will show that, in contrast, a refute operator which is stricter than Dung's classical refute will always be polished.

Proposition 5.18. *Let $F = (A, R)$ be an AF and δ a refute operator. If $\delta(E) \subseteq E^+$ for each $E \subseteq A$, then δ is polished.*

	δ_c	δ_{self}	δ_{wu}	δ_w
Induces fundamental lemma	✓	✓	✓	✗
cf-monotonicity	✓	✓	✗	✗
Conflict-free transfer	✓	✓	✓	✓
Admissibility transfer	✓	✓	✓	✗
Polished	✓	✗	✗	✗
Requiring attacks	✓	✓	✓	✓
Refute possible	✓	✓	✓	✓
Monotonicity	✓	✓	✗	✗
Self-attack refutal	✗	✓	✗	✓
Self-attack neglection	✗	✓	✗	✓
Additivity	✓	✓	✗	✗
Context-free	✓	✗	✗	✗
Generalize classical defense	✓	✓	✓	✓
Tractable	✓	✓	✓	✗

Table 1: Refute Operator Principles

6 Principles and Their Relationships

Within the last decade, AF semantics have been extensively studied w.r.t. certain *principles*, i.e. properties which are satisfied (or not) by certain semantics. This line of research helps establishing a comprehensive understanding of their behavior. In view of this, the present paper gives rise to approaching this endeavor in terms of the underlying refute operator, i.e. this section is driven by the following research questions.

- Can we investigate principles which are satisfied by semantics by considering the underlying refute operator?
- What are the principles a refute operator should satisfy?

As we already mentioned, the non-classical semantics considered in this paper all serve the purpose of minimizing the damage caused by self-defeating arguments. Inspired by this, we also want to consider some specific principles devised to compare semantics of this kind, i.e. we study:

- What is a “good” weak version of admissibility?

Throughout this section we assume an arbitrary but fixed AF $F = (A, R)$. The attentive reader will observe that we only discuss a small selection of the usual principles mentioned in the literature. This choice is made due to space restrictions; a comprehensive study of this matter is left for future work. A summary of all principles and which refute operators satisfy them is reported in Table 1.

6.1 Defense and Defeat

We start by some basic properties of refute and defense operators. Our first principles establish some basic ideas of refute operators we intuitively expect. For classical refute $\delta_c(E) = E^+$ these are all clear, but since in our general setting δ is an arbitrary mapping, we want to mention them here in our general investigation.

Principle 6.1 (Requiring Attacks)**.** *A refute operator δ requires attacks if $a^- = \emptyset$ implies $a \notin \delta(E)$ for each $E \subseteq A$.*

Principle 6.2 (Refute Possible)**.** *A refute operator δ satisfies the* refute possible *principle if for each argument $a \in \mathcal{U}$ there is some AF $F = (A, R)$ and $E \subseteq A$ s.t. $a \in \delta(E)$.*

So these principles simply formalize that refutal of an argument should in general be possible, but not without presence of another argument questioning it. We want to emphasize that requiring attacks does not distinguish whether some attack stems from E, i.e. it is possible for a refute operator to satisfy it, but we still have $a^- \cap E = \emptyset$ for an argument $a \in \delta(E)$. Full absence of any attack within the given AF however renders δ trivial.

Fact 6.3. *If $R = \emptyset$, i.e. $F = (A, \emptyset)$, and δ requires attacks, then $\delta(E) = \emptyset$ for each $E \subseteq A$. Moreover, $gr_\delta(F) = \{A\}$.*

Next we consider monotonicity of δ. It is clear that this is an important property, but as we saw already in the last sections, not every reasonable operator satisfies it.

Principle 6.4 (Monotonicity)**.** *A refute operator δ satisfies* monotonicity *if $E \subseteq E' \subseteq A$ implies $\delta(E) \subseteq \delta(E')$.*

The next principles are concerned with self-attackers and how they influence the refuted sets of arguments. This aspect is interesting for applications where the presence of such arguments shall not influence the outcome of a discussion, as is the case for our examples from Section 4.

