Argument & Computation 9 (2018) 249–267
DOI 10.3233/AAC-180427
IOS Press

Representing the semantics of abstract dialectical frameworks based on arguments and attacks

Phan Minh Dung [a,*] and Phan Minh Thang [b]

[a] *Department of Communication and Information Technology, Asian Institute of Technology, Thailand*
[b] *Burapha University International College, Thailand*

Abstract. Abstract dialectical frameworks have been proposed as a generalization of the abstract argumentation frameworks. The semantics of abstract dialectical frameworks is defined by identifying different classes of models. In this paper, we show that the semantics of abstract dialectical frameworks could naturally be defined based on simple notions of arguments and attacks like in abstract argumentation. This insight allows us to adapt directly the semantical concepts in abstract argumentation to abstract dialectical frameworks that not only capture the standard semantics of abstract dialectical frameworks, but also suggest other new semantics based on the idea of "rejection as assumption" (raa) (similar to the concept of "negation as assumption" in assumption-based argumentation and logic programming) like the well-founded semantics or the raa-preferential semantics.

Keywords: Abstract argumentation, abstract dialectical frameworks

1. Introduction

There are many generalizations of the abstract argumentation frameworks [9]. Cayrol and Lagasquie-Schiex [8] presented bipolar argumentation frameworks in which arguments can also support each other. Modgil [16], Baroni, Cerutti, Giacomin and Guida [2], Hanh, Dung, Hung and Thang [14], Gabbay [11] introduced attack on attacks on attacks. Nielsen and Parson [17] studied attacks from sets of arguments. Amgoud and Cayrol [1], Bench-Capon and Atkinson [3] introduced preferences between arguments. A prominent generalization of abstract argumentation is the abstract dialectical frameworks introduced by Brewka and Woltran [7]. There have been very active research on the semantics of ADFs [5–7,20–22]. The semantics of abstract dialectical frameworks are defined by identifying different classes of models that are fixed points of the Brewka and Woltran operator [6,7]. A closer look at the fixed-point-model-semantics of ADFs reveals that they could be characterized by how a justification (or argument) for the acceptance of a statement is viewed.

Example 1. For illustration, consider the following ADF D_0

$a[a] \qquad b[\neg a]$

stating that

*Corresponding author. E-mail: dung.phanminh@gmail.com.

1946-2166/18/$35.00

– a is accepted (resp. rejected) if a is accepted (resp. rejected), and
– b is accepted (resp. rejected) if a is rejected (resp. accepted).

The semantics of D_0 depends on whether the condition "*a is accepted (resp. rejected) if a is accepted (resp. rejected)*" is considered vacuous or not.

If such condition is considered vacuous (as it is the case according to the stable semantics [6]), $\{a\}$ therefore does not provide any justification for accepting a. As there is no other justification for a, a is considered rejected and consequently b accepted.

In contrast, according to the semantics based on the fixed points of the Brewka–Woltran operator [6,7], the condition "*a is accepted (resp. rejected) if a is accepted (resp. rejected)*" is not viewed as vacuous and hence $\{a\}$, $\{\neg a\}$ provide justifications for a and $\neg a$ respectively. Therefore there are two preferred models $\{a\}$, $\{\neg a\}$ while the grounded model is empty.

The above discussion suggests that the semantics of ADFs is characterized by how the notion of a justification (or argument) for a statement is viewed.

In logic programming, stable models [13] arguably represent the most prominent approach to the negation-as-assumption view where negation-as-failure literals are viewed as assumptions [4,9,10]. Other approaches are the partial stable models, the three-valued stable models and the well-founded model [12,18,19]. It is well-known that all the four approaches could be captured by the extensions of argumentation frameworks whose arguments are proof trees constructed from the logic program rules where negation-as-failure literals are considered as assumptions [9,15,23].

As the stable models of ADFs [6] originate from stable models in logic programming, it indeed also adopts a similar view of "rejection as assumption" (raa) where rejected statements are viewed as assumptions. Formally, we will show that semantically, ADFs could be represented by argumentation frameworks referred to as normal argumentation frameworks, whose arguments are support trees for statements in ADFs where rejected statements are viewed as assumptions and the stable models of ADFs are captured by the stable extensions of normal argumentation frameworks. This insight sugests that the other extensions of the normal argumentation frameworks could be viewed as representing new semantics of ADFs where the grounded extension could be viewed as the most skeptical one that we refer to as well-founded semantics of ADFs to distinguish it from the BW-grounded model defined in [6], while the preferred extensions are referred to as "rejection-as-assumption" (raa)-preferential semantics.

It turns out that both the BW-grounded model defined in [7] as well as the BESWW-preferred models introduced in [6] are also naturally captured by the extensions of simple argumentation frameworks whose arguments are supports (or justifications) constructed directly from the acceptance conditions of statements in ADFs.

Example 2. Consider the ADF D_1

$$a[a] \qquad b[\neg c, \neg a] \qquad c[\neg b, \neg a] \qquad d[\neg d]$$

The BW-grounded model [7] is empty while the BESWW-preferred models [6] are $\{a, \neg b, \neg c\}$, $\{\neg a, \neg b, c\}$, $\{\neg a, b, \neg c\}$.

There is no stable model. The well-founded semantics gives $\{\neg a\}$. The raa-preferred semantics gives $\{\neg a, \neg b, c\}$, $\{\neg a, b, \neg c\}$.

Our results are represented in Fig. 1.

	Simple AF	Normal AF
Grounded Extensions	BW - grounded Model	Well-founded Semantics (new)
Preferred Extensions	BESWW - preferred Models	Rejection-as-Assumptions (RAA)- preferrential Semantics (new)
Stable Extensions	two-valued BESWW- preferred Models	BESWW - stable Models

Fig. 1. Classification of ADF semantics.

The paper is organized in 7 sections including this introduction. In the next section, we recall the key concepts of the AFs and ADFs. In the following section, we present the simple argumentation frameworks whose extensions capture the classes of BW-grounded and BESWW-complete models of ADFs. In Section 4, we first argue that the stable models of ADFs are based on a view of "rejection as assumption" by giving an equivalent characterization of them that reflect the view of "rejection as assumption" in a more direct way. We then proceed to show that stable models of ADFs are captured by stable extensions of normal argumentation frameworks with arguments being support trees whose leaves are labelled by assumptions represented by rejected statements. As an immediate consequence, we present two new semantics based on the grounded and preferred extensions of normal argumentation frameworks.We conclude in Section 5. We give the detailed proofs of the theorems and lemmas in the Appendix. We acknowledge the supports we got from the reviewers and colleagues in the Acknowledgements section.

2. Preliminaries

2.1. Argumentation framework

An abstract *argumentation framework* (AF) [9] is a pair (AR, att) where AR is a set of arguments and $att \subseteq AR \times AR$ is the attack relation between arguments. An argument $A \in AR$ attacks an argument $B \in AR$ if $(A, B) \in att$. A is called an attacker of B if A attacks B.

A set of arguments $S \subseteq AR$ *attacks* B if there exists $A \in S$ such that A attacks B.

S *defends* A if S attacks each attacker of A.

S is *conflict-free* if it does not attack any of its own arguments.

S is *admissible* if it is conflict-free and defends each of its arguments.

The characteristic function of AF is defined by $F_{AF}(S) = \{A \in AR \mid S \text{ defends } A\}$.

Given an $AF = (AR, att)$, a set of arguments $S \subseteq AR$ is

a *stable extension* of AF if it is conflict-free and attacks each argument $A \notin S$;
a *preferred extension* of AF if it is a maximal (wrt set inclusion) admissible set of arguments;
a *complete extension* of AF if it is admissible and contains each argument it defends (or equivalently a conflict-free fixed point of F_{AF});

a *grounded extension* of AF if it is the least complete extension (or equivalently the least fixed point of F_{AF}).

It is well-known that stable extensions are preferred extensions but not vice versa. While stable extensions may not exist, grounded extension and preferred extensions always exist.

