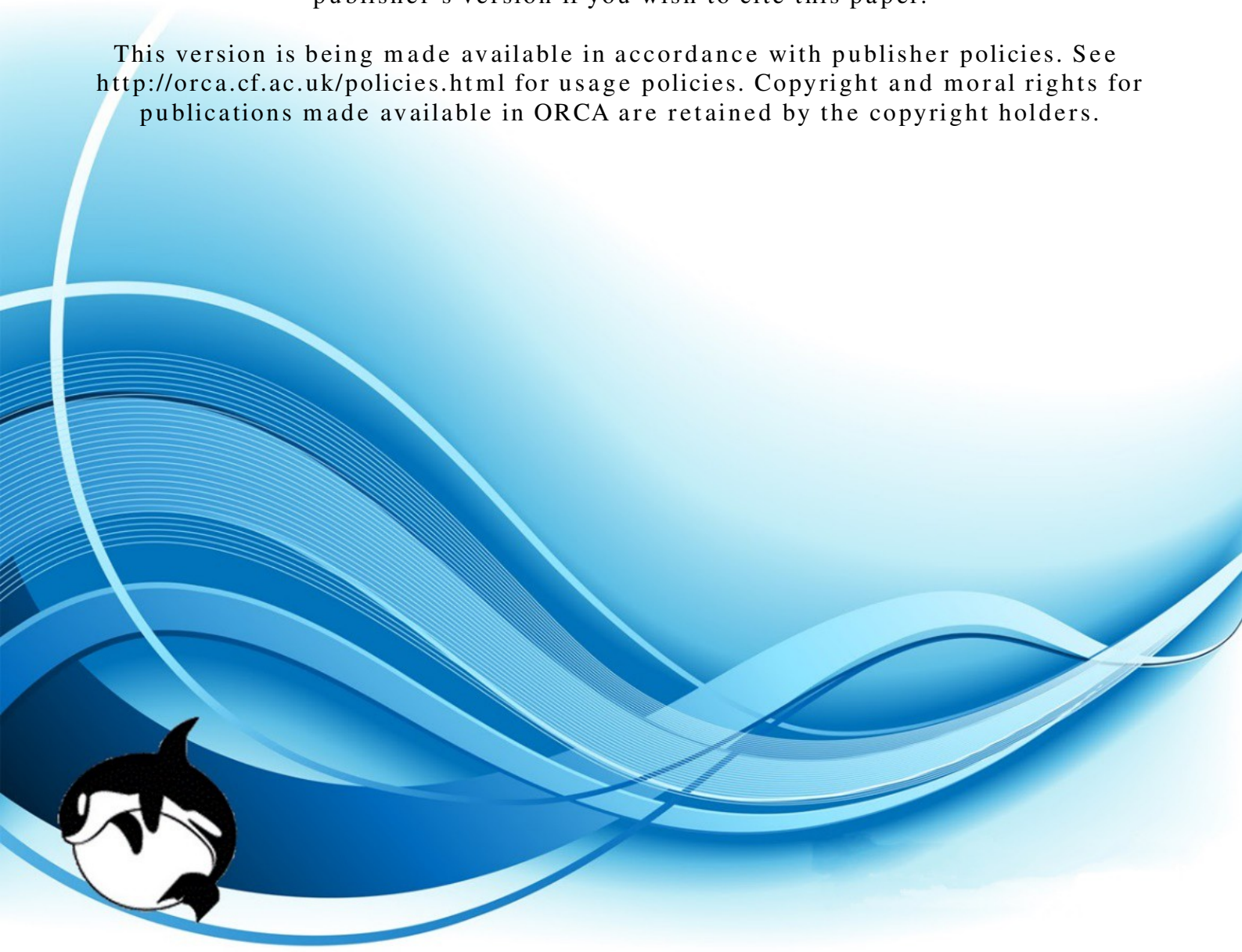

Admissibility in the Abstract Dialectical Framework

Sylwia Polberg, Johannes Peter Wallner, and Stefan Woltran

Vienna University of Technology
Institute of Information Systems
Favoritenstraße 9-11, 1040 Vienna, Austria

Abstract. The aim of this paper is to study the concept of admissibility in abstract dialectical frameworks (ADFs). While admissibility is well-understood in Dung-style frameworks, a generalization to ADFs is not trivial. Indeed, the original proposal turned out to behave unintuitively at certain instances. A recent approach circumvented this problem by using a three-valued concept. In this paper, we propose a novel two-valued approach which more directly follows the original understanding of admissibility. We compare the two approaches and show that they behave differently on certain ADFs. Our results imply that for generalizations of Dung-style frameworks, establishing a precise correspondence between two-valued (i.e. extension-based) and three-value (i.e. labeling-based) characterizations of argumentation semantics is not easy and requires further investigations.

Keywords: abstract argumentation, abstract dialectical framework, admissible semantics

1 Introduction

Dung's abstract argumentation frameworks [1] have proven successful in many applications related to multi-agent systems (see, e.g. [2]). These frameworks are conceptually simple and appealing: arguments are viewed only on an abstract level and a binary attack relation models conflicts between arguments. In several domains this simplicity however leads to certain limitations. Therefore, several enrichments of Dung's approach were proposed [3–9], with abstract dialectical frameworks (ADFs) [10] being one of the most general of these concepts.[1] Simply speaking, in ADFs it is not only the arguments that are abstract but also the relations between them. This is achieved by associating a propositional formula with each argument describing its relation to the other arguments. A

[1] A different approach to model relations between arguments which are beyond attack is meta-argumentation [11]. Here additional (artificial) arguments are added together with certain gadgets to capture the functioning of relations which cannot be modeled with binary attacks.

common problem in applications of abstract argumentation concerns instantiation. Preliminary results on this matter in the case of ADFs can be found in this volume [12]. Moreover, the application of ADFs in the context of the Carneades system [13] and proof standards [10] have been studied in the literature, advising that ADFs might also be applicable to certain problems from the domain of multi-agent systems.

One of the most central concepts in Dung's frameworks is the notion of admissibility which is based on defense. In a nutshell, an argument a is defended (in a given framework) by a set S of arguments if all arguments attacking a are counter-attacked by S. A (conflict-free) set S of arguments is called admissible if each $a \in S$ is defended by S. In fact, many semantics for abstract argumentation are based on admissibility, and in the context of instantiation-based argumentation, admissibility plays an important role w.r.t. rationality postulates, see e.g. [14].