Principle 6.5 (Self-Attack Refute)**.** *A refute operator δ satisfies* self-attack refutal *if $A^\circ \subseteq \delta(E)$ for each E.*

Let us also consider an even stronger version of this principles formalizing that self-attackers do not have any influence on the attacked arguments whatsoever (at least if E itself does not contain a self-attacker). For this, we let F° be the AF after removing self-attackers, i.e. $F^\circ = F{\downarrow}_{A \setminus A^\circ}$.

Principle 6.6 (Self-Attack Neglection)**.** *A refute operator δ satisfies* self-attack neglection *if for each $E \subseteq A(F^\circ)$, $a \in A$ it holds that $a \in \delta(F, E)$ iff $(a, a) \in R$ or $a \in \delta(F^\circ, E)$.*

Observe that for refute operators which behave in a certain sense intuitively, we would expect that self-attack refutal implies self-attack neglection. For this, suppose E does not contain any self-attacker. Then any self-attacker a is refuted due to self-attack refutal. For arguments which are no self-attackers, the condition $a \in \delta(F, E)$ iff $a \in \delta(F^\circ, E)$ simply implies that the present self-attackers do not influence whether or not E refutes a. Since E does not contain any self-attacker, we would expect this behavior.

More broadly, the following principles consider the interaction of arguments with each other in E as well as the sensitivity to the remaining AF.

Principle 6.7 (Additivity)**.** *A refute operator δ satisfies* additivity *if $\delta(E) = \bigcup_{e \in E} \delta(\{e\})$ for each $\emptyset \neq E \subseteq A$.*

Principle 6.8 (Context-free)**.** *A refute operator δ is* context-free *if $\delta(F, E) = \delta(F', E)$ where $F' = (A, R')$ is an AF satisfying $(a, b) \in R$ iff $(a, b) \in R'$ for each $a \in E$.*

The latter condition formalizes that E has the same outgoing attacks in F'. For example, classical refute based on out-going attacks satisfies both additivity and context-free since $E^+ = \bigcup_{e \in E} \{e\}^+$ and E^+ depends solely on E and its out-going attacks. On the other hand, we can formalize that under mild conditions a refute operator which disregards self-attacking arguments cannot be context-free.

Proposition 6.9. *There is no refute operator satisfying requiring attacks, self-attack refutal, and context-free.*

With the addition of the additivity principle one might now expect that there should be some quite natural characterization of the classical refute operator $\delta_c(E) = E^+$. However, it is still surprisingly difficult to capture δ_c in terms of our abstract properties. At least for semantics which are more liberal than ad – i.e. we have that $\bigcup ad(F) \subseteq \bigcup ad_\delta(F)$ – we can achieve this as follows.

Proposition 6.10. *Let δ be a refute operator satisfying requiring attacks, context-free, and additivity. If $\bigcup ad(F) \subseteq \bigcup ad_\delta(F)$, then $\delta(E) = E^+$ for each set E of arguments satisfying $E \cap A^\circ = \emptyset$.*

For our final principle regarding refute operators, let us have a closer look at the notion of defense. The following relates χ_δ to the classical operator Γ, formalizing that χ_δ contains at least the classically defended arguments.

Principle 6.11. *A refute operator δ* generalizes classical defense *if for each AF $F = (A, R)$ and $E \in cf(F)$ it holds that $\Gamma(E) \subseteq \chi_\delta(E)$.*

In particular refute operators designed to yield more liberal versions of ad –like the ones we consider in the present paper– should adhere to this property since it would not make sense for these to credulously accept fewer arguments than Γ. To conclude our discussion on principles for refute operators we turn our attention to Table 1 summarizing which operators satisfy which of our novel principles.

Theorem 6.12. *Satisfaction of principles by the refute operators considered in this paper is as depicted in Table 1.*

6.2 Weak Admissibility Notions

In this section we discuss a selection of principles which are tailored to compare semantics devised to yield more liberal notions of defense. As is the case for our running example semantics, we will assume that the underlying motivation is to limit the effect of contradictory or self-attacking arguments whose status is undecided. We start with a simple property, stating that the given semantics is more liberal than ad which formalizes that we indeed accept more arguments.