2.2. Abstract dialectical framework (ADF)

An *abstract dialectical framework (ADF)* [7] is a triple $D = (S, L, C)$ where

- S is a finite set of statements (positions, nodes),
- $L \subseteq S \times S$ is a set of links,
- $C = \{C_s\}_{s \in S}$ is a set of total functions $C_s : 2^{\text{par}(s)} \rightarrow \{in, out\}$, one for each $s \in S$.[1]
  C_s is called the acceptance condition of s.

The intuition of the acceptance condition $C_s(X) = in$ (resp. $C_s(X) = out$) is that when the statements in X are accepted (i.e. true) and those in $\text{par}(s) \setminus X$ are rejected (i.e. false) then s should be accepted (resp. rejected).

When s is rejected, we often say that the complement of s denoted by $\neg s$, is accepted.

Let X be a set of statements.

A *partial interpretation* of X is a set of assertions of the form $\Pi \cup \neg\Omega$ such that $\Pi, \Omega \subseteq X$ and $\Pi \cap \Omega = \emptyset$ where $\neg\Omega = \{\neg q \mid q \in \Omega\}$.

A *full interpretation* I of X is a partial interpretation of X such that for each statement $s \in X$, either s or $\neg s$ belongs to I.

The set of all partial interpretations of X is denoted by PI_X. Similarly, the set of all full interpretations of X is denoted by FI_X.

A set $K \subseteq X \cup \neg X$ is said to be *inconsistent* if $\exists s \in X$ such that $\{s, \neg s\} \subseteq K$.

It is obvious that the acceptance function C_s of any ADF $D = (S, L, C)$ could equivalently be defined as a total function $C_s : FI_{\text{par}(s)} \rightarrow \{in, out\}$.

Often it is convenient to represent the acceptance conditions as propositional formulas. For this reason and from now on, an ADF is represented as a pair $(S, \{\sigma_s\}_{s \in S})$ with σ_s being a propositional formula where there exists a link from a node r to s if r appears in σ_s [6].

Definition 1 ([6]). An *abstract dialectical framework (ADF)* is a pair $D = (S, C)$ where

- S is a finite set of statements (positions,nodes),
- $C = \{\sigma_s\}_{s \in S}$ is a set of propositional formulas over S where acceptance function C_s is defined by:

$$\forall I \in FI_{\text{par}(s)}: \quad C_s(I) = in \quad \text{iff} \quad I \models \sigma_s$$

Remark 1. From now on, whenever we mention an ADF, we refer to the above Definition 1.

Remark 2. We often present an ADF as a collection of expressions of the form $s[\sigma_s]$, one for each $s \in S$ like in Examples 1, 2.

Let $X \subseteq S$. The restriction on X of a partial interpretation I of S is defined by:

$$I \downarrow X = I \cap (X \cup \neg X)$$

[1] $\text{par}(s)$ is the set of all parents of s where r is a parent of s if there is a direct link from r to s.

Definition 2. Let $D = (S, C)$, $C = \{\sigma_s\}_{s \in S}$ be an ADF and $I = \Pi \cup \neg\Omega$, $\Pi, \Omega \subseteq S$, be a full interpretation of S.

I is said to be a *model* of D iff for each $s \in S$, $I \models \sigma_s$ iff $s \in \Pi$.

The semantics of ADFs are defined by identifying classes of models based on an operator Γ_D, referred to as the BW-operator in this paper, defined on partial interpretation I as follows [6,7]:

$$\Gamma_D(I) = \Pi \cup \neg\Omega \quad \text{where}$$

$$\Pi = \{s \mid \forall J \in FI_S : \text{if } J \supseteq I \text{ then } J \models \sigma_s\}$$

$$\Omega = \{s \mid \forall J \in FI_S : \text{if } J \supseteq I \text{ then } J \models \neg\sigma_s\}$$

Let $I \in PI_S$, i.e. I be a partial interpretation of S, and σ be a propositional formula over S. We write

$$I \models \sigma \quad \text{iff} \quad \forall J \in FI_S: \quad \text{if } J \supseteq I \quad \text{then } J \models \sigma$$

Remark 3. It is obvious that for each $I \in PI_S$, for each $s \in S$, $I \models \sigma_s$ iff $I \downarrow \text{par}(s) \models \sigma_s$.

Hence for each $I \in PI_{\text{par}(s)}$ and each $s \in S$, $I \models \sigma_s$ iff $\forall J \in FI_{\text{par}(s)}$: if $J \supseteq I$ then $J \models \sigma_s$.

The following lemma follows immediately from the definition of Γ_D.

Lemma 1. *Let $D = (S, \{\sigma_s\}_{s \in S})$ be an ADF and I be a partial interpretation of S. It holds that*

$$\Gamma_D(I) = \{s \mid I \models \sigma_s\} \cup \{\neg s \mid I \models \neg\sigma_s\}$$

As for any ADF D, Γ_D is monotonic (wrt set inclusion), it has a least fixed point representing the *BW-grounded model* of D [7].

The BW-grounded model of D represents the most skeptical semantics of ADFs. More creduluous semantics are represented by the *BESWW-complete models* of D defined as the fixed points of Γ_D [6]. The *BESWW-preferred models* of D are then defined as the maximal fixpoints of Γ_D.

Stable semantics of ADFs is defined in [6] and will be recalled later.

3. Simple argumentation frameworks and fixed points of BW-operators

We present in this section the simple argumentation frameworks whose extensions capture the classes of BW-grounded and BESWW-complete models of ADFs.

We first introduce the concept of immediate supports of a statement.

Definition 3 (i-supports)**.** A partial interpretation $M \in PI_{\text{par}(s)}$ is said to be an immediate support (or just i-support for short) for s wrt ADF $D = (S, \{\sigma_s\}_{s \in S})$ iff $M \models \sigma_s$.

M is said to be an i-support for $\neg s$ wrt D iff $M \models \neg\sigma_s$.

Remark 4. For convenience, we refer to statements or their complements (also often referred to as their negation) as assertions. A positive assertion about a statement s is s itself while a negative assertion about s is the negation of s. The complement of an assertion α is denoted by $\neg\alpha$.[2]

[2]Note again that the complement of $\neg s$ (resp. s) is s (resp. $\neg s$).

It follows immediately that

Lemma 2. *Let M be an i-support for an assertion α about a statement s. Then any partial interpretation $I \in PI_{\text{par}(s)}$ such that $M \subseteq I$, is also an i-support for α.*

The following simple lemma explains the interaction between i-supports for a statement and its complement.

Lemma 3. *Let $D = (S, \{\sigma_s\}_{s \in S})$ be an ADF and α be an assertion about a statement s and M be a partial interpretation over* par(s).
M is an i-support for α iff for each i-support N of $\neg\alpha$, $M \cup N$ is inconsistent.

Proof. The "only-if-direction" is obvious. We only need to prove the other direction.
Suppose for each i-support N of $\neg\alpha$, $M \cup N$ is inconsistent. Therefore, for each full interpretation $I \in FI_{\text{par}(s)}$ such that $M \subseteq I$, I can not be an i-support of $\neg\alpha$ (otherwise $M \cup I = I$ is inconsistent, contradicting the fact that I is an interpretation).
Let $\alpha = s$. Since I can not be an i-support of $\neg\alpha$ and $\neg\alpha \equiv \neg s$, $I \not\models \neg\sigma_s$. Hence $I \models \sigma_s$. Therefore $M \models \sigma_s$, i.e. M is an i-support of α.
Let $\alpha = \neg s$. Since I can not be an i-support of $\neg\alpha$ and $\neg\alpha \equiv s$, $I \not\models \sigma_s$. Hence $I \models \neg\sigma_s$. Therefore $M \models \neg\sigma_s$, i.e. M is an i-support of $\neg s$ and hence M is an i-support of α. $\square$

We can view an i-support J of an assertion α as an "argument" (J, α) for α. Lemma 3 allows us to establish the attack relation between arguments.
Let $I = \{\alpha_1, \ldots, \alpha_n\}$ be a partial interpretation. Suppose we have accepted "arguments" $(J_1, \alpha_1), \ldots, (J_n, \alpha_n)$. Further let s be some statement such that each "argument" (N, s) supporting s, is "attacked" by some argument (J_i, α_i) (i.e. $\neg\alpha_i \in N$). Therefore $I \cup N$ is inconsistent. From Lemma 3, it follows that I is an i-support of $\neg s$. Hence we could conclude $\neg s$.

