

On the Decomposition of Abstract Dialectical Frameworks and the Complexity of Naive-based Semantics

Sarah Alice Gaggl

*Logic Programming and Argumentation Group
Faculty of Computer Science
TU Dresden, Germany*

SARAH.GAGGL@TU-DRESDEN.DE

Sebastian Rudolph

*Computational Logic Group
Faculty of Computer Science
TU Dresden, Germany*

SEBASTIAN.RUDOLPH@TU-DRESDEN.DE

Hannes Strass

*Intelligent Systems Group
Computer Science Institute
Leipzig University, Germany*

STRASS@INFORMATIK.UNI-LEIPZIG.DE

Abstract

Abstract dialectical frameworks (ADFs) are a recently introduced powerful generalization of Dung's popular abstract argumentation frameworks (AFs). Inspired by similar work for AFs, we introduce a decomposition scheme for ADFs, which proceeds along the ADF's strongly connected components. We find that, for several semantics, the decomposition-based version coincides with the original semantics, whereas for others, it gives rise to a new semantics. These new semantics allow us to deal with pertinent problems such as odd-length negative cycles in a more general setting, that for instance also encompasses logic programs. We perform an exhaustive analysis of the computational complexity of these new, so-called naive-based semantics. The results are quite interesting, for some of them involve little-known classes of the so-called Boolean hierarchy (another hierarchy in between classes of the polynomial hierarchy). Furthermore, in credulous and sceptical entailment, the complexity can be different depending on whether we check for truth or falsity of a specific statement.

1. Introduction

Over the last decade, argumentation theory emerged as one of the major fields in artificial intelligence and non-monotonic reasoning. In this area, abstract argumentation frameworks (AFs) as introduced by Dung (1995) became a key formalism with applications to a variety of non-monotonic reasoning problems such as logic programming, inconsistency handling, legal reasoning and many others (Rahwan, Simari, & van Benthem, 2009; Toni & Sergot, 2011; Amgoud & Vesic, 2010; Zhang & Lin, 2009).

The basic Dung-style framework only comprises a set of *abstract arguments* and a binary relation between them, denoted as *attacks*. The evaluation of such an AF is then based on model-theoretic semantics, by means of which it is specified which sets of arguments can “stand together”. The need to represent more complex relations between the abstract entities led to various extensions of the original AFs, allowing to handle preferences and values

on arguments (Amgoud & Cayrol, 2002; Bench-Capon, 2003), weights (Dunne, Hunter, McBurney, Parsons, & Wooldridge, 2011), probabilities (Li, Oren, & Norman, 2011) or introducing a positive relation between arguments, so-called *supports* (Amgoud, Cayrol, Lagasque-Schieux, & Livet, 2008; Nouioua & Risch, 2011). Recently, abstract dialectical frameworks (ADFs) have been introduced (Brewka & Woltran, 2010; Brewka, Strass, Ellmauthaler, Wallner, & Woltran, 2013) as a powerful generalisation of Dung’s framework. ADFs allow for more general interactions between arguments, for example support, joint attack, joint support and mixed combinations. Furthermore, ADFs can handle preferences, on both the statements and the links (Brewka & Woltran, 2010; Brewka et al., 2013), as well as probabilities (Polberg & Doder, 2014). Finally, ADFs can not only be seen as an extension of Dung’s AFs but also as a target language for compilation from more concrete and application-based languages, and thus, serve as “argumentation middleware” (Brewka et al., 2013), e.g. Carneades (Brewka & Gordon, 2010), simple defeasible theories (Strass, 2018), and recent applications to legal reasoning and reasoning with cases by Al-Abdulkarim et al. (Al-Abdulkarim, Atkinson, & Bench-Capon, 2014, 2015, 2016; Al-Abdulkarim, Atkinson, Bench-Capon, Whittle, Williams, & Wolfenden, 2017).