Principle 6.13. *A semantics σ satisfies* ad-liberalization *if for each AF F, $\bigcup ad(F) \subseteq \bigcup \sigma(F)$ and there is some AF F_0 s.t. $\bigcup ad(F_0) \subsetneq \bigcup \sigma(F_0)$.*

This is a basic requirement for the semantics sharing our underlying motivation, with the exception of the "grounded" version. Next we recall a property from (Baumann, Brewka, and Ulbricht 2020a) formalizing that a semantics disregards self-attackers.

Principle 6.14. *A semantics σ satisfies* neglection of self-attackers *if for any AF F, $\sigma(F) = \sigma(F^\circ)$.*

Interestingly, the way semantics are defined based on a general refute operator allows us to infer satisfaction in an intuitive way by means of the behavior of δ.

Proposition 6.15. *If δ satisfies self-attack neglection, then σ satisfies neglection of self-attackers for each $\sigma \in \{ad_\delta, co_\delta, gr_\delta, pr_\delta, stb_\delta\}$.*

As a corollary, we obtain that the considered semantics based on δ_{self} and δ_w satisfy neglection of self-attackers.

Next we introduce a property which can be viewed as a generalization of neglection of self-attackers where we make use of unattacked arguments, i.e. a set U of arguments s.t. $(a, b) \in R$ with $b \in U$ implies $a \in U$. The underlying idea is as follows. Given some unattacked set U of arguments where $\sigma(F{\downarrow_U}) = \{\emptyset\}$. Then, as in the case of self-attackers, no argument in U is acceptable. A semantics following the intuition that arguments which can never be accepted shall not be able to cause any harm can now be expected to behave as if U was not part of the AF at all. We call this property the *separation property*.

Principle 6.16. *A semantics σ satisfies the* separation property *if for any AF F and any unattacked $U \subseteq A$ the following holds: If $\sigma(F{\downarrow_U}) = \{\emptyset\}$, then $\sigma(F) = \sigma(F{\downarrow_{A\setminus U}})$.*

The separation property can grasp, among others, the idea that an unresolved odd cycle can be removed from the AF.

Example 6.17. *Consider F, G, and H from Example 4.11. Semantics which require extensions to be conflict-free as well as satisfying the separation property treat F like the even cycle consisting of b and c. If in addition no argument in an isolated odd cycle is accepted, G is equivalent to the isolated argument b. In case of H, however, the separation property will most likely not make any difference since the even cycle a_1 and a_2 yields two accepted arguments for most semantics.*

6.3 On the Relation to Classical Principles

We conclude our discussion on principles by recalling some of the most basic ones from the literature and investigating them by means of our general framework.

Definition 6.18. *A semantics σ satisfies*

- defense *if $E \subseteq \Gamma(E)$ for each $F \in \mathcal{F}$, $E \in \sigma(F)$;*
- reinstatement *if $\Gamma(E) \subseteq E$ for each $F \in \mathcal{F}$, $E \in \sigma(F)$;*
- I-maximality *if $\sigma(F)$ forms an antichain for each $F \in \mathcal{F}$.*

We can ensure satisfaction of defense whenever δ is more cautious than δ_c.

Proposition 6.19. *Let δ be a refute operator s.t. $\delta \subseteq \delta_c$. Then for each $\sigma \in \{ad_\delta, co_\delta, pr_\delta, gr_\delta, stb_\delta\}$, the semantics σ satisfies defense.*

The reinstatement property holds for complete-based semantics whenever δ generalizes classical defense. Clearly, this cannot be ensured for δ-admissible extensions in general ad is exclusively concerned with defense of the set E.

Proposition 6.20. *Let δ be a refute operator generalizing classical defense. Then for each $\sigma \in \{co_\delta, pr_\delta, gr_\delta, stb_\delta\}$, the semantics σ satisfies reinstatement.*

Moreover, I-maximality holds by definition for any version of grounded and preferred semantics. Observe that, at least if δ is defined in a somewhat counter-intuitive way, we cannot be certain that stb_δ adheres to I-maximality.