Example 3. Consider an ADF $D = (S, \{\sigma_s\}_{s \in S})$, and $\sigma_s = \neg a \vee \neg b$ for some $s \in S$. It is obvious that any i-support for s contains either $\neg a$ or $\neg b$. Suppose you have accepted a set $\mathcal{A}$ of arguments such that both a and b are supported by some arguments in $\mathcal{A}$. Since any i-support for s contains either $\neg a$ or $\neg b$, it is "attacked" by some argument in $\mathcal{A}$. We hence would expect $\mathcal{A}$ to sanction the conclusion $\neg s$.

In other words, if each possible "argument" supporting s is "attacked" by accepted "arguments" then $\neg s$ should be accepted. This insight allows us to give a rather simple argumentation frameworks whose extensions capture the BW-grounded and BESWW-complete models.

Definition 4 (Simple argumentation frameworks)**.** Let $D = (S, C)$ be an ADF. The *simple argumentation framework of* D, *denoted by* $SAF_D = (SAR_D, satt_D)$, is defined as follows:

1. Each argument $A \in SAR_D$ has one of the following forms:
   (a) $A = (M, s)$ where s is a statement from S and M is an i-support for s.
   (b) $A = (\{\neg s\}, \neg s)$ where s is a statement from S.
   Note that arguments of the form $(\{\neg s\}, \neg s)$ are often written as $[\neg s]$.

   For any argument $A = (B, \alpha)$ from SAR_D, the conclusion and base of A, denoted by cnl(A) and base(A) respectively, are defined by cnl$(A) = \alpha$ and base$(A) = B$.

2. An argument A from SAR_D attacks an argument B from SAR_D (i.e. $(A, B) \in satt_D$) iff $\neg\,\mathrm{cnl}(A) \in \mathrm{base}(B)$.

Remark 5. For a set $\mathcal{A} \subseteq SAR_D$, $\mathrm{cnl}(\mathcal{A})$ denotes the set of all conclusions of arguments in $\mathcal{A}$.

Example 4. Let D_0 be the ADF in Example 1.

- $SAF_D = (SAR_D, satt_D)$ where
  - $SAR_D = \{A_0, A_1, B_0, B_1\}$ where $A_0 = [\neg a]$, $A_1 = (\{a\}, a)$, $B_0 = [\neg b]$, $B_1 = (\{\neg a\}, b)$.
  - $satt_D = \{(A_0, A_1), (A_1, A_0), (A_1, B_1), (B_1, B_0)\}$.
- The grounded extension of SAF_D is empty. The BW-grounded model is also empty.
- There are two preferred extensions $\{A_0, B_1\}$ and $\{A_1, B_0\}$ whose conclusions correspond to the BESWW-preferred models $\{\neg a, b\}$ and $\{a, \neg b\}$ respectively.

Let I be a partial interpretation over S. Define

$$\mathcal{A}_I = \big\{A \in SAR_D \mid \mathrm{base}(A) \subseteq I\big\}$$

Example 5. Let us continue Example 4.
Let $I_0 = \emptyset$. Then $\mathcal{A}_{I_0} = \emptyset$.
Let $I_1 = \{\neg a, b\}$. Then $\mathcal{A}_{I_1} = \{[\neg a], (\{\neg a\}, b)\} = \{A_0, B_1\}$.
Let $I_2 = \{a, \neg b\}$. Then $\mathcal{A}_{I_2} = \{(\{a\}, a), [\neg b]\} = \{A_1, B_0\}$.

The following theorem shows that the BESWW-complete models are captured by the complete extensions of the simple argumentation frameworks.

Theorem 1.

1. *Let $\mathcal{A}$ be a complete extension of SAF_D. Then $\mathrm{cnl}(\mathcal{A})$ is a fixed-point of Γ_D (and hence a BESWW-complete model of D).*
2. *Let I be a BESWW-complete model of D (and hence a fixed-point of Γ_D). Then $\mathcal{A}_I$ is a complete extension of SAF_D and $\mathrm{cnl}(\mathcal{A}_I) = I$.*

Proof. See Appendix A.3. □

The following theorem shows that the grounded (resp. preferred) extensions of the simple argumentation frameworks capture the BW-grounded model (resp. BESWW-preferred models).

Theorem 2. *Let D be an ADF.*

1. *Let M be the BW-grounded model of D and G be the grounded extension of SAF_D. It holds that*

$$\mathrm{cnl}(G) = M.$$

2. *Let M be the BESWW-preferred model of D. Then $\mathcal{A}_M$ is a preferred extension of SAF_D.*
3. *Let E be a preferred extension of SAF_D. Then $\mathrm{cnl}(E)$ is a BESWW-preferred model of D.*

Proof. See Appendix A.3. □

4. Rejection as assumption and normal argumentation frameworks

We first argue that the stable models of ADFs are based on a view of "rejection as assumption" by giving an equivalent characterization of them that reflect the view of "rejection as assumption" in a more direct way.

We then proceed to show that stable models of ADFs are captured by stable extensions of normal argumentation frameworks with arguments being support trees whose leaves are labelled by assumptions represented by rejected statements. The insight suggests two new semantics based on the grounded and preferred extensions of normal argumentation frameworks.

4.1. Stable models of ADFs

Let $D = (S, C)$, $C = \{\sigma_s\}_{s \in S}$ be an ADF and $I = \Pi \cup \neg\Omega$, $\Pi, \Omega \subseteq S$, be a full interpretation of S. The *reduct of D* wrt I [6] is the ADF $D^I = (\Pi, C^I)$, $C^I = \{\sigma_s[\mathit{false}/x : x \in \Omega]\}_{s \in \Pi}$.[3]

Remark 6. Note that C^I contains a formula $\sigma_s[\mathit{false}/x : x \in \Omega]$ only for $s \in \Pi$.

Definition 5 ([6])**.** I is said to be a *stable model* of D iff

1. I is a model of D, and
2. Π is the BW-grounded model of the reduct D^I.

Example 6. Consider the ADF D_0 in Example 1 recalled below for ease of reference.

$$b[\neg a] \qquad a[a]$$

Let $I = \{\neg a, b\}$. It is clear that I is a model of D. Further it is not difficult to see that $C_b[\mathit{false}/a] \equiv \mathit{true}$. Therefore $D^I = \{b[\mathit{true}]\}$. Obviously $\{b\}$ is the BW-grounded model of D^I. Hence I is a stable model of D.

Looking at Definition 5, one may wonder whether the condition that I is a model of D could be dropped. The following example shows that the answer is no.

Example 7. Let D be the ADF defined by

$$a[\neg b] \qquad b[a]$$

Let $I = \{a, \neg b\}$. It is not difficult to see that $C_a[\mathit{false}/b] \equiv \mathit{true}$. Therefore $D^I = \{a[\mathit{true}]\}$. Obviously $\{a\}$ is the BW-grounded model of D^I. But I is not a model of D and hence not a stable model of D.

The intuition of the stable models is rather simple: An interpretation $I = \Pi \cup \neg\Omega$ is stable iff assuming that the statements in Ω are rejected (i.e. false) would lead to the acceptance of the statements in Π.

This idea can be formalized in two steps:

– Construct a revised reduct of the ADF in which the statements in Ω are replaced by *false*.
– Show that the revised reduct derives exactly the statements in Π.

[3] $\sigma_s[\mathit{false}/x : x \in \Omega]$ is the formula obtained from σ_s by replacing each occurrence of statement $x \in \Omega$ by the value false.

Let $\Omega \subseteq S$ and $D = (S, C)$.
The *Ω-reduct* of D is the ADF $D_\Omega = (S, C_\Omega)$, $C_\Omega = \{\sigma_s[false/r : r \in \Omega]\}_{s \in S}$.