While the concept of admissibility is very intuitive in the Dung setting it is not easy to be generalized to extensions of the Dung–style framework where relations between arguments are not restricted to attacks. As a minimal requirement for such generalized notions of admissibility one would first state "downward-compatibility". Basically speaking, if a given object F in an extended formalism corresponds to a standard Dung framework F', then the admissible (in its generalized form) sets of F should match the admissible sets of F'. In the world of ADFs, the original proposal for admissibility, albeit satisfying this minimal requirement, turned out to behave unintuitively at certain instances. A recent approach first presented in [15] and slightly simplified in [16] is based on (post) fixed points in three–valued interpretations. However, the original intuition that arguments in the set have to "stand together" against the arguments outside the set is somehow lost in that approach (nonetheless, there is a certain correspondence to the characteristic function of Dung-style frameworks).

In this work, we propose a novel two–valued approach which more directly follows the original understanding of admissibility. We call our approach the decisive outing formulation, reflecting its definition which iteratively decides of the status of the arguments. We compare our approach with the three–valued approach from [16] and show that the two semantics can consider different sets of arguments admissible. Since both approaches are downward–compatible, they clearly coincide on Dung-style ADFs; in the paper, we define another class of ADFs where this relation is also preserved. Finally, we further elaborate on these two approaches by showing that each decisive outing admissible extension has a counterpart in the three–valued setting, but not vice versa.

Our results not only show that admissibility can be naturally generalized in different ways, they also imply that for descendants of Dung–style frameworks, establishing one–to–one correspondences between two–valued (i.e. extension–based) and three–valued (i.e. labeling–based) characterizations of argumentation semantics is not necessarily granted. This could stipulate further investigations towards a better understanding of admissibility for more expressive formalisms taking into account also the work of Kakas et al. [17] in logic programming.

The structure of this paper is as follows. In Section 2 we present the theoretical background and notations. Section 3 is dedicated to describing and comparing the three formulations of admissibility and Section 4 is focused on discussion and some pointers for future work.

2 Background

2.1 Dung's Abstract Argumentation Frameworks

The argumentation framework developed by Phan Minh Dung is the simplest, yet quite powerful, formalism for abstract argumentation [1].

Definition 1. *A **Dung abstract argumentation framework**, or a **Dung Framework** is a pair (A, R), where A is a set of arguments and $R \subseteq A \times A$ represents the attack relation.*

Due to the great interest it has received, many semantics have been developed. Semantics define the properties or methods of obtaining framework extensions, i.e. sets of arguments we can accept. Nevertheless, it is generally agreed that any rational opinion should be consistent. This minimal property is expressed with the conflict–free semantics, a common root for all other developed approaches.[2]

Definition 2. *Let $AF = (A, R)$ be a Dung framework. A set $S \subseteq A$ is a **conflict–free extension** of AF, if for each $a, b \in S, (a, b) \notin R$.*

Admissibility is another fundamental requirement in argumentation. It comes from the fact that regardless of the presented point of view, we should be able to defend it. In the Dung setting, due to only one type of relation, it boils down to the following definitions.

Definition 3. *Let $AF = (A, R)$ be a Dung framework. An argument $a \in A$ is **defended** by a set S in AF, if for each $b \in A$ s.t. $(b, a) \in R$, there exists $c \in S$ s.t. $(c, b) \in R$. A conflict–free extension S is an **admissible extension** of AF if each $a \in S$ is defended by S in AF.*

With this at hand, we can start describing the stronger semantics. They can be roughly grouped by varying concepts of maximality or skepticism. Prominent examples are the stable and preferred semantics.

Definition 4. *Let $AF = (A, R)$ be a Dung framework. A conflict–free extension S is a **stable extension** of AF iff for each $a \in A \setminus S$ there exists an argument $b \in S$ s.t. $(b, a) \in R$.*

[2] Conflict–freeness and admissibility can also be treated as some basic properties, rather than very weak semantics. Due to the fact that in some approaches of argumentation frameworks additional types of conflict–freeness have been introduced, we have chosen the latter.

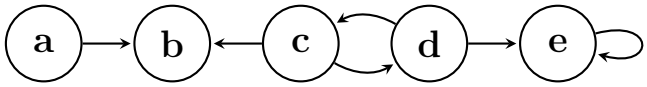

Fig. 1. Sample Dung framework

Definition 5. *Let $AF = (A, R)$ be a Dung framework. A **preferred extension** of AF is a maximal admissible extension of AF w.r.t. subset inclusion.*

We close the list with a semantics belonging to the unique–state approach class, i.e. a semantics producing only a single extension. To this end, we first need to introduce the characteristic function of a framework.

Definition 6. *Let $AF = (A, R)$ be a Dung framework. Its characteristic function $F_{AF} : 2^A \to 2^A$ is defined as follows:*

$$F_{AF}(S) = \{a \mid a \text{ is defended by } S \text{ in } AF\}$$

Definition 7. *Let $AF = (A, R)$ be a Dung framework. The **grounded extension** of AF is the least fixed point of F_{AF}.*

Please note that further semantics can be described via the characteristic function [1, 15]. For our purposes, the most important is the alternative formulation of admissibility as already presented in [1].

Lemma 1. *Let $AF = (A, R)$ be a Dung framework and F_{AF} its characteristic function. A set $S \subseteq A$ is an **admissible extension** of AF iff it is conflict–free and $S \subseteq F_{AF}(S)$.*

Example 1. Consider the Dung framework $AF = (A, R)$ with $A = \{a, b, c, d, e\}$ and the attack relation $R = \{(a, b), (c, b), (c, d), (d, c), (d, e), (e, e)\}$, as depicted in Figure 1. It has eight conflict–free extensions in total, namely $\{a, c\}$,$\{a, d\}$, $\{b, d\}$, $\{a\}$, $\{b\}$, $\{c\}$, $\{d\}$ and $\emptyset$. As b is attacked by an unattacked argument, it cannot be defended against it and will not be in any admissible extension. We end up with two preferred extensions, $\{a, c\}$ and $\{a, d\}$. However, only $\{a, d\}$ is stable, and $\{a\}$ is the grounded extension.

2.2 Abstract Dialectical Frameworks

The main goal of abstract dialectical frameworks (ADFs) [10] is to overcome the limitations of the pure attack relation in the Dung frameworks and its descendants. They assume some predefined set of connection types – attacking, attacking or supporting, and so on – which affects what can be expressed in a framework naturally, and what requires some semantics–dependent modifications. In ADFs relation abstractness is achieved by the introduction of the acceptance conditions instead of adding new elements to the set of relations. They define what (sets of) arguments related to a given argument should be present for it to be included/excluded from an extension.