Proposition 6.21. *Let δ be a refute operator. Then for each $\sigma \in \{pr_\delta, gr_\delta\}$, the semantics σ satisfies I-maximality.*

To concludes this section, let us consider Table 2 reporting satisfaction of principles formulated for semantics.

Theorem 6.22. *Satisfaction of principles by the semantics considered in this paper is as depicted in Table 2.*

	δ_c	δ_{self}	δ_{wu}	δ_w
Ad-liberal.	✗	ad	ad	ad
Self-attack neg	✗	✓	✗	✓
Sep prop.	✗	✗	✗	✓
Defense	✓	✗	gr	✗
Reinstatement	$\sigma \neq ad_\delta$	$\sigma \neq ad_\delta$	$\sigma \neq ad_\delta$	$\sigma \neq ad_\delta$
I-maximality	gr, pr, stb	gr, pr, stb	gr, pr	gr, pr, stb

Table 2: Semantics Principles. Here a checkmark means the property holds for each induced σ. Similarly, $\sigma \neq ad_\delta$ indicates that the property holds for all except $_\delta$. A crossmark indicates satisfaction for none of them. Gray entries are known from the literature

7 Computational Complexity

The computational complexity of reasoning in AFs is well-studied for the usual semantics (Dvorák and Dunne 2018). In this section we discuss generic membership results we can infer from the properties of the refute operator. We assume the reader to be familiar with the polynomial hierarchy.

Definition 7.1. *A refute operator is* tractable *if for each AF $F = (A, R)$ and $E \subseteq A$, the set $\delta(E)$ can be computed in polynomial time.*

If we are given a tractable refute operator, we can infer the following upper bounds for the standard reasoning tasks.

Proposition 7.2. *Let δ be a tractable refute operator. Deciding whether a set $E \subseteq A$ of arguments in an AF $F = (A, R)$ satisfies $E \in \sigma_\delta(F)$ is tractable for each $\sigma \in \{ad, co, stb\}$ and in* coNP *for $\sigma \in \{gr, pr\}$.*

Proposition 7.3. *Let δ be a tractable refute operator. Deciding whether an argument $a \in A$ is credulously accepted w.r.t. σ_δ in an AF $F = (A, R)$ is in* NP *for $\sigma \in \{ad, co, pr, stb\}$ and in Σ_2^P for $\sigma = gr$.*

Regarding skeptical reasoning, we need to check $\sigma(F) \neq \emptyset$ in addition to the actual reasoning task. So we reduce our attention to cases where this is unnecessary. From our results from Section 5 we know that refute operators inducing the fundamental lemma ensure $co_\delta(F) \neq \emptyset$.

Proposition 7.4. *For δ tractable and inducing the fundamental lemma, deciding whether $a \in A$ is skeptically accepted w.r.t. σ_δ in an AF $F = (A, R)$ is trivial for $\sigma = ad$, in* coNP *for $\sigma \in \{co, gr\}$ and in Π_2^P for $\sigma = pr$.*

The computational complexity of the grounded extension is surprisingly high in general considering that reasoning with Dung's classical variant is tractable. The latter cannot be attributed to the fact that reasoning with the minimal complete extension(s) is an easy task in general. Rather, this stems from the fact that gr can be computed by iterating the characteristic function Γ. This observation yields:

Theorem 7.5. *Let δ be a tractable refute operator s.t. χ_δ satisfies cf-monotonicity and admissibility transfer. Then the following reasoning tasks are tractable. Deciding $a \in A$ is i) credulously accepted w.r.t. gr_δ, ii) skeptically accepted w.r.t. gr_δ, iii) skeptically accepted w.r.t. co_δ.*

Let us now come to the operators considered in this paper. Clearly, the classical refute operator $\delta_c(E) = E^+$ is tractable. The same is true for δ_{cog} and δ_{wu}.