Remark 7. Note that in contrast to reducts, the Ω-reducts have the same set of statements like the original ADFs and hence the acceptance function C_Ω contains a formula $\sigma_s[false/r : r \in \Omega]$ for each statement in S.

We next introduce a generalization of a well-known immediate-consequence operator in definite logic programming:

$$\text{for } \Pi \subseteq S: \quad T_D(\Pi) = \{s \in S \mid \Pi \models \sigma_s\}$$

It is clear that T_D is monotonic.[4]
We next present an obvious but helpful lemma.

Lemma 4. *Let $D = (S, C)$, $C = \{\sigma_s\}_{s \in S}$ be an ADF and $\Pi, \Omega \subseteq S$ such that $\Pi \cup \Omega = S$ and $\Pi \cap \Omega = \emptyset$. Further let σ be a propositional formula over S and I be a partial interpretation over Π. The following property holds:*

$$I \models \sigma[false/x : x \in \Omega] \quad \textit{iff} \quad I \cup \neg\Omega \models \sigma$$

The following theorem captures the intuition of stable models explained shortly above.

Theorem 3. *Let $D = (S, C)$, $C = \{\sigma_s\}_{s \in S}$ be an ADF and $M = \Pi \cup \neg\Omega$, $\Pi, \Omega \subseteq S$, be a full interpretation of S. Let $D_\Omega = (S, C_\Omega)$.*
M is a stable model of D iff Π is the least-fixed point of T_{D_Ω}.

Proof. See Appendix A.1. □

Convention 1. *For ease of reference and understanding, from now on, whenever we refer to a stable model of an ADF D, we mean a full interpretation $M = \Pi \cup \neg\Omega$ s.t. Π is the least fixed point of T_{D_Ω}.*

Example 8. Consider again the ADF D_0 in Example 1 recalled below for ease of reference.

$$b[\neg a] \qquad a[a]$$

Let $M = \{\neg a, b\}$ and $\Omega = \{a\}$.
It is not difficult to see that $C_b[false/a] \equiv true$ and $C_a[false/a] \equiv false$. Therefore $D_\Omega = \{b[true], a[false]\}$. Obviously $\{b\}$ is the least fixed point of D_Ω. Hence M is a stable model of D.

[4] I.e. for $\Pi \subseteq \Pi' \subseteq S : T_D(\Pi) \subseteq T_D(\Pi')$.

4.2. Support trees

The intuition of the "rejection-as-assumption" view is captured by considering arguments as support trees where rejected assignments label the leaves of the trees.

Definition 6 (Support trees)**.** A *support tree* for an assertion α w.r.t. an ADF $D = (S, \{\sigma_s\}_{s \in S})$ is a finite tree τ with nodes labeled by assertions from $S \cup \neg S$ such that

1. the root is labeled by α;
2. every non-leaf node N of τ is labeled by some statement $s \in S$ such that if N has n children labeled by $\varphi_1, \ldots, \varphi_n$ then $\{\varphi_1, \ldots, \varphi_n\}$ is an i-support of s;
3. every leaf-node of τ is labeled with some negative assertion $\neg s \in \neg S$ or a statement s with $\sigma_s \equiv$ *true*.

α is often referred to as the *conclusion* of τ and denoted by $\text{cnl}(\tau)$. Furthermore the set of all negative assertions labeling the leaves of τ is called the *base* of τ and denoted by $\text{base}(\tau)$.

Remark 8. It is easy to see that if the conclusion of a support tree τ is a negative assertion $\neg s$, $s \in S$, then τ consists only of its root that is labelled by $\neg s$. Abusing the notation for simplicity, we also denote such trees by $[\neg s]$.

Remark 9. The set of the conclusions of support trees belonging to a set E of support trees is denoted by $\text{cnl}(E)$.

For illustration, Fig. 2 gives all support trees of the ADF D_1 in Example 2.

Let $\Omega, \Pi \subseteq S$ s.t. $\Pi \cap \Omega = \emptyset$. From Lemma 4 and the definition of the T-operator, it follows immediately that

$$T_{D_\Omega}(\Pi) = \{s \in S \mid \text{there is an i-support } J \text{ of } s \text{ wrt } D \text{ such that } J \subseteq \neg\Omega \cup \Pi\}^5$$

There is a close connection between the least fixed point of T_{D_Ω} operator and the notion of support tree.

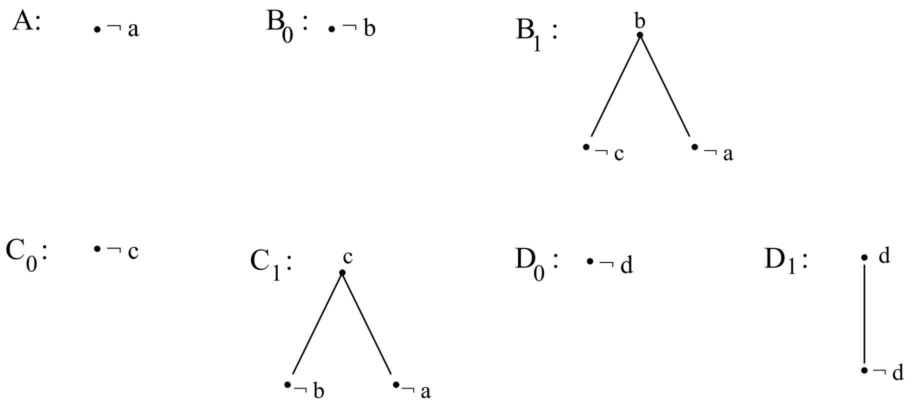

Fig. 2. Support trees.

[5] It is obvious to see that $s \in T_{D_\Omega}(\Pi)$ iff $\Pi \models \sigma_s[\mathit{false}/r : r \in \Omega]$ iff $\Pi \cap \text{par}(s) \models \sigma_s[\mathit{false}/r : r \in \Omega]$ iff $\Pi \cap \text{par}(s) \cup \neg\Omega \models \sigma_s$ iff $\Pi \cap \text{par}(s) \cup \neg(\Omega \cap \text{par}(s)) \models \sigma_s$. It is clear that $\Pi \cap \text{par}(s) \cup \neg(\Omega \cap \text{par}(s))$ is an i-support of s.

Lemma 5. *Let $\Omega \subseteq S$. The following equation holds:*

$$\mathrm{lfp}(T_{D_\Omega}) = \{s \in S \mid \textit{there is a support tree } \tau \textit{ for } s \textit{ wrt } D \textit{ such that } \mathrm{base}(\tau) \subseteq \neg\Omega\}$$

Proof. See Appendix A.2. □

4.3. Normal argumentation frameworks

Definition 7 (Normal AFs)**.** Let $D = (S, \{\sigma_s\}_{s \in S})$ be an ADF. Define the normal argumentation framework of D, denoted by $AF_D = (AR_D, att_D)$, as follows:

1. AR_D consists of all support trees wrt D.
2. An argument A from AR_D attacks an argument B from AR_D (i.e. $(A, B) \in att_D$) iff $\neg\,\mathrm{concl}(A) \in \mathrm{base}(B)$.

Remark 10. Note that there are no positive assertions in the base of any argument in AR_D. Therefore arguments of the form $[\neg s]$ never attack any other arguments in AR_D.

Example 9. The normal argumentation framework $AF_D = (AR_D, att_D)$ of the ADF D_1 in Example 2 is given by:

- The arguments are given in Fig. 2.
- $att_D = \{(B_1, B_0), (C_1, C_0), (B_1, C_1), (C_1, B_1), (D_1, D_0), (D_1, D_1)\}$.

The grounded extension is $\{A\}$.
There are two preferred extensions: $\{A, B_1, C_0\}$, $\{A, C_1, B_0\}$.
There is no stable extension.

We will show next that for any ADF D the stable extensions of the normal argumentation frameworks AF_D capture the BESWW-stable models.