Definition 8. *An **abstract dialectical framework** (ADF) is a tuple (S, L, C), where S is a set of abstract **arguments** (nodes, statements), $L \subseteq S \times S$ is a set of **links** (edges) and $C = \{C_s\}_{s \in S}$ is a set of **acceptance conditions**, one condition per each argument.*

Originally, the acceptance conditions were defined in terms of functions:

Definition 9. *Let (S, L, C) be an ADF. The set of **parents** of an argument s, denoted par(s), consists of those $p \in S$ for which (p,s) $\in L$. An **acceptance condition** is given by a total function $C_s : 2^{par(s)} \to \{in, out\}$.*

Alternatively, one can also use the propositional formula representation, described in detail in [18]. These two forms are equivalent, and we will be referring to both of them in the rest of this paper.

Definition 10. *Let (S, L, C) be an ADF. **Propositional acceptance conditions** are formulas of the form:*

$$\varphi ::= a \in S \mid \bot \mid \top \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid (\varphi \to \varphi)$$

All and only parents of an argument appear as atoms in the acceptance condition of this argument. In what follows, we use $a : \varphi$ as shorthand for $C_a = \varphi$.

Note that the set L of links can be extracted from the acceptance conditions (more on this matter can be found in [18]). Hence, making it explicit is not necessary. We have decided to keep L in its current form in order to have a consistent representation when weights or more advanced relation properties are added to ADFs.

In the original setting, the truth value of a formula is based on the standard propositional valuation function (i.e. truth tables). However, in [16] Kleene's strong three–valued logic has been used. We will come back to this approach in Section 3.

Due to the abstractness of ADFs, redefining the semantics in an intuitive manner is still an ongoing work and one of the main topics of this paper. In order to take the research step by step, a subclass of ADFs called bipolar was identified in the original paper [10]:

Definition 11. *Let $D = (S, L, C)$ be an ADF. A link $(r, s) \in L$ is*

1. ***supporting**: for no $R \subseteq par(s)$ we have $C_s(R) = in$ and $C_s(R \cup \{r\}) = out$,*
2. ***attacking**: for no $R \subseteq par(s)$ we have $C_s(R) = out$ and $C_s(R \cup \{r\}) = in$.*

*D is **bipolar** iff all links in L are supporting or attacking and we can write it as $D = (S, (L^+ \cup L^-), C)$. The links L^+ denote the supporting links and L^- denote the attacking links. The set of **parents supporting** an argument x is defined as $supp_D(x) = \{y \mid (y, x) \in L^+\}$. The set of **parents attacking** an argument x is defined as $att_D(x) = \{y \mid (y, x) \in L^-\}$.*

Along with the support relations came the problem of the support cycles. We will discuss it further in Section 4.

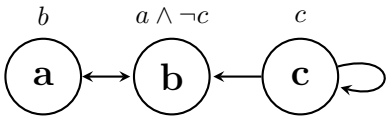

Fig. 2. Example of support cycles

Definition 12. *Let $D = (S, (L^+ \cup L^-), C)$ be a bipolar ADF. D is a bipolar ADF* ***without support cycles*** *if L^+ is acyclic.*

Example 2. Let us look at the ADF depicted in Figure 2: $D = (\{a, b, c, \}, \{(a, b), (b, a), (c, b), (c, c)\}, \{a : b, b : a \wedge \neg c, c : c\})$. In this case c self–supports itself, and a and b exchange supports.

We continue with several semantics that have already been developed for the general class of ADFs.

Definition 13. *Let $D = (S, L, C)$ be an ADF. $M \subseteq S$ is a* ***conflict–free extension*** *of D if for all $s \in M$ we have $C_s(M \cap par(s)) = in$.*

The model semantics follows the 'what can be accepted, should be accepted' intuition. It coincides with the stable semantics in the Dung setting.

Definition 14. *Let $D = (S, L, C)$ be an ADF. $M \subseteq S$ is a* ***model*** *of D if M is conflict–free and $\forall\, s \in S,\ \ C_s(M \cap par(s)) = in$ implies $s \in M$.*

Finally, we also have the grounded semantics (here referred to as well–founded). Just like in the Dung framework, it is obtained by the means of a special function:

Definition 15. *Let $D = (S, L, C)$ be an ADF. Consider the operator*

$$\Gamma_D^W(A, R) = (acc(A, R), reb(A, R))$$

where:

$$acc(A, R) = \{r \in S \mid A \subseteq S' \subseteq (S \backslash R) \Rightarrow C_r(S' \cap par(s)) = in\}$$
$$reb(A, R) = \{r \in S \mid A \subseteq S' \subseteq (S \backslash R) \Rightarrow C_r(S' \cap par(s)) = out\}$$

Γ_D^W is monotonic in both arguments and thus has a least fix–point. E is the ***well–founded model*** *of D iff for some $E' \subseteq S, (E, E')$ is the least fix–point of Γ_D^W.*

Example 3. Let us transform the Dung framework $F = (A, R)$ from Example 1 into an ADF $D = (S, L, C)$. The set of arguments does not change: $A = S$. The same goes for the set of links, please note however, that L loses its meaning – it now represents the connections only, without any information as to their nature. Argument a is unattacked and can always be accepted, hence its acceptance condition is $\top$. b can only be accepted when both a and c are not present ($\neg a \wedge \neg c$). Next, c and d mutually exclude one another (respectively $\neg d$ and $\neg c$). Finally, e is attacked not only by d, but also by itself, and its acceptance condition is $\neg d \wedge \neg e$. Therefore in total we obtain an abstract dialectical framework $D = (A, R, \{a : \top, b : \neg a \wedge \neg c, c : \neg d, d : \neg c, e : \neg d \wedge \neg e\})$.

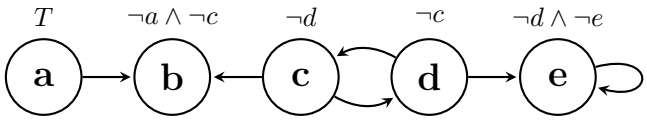

Fig. 3. Sample Dung–style ADF

2.3 Kleene's Three–Valued Logic and Interpretations

In order to be able to explain one of the approaches to the admissibility in ADFs, we need to provide a short recap on the three–valued interpretations and lattices. A more detailed background can be found in [16].

Given a set of arguments S, a three–valued interpretation is a mapping $v : S \to \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$. The truth tables for the basic connectives are given in Figure 4.