Proposition 7.6. *The operators δ_{cog} and δ_{wu} are tractable.*

The same is not true for δ_w as the complexity analysis provided in (Dvorák, Ulbricht, and Woltran 2021) shows. By applying the reduction given in this paper, we can infer that the underlying decision problem is PSPACE-complete.

Proposition 7.7. *Deciding whether two input $E, X \subseteq A$ of arguments in an input AF $F = (A, R)$ satisfy $E = \delta_w(X)$ is* PSPACE*-complete.*

8 Discussion

Generalizing the underlying defense or refute notion in AFs is conceptually not novel. A recent example is the paper (Vassiliades et al. 2021) where the authors use different kinds of attacks, some of which are capable of inducing conflicts, whereas others can contribute to defense. The paper (Yu et al. 2021) uses an adjusted notion of defense for a multi-agent setting. In (Fan and Toni 2015) so-called related admissibility is the theoretical foundation for computing dialectical explanations for acceptance of arguments.

As mentioned in the introduction, several recent proposals attempt to find a middle ground between naivity-based and admissibility-based semantics (Bodanza and Tohmé 2009; Dondio 2018; Dondio and Longo 2021; Baumann, Brewka, and Ulbricht 2020b; Dauphin, Rienstra, and van der Torre 2021), but also semantics defined by the SCC-recursive scheme (Baroni, Giacomin, and Guida 2005), in particular $cf2$ and $stage2$ (Dvořák and Gaggl 2014), can be seen as semantics of this kind. The recently introduced weak admissibility (Baumann, Brewka, and Ulbricht 2020b) is already well-studied (Baumann, Brewka, and Ulbricht 2020a; Dauphin, Rienstra, and van der Torre 2020; Dvorák, Ulbricht, and Woltran 2021) and interestingly, the approach discussed in (Dondio and Longo 2021) also contributes a principle-based comparison. The paper (Dauphin, Rienstra, and van der Torre 2020) also proposes novel SCCs-based semantics.

Our research induces several interesting future work directions. First, our Theorem 4.16 does not hold for the alternatives proposed in (Dauphin, Rienstra, and van der Torre 2021), thus finding suitable refute operators would broaden the results. The same is true for SCC-based semantics like the ones introduced in (Dauphin, Rienstra, and van der Torre 2020). As a general observation, it is probably not straightforward how our approach could capture the SCC-recursive scheme. The present paper generalizes Dung's classical semantics, but one could also consider generalized versions of e.g. ideal (Dung, Mancarella, and Toni 2007), eager (Caminada 2007), or semi-stable (Caminada 2006) semantics. Our analysis in Section 5 was focused on the existence of complete extensions, so it would be interesting to find conditions ensuring the existence of δ-stable extensions, as done in (Dung 1995) for the classical semantics. One could also propose and investigate further refute operators, in particular notions which are more cautious than admissibility, similar in spirit to strong admissibility (Caminada 2014). Finally the investigation of further principles would be conceivable, in particular generalized versions of those discussed in Section 6.3, by phrasing them in terms of χ_δ.

Acknowledgements

This work was supported by the German Research Foundation (DFG, BA 6170/2-1) and 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

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

Baroni, P.; Caminada, M.; and Giacomin, M. 2018. Abstract argumentation frameworks and their semantics. In *Handbook of Formal Argumentation*. College Publications.

Baroni, P.; Giacomin, M.; and Guida, G. 2005. SCC-recursiveness: a general schema for argumentation semantics. *Artif. Intell.* 168(1-2):162–210.

Baroni, P.; Giacomin, M.; and Liao, B. 2018. Locality and modularity in abstract argumentation. In *Handbook of Formal Argumentation*. College Publications. chapter 19.

Baumann, R., and Spanring, C. 2015. Infinite argumentation frameworks - on the existence and uniqueness of extensions. In *Essays Dedicated to Gerhard Brewka on the Occasion of His 60th Birthday*, volume 9060, 281–295. Springer.