Theorem 4. *Let AF_D be the normal argumentation framework of an ADF $D = (S, C)$.*

1. *Let $M = \Pi \cup \neg\Omega$ be a stable model of D. Then*

$$E = \{\tau \mid \tau \textit{ is a support tree wrt } D \textit{ such that } \mathrm{base}(\tau) \subseteq \neg\Omega\}$$

   is a stable extension of AF_D.
2. *Let E be a stable extension of AF_D. Then $\mathrm{cnl}(E)$ is a stable model of D.*

Proof. See Appendix A.2. □

4.4. New "Rejection as assumption" (RAA) – Semantics

As stable models represent a credulous approach to semantics of ADFs based on the intuition of "rejected statements as assumptions", there are other approaches based on different classes of extensions of the normal argumentation frameworks of the respective ADFs. For example, the set of preferred extensions define a new kind of credulous semantics generalizing the partial stable models in logic programming [15,19] while a new skeptical semantics for ADFs is defined by the grounded extension of their normal argumentation frameworks generalizing the well-founded semantics of logic programming [12].

Definition 8 (New semantics)**.** Let D be an ADF.

- The *well-founded semantics* of D is defined by the grounded extension of the normal argumentation framework AF_D.
- The *raa-preferential semantics* of D is defined by the set of preferred extensions of the normal argumentation framework AF_D.

Example 10. Let us continue with Example 9. The well-founded semantics is represented by the grounded extension $\{A\}$ while the raa-preferential semantics is represented by the preferred extensions $\{A, B_1, C_0\}$, $\{A, C_1, B_0\}$.

Note that the stable semantics is not defined as there is no stable extension.

5. Discussion and conclusion

We have showed that the semantics of ADFs could naturally be based on arguments and attacks. In other words, semantically, ADFs could be viewed as instances of abstract argumentation. The new insight allows us to adapt the standard concepts of abstract argumentation to ADFs in a straightforward and intuitive way. It also suggests new natural semantics for ADFs like the well-founded semantics or the rejection-as-assumptions (raa)-preferential semantics. This is not unlike the situation in logic programming where the semantical concepts in abstract argumentation help to explain and unify the semantics of logic programming.

Acknowledgements

Many thanks to Gerhard Brewka and Hannes Strass for many very helpful comments, especially for pointing out a mistake in an earlier version of this paper. Many thanks to the anonymous reviewers 1, 2 for the critical and constructive comments and suggestions. Thanks also to Sarah Gaggl for her cooperative spirit.

Appendix

A.1. Revised definition of stable models

Theorem 3. *Let $D = (S, C)$, $C = \{\sigma_s\}_{s\in S}$ be an ADF and $M = \Pi \cup \neg\Omega$, $\Pi, \Omega \subseteq S$, be a full interpretation of S. Let $D_\Omega = (S, C_\Omega)$.*

M is a stable model of D iff Π is the least-fixed point of T_{D_Ω}.

Proof.

1. Suppose Π is the least-fixed point of T_{D_Ω}.
   Let $x \in \Omega$. Since $C_\Omega(x)$ contains only statements from Π, it is clear that either $\Pi \models C_\Omega(x)$ or $\Pi \models \neg C_\Omega(x)$ holds.
   Since Π is a fixed point of T_{D_Ω}, it follows that $\Pi \models \neg C_\Omega(x)$ holds. Thus, the following holds:

$$\forall s \in S:\quad \Pi \models C_\Omega(s) \quad \text{iff} \quad s \in \Pi \tag{EQ}$$

(a) We first show that M is a model of D.
From the definition of $C_\Omega(s)$ and Lemma 4, it follows that for each $s \in S$, $s \in \Pi$ iff $\Pi \models C_\Omega(s)$ iff $M \models C(s)$.
M is hence a model of D.

(b) Let $M_0 = \emptyset$ and $M_{k+1} = \Gamma_{D^M}(M_k)$ and $\Pi_0 = \emptyset$ and $\Pi_{k+1} = T_{D_\Omega}(\Pi_k)$.
We show by induction that for all $k \geqslant 0$, $M_k = \Pi_k$.
It is clear that $M_0 = \Pi_0$. Suppose $M_k = \Pi_k$. We show that $M_{k+1} = \Pi_{k+1}$. From $\Pi_i \subseteq \Pi$ for all $i \geqslant 0$, it follows immediately that $\Pi_{k+1} = \{s \in \Pi \mid \Pi_k \models C_\Omega(s)\}$.
From the assertion (EQ) and $M_k = \Pi_k \subseteq \Pi$ and $C^M(s) = C_\Omega(s)$ for $s \in \Pi$, it follows that $\forall s \in \Pi$, $M_k \not\models \neg C_\Omega(s)$.
From $M_{k+1} = \{s \in \Pi \mid M_k \models C^M(s)\} \cup \{\neg s \in \neg\Pi \mid M_k \models \neg C^M(s)\} = \{s \in \Pi \mid M_k \models C_\Omega(s)\} \cup \{\neg s \in \neg\Pi \mid M_k \models \neg C_\Omega(s)\} = \{s \in \Pi \mid \Pi_k \models C_\Omega(s)\} = \Pi_{k+1}$.
Therefore $\Pi = \bigcup_{k\geqslant 0} M_k$.

We have proved that M is a stable model of D.

2. Suppose M is a stable model of D.
Let $M_0 = \emptyset$ and $M_{k+1} = \Gamma_{D^M}(M_k)$ and $J_0 = \emptyset$ and $J_{k+1} = T_{D_\Omega}(J_k)$.
Since M is a stable model of D, $\Pi = \bigcup_{k\geqslant 0} M_k$.
We show by induction that for all $k \geqslant 0$, $M_k = J_k$.
It is clear that $M_0 = J_0$. Suppose $M_k = J_k$.
$M_{k+1} = \{s \in \Pi \mid M_k \models C^M(s)\}$.
Since $C^M(s)$, $s \in \Pi$, contains only statements from Π and $C^M(s) = C_\Omega(s)$, it follows immediately that $M_{k+1} = \{s \in \Pi \mid M_k \models C^M(s)\} = \{s \in \Pi \mid J_k \models C_\Omega(s)\} \subseteq J_{k+1}$.
Suppose $M_{k+1} \neq J_{k+1}$. Let $s \in J_{k+1} \setminus M_{k+1}$. Therefore, $J_k \models C_\Omega(s)$. From $J_k = M_k \subseteq \Pi$, it follows that $\Pi \models C_\Omega(s)$ and hence $M \models C(s)$ (from Lemma 4). Since M is a model of D, it follows that $s \in \Pi$. From $M_k = J_k$, and $C^M(s) = C_\Omega(s)$, it follows that $M_k \models C^M(s)$. Thus $s \in M_{k+1}$ (contradiction).
From $\Pi = \bigcup_{k\geqslant 0} M_k$, it follows that $\Pi = \bigcup_{k\geqslant 0} J_k$. Hence Π is the least-fixed point of T_{D_Ω}. □

A.2. *Stable models are stable extensions*

Lemma 5. *Let $\Omega \subseteq S$. The following equation holds:*

$$\mathrm{lfp}(T_{D_\Omega}) = \big\{s \in S \mid \textit{there is a support tree } \tau \textit{ for } s \textit{ wrt } D \textit{ such that } \mathrm{base}(\tau) \subseteq \neg\Omega\big\}$$

Proof.

1. Let $\Pi_k = T^k_{D_\Omega}(\emptyset)$, $k \geqslant 0$. We show by induction that for each $k \geqslant 0$,

$$\Pi_k \subseteq \big\{s \in S \mid \text{there is a support tree } \tau \text{for } s \text{ wrt } D \text{ such that } \mathrm{base}(\tau) \subseteq \neg\Omega\big\}.$$

It is obvious that the inequality holds for $k = 0$.

$$\Pi_{k+1} = T_\Omega(\Pi_k) = \{s \in S \mid \text{there is an i-support } J \text{ of } s \text{ wrt } D \text{ such that } J \subseteq \neg\Omega \cup \Pi_k\}.$$

Let $s \in \Pi_{k+1}$ and J be an i-support of s such that $J = \{s_1, \ldots, s_n, \neg\omega_1, \ldots, \neg\omega_m\}$ such that $\{s_1, \ldots, s_n\} \subseteq \Pi_k$ and $\{\neg\omega_1, \ldots, \neg\omega_m\} \subseteq \neg\Omega$.