¬	
t	f
f	t
u	u

∨	t	u	f
t	t	t	t
u	t	u	u
f	t	u	f

∧	t	u	f
t	t	u	f
u	u	u	f
f	f	f	f

→	t	u	f
t	t	u	f
u	t	u	u
f	t	t	t

Fig. 4. Truth tables for the three–valued logic of Kleene

Let us assume the following partial order $\leq_i$ according to information content: $\mathbf{u} <_i \mathbf{t}$ and $\mathbf{u} <_i \mathbf{f}$. The pair $(\{\mathbf{t}, \mathbf{f}, \mathbf{u}\}, \leq_i)$ forms a complete meet–semilattice with the meet operation $\sqcap$ assigning values in the following way: $\mathbf{t} \sqcap \mathbf{t} = \mathbf{t}$, $\mathbf{f} \sqcap \mathbf{f} = \mathbf{f}$ and $\mathbf{u}$ in all other cases. Given two valuations v and v', we say that v' **contains more information** than v, denoted $v \leq_i v'$ iff $\forall_{s \in S}\ v(s) \leq_i v'(s)$; in case v is three–valued and v' two–valued, then we say that v' **extends** v. This means that elements mapped originally to $\mathbf{u}$ are now assigned either $\mathbf{t}$ or $\mathbf{f}$. The set of all two–valued interpretations extending v is denoted $[v]_2$.

Given a set A, we say that an interpretation v is **partial** if it is defined for a nonempty $B \subseteq A$. Let v' be some interpretation on A. We define a shorthand $v \subseteq v'$ meaning that $\forall b \in B,\ v(b) = v'(b)$. We say that v' is **completion** of v to A. v' is respectively a $\mathbf{t}/\mathbf{f}/\mathbf{u}$**–completion**, if it maps all elements from $A \setminus B$ to respectively $\mathbf{t}/\mathbf{f}/\mathbf{u}$.

It is often very handy to be able to talk about the set of arguments mapped to a certain value by a given interpretation:

Definition 16. *Let v be an interpretation. Then $v^x = \{s \mid v(s) = x\}$ for $x \in \{\mathbf{t}, \mathbf{f}\}$ in case v is two–valued and $x \in \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ if v is three–valued.*

When it comes to the two–valued setting, we can use interpretations and sets of accepted arguments as extensions interchangeably as they uniquely define one another. Unfortunately, this is not the case with the three–valued interpretations. In order to compare both of these settings we need to focus on arguments accepted in both of them. Therefore, sometimes we may refer to a family of the

three–valued interpretations using a set of arguments they map to **t**. Finally, we define a shorthand $v(\varphi)$ for evaluation of a propositional formula φ under an interpretation v.

3 Admissible Semantics for ADFs

In this section we will recall some work on argumentation semantics and discuss several approaches to defining the admissibility for ADFs. We will start with the original definition from [10] and recall some objections raised on it. Then, we introduce two recent formulations (one from [16] and our own novel approach) that are different both in spirit and resulting extensions. At the end of this section we will compare the two in a formal way.

3.1 Related Work on Semantics Rationalities

Throughout the time, many different argumentation semantics have been developed [19]. Very often a new semantics is an improvement of an already existing one by introducing further restrictions on the set of accepted arguments or possible attackers. One of the most important semantical problems is concerned with the cycles in a framework. A thorough study of attack cycles and self–attackers in the Dung setting can be found in [20]. In the bipolar setting, the situation is not yet analyzed this well and approaches differ between available frameworks [5, 9, 21]. The moment we introduce a new type of relation, the situation gets more complicated and every Dung semantics gives rise to several further specializations. Currently, our focus is on whether arguments taking part in support cycles can be in an extension and if they should be considered valid attackers. We will discuss the validity of support cycles further in Section 4. The two recent definitions of admissibility we are going to present differ in the treatment of cycles. The explanation will be provided in Section 3.5.

3.2 Original Formulation

The main motivation behind the original formulation of admissibility in [10] was to create a definition that would not explicitly use the notion of defense. Unfortunately, it was only applicable for the bipolar ADFs. The admissible extensions were obtained via the stable models as proposed in [10].

Definition 17. *Let $D = (S, L, C)$ be a bipolar ADF. A model M of D is a **stable model** of D if M is the least model of the reduced ADF D^M obtained from D by:*

1. *eliminating all nodes not contained in M together with all links in which any of these nodes appear,*
2. *eliminating all attacking links,*
3. *replacing in each acceptance condition C_s of a node s in D^M each occurrence of a statement $t \notin M$ with $\bot$.*

With this at hand, admissible semantics are defined as follows.

Definition 18. *Let $D = (S, L, C)$ be a bipolar ADF. $M \subseteq S$ is **admissible** in D iff there is $R \subseteq S$ such that no element in R attacks an element in M and M is a stable model of $D \setminus R$.*

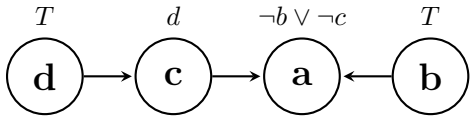

Fig. 5. Counterexample for the original formulation of admissibility.

However, this definition has been proved to give undesired extensions [16]. Take for example the framework depicted in Figure 5. In this setting we have the following admissible extensions: $\emptyset$, $\{b\}$, $\{d\}$, $\{b, d\}$, $\{a, b\}$, $\{c, d\}$, $\{c, b, d\}$. $\{a, b\}$ is not a desired answer as we have no way of preventing our opponent from uttering c since the acceptance condition of d is always *in*. Therefore, the need for a more appropriate definition arises.

3.3 Lattice Formulation

In abstract argumentation, semantics can usually be described in more than one way. The main idea behind it is to provide a relatively constructive formulation that would give us a hint on how to create extensions in a more systematic manner. For example, in case of Dung's frameworks grounded and admissible extensions can be obtained via the characteristic function (see Section 2.1). For ADFs, the original definition has been revised and a new, constructive variant, based on (post) fixed-points, is presented in [15]. A simplified approach published in [16] is based on three–valued interpretations, which we will use in this paper.

The semantics are defined via the following operator, which is similar to the characteristic function of Dung's frameworks. Based on a three-valued interpretation a new one is returned by the function, which accepts or rejects arguments based on the given interpretation. For convenience we will slightly abuse our notation and identify *in* with $\mathbf{t}$ and *out* with $\mathbf{f}$.

Definition 19. *Let $D = (S, L, C)$ be an ADF, v a three–valued interpretation defined over S, $s \in S$ and $\Gamma_D : (S \to \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}) \to (S \to \{\mathbf{t}, \mathbf{f}, \mathbf{u}\})$ a function from three-valued interpretations to three-valued interpretations. Then $\Gamma_D(v) = v'$ with*

$$v'(s) = \bigsqcap_{w \in [v]_2} C_s(par(s) \cap w^{\mathbf{t}})$$

That is, given a three-valued v interpretation a new one is returned by Γ_D for an ADF D. The new truth value for each argument s is given by considering all two-valued interpretations that extend v, i.e. all interpretations that assign either $\mathbf{t}$ or $\mathbf{f}$ to an argument, which is assigned $\mathbf{u}$ by v. Now we evaluate the acceptance

condition of each argument under all these two–valued interpretations. If all of them agree on the truth value, i.e. all of them evaluate to *in* (**t**) or respectively *out* (**f**) , then this is the result or the overall consensus. Otherwise, if there is a disagreement, i.e. we have **t** for one evaluation and **f** for another, then the result is undecided, i.e. **u**.