This generality is achieved by using versatile acceptance conditions for the statements – Boolean functions determining the acceptance of a statement s depending on the acceptance of its parents, the statements with a link to s . These acceptance functions can also be represented as propositional formulas φ_s , thus the status of a statement s can be obtained by the evaluation of φ_s . For example, the AF-like relationship where statements a and b *individually attack* c can then be expressed by $\varphi_c = \neg a \wedge \neg b$. That is, c is accepted (true) if neither of its attackers is accepted (true). A *set attack* from a and b to c is written as $\varphi_c = \neg a \vee \neg b$ where c is only rejected if both a and b are accepted. The same works for *support*: $\varphi_c = a \wedge b$ means that c needs support from both a and b , and $\varphi_c = a \vee b$ says that c can be accepted if at least one of a or b is accepted. Clearly these different “acceptance patterns” can be arbitrarily combined, leading to an expressive language for abstract argumentation (Strass, 2015). Most of the semantics of ADFs are defined over the acceptance conditions, however additionally the links between the statements are explicitly represented in the same way as it is done in AFs. This does not only have the advantage of the handy representation as a directed graph, it also provides information about the structure of an ADF, like cycles and strongly connected components (SCCs).

As usual, the greater expressiveness of a formalism comes with a price. In our case the computational complexity of semantics for ADFs is in general higher than for AFs (Strass & Wallner, 2015). A successful way of dealing with big or complex problems is to split them into smaller sub-problems where it is easier to find a solution. The overall solution then consists of a combination of the solutions of all sub-problems.¹ Here, we propose an approach to decompose ADFs along their SCCs. While our approach is inspired by similar work on AFs by Baroni, Giacomin, and Guida (2005), there are important differences.

First, the SCC-recursive schema for AFs is based on a recursive decomposition of an AF along its SCCs, where in each step the semantics are computed for sub-frameworks consisting of single SCCs. The SCCs of an AF can change during the computation, depending on

1. For AFs, there is already some evidence that decomposition benefits computational efficiency (Baumann, Brewka, & Wong, 2011; Liao, 2013, 2014).

the outcome of the semantics from the previous SCCs. In particular, arguments which are attacked from outside their SCCs from an accepted argument are eliminated, which can change the remaining SCCs of the framework. As ADFs allow complex acceptance conditions for statements one needs a way to pass on the outcome of preceding components to all acceptance conditions of statements depending on them and additionally handle the change of SCCs.

Second, when acceptance conditions are represented as propositional formulas, there might be redundancies in the representation. For instance, the formula $\varphi_s = a \vee \top$ always evaluates to true, hence a is redundant in φ_s and can be removed. However this redundant information would also be given in the links of the ADF and may lead to dependencies in the graph that are not actually present. Hence, a pure decomposition along SCCs would not work correctly.

Third, in the AF case, some semantics are defined in a simplified version of the general SCC-recursive schema, namely for *stable*, *cf2* and *stage2* the notion of defence is weakened (Dvořák & Gaggl, 2016). The recursive procedure we propose allows to propagate already obtained information on the acceptance state of statements to others that depend on them. Within these propagation steps, redundant information is identified and eliminated. It turns out that our approach is indeed a generalisation of the SCC-recursive schema for AFs, as it allows to compute all standard admissible-based semantics and the naive-based ones within the same procedure. Hence, it can also be seen as an alternative characterisation of the general SCC-recursive schema like the one for *cf2* and *stage2* semantics presented by Gaggl and Woltran (2013) and Dvořák and Gaggl (2016).²

As is the case for AFs, our decomposition technique is able to reconstruct the standard semantics (admissible, complete, preferred, model), and leads to new semantics when using semantics based on conflict-freeness (such as naive or stage) in each strongly connected component. This leads to the new ADF semantics *nai₂* and *stg₂* which are the ADF counterparts to the AF semantics *cf2* (Baroni et al., 2005) and *stage2* (Dvořák & Gaggl, 2016). The notion of conflict-freeness we use here is three-valued (Strass, 2013), and a direct generalisation of the one used by Baroni, Caminada, and Giacomin (2011) (Definition 16). Basically, it says that in order to be conflict-free, a three-valued interpretation must adhere to certain “justification standards” when accepting or rejecting statements: If a statement s is assigned truth value true (is accepted), then this must be justified by statement s being possibly acceptable, that is, there must be an assignment to the remaining statements such that the acceptance condition of s is fulfilled. If a statement is assigned truth value false (is rejected), then this must be justified by statement s *not* being possibly acceptable, that is, a satisfying assignment of the acceptance condition must not exist. We opted for this version of three-valued conflict-freeness to preserve downward compatibility with AFs; there is also a recent alternative definition of three-valued conflict-free interpretations by Strass and Wallner (2015). Their definition is more permissive; it has the same requirement for accepted statements, and stipulates that for rejected arguments s , there must be an assignment to the remaining statements such that the acceptance condition of s is *not* fulfilled. So where our definition uses satisfiability and unsatisfiability, the definition of

2. Interestingly, Linsbichler (2014) independently presented a methodology for directionally *splitting* abstract dialectical frameworks. Applying his splitting technique along the strongly connected components of an ADF exactly yields our procedure.