From the induction hypothesis, there are support trees $\tau_1, \dots, \tau_n$ s.t. $\mathrm{cnl}(\tau_i) = s_i$ and $\mathrm{base}(\tau_i) \subseteq \neg\Omega$, $1 \leqslant i \leqslant n$.
Let τ be the tree with root labelled by s and let the children of the root be labelled by $s_1, \dots, s_n$, $\neg\omega_1, \dots, \neg\omega_m$ where $s_1, \dots, s_n$ are themself the roots of support trees $\tau_1, \dots, \tau_n$. It is obvious that τ is a support tree for s such that $\mathrm{base}(\tau) \subseteq \neg\Omega$.

2. We show $\mathrm{lfp}(T_\Omega) \supseteq \{s \in S \mid$ there is a support tree τ for s wrt D such that $\mathrm{base}(\tau) \subseteq \neg\Omega\}$.
Let τ be a support tree for $s \in S$ such that $\mathrm{base}(\tau) \subseteq \neg\Omega$.
The height of τ denoted by $h(\tau)$, is the number of nodes of the longest path from the root to a node in τ.
We prove by induction on $h(\tau)$ that $s \in \mathrm{lfp}(T_\Omega)$.
Let τ be a tree of height 0 and $\mathrm{base}(\tau) \subseteq \neg\Omega$. It follows immediately that $C(s) \equiv \mathit{true}$. It is obvious that $s \in T_\Omega(\emptyset)$.
Let τ be a tree of height $k+1$ and $\mathrm{base}(\tau) \subseteq \neg\Omega$.
Let $s_1, \dots, s_n, \neg\omega_1, \dots, \neg\omega_m$ be the labels of the children of the root of τ. Let $\tau_1, \dots, \tau_n$ be the subtrees of τ whose roots are the children $s_1, \dots, s_n$ of the root of τ. It is clear that the heights of $\tau_1, \dots, \tau_n$ are less than or equal to k and their bases are subsets of $\neg\Omega$. It follows from the induction hypothesis that $s_i \in \mathrm{lfp}(T_\Omega)$, $1 \leqslant i \leqslant n$. Therefore for each i, there is j_i s.t. $s_i \in T_\Omega^{j_i}(\emptyset)$.
Let $j = \max\{j_1, \dots, j_n\}$. Thus for each i, $s_i \in T_\Omega^{j}(\emptyset)$.
Since $\{s_1, \dots, s_n, \neg\omega_1, \dots, \neg\omega_m\}$ is an i-support of s, it follows that $s \in T_\Omega^{j+1}(\emptyset) \subseteq \mathrm{lfp}(T_\Omega)$. $\square$

Theorem 4. *Let AF_D be the normal argumentation framework of an ADF $D = (S, C)$.*

1. *Let $M = \Pi \cup \neg\Omega$ be a stable model of D. Then*

$$E = \{\tau \mid \tau \textit{ is a support tree wrt } D \textit{ such that } \mathrm{base}(\tau) \subseteq \neg\Omega\}$$

is a stable extension of AF_D.

2. *Let E be a stable extension of AF_D. Then $\mathrm{cnl}(E)$ is a BESWW-stable model of D.*

Proof. We apply Convention 1 in both parts of the proof.

1. Let $D_\Omega = (S, C_\Omega)$ and $\Pi_k = T_{D_\Omega}^k(\emptyset)$.
From Convention 1 and Lemma 5, it follows that the following equations hold:

$$\begin{aligned} M = \Pi \cup \neg\Omega &= \mathrm{lfp}(T_\Omega) \cup \neg\Omega \\ &= \{s \in S \mid \text{there is a support tree } \tau \text{ for } s \text{ such that } \mathrm{base}(\tau) \subseteq \neg\Omega\} \cup \neg\Omega \\ &= \{\mathrm{cnl}(\tau) \mid \tau \text{ is a support tree such that } \mathrm{base}(\tau) \subseteq \neg\Omega\}. \end{aligned}$$

It is clear that $M = \mathrm{cnl}(E)$.
We show that E is a stable extension of AF_D. From $M = \mathrm{cnl}(E)$, it is clear that E is conflict-free.
Let $X \in AR_D \setminus E$. Therefore $\mathrm{base}(X) \setminus \neg\Omega \neq \emptyset$. Thus there exists $s \in \Pi$ such that $\neg s \in \mathrm{base}(X)$.
Since $M = \mathrm{cnl}(E)$, there is an argument $A \in E$ with $\mathrm{cnl}(A) = s$. Therefore A attacks X.
We have proved that E is a stable extension of AF_D.

2. Let $\neg\Omega = \{\neg s \mid [\neg s] \in E\}$.
   We first show that for each support tree τ, the following assertion holds:

$$\tau \in E \quad \text{iff} \quad \text{base}(\tau) \subseteq \neg\Omega \tag{AS}$$

   Let β be a support tree β such that base$(\beta) \subseteq \neg\Omega$. Suppose $\beta \notin E$. Therefore E attacks β. Hence $\exists\tau \in E$ such that τ attacks β. Therefore $\neg\,\text{cnl}(\tau) \in \text{base}(\beta) \subseteq \neg\Omega$. Hence $[\neg\,\text{cnl}(\tau)] \in E$. Thus E is not conflict-free (contradiction!). Therefore we have proved that if base$(\beta) \subseteq \neg\Omega$, then $\beta \in E$.
   Suppose $\beta \in E$. It is easy to see that each attack against a subargument of β is an attack against β. Therefore all subarguments of the form $[\neg s]$ of β belong to E. Hence base$(\tau) \subseteq \neg\Omega$.
   Let $M = \Pi \cup \neg\Omega$ where $\Pi = \{s \in S \mid \exists\tau \in E : \text{cnl}(\tau) = s\}$. It follows immediately from assertion (AS) that

$$\Pi = \big\{s \in S \mid \text{there is support tree } \tau \text{ of } s \text{ wrt } D \text{ such that } \text{base}(\beta) \subseteq \neg\Omega\big\}.$$

   From Lemma 5, lfp$(T_\Omega) = \Pi$. M is hence a stable model. □

A.3. BW-grounded and BESWW-preferred models are grounded and preferred extensions

Lemma 6. *Let I be a partial interpretation over S that is a fixed-point of Γ_D. Furthermore let $\alpha \in I$. Then there exists an argument $(B, \alpha) \in SAR_D$ such that $B \subseteq I$.*

Proof. Let $\alpha = s$, $s \in S$. Therefore $I \models \sigma_s$. Let $M = I \downarrow \text{par}(s)$. It is clear that M is an i-support of s. Hence $(M, s) \in SAR_D$.

Let $\alpha = \neg s$. Obviously $\{\neg s\} \subseteq I$ and $(\{\neg s\}, s) \in SAR_D$. □

Lemma 7. *Let $\mathcal{A}$ be a complete extension of SAF_D. The following properties hold:*

1. cnl$(\mathcal{A})$ *is consistent.*
2. $[\neg s] \in \mathcal{A}$ *iff* cnl$(\mathcal{A}) \downarrow \text{par}(s)$ *is an i-support of* $\neg s$.
3. *Let* $(M, s) \in SAR_D$, $s \in S$. *It holds that* $(M, s) \in \mathcal{A}$ *iff* $M \subseteq \text{cnl}(\mathcal{A})$.

Proof.

1. Suppose cnl$(\mathcal{A})$ is inconsistent. Hence $\exists s \in S : \{s, \neg s\} \subseteq \text{cnl}(\mathcal{A})$. Therefore there are arguments of both form (M, s), $[\neg s]$ in $\mathcal{A}$. Since (M, s) attacks $[\neg s]$, $\mathcal{A}$ is not conflict-free. Contradiction, as $\mathcal{A}$ is a complete extension.
2. Let $s \in S$. Since $\mathcal{A}$ is complete, it holds that:
   $[\neg s] \in \mathcal{A}$ iff

   for each $B = (X, s) \in SAF_D$, there exists $X_B \in \mathcal{A}$ attacking B iff
   for each $B = (X, s) \in SAF_D$, there exists $\alpha \in \text{cnl}(\mathcal{A})$ such that $\neg\alpha \in X$ iff
   for each $B = (X, s) \in SAF_D$, cnl$(\mathcal{A}) \cup X$ is inconsistent iff
   for each $B = (X, s) \in SAF_D$, $M \cup X$ is inconsistent where $M = \text{cnl}(\mathcal{A}) \downarrow \text{par}(s)$ (since X is a partial interpretation of par(s)) iff
   M is an i-support of $\neg s$ (from Lemma 3).