The new definition of admissibility resembles the one for AFs. We apply Γ_D similarly as the characteristic function and just use the information ordering instead of the subset relation. Please note that conflict–freeness is already incorporated in this definition.

Definition 20. *A three-valued interpretation v for an ADF $D = (S, L, C)$ is* ***admissible*** *in D iff $v \leq_i \Gamma_D(v)$.*

The following example illustrates this definition.

Example 4. Let us go back to the framework in Figure 5. The following three-valued interpretations are then admissible $v_1 = \{d \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}, a \mapsto \mathbf{u}\}$, $v_2 = \{d \mapsto \mathbf{t}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}, a \mapsto \mathbf{u}\}$, $v_3 = \{d \mapsto \mathbf{t}, b \mapsto \mathbf{u}, c \mapsto \mathbf{t}, a \mapsto \mathbf{u}\}$, $v_4 = \{d \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{u}, a \mapsto \mathbf{u}\}$, $v_5 = \{d \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{u}, a \mapsto \mathbf{u}\}$, $v_6 = \{d \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}, a \mapsto \mathbf{u}\}$, $v_7 = \{d \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}, a \mapsto \mathbf{f}\}$.

Let us inspect closer why v_7 is admissible in this ADF. The three-valued interpretation v_7 is already two–valued, i.e. no argument is assigned the value **u**. This means that $[v_7]_2 = \{v_7\}$. Now if we evaluate for each argument its acceptance condition under v_7, then the result is the same as the assigned value by v_7. Consider for instance argument a with the acceptance condition $\neg b \vee \neg c$ as a propositional formula. This formula evaluates to **f** under v_7, which is the same value assigned by v_7, i.e. $v_7(a) = C_a(par(a) \cap v_7^{\mathbf{t}}) = \mathbf{f}$.

Considering a slightly more complex example, let us look at v_6. Here $[v_6]_2 = \{v, v'\}$ with $v = \{d \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}, a \mapsto \mathbf{t}\}$ and $v' = \{d \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}, a \mapsto \mathbf{f}\}$. This means we have to consider both evaluations, one assigning the argument a true and one false. Now the acceptance condition of a evaluates under both v and v' to **f**. This means that $\Gamma_D(v_6) = v_6'$ and $v_6(a) = \mathbf{u} \leq_i v_6'(a) = \mathbf{f}$, since $\mathbf{f} \sqcap \mathbf{f} = \mathbf{f}$. Similarly for the other arguments and hence v_6 is admissible.

Let us check if there exists an admissible three–valued interpretation v, which assigns **f** to d, i.e. $v(d) = \mathbf{f}$. Since the acceptance condition of d always evaluates to true, we know that for any two-valued interpretation w we have $C_a(par(a) \cap w^{\mathbf{t}}) = \mathbf{t}$. This in particular holds for for all $v' \in [v]_2$. Hence $\Gamma_D(v) = v'$, with $v(d) = \mathbf{f} \not\leq_i \mathbf{t} = v'(d)$ and v is not admissible.

3.4 Decisive Outing Formulation

We now introduce an alternative definition of admissibility that comes back to the intuition behind the semantics. An admissible extension is supposed to be able to 'stand on its own' [22], i.e. discard any argument that would render any of the set's elements unacceptable. In the Dung setting, a set defends an argument if it attacks all of its attackers. In the ADF setting, which is more abstract, this is not enough. We can discard an argument in more ways than just a direct attack

– overall we want to make sure that the acceptance of a different argument will not make the 'bad' one acceptable via a chain reaction. Moreover, due to the various types of relations available in ADF, it might be the case that to discard one argument, more counterarguments may be required (in the Dung case, one 'attacker' per 'attacker' was sufficient).

This intuition is enough to create a definition of admissibility that does not make use of the notion of attack or defense, which is quite appropriate for this abstract setting. Our approach is based on iteratively building a set of arguments that our candidate for admissibility has the power to permanently set to *out*. Important in this construction is the notion of decisiveness:

Definition 21. *Let $D = (S, L, C)$ be an ADF and $s \in S$. Let v_Z be a two or three–valued interpretation defined on a set $Z \subseteq par(s)$, We say that v_Z is **decisive** for s iff for any two (respectively two or three–valued) completions $v_{par(s)}$ and $v'_{par(s)}$ of v to $par(s)$, it holds that $v_{par(s)}(C_s) = v'_{par(s)}(C_s)$.*
*We say that s is **decisively out/in/undecided** wrt v_Z if v_Z is decisive and all of its completions map s to respectively $out, in, undec$.*

Example 5. The idea behind this formulation is to identify the partial interpretations that are "enough" to know the final value of an acceptance condition. Assume an ADF $D = (\{a, b\}, \{(a, a), (a, b), (b, a), (b, b)\}, \{a : a \rightarrow b, b : a \wedge b\})$. Let v be a partial two–valued interpretation s.t. $v(b) = \mathbf{t}$. Then $a \rightarrow b$ will always evaluate to $\mathbf{t}$ no matter the assignment of a and we can say that a is decisively in wrt to v. It is of course not decisive for b.

With this at hand, we can define the set of arguments permanently excluded by a given set. The idea behind it corresponds to identifying all the arguments attacked by an extension E in the Dung setting and is known as the E^+ set. Due to its abstractness, ADFs also give us indirect ways of discarding an argument and such a straightforward check would be inadequate.

Definition 22. *Let $D = (S, L, C)$ be an ADF and $A \subseteq S$ a conflict–free extension of D. Let v be a partial two–valued interpretation built as follows:*

1. *Let $M = A$. For every $a \in M$ set $v(a) = \mathbf{t}$.*
2. *For every argument $b \in S \setminus M$ that is decisively out in v, set $v(b) = \mathbf{f}$ and add b to M.*
3. *Repeat the previous step until there are no new elements added to M.*

*By A^+ we understand the set of arguments $v^{\mathbf{f}}$. The **range** of A, denoted A^R is defined as $A \cup A^+$. We refer to v as **range interpretation** of A.*[3]

We can now naturally proceed to admissibility:

Definition 23. *Let $D = (S, L, C)$ be an ADF, $A \subseteq S$ a conflict–free extension of D and A^+ its discarded set. A is **admissible** in D iff for any $F \subseteq S \setminus A$ $(F \neq \emptyset)$, if there exists an $a \in A$ s.t. $C_a(par(a) \cap (F \cup A)) = out$ then $F \cap A^+ \neq \emptyset$.*

[3] Please note that although these notions were originally defined for arbitrary sets, in practice they were always used for at least conflict–free ones and this assumption allowed us to create a cleaner formulation.