Strass and Wallner (2015) uses satisfiability and refutability. Some more discussion of these two variants of three-valued conflict-freeness in ADFs was given by Strass (2017).

As hinted above, our decomposition procedure defines new semantics for ADFs. In this paper, we also study the computational complexity of these newly defined naive-based semantics for abstract dialectical frameworks. Generally speaking, typical reasoning problems for ADFs are model verification, sceptical and credulous reasoning and existence of a non-trivial interpretation. Analysing the computational complexity of these reasoning problems is a crucial topic for theoretical and practical reasons. First, complexity results often serve as an indicator for how difficult and how expressive (in a problem-encoding sense) a reasoning task can be. Second, knowing about the complexity of a reasoning problem is essential for the development of adequate algorithms and systems. A comprehensive complexity analysis of the ADF semantics defined by Brewka et al. (2013) has been given by Strass and Wallner (2015). Consequently, we “only” need to analyse all reasoning tasks mentioned earlier (model verification, non-trivial existence, and credulous and sceptical reasoning) for the conflict-free, naive, stage, *nai*₂ and *stg*₂ semantics. The results show that these tasks are (sometimes considerably) more difficult than their counterparts in AFs. While for the standard Dung semantics (admissible, preferred, complete, stable), their ADF generalisations are mildly more complex (one level up in the polynomial hierarchy; Strass and Wallner, 2015), for the naive-based semantics, the differences can be far more significant. For example, deciding whether an argument is true in every naive extension can be done in logarithmic space for AFs (Spanring, 2012, Lemma 2.5.10), while it is hard at least for the second level of the polynomial hierarchy in the case of ADFs. The complexity becomes even higher (completeness for the third level) if we want to check whether a statement is *false* in every naive interpretation of an ADF. In general, different complexities for entailment of truth and entailment of falsity seems to be quite uncommon in logic-based formalisms. We can trace the reason for this difference in naive-based semantics for ADFs back to the definition of a conflict-free interpretation, which basically requires different strengths of justification depending on which truth value is assigned to a statement as discussed above. Quite possibly even more interesting (and the hardest proof of all our results) is the complexity of deciding existence of non-trivial conflict-free interpretations. We show that the problem is complete for the second level of the Boolean hierarchy (Wechsung, 1985). The Boolean hierarchy consists of classes that are composed of Boolean combinations of problems from NP and complements thereof. A somewhat better-known example is the class DP, a logical “and” of one NP- and one coNP-problem.

In addition to analysing the computational complexity of decision problems related to general ADFs, we also study the same complexities for a proper subclass, *bipolar* abstract dialectical frameworks. Bipolar ADFs have been recognised as representationally expressive while computationally attractive (Strass, 2015): they retain many of the means of expression of general ADFs and are thus strictly more expressive than AFs, while computation with all the standard semantics for bipolar ADFs is of the same complexity as for AFs. Roughly, from AFs to general ADFs, complexities increase by one level of the polynomial hierarchy; from general ADFs to bipolar ADFs, complexities decrease by one level of the polynomial hierarchy. The same drop (by one level of the polynomial hierarchy when going from general to bipolar ADFs) can be observed for the semantics newly defined in this paper, providing

further support to the idea that bipolar ADFs are an attractive formalism for representing knowledge about arguments and their interrelationships.