3. (a) Let $(M, s) \in \mathcal{A}$, $s \in S$. Let $\alpha \in M$. We show that $\alpha \in \mathrm{cnl}(\mathcal{A})$.
   There are two cases:

   – $\alpha \in S$. Therefore $[\neg\alpha]$ attacks (M, s). Since $\mathcal{A}$ is complete, there exists $B \in \mathcal{A}$ attacking $[\neg\alpha]$. It is obvious that $\mathrm{cnl}(B) = \alpha$. Hence $\alpha \in \mathrm{cnl}(\mathcal{A})$.
   – $\alpha = \neg t$, $t \in S$. Let $\mathcal{B}$ be the set of all arguments of the form (X, t). It is clear that $\mathcal{B}$ contains all arguments attacking (M, s) at α. $\mathcal{B}$ also contains all arguments attacking $[\neg t]$. Since $\mathcal{A}$ is complete, $\mathcal{A}$ attacks each $B \in \mathcal{B}$. Therefore $[\neg t]$ is defended by $\mathcal{A}$. Since $\mathcal{A}$ is complete, $[\neg t] \in \mathcal{A}$. Therefore $\alpha = \neg t \in \mathrm{cnl}(\mathcal{A})$.

   Therefore $M \subseteq \mathrm{cnl}(\mathcal{A})$.

   (b) Let $M \subseteq \mathrm{cnl}(\mathcal{A})$. Let C be an attack against (M, s). Let $\beta = \mathrm{cnl}(C)$. There are two cases:

   – $\beta \in S$. Therefore $\neg\beta \in M \subseteq \mathrm{cnl}(\mathcal{A})$. Therefore $[\neg\beta] \in \mathcal{A}$. Hence C attacks $[\neg\beta]$. Since $\mathcal{A}$ is complete, there exists $B \in \mathcal{A}$ s.t. B attacks C.
   – $\beta = \neg t$, $t \in S$ (i.e. $C = [\neg t]$). Therefore $t \in M \subseteq \mathrm{cnl}(\mathcal{A})$. Therefore there exists $B \in \mathcal{A}$ s.t. $\mathrm{cnl}(B) = t$. Hence B attacks C.

   We have proved that there exists $B \in \mathcal{A}$ s.t. B attacks C. Therefore $\mathcal{A}$ defends (M, s) against each attack against (M, s). Since $\mathcal{A}$ is complete, $(M, s) \in \mathcal{A}$. $\square$

Theorem 1.

1. *Let $\mathcal{A}$ be a complete extension of SAF_D. Then $\mathrm{cnl}(\mathcal{A})$ is a fixed-point of Γ_D.*
2. *Let I be a partial interpretation over S that is a fixed-point of Γ_D. Then $\mathrm{cnl}(\mathcal{A}_I) = I$ and $\mathcal{A}_I$ is a complete extension of SAF_D.*

Proof.

1. Let $I = \mathrm{cnl}(\mathcal{A})$. We show $I = \Gamma_D(I)$. Let $\alpha \in I$.

   – $\alpha = \neg s$ and $s \in S$. From Lemma 7, it follows immediately that:
   $\neg s \in I$ iff $[\neg s] \in \mathcal{A}$ iff $I \downarrow \mathrm{par}(s)$ is an i-support of $\neg s$ iff $\neg s \in \Gamma_D(I)$.
   – $\alpha = s$ and $s \in S$. From Lemma 7, it follows immediately that:
   $s \in I$ iff $\exists (M, s) \in \mathcal{A}$ iff $M \subseteq I$ iff $I \downarrow \mathrm{par}(s)$ is an i-support of s iff $s \in \Gamma_D(I)$.

2. Since I is a fixed point of Γ_D, it is obvious from Lemma 6 that $I \subseteq \mathrm{cnl}(\mathcal{A}_I)$.
   We show $\mathrm{cnl}(\mathcal{A}_I) \subseteq I$.
   Let $\neg s \in \mathrm{cnl}(\mathcal{A}_I)$, $s \in S$. Hence $[\neg s] \in \mathcal{A}_I$. Thus $\neg s \in I$ (from the definition of $\mathcal{A}_I$).
   Let $s \in \mathrm{cnl}(\mathcal{A}_I)$, $s \in S$. Hence $\exists (M, s) \in \mathcal{A}_I$. Hence $M \models \sigma_s$.
   From $M \subseteq I$, it follows $I \models \sigma_s$. Thus $s \in I$.
   We have proved that $I = \mathrm{cnl}(\mathcal{A}_I)$.

   (a) We show that $\mathcal{A}_I$ is admissible.
   Since I is consistent, $\mathcal{A}_I$ is conflict-free.
   Let $B \in SAR_D$ such that B attacks $\mathcal{A}_I$. We show that $\mathcal{A}_I$ attacks B. There are two cases:

   – $B = [\neg r]$ for $r \in S$. Therefore B attacks an argument $(M, s) \in \mathcal{A}_I$. Thus $r \in M$. From the definition of $\mathcal{A}_I$, it follows $M \subseteq I$. Hence $r \in I$. Since I is a fixed point of Γ_D, it follows from Lemma 6 that there exists an argument $C = (X, r) \in \mathcal{A}_I$. Therefore C attacks B. Hence $\mathcal{A}_I$ attacks B.

– $B = (N, r)$, $r \in S$. Therefore B attacks the argument $[\neg r] \in \mathcal{A}_I$. From the definition of $\mathcal{A}_I$, it follows that $\neg r \in I$. Let $Y = I \downarrow \text{par}(r)$. Since I is a fixed point of Γ_D, it is clear that $Y \models \neg\sigma_r$. Therefore for any i-support J for r, $J \cup Y$ is inconsistent. Since N is an i-support for r, $N \cup Y$ is inconsistent. Let t be an assertion such that $t \in Y$ and $\neg t \in N$. From $Y \subseteq I$, it follows that $t \in I$. From Lemma 6 it follows that there exists argument $A_t \in \mathcal{A}_I$ such that $\text{cnl}(A_t) = t$. Hence A_t attacks B. Thus $\mathcal{A}_I$ attacks B.

(b) Let A be an argument defended by $\mathcal{A}_I$. We show that $A \in \mathcal{A}_I$.
There are two cases:

i. $A = [\neg z]$.
Let $\mathcal{B}$ be the set of all arguments of the form (Z, z). It is clear that each argument in $\mathcal{B}$ attacks A. For each $B \in \mathcal{B}$, there exists $X_B \in \mathcal{A}_I$ attacking B. Let $N = \{\text{cnl}(X_B) \mid B \in \mathcal{B}\} \subseteq \text{cnl}(\mathcal{A}_I) = I$. It is clear that for each i-support J for z, $N \cup J$ is inconsistent. Therefore $N \models \neg\sigma_z$. From $N \subseteq I$ and I is a fixed point of Γ_D, it follows $\neg z \in I$. Therefore $A = [\neg z] \in \mathcal{A}_I$.

ii. $A = (M, s)$.
Let $\alpha \in M$. We show that $\alpha \in \text{cnl}(\mathcal{A}_I)$. There are two cases:

– $\alpha \in S$. Therefore $[\neg\alpha]$ attacks A. Therefore there exists an argument $B \in \mathcal{A}_I$ s.t. $\text{cnl}(B) = \alpha$. Hence $\alpha \in \text{cnl}(\mathcal{A}_I)$.
– $\alpha = \neg t$ and $t \in S$. Let $A' = [\neg t]$. It is clear that each attack against A' is an attack against A. Hence A' is defended by $\mathcal{A}_I$. From the previous case 2(b-i), we have $A' = [\neg t] \in \mathcal{A}_I$. Hence $\neg t \in \text{cnl}(\mathcal{A}_I)$. We have proved $\alpha \in \text{cnl}(\mathcal{A}_I)$.