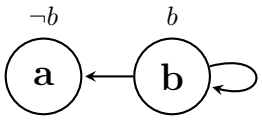

Fig. 6. Example of nonequivalence between the formulations of admissibility.

Example 6. Let us come back to the counterexample from Figure 5. Clearly $\emptyset, \{b\}, \{d\}, \{b, d\}, \{c, d\}$ and $\{b, c, d\}$ are admissible; they are not attacked in any way and hence implication is always true. Let us now check $\{a\}$: its discarded set is empty, while the set $\{c, d\}$ has the power to out the acceptance condition of a. The same situation can be observed for $\{a, c, d\}$ and $\{a, b\}$: the discarded sets are both empty, while we need to be able to counter $\{b\}$ and $\{c\}$ respectively. Thus, none of these sets is admissible.

3.5 Comparison

Moving from the two–valued to the three–valued approach is more than just a structural change. This was the case also in the Dung setting, even though both approaches were strongly related [19]. When computing classical extensions, we focus on what arguments we can accept. In the three–valued setting, a discarded argument is also important and **f** means something more than just a lack of acceptance. In this setting **u** represents the lack of either a proper reason to accept or discard an argument or will to commit to a value (i.e. we decide not to assign **t** or **f** even though we have sufficient basis for that). As a result, a semantics truly exploiting a three–valued setting has naturally different assumptions than a two–valued one. For example it would rather maximize on the arguments that are not left undecided, rather than just on the ones we are ready to accept. Therefore if one decides to treat the three–valued setting as the means of computing extensions in the two–valued one, he or she should take special care when choosing semantics.

Let us compare the decisive outing and lattice formulations of admissibility. The main difference lies in the treatment of the self–support and support cycles. The first one admits both and treats attacks generated by them as valid. The latter also admits both, however, attacks coming from them do not need to be defended from. Take the framework depicted in Figure 6. According to the outing formulation, $\emptyset$ and $\{b\}$ are the only admissible extensions. This comes from the fact that if we were to utter $\{a\}$, an opponent could always respond with $\{b\}$, which we cannot counter. In the lattice setting, if we collect just the arguments set to **t**, i.e. the arguments accepted in an admissible three-valued interpretation, then we obtain the following sets: $\emptyset$, $\{a\}$ and $\{b\}$.

The fact that the outing definition admits arguments forming support cycles as valid attackers has some side effects. Most importantly, it breaks the relation between the stable and the admissible semantics – in this example $\{a\}$ is a stable extension, but not (outing) admissible. This does not occur in the lattice approach.

As we have mentioned before, there is a difference in motivation behind the two and three–valued semantics. Take for example the preferred extensions, which in general do not have to agree even if their admissible bases do. Let us assume a framework consisting of a single self–supporting argument $D = (\{a\}, \{(a, a)\}, \{a : a\})$. If we were to follow the standard set inclusion maximality definition then $\{a\}$ would be the preferred extension according to the outing formulation. However, the lattice version follows information maximality and both $\emptyset$ and $\{a\}$ would be considered preferred.

Let us close this section with some formal results on how and when can extensions under both approaches coincide. We will start with the outing to lattice direction. Please note that it recreates the relation between extensions from the two to three–valued setting that held in the Dung framework [19].

Theorem 1. *Let $D = (S, L, C)$ be an ADF. For any (decisive outing) admissible extension E of D there exists a lattice admissible three–valued interpretation v_3 s.t. $v_3^{\mathbf{t}} = E$.*

Proof. We will prove this theorem by constructing an appropriate interpretation (please note there may be more than one per extension). Let v_3 be a u–completion to S of the range interpretation v of E. Assume that v_3 is not lattice admissible, i.e. it is not the case that $v_3 \leq_i \Gamma_D(v_3)$. This means that the new interpretation "loses" information, i.e. at least one element formerly mapped to $\mathbf{t}$ or $\mathbf{f}$ is now $\mathbf{u}$, or becomes incomparable (some element formerly mapped to $\mathbf{t}/\mathbf{f}$ goes to $\mathbf{f}/\mathbf{t}$).

Let us first take a look at the case when $v_3(a) = \mathbf{f}$ and $\Gamma_D(v_3)(a) \neq \mathbf{f}$. This means that for at least one $w \in [v_3]_2$, $w(C_a) = \mathbf{t}$. Consequently, a is not decisively out in v_3 and could not have been decisively out in v. Contradiction.

Now let us consider the case when $v_3(a) = \mathbf{t}$ and $\Gamma_D(v_3)(a) \neq \mathbf{t}$. From this follows that there is at least one $w \in [v_3]_2$ s.t. $w(C_a) = \mathbf{f}$. Let F be the set of all arguments originally mapped to $\mathbf{u}$ that are now assigned $\mathbf{t}$, i.e. $F = \{f \in S$ s.t. $v_3(f) = \mathbf{u}$ and $w(f) = \mathbf{t}\}$. If the set F is empty, then w is a $\mathbf{f}$–completion of v and therefore failure for a means E cannot be two–valued conflict–free. Contradiction. If set F is not empty, it means that $C_a(par(a) \cap E) = in$ (by conflict–freeness) and $C_a(par(a) \cap (E \cup F)) = out$ (coming from w). Moreover, $F \cap E^+ = \emptyset$ by construction – no element from E^+ is assigned $\mathbf{u}$, which is the requirement for adding to F. Conclusion is that E cannot be outing admissible. Contradiction.

In what follows we show that the two notions of admissibility, the lattice and decisive outing formulation, coincide on a special class of ADFs, namely the bipolar ADFs without support cycles. Although we do not claim that this class is the maximal one where the semantics agree, it appears natural to consider, since the semantics can differ when support cycles are present. Note that we assume finite ADFs, i.e. the set of arguments S is finite.

We prove a technical lemma, which intuitively states that every argument that is set to $\mathbf{f}$ in a three-valued admissible interpretation is rejected either because the set of accepted arguments together are enough reason to reject it, or it requires supporters, which are rejected.