The remainder of the paper is structured as follows. We introduce the necessary background on ADFs, existing semantics, SCC-recursiveness, and complexity theory in Section 2. In Section 3, we generalise the SCC-recursive decomposition procedure of Baroni et al. (2005) to ADFs and analyse the resulting semantics. Then the complexity analysis is performed in Section 4, where we grouped our results according to the two analysed ADF classes, general (Section 4.1) and bipolar (Section 4.2), that are further compartmentalised into sections corresponding to the four decision problems verification (Sections 4.1.1 and 4.2.1), existence of a non-trivial interpretation (Sections 4.1.2 and 4.2.2), and entailment (Sections 4.1.3 and 4.2.3). We conclude the paper in Section 5.

This article combines and significantly extends the results of Gaggl and Strass (2014) as well as Gaggl, Rudolph, and Strass (2015). We introduce the decomposition scheme and its semantics in more detail and with more examples. We provide detailed proofs of all previously published results as well as a completely new section on the computational complexity of *bipolar* ADFs for the mentioned semantics containing 30 novel complexity results, all of which are previously unpublished.

2. Background

For functions $f : A \rightarrow B$ and $g : C \rightarrow D$, we denote the *update of f with g* by $f \circ g$, where

$$f \circ g : A \cup C \rightarrow B \cup D \quad \text{with} \quad x \mapsto \begin{cases} g(x) & \text{if } x \in C \\ f(x) & \text{otherwise} \end{cases}$$

So even if $x \in A \cap C$ and $f(x)$ is defined, we have $(f \circ g)(x) = g(x)$. For a function $f : A \rightarrow B$ and $b \in B$ we denote $f^{-1}(b) = \{a \in A \mid f(a) = b\}$. For $A' \subseteq A$ the function $f|_{A'} : A' \rightarrow B$ is the restriction of f 's domain to A' . To define a constant function $f : A \rightarrow B$, we sometimes write $f = (A \mapsto b)$ for one $b \in B$ and take it to mean that $f(a) = b$ for all $a \in A$. The disjoint union of sets A and B is denoted by $A \uplus B$.

We will make use of many standard concepts of classical propositional logic in this paper, including the usual notions of formulas, interpretations and models, as well as satisfiability and refutability. Our analysis in this paper will be based on three-valued interpretations, mappings $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ that assign one of the truth values true (\mathbf{t}), false (\mathbf{f}) or unknown (\mathbf{u}) to each statement. A comparable treatment for AFs was given by the three-valued argumentation stages of Verheij (1996). For uniformity among logic-based and argumentation-based formalisms, in this paper we use standard notation and terminology from mathematical logic.³

The three truth values are partially ordered by \leq_i according to their information content: we have $\mathbf{u} <_i \mathbf{t}$ and $\mathbf{u} <_i \mathbf{f}$ and no other pair in $<_i$, in particular \mathbf{t} and \mathbf{f} are incomparable, which intuitively means that the classical truth values contain more information than the truth value unknown. The information ordering \leq_i extends in a straightforward way to

3. This does not mean we use any particular three-valued logic with its truth tables; rather, \mathbf{u} serves as a placeholder that can be replaced by any of the two classical truth values before the formula is actually evaluated.

valuations v_1, v_2 over S in that $v_1 \leq_i v_2$ iff $v_1(s) \leq_i v_2(s)$ for all $s \in S$. The \leq_i least element of the set of all valuations is the valuation mapping all statements to unknown – the least informative interpretation. Obviously, a three-valued interpretation v is two-valued if all statements are mapped to either true or false. Such two-valued interpretations are \leq_i -maximal.

A particular non-standard notion we use is that of the partial evaluation of a formula. Given a three-valued interpretation v and a formula φ , the partial evaluation of φ with v takes the two-valued part of v and replaces the evaluated variables by their truth values.

Definition 2.1. Let φ be a propositional formula over vocabulary S and for an $M \subseteq S$ let $v : M \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ be a three-valued interpretation. The *partial valuation of φ by v* is

$$\varphi^v = \varphi[p/\top : v(p) = \mathbf{t}][p/\perp : v(p) = \mathbf{f}]$$

For example, consider the propositional formula $\varphi = a \vee (b \wedge c)$ and the interpretation $v_1 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{u}\}$. Statement c with $v_1(c) = \mathbf{u}$ will remain in φ , while a and b are replaced, and we get $\varphi^{v_1} = \perp \vee (\top \wedge c)$. This formula is equivalent to c and thus both satisfiable (by $\{c \mapsto \mathbf{t}\}$) and refutable (by $\{c \mapsto \mathbf{f}\}$). In contrast, for $v_2 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}\}$ the formula $\varphi^{v_2} = \top \vee (b \wedge c)$ is irrefutable; for $v_3 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}\}$ the formula $\varphi^{v_3} = \perp \vee (\perp \wedge c)$ is unsatisfiable.