Hence $M \subseteq \text{cnl}(\mathcal{A}_I) = I$. Hence $A \in \mathcal{A}_I$. $\square$

Lemma 8. *Let E be a complete extension of SAF_D. It holds that:*

$$\mathcal{A}_{\text{cnl}(E)} = E$$

Proof. Let $s \in S$.
$[\neg s] \in \mathcal{A}_{\text{cnl}(E)}$ iff $\neg s \in \text{cnl}(E)$ iff $[\neg s] \in E$.
$(M, s) \in \mathcal{A}_{\text{cnl}(E)}$ iff $M \subseteq \text{cnl}(E)$ iff (from Lemma 7) $(M, s) \in E$. $\square$

Theorem 2. *Let D be an ADF.*

1. *Let M be the BW-grounded model of D and G be the grounded extension of SAF_D. It holds that:*

$$\text{cnl}(G) = M.$$

2. *Let M be the BESWW-preferred model of D. Then $\mathcal{A}_M$ is a preferred extension of SAF_D.*
3. *Let E be a preferred extension of SAF_D. Then $\text{cnl}(E)$ is a BESWW-preferred model of D.*

Proof.

1. Let G be the grounded extension of SAF_D. From Theorem 1, $\text{cnl}(G)$ is a fixed point of Γ_D.
Since M is the least fixed point of the Γ_D, it follows that $M \subseteq \text{cnl}(G)$. Therefore $\mathcal{A}_M \subseteq \mathcal{A}_{\text{cnl}(G)}$. From Lemma 8, $\mathcal{A}_{\text{cnl}(G)} = G$. From Theorem 1, it follows that $\mathcal{A}_M$ is a complete extension. Since

G is grounded, it follows that $\mathcal{A}_M = G$. From Theorem 1, it follows that $\mathrm{cnl}(\mathcal{A}_M) = M$. Hence $M = \mathrm{cnl}(G)$.

2. From Theorem 1, $\mathcal{A}_M$ is a complete extension of SAF_D. Let E be a preferred extension of SAF_D s.t. $\mathcal{A}_M \subseteq E$. Therefore $\mathrm{cnl}(\mathcal{A}_M) \subseteq \mathrm{cnl}(E)$. Since $M = \mathrm{cnl}(\mathcal{A}_M)$ (Theorem 1), and $\mathrm{cnl}(E)$ is a fixed point of Γ_D (Theorem 1), it follows that $M = \mathrm{cnl}(E)$. Therefore $\mathcal{A}_M = \mathcal{A}_{\mathrm{cnl}(E)} = E$ (Lemma 8) implying that $\mathcal{A}_M$ is a preferred extension of SAF_D.
3. Let E be a preferred extension of SAF_D. Then $\mathrm{cnl}(E)$ is a fixed point of Γ_D. Let M be a preferred model of D s.t. $\mathrm{cnl}(E) \subseteq M$. Therefore $\mathcal{A}_{\mathrm{cnl}(E)} \subseteq \mathcal{A}_M$. From $\mathcal{A}_{\mathrm{cnl}(E)} = E$ (Lemma 8), it follows that $E = \mathcal{A}_M$. From $\mathrm{cnl}(\mathcal{A}_M) = M$ (Theorem 1), it follows that $\mathrm{cnl}(E) = \mathrm{cnl}(\mathcal{A}_M)$ is a preferred model of D. $\square$

References

[1] L. Amgoud and C. Cayrol, Inferring from inconsistency in preference-based argumentation frameworks, *J. Autom. Reasoning* **29**(2) (2002), 125–169. doi:10.1023/A:1021603608656.

[2] P. Baroni, F. Cerutti, M. Giacomin and G. Guida, Encompassing attacks to attacks in abstract argumentation frameworks, in: *ECSQARU, 2009*, 2009, pp. 83–94.

[3] T.J.M. Bench-Capon and K. Atkinson, Abstract argumentation and values, in: *Argumentation in AI*, Springer Verlag, 2009.

[4] A. Bondarenko, P.M. Dung, R.A. Kowalski and F. Toni, An abstract, argumentation-theoretic approach to default reasoning, *Artif. Intell.* **93** (1997), 63–101. doi:10.1016/S0004-3702(97)00015-5.

[5] G. Brewka, P.E. Dunne and S. Woltran, Relating the semantics of abstract dialectical framework and standard AFs, in: *Proc. IJCAI 2011*, 2011.

[6] G. Brewka, S. Ellmauthaler, H. Strass, J.P. Wallner and S. Woltran, Abstract dialectical frameworks revisited, in: *Proc. IJCAI/AAAI 2013*, 2013.

[7] G. Brewka and S. Woltran, Abstract dialectical frameworks, in: *Proc. of KR 2010*, AAAI Press, 2010, pp. 102–111.

[8] C. Cayrol and M. Lagasquie-Schiex, Bipolar abstract argumentation system, in: *Argumentation in AI*, Springer Verlag, 2009.

[9] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games, *Artif. Intell.* **77**(2) (1995), 321–358. doi:10.1016/0004-3702(94)00041-X.

[10] K. Eshghi and R.A. Kowalski, Abduction compared with negation by failure, in: *Logic Programming: Proceedings of the Sixth International Conference*, The MIT Press, Cambridge, and Massachusetts, 1989, pp. 234–254.

[11] D.M. Gabbay, Equational approach to argumentation networks, *Argument and Computation* **3**(2–3) (2012), 87–142. doi:10.1080/19462166.2012.704398.

[12] A.V. Gelder, K.A. Ross and J.S. Schlipf, The well-founded semantics for general logic programs, *J. of ACM* **38**(3) (1991), 630–650.

[13] M. Gelfond and V. Lifschitz, The stable model semantics for logic programming, in: *Proc. of 5th ICLP*, MIT Press, 1988, pp. 1070–1080.

[14] D.D. Hanh, P.M. Dung, N.D. Hung and P.M. Thang, Inductive defence for skeptical semantics of extended argumentation, *J. of Logic and Computation* **21**(2) (2011), 307–349. doi:10.1093/logcom/exq018.

[15] A.C. Kakas and P. Mancarella, Preferred extensions are partial stable models, *Journal of Logic Programming* **14** (1992), 341–348. doi:10.1016/0743-1066(92)90015-U.

[16] S. Modgil, Reasoning about preferences in argumentation frameworks, *Artif. Intell.* **173**(9–10) (2009), 901–934. doi:10.1016/j.artint.2009.02.001.

[17] S. Nielsen and S. Parsons, A generalization of Dung's abstract framework for argumentation: Arguing with sets of attacking argments, in: *Argumentation in Multi-Agent Systems*, LNCS, Vol. 4766, 2006, pp. 54–73. doi:10.1007/978-3-540-75526-5_4.

[18] T.C. Przymunsinski, The well-founded semantics coincides with the three-valued stable semantics, *Fundamenta Informaticae* **13**(4) (1990), 445–463.

[19] D. Sacca and C. Zaniolo, Stable models and nondeterminism for logic programs with negation, in: *Proc. of the ACM SIGMOD-SIGACT Symposium on Principles of Database Systems*, 1990, pp. 205–217.

[20] H. Strass, Approximating operators and semantics of abstract dialectical frameworks, *Artificial Intelligence* **205** (2013), 39–70.

[21] H. Strass, Expressiveness of two-valued semantics for ADF, *JAIR* (2015).

[22] H. Strass and P.W. Wallner, Analyzing the computational complexity of abstract dialectical frameworks via approximation fixpoint theory, *Artificial Intelligence* **226** (2015), 34–74. doi:10.1016/j.artint.2015.05.003.

[23] Y. Wu, M. Caminada and D.M. Gabbay, Complete extensions in argumentation coincide with 3-valued stable models in logic programming, *Studia Logica* **93** (2009), 383–403. doi:10.1007/s11225-009-9210-5.