Lemma 2. *Let $D = (S, L, C)$ be a bipolar ADF without support cycles and v a lattice admissible three-valued interpretation in D and $a \in v^{\mathbf{f}}$. Then at least one of the following statements is true.*

1. *For any $M \supseteq v^{\mathbf{t}}$ we have $C_a(par(a) \cap M) = out$, or*
2. *$supp_D(a) \neq \emptyset$ and for any $M \supseteq v^{\mathbf{t}}$ with $M \cap supp_D(a) \cap v^{\mathbf{f}} = \emptyset$ we have $C_a(par(a) \cap M) = out$.*

Proof. Assume that v is admissible in D and $a \in v^{\mathbf{f}}$. Assume that statement 1 does not hold. This means there exists a $M' \supseteq v^{\mathbf{t}}$ s.t. $C_a(par(a) \cap M') = in$. Since v is admissible we have that $C_a(par(a) \cap v^{\mathbf{t}}) = out$. This follows from the fact that there exists a $w \in [v]_2$ with $w^{\mathbf{t}} = v^{\mathbf{t}}$ and $w^{\mathbf{f}} = v^{\mathbf{f}} \cup v^{\mathbf{u}}$. Since v is admissible, it follows that $C_a(par(a) \cap w^{\mathbf{t}}) = out$, since otherwise $v(a) \not\leq_i w(a)$. Hence, there exists a $x \in (M' \setminus v^{\mathbf{t}})$ which is supporting a.

Now let $M \supseteq v^{\mathbf{t}}$ and $M \cap supp_D(a) \cap v^{\mathbf{f}} = \emptyset$. Let further $M^p = M \cap par(a)$, i.e. M^p is restricted to the parents of a. Suppose $C_a(M^p) = in$, let $X = (M^p \setminus (att_D(a) \setminus v^{\mathbf{t}}))$, i.e. X is a subset of M^p, without the attackers of a, which are not in $v^{\mathbf{t}}$. Then we have that also $C_a(X) = in$. Suppose the contrary, i.e. $C_a(X) = out$, but since $C_a(M^p) = in$ this means that there exists a $b \in (M^p \setminus X)$ with $b \in supp_D(a)$, which is a contradiction. This in turn implies $X \cap v^{\mathbf{f}} \cap par(a) = \emptyset$. This is a contradiction to admissibility of v, since also $(v^{\mathbf{t}} \cap par(a)) \subseteq X$ holds and admissibility requires that in this case $C_a(X) = out$, by a similar reasoning as above.

Now we can show the coincidence of the admissible semantics on the bipolar ADFs without support cycles.

Theorem 2. *Let $D = (S, L, C)$ be a bipolar ADF without support cycles and v a lattice admissible three-valued interpretation in D, then $A = v^{\mathbf{t}}$ is (decisive outing) admissible in D.*

Proof. Assume there exists a non-empty set $F \subseteq (S \setminus A)$ and $M = F \cup A$, s.t. there exists an argument $a \in A$ with $C_a(par(a) \cap M) = out$. We first show that $M \cap v^{\mathbf{f}} \neq \emptyset$. Suppose the contrary, i.e. $M \cap v^{\mathbf{f}} = \emptyset$. It is straightforward to see that $M \supsetneq v^{\mathbf{t}}$, since $C_a(par(a) \cap v^{\mathbf{t}}) = in$, otherwise v would not be admissible in D. Suppose all elements in M which are not in $v^{\mathbf{t}}$ are undecided in v, i.e. $(M \setminus v^{\mathbf{t}}) \subseteq v^{\mathbf{u}}$. But this implies that the corresponding two-valued interpretation of M, namely $v'(s) = \mathbf{t}$ if $s \in M$ and $v'(s) = \mathbf{f}$ otherwise, must be in $[v]_2$ and hence v would not be admissible, since then $\Gamma_D(v)(a) \neq \mathbf{t}$.

Now we show that for every $r \in v^{\mathbf{f}}$ it holds that $r \in A^+$, hence $v^{\mathbf{t}}$ is decisive outing admissible in D. Let L^+ and L^- be the supporting and attacking links of D and $L = L^+ \cup L^-$. Since the graph $G = (S, L^+)$ is an acyclic directed graph (DAG), we can construct a topological ordering, represented by the function $f : S \to \mathbb{N}$, on the vertices S such that if $(a, b) \in L^+$ we have $f(a) < f(b)$. This means if a supports b, then the former is ordered lower than the latter. We now show the claim by induction on $f(s)$ for arguments in S.
(IH): Let $r \in v^{\mathbf{f}}$, $f(r) = i$, if $\forall r' \in v^{\mathbf{f}}$, s.t. $f(r') < f(r)$ we have $r' \in A^+$, then $r \in A^+$.

(IB): The claim holds for all $s \notin v^{\mathbf{f}}$, hence we look at the smallest element in $r \in v^{\mathbf{f}}$ w.r.t. the ordering induced by f. We know that one of the two statements of Lemma 2 must hold for r. If the first one holds, then clearly $r \in A^+$. Otherwise we have $supp_D(r) \cap v^{\mathbf{f}} = \emptyset$, since r must be the minimal element of the order induced by f. But then we know that for any $M \supseteq A$ we have that $M \cap supp_D(r) \cap v^{\mathbf{f}} = \emptyset$ and thus $C_r(par(r) \cap M) = out$. Hence $r \in A^+$.
(IS): Let $r \in v^{\mathbf{f}}$ with $f(r) = i$. We assume that $\forall r' \in v^{\mathbf{f}}$ with $f(r') < f(r)$ it holds that $r' \in A^+$. Again, since $r \in v^{\mathbf{f}}$ we know that one of the statements of Lemma 2 is true. Furthermore, if the first one is true, then clearly $r \in A^+$. Suppose only the second statement is true. By assumption, we know that $\forall x \in supp_D(r) \cap v^{\mathbf{f}}$ we have that $x \in A^+$, since all of the elements in this set are in $v^{\mathbf{f}}$ and have a lower order w.r.t. f. This means $(supp_D(r) \cap v^{\mathbf{f}}) \subseteq A^+$. But then r must be in A^+, since r is decisively out for the partial two-valued interpretation v', which sets all elements in A^+, in particular $supp_D(r) \cap v^{\mathbf{f}}$ to $\mathbf{f}$ and all elements in A to true. Indeed for all $M \supseteq A$, s.t. $M \cap supp_D(r) \cap v^{\mathbf{f}} = \emptyset$ we have that $C_r(par(r) \cap M) = out$.

4 Discussion

Notes on defense. Strongly tied to the notion of admissibility is the concept of defense. Although we have managed to formulate admissibility without making the defense explicit, giving a proper account of it is required for redefining some of the stronger semantics. The current definition of the discarded set (A^+) can be a base for detecting defense known from the conflict–based setting (i.e. counterattacking) and one arising in the bipolar setting (e.g. cutting off the support of an attacker). However, in ADFs, one can defend in one more way. Due to the fact that the framework (mostly via disjunction in acceptance conditions) has the possibility to express some weak notion of preference between incoming relations, we have a case of *overpowering defense*. Instead of responding to a discard with another discard, we overpower it. A simple example of it would be an acceptance condition of the form $\neg a \vee b$. As long as b is present in the framework, accepting a has no effect. It does not require the "defender" and the "attacker" to be connected by a link, and hence cannot be detected by the discarded set. This type of defense in ADFs is also problematic as often a conflict–free extension possessing it simply does not "react" to incoming conflicts. Therefore, verifying whether a set has the power to defend an argument not belonging to it in this particular way is challenging.