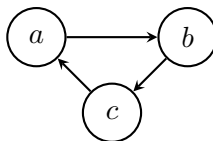
2.1 Abstract Argumentation Frameworks

In this section we introduce the basics of abstract argumentation and the semantics we need for further investigations. We first give the formal definition of abstract argumentation frameworks as introduced by Dung (1995).

Definition 2.2. An *argumentation framework (AF)* is a pair $F = (A, R)$, where A is a finite set of arguments and $R \subseteq A \times A$ is the attack relation. The pair $(a, b) \in R$ means that a attacks b . A set $S \subseteq A$ of arguments attacks $b \in A$ (in F), if there is an $a \in S$ such that $(a, b) \in R$.

Such an AF is typically represented as a directed graph as shown in the following example.

Example 2.3. Consider the AF $F = (A, R)$ where the set of arguments is given by $A = \{a, b, c\}$ and the set of attacks is defined as $R = \{(a, b), (b, c), (c, a)\}$ as depicted below:



Conflicts between arguments are resolved on a semantical level. An argument can either be accepted, rejected or it is undecided whether to accept or reject the argument. Here we will use the notion of labellings, as they directly correspond to three-valued interpretations of ADFs. For an overview about labellings for most argumentation semantics we refer to Baroni et al. (2011). Thus, accepted arguments are labelled with \mathbf{t} (true), rejected ones with \mathbf{f} (false) and undecided ones with \mathbf{u} .

For an AF $F = (A, R)$, a *labelling* is a total function $v : A \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$. Then, a labelling can be denoted as a triple $v = (v_{\mathbf{t}}, v_{\mathbf{f}}, v_{\mathbf{u}}) = (v^{-1}(\mathbf{t}), v^{-1}(\mathbf{f}), v^{-1}(\mathbf{u}))$ corresponding to a partition of A . Following Baroni et al. (2011), conflict-free and naive labellings are given as follows.

Definition 2.4. Let $F = (A, R)$ be an AF.

- v is a *conflict-free labelling* of F , i.e. $v \in cfi(F)$, iff
 - for all $a \in v_{\mathbf{t}}$ there is no $b \in v_{\mathbf{t}}$ such that $(a, b) \in R$,
 - for all $a \in v_{\mathbf{f}}$ there exists a $b \in v_{\mathbf{t}}$ such that $(b, a) \in R$.
- Then, v is a *naive labelling* of F , i.e. $v \in nai(F)$, iff $v \in cfi(F)$ and there is no $v' \in cfi(F)$ with $v <_i v'$.
- Finally, v is a *stage labelling* of F , i. e. $v \in stg(F)$, iff $v \in cfi(F)$ and there is no $v' \in cfi(F)$ with $v'_{\mathbf{u}} \subsetneq v_{\mathbf{u}}$.

Example 2.5. Let $F = (A, R)$ be the AF from Example 2.3. We obtain the following conflict-free labellings.

$$cfi(F) = \{ \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}\}, \{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}\}, \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{u}\}, \\ \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{t}\}, \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}\}, \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}\}, \\ \{a \mapsto \mathbf{f}, b \mapsto \mathbf{u}, c \mapsto \mathbf{t}\} \}$$

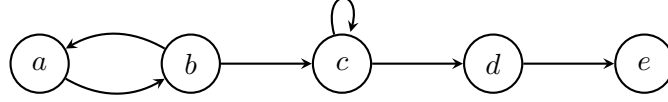
The *cf2* semantics is based on a decomposition along the SCCs of an AF. Hence, we require some further formal machinery. By $SCCs(F)$, we denote the set of *strongly connected components* of an AF $F = (A, R)$, i.e. sets of vertices of the maximal strongly connected⁴ sub-graphs of F . Moreover, for an $a \in A$, we denote by $C_F(a)$ the component of F where a occurs in, i.e. the (unique) set $C \in SCCs(F)$, such that $a \in C$. It turns out to be convenient to use two different concepts to obtain sub-frameworks of AFs. Let $F = (A, R)$ be an AF and $S \subseteq A$. Then, $F|_S = ((A \cap S), R \cap (S \times S))$ is the *sub-framework* of F w.r.t. S , and we also use $F - S = F|_{A \setminus S}$. We note the following relation (which we use implicitly later on), for an AF F and sets S, S' : $F|_{S \setminus S'} = F|_S - S' = (F - S')|_S$. We now give the definition of the *cf2* semantics in form of labellings (Baroni et al., 2011).