Baumann, R.; Brewka, G.; and Ulbricht, M. 2020a. Comparing weak admissibility semantics to their dung-style counterparts - reduct, modularization, and strong equivalence in abstract argumentation. In *Proc. KR 2020, September 12-18, 2020*, 79–88.

Baumann, R.; Brewka, G.; and Ulbricht, M. 2020b. Revisiting the foundations of abstract argumentation: Semantics based on weak admissibility and weak defense. In *Proc. AAAI 2020*, 2742–2749. AAAI Press.

Baumann, R. 2011. Splitting an argumentation framework. In *Proc. LPNMR 2011*, 40–53. Springer.

Baumann, R. 2016. Characterizing equivalence notions for labelling-based semantics. In *Proc KR 2016*, 22–32.

Bodanza, G. A., and Tohmé, F. A. 2009. Two approaches to the problems of self-attacking arguments and general odd-length cycles of attack. *Journal of Applied Logic* 7(4):403–420. Special Issue: Formal Models of Belief Change in Rational Agents.

Caminada, M. 2006. Semi-stable semantics. *COMMA* 144:121–130.

Caminada, M. 2007. Comparing two unique extension semantics for formal argumentation: ideal and eager. In *Proc. BNAIC 2007*, 81–87. Utrecht University Press.

Caminada, M. 2014. Strong admissibility revisited. In *Computational Models of Argument - Proceedings of COMMA 2014*, 197–208.

Dauphin, J.; Rienstra, T.; and van der Torre, L. 2020. A principle-based analysis of weakly admissible semantics. In *Proc. COMMA 2020*, volume 326 of *Frontiers in Artificial Intelligence and Applications*, 167–178. IOS Press.

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

Dondio, P., and Longo, L. 2021. Weakly complete semantics based on undecidedness blocking. *arXiv preprint arXiv:2103.10701*.

Dondio, P. 2018. A proposal to embed the in dubio pro reo principle into abstract argumentation semantics based on topological ordering and undecidedness propagation. In *Proceedings of the 2nd Workshop on Advances In Argumentation In Artificial Intelligence, 2018*, volume 2296 of *CEUR Workshop Proceedings*, 42–56. CEUR-WS.org.

Dung, P. M.; Mancarella, P.; and Toni, F. 2007. Computing ideal sceptical argumentation. *Artificial Intelligence* 171(10):642–674.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77(2):321–357.

Dunne, P. E.; Dvořák, W.; Linsbichler, T.; and Woltran, S. 2015. Characteristics of multiple viewpoints in abstract argumentation. *Artificial Intelligence* 228:153–178.

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

Dvořák, W., and Gaggl, S. A. 2014. Stage semantics and the scc-recursive schema for argumentation semantics. *Journal of Logic and Computation* 26(4):1149–1202.

Dvorák, W.; Ulbricht, M.; and Woltran, S. 2021. Recursion in abstract argumentation is hard - on the complexity of semantics based on weak admissibility. In *Proc AAAI 2021*, 6288–6295. AAAI Press.

Fan, X., and Toni, F. 2015. On computing explanations in argumentation. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA*, 1496–1502. AAAI Press.

Oikarinen, E., and Woltran, S. 2011. Characterizing strong equivalence for argumentation frameworks. *Artificial Intelligence* 175:1985–2009.

van der Torre, L., and Vesic, S. 2018. The principle-based approach to abstract argumentation semantics. In *Handbook of Formal Argumentation*. College Publications.

Vassiliades, A.; Flouris, G.; Patkos, T.; Bikakis, A.; Bassiliades, N.; and Plexousakis, D. 2021. A multi attack argumentation framework. In *International Conference on Logic and Argumentation*, 417–436. Springer.

Weydert, E. 2011. Semi-stable extensions for infinite frameworks. In *Benelux Conference on Artificial Intelligence*, 336–343.

Yu, L.; Chen, D.; Qiao, L.; Shen, Y.; and van der Torre, L. 2021. A principle-based analysis of abstract agent argumentation semantics. In *Proc. KR 2021, Online event, November 3-12, 2021*, 629–640.