Revisiting support cycles. From the point of view of ADFs, the ongoing research on bipolarity in argumentation is very important. A thorough overview can be found in [5]. Although the acceptance conditions allow us to express support in several ways, we do not yet take into account all of its side effects. In this section we would like to discuss the problem of support cycles in argumentation. Although discarded in logic programming and some frameworks [9], they do not always represent an error in our thinking. A very simple example, yet common in every day life and, for instance, game theory is the case of mutual agreement.

An agent can decide to cooperate as long as his opponent agrees to do the same. This rule 'I play nice as long as you play nice' is not something irrational or rare. The 'good will' mutual agreement is in our opinion a very important example of reasoning that is not only defeasible (we just 'assume' everyone else is following their commitment, and this assumption can very well be withdrawn when it turns out it is not the case) but also has a support cycle in it. And yet, it is very reasonable and, be it good or not, unavoidable.

Nevertheless, there are support cycles that are clearly erroneous and need to be avoided. Unfortunately, there is not much intuition on how to distinguish between the 'good' and the 'bad' ones. For these reasons, in future we would like to admit the semantics both with and without the support cycles and use them according to a given situation. We hope that further research will shed more light on this case.

Future work. Throughout this paper we have mentioned several open questions and problems concerning not only the ADFs, but also argumentation overall. First, we see a need for a discussion on the rationality of arguments, i.e. how should self–attackers, self–supporters and support cycles be treated. Addressing the rationality issue would give rise to stricter notions of semantics. Another task for the future is moving the logic programming style acceptability [17] to ADFs. In order to give an intuitive definition, a proper account of support cycles in ADFs is required, which were so far informally described in Section 2.2.

Finally, we would like to formalize the concept of defense in ADFs and provide a tool for an efficient detection of overpowering. With this concept at hand, moving over to other well known semantics in this abstract framework is a next natural step. In particular, complete and preferred semantics can be based on our notion of admissibility. In case of the latter, this could circumvent certain problems of the formulation introduced in [16], where three–valued preferred extensions are not necessarily incomparable on the sets of accepted arguments.

5 Conclusion

In this paper we have reviewed the existing definitions of admissibility in abstract dialectical frameworks — one of the most general enhancements of Dung's abstract frameworks — and introduced a novel two–valued approach reflecting the original formulation of admissibility in a more direct way. Besides a thorough discussion on the conceptual level, we have also compared the approaches on a formal one. The results show that each new two–valued admissible extension is also admissible in the three–valued setting of [16], but that the other direction does not hold in general.

Acknowledgements

This work has been funded by FWF through project I1102. Sylwia Polberg is financially supported by the Vienna PhD School of Informatics. We would like to thank Martin Riener for his help and valuable suggestions.

References

1. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artif. Intell. **77** (1995) 321–357
2. McBurney, P., Parsons, S., Rahwan, I., eds.: Argumentation in Multi-Agent Systems - 8th International Workshop, ArgMAS 2011, Revised Selected Papers. Volume 7543 of LNCS. Springer (2012)
3. Amgoud, L., Vesic, S.: A new approach for preference-based argumentation frameworks. Annals of Mathematics and Artificial Intelligence **63** (2011) 149–183
4. Baroni, P., Cerutti, F., Giacomin, M., Guida, G.: AFRA: Argumentation framework with recursive attacks. Int. J. Approx. Reasoning **52**(1) (2011) 19–37
5. Cayrol, C., Lagasquie-Schiex, M.C.: Bipolarity in argumentation graphs: Towards a better understanding. Int. J. Approx. Reasoning (2013) In Press.
6. Bench-Capon, T.J.M.: Persuasion in practical argument using value-based argumentation frameworks. J. Log. Comput. **13**(3) (2003) 429–448
7. Modgil, S.: Reasoning about preferences in argumentation frameworks. Artif. Intell. **173**(9-10) (2009) 901–934
8. Nielsen, S., Parsons, S.: A generalization of Dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In: ArgMAS. Volume 4766 of LNCS. Springer (2007) 54–73
9. Nouioua, F., Risch, V.: Argumentation frameworks with necessities. In: SUM. Volume 6929 of LNCS. Springer (2011) 163–176
10. Brewka, G., Woltran, S.: Abstract dialectical frameworks. In: KR. (2010) 102–111
11. Boella, G., Gabbay, D.M., van der Torre, L., Villata, S.: Meta-argumentation modelling I: Methodology and techniques. Studia Logica **93**(2-3) (2009) 297–355
12. Strass, H.: Instantiating knowledge bases in abstract dialectical frameworks. In: CLIMA XIV. (2013) In Press.
13. Brewka, G., Gordon, T.F.: Carneades and abstract dialectical frameworks: A reconstruction. In: COMMA, IOS Press (2010) 3–12
14. Caminada, M., Amgoud, L.: On the evaluation of argumentation formalisms. Artif. Intell. **171**(5-6) (2007) 286–310
15. Strass, H.: Approximating operators and semantics for Abstract Dialectical Frameworks. Technical Report 1, Institute of Computer Science, Leipzig University (2013)
16. Brewka, G., Ellmauthaler, S., Strass, H., Wallner, J.P., Woltran, S.: Abstract dialectical frameworks revisited. In: IJCAI. (2013) In Press.
17. Kakas, A.C., Mancarella, P., Dung, P.M.: The acceptability semantics for logic programs. In: ICLP. (1994) 504–519
18. Ellmauthaler, S.: Abstract dialectical frameworks: properties, complexity, and implementation. Master's thesis, Vienna University of Technology (2012)
19. Baroni, P., Caminada, M., Giacomin, M.: An introduction to argumentation semantics. Knowledge Eng. Review **26**(4) (2011) 365–410
20. Baroni, P., Giacomin, M., Guida, G.: SCC–recursiveness: a general schema for argumentation semantics. Artif. Intell. **168**(12) (2005) 162 – 210
21. Oren, N., Norman, T.J.: Semantics for evidence-based argumentation. In: COMMA. (2008) 276–284
22. Baroni, P., Giacomin, M.: Semantics of abstract argument systems. In Simari, G., Rahwan, I., eds.: Argumentation in Artificial Intelligence. Springer (2009) 25–44