Definition 2.6. Let $F = (A, R)$ be an AF and v be a labelling of F . Any $b \in A$ is *component-defeated* by $v_{\mathbf{t}}$ (in F), if $\exists a \in v_{\mathbf{t}}$, s.t. $(a, b) \in R$ and $a \notin C_F(b)$. The set of arguments component-defeated by $v_{\mathbf{t}}$ in F is denoted by $D_F(v_{\mathbf{t}})$. Then, v is a *cf2 labelling* of F , i.e. $v \in cf2(F)$, iff

- $v \in nai(F)$, in case $|SCCs(F)| = 1$;
- otherwise, $\forall C \in SCCs(F), v|_{C \setminus D_F(v_{\mathbf{t}})} \in cf2(F|_C - D_F(v_{\mathbf{t}})), D_F(v_{\mathbf{t}}) \subseteq v_{\mathbf{f}}$.

4. A directed graph is called *strongly connected* if there is a directed path from each vertex in the graph to every other vertex of the graph.

Example 2.7. To illustrate the difference between naive, stage and *cf2* labellings we consider the AF shown below:



This AF possesses the following labellings.

$$\begin{aligned}
 nai(F) &= \{ \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, d \mapsto \mathbf{t}, e \mapsto \mathbf{f}\}, \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}, d \mapsto \mathbf{t}, e \mapsto \mathbf{f}\}, \\
 &\quad \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, d \mapsto \mathbf{u}, e \mapsto \mathbf{t}\}, \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}, d \mapsto \mathbf{u}, e \mapsto \mathbf{t}\} \} \\
 stg(F) &= \{ \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, d \mapsto \mathbf{t}, e \mapsto \mathbf{f}\} \} \\
 cf2(F) &= \{ \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, d \mapsto \mathbf{t}, e \mapsto \mathbf{f}\}, \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}, d \mapsto \mathbf{t}, e \mapsto \mathbf{f}\} \}
 \end{aligned}$$

The naive labellings are clearly conflict-free according to Definition 2.4. For instance in the labelling $\{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}, d \mapsto \mathbf{u}, e \mapsto \mathbf{t}\}$, a is set to \mathbf{t} and the only attacker of a is the argument b which is set to \mathbf{f} . Regarding the argument c , it is attacked by b and c , although b is set to \mathbf{f} , c cannot be set to \mathbf{t} , as it is self-attacking. The only choice left is to make c undecided. Argument e is set to \mathbf{t} , which is valid, as d is not set to \mathbf{t} as well. To show that the labelling is also naive, we need to demonstrate that there is no other conflict-free labelling higher up in the information ordering. The only possible changes would be for arguments c and d as they are undecided, but both cannot be set to either \mathbf{t} nor \mathbf{f} without violating the definition of conflict-free labellings. Thus, $\{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}, d \mapsto \mathbf{u}, e \mapsto \mathbf{t}\}$ is a naive labelling of F .

For the only stage labelling, it is easy to see that it is also a naive labelling (thus conflict-free) and as it does not have any undecided arguments, it clearly satisfies the requirement for stage labellings. As all other naive labellings contained undecided arguments, it is the only stage labelling for the AF F .

Let us now have a closer look at the *cf2* labelling $v = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, d \mapsto \mathbf{t}, e \mapsto \mathbf{f}\}$. As the AF F has the four SCCs $C_1 = \{a, b\}$, $C_2 = \{c\}$, $C_3 = \{d\}$ and $C_4 = \{e\}$ we have the following cases, where $D_F(v_{\mathbf{t}}) = \{c, e\}$ and $\{c, e\} \subseteq v_{\mathbf{f}}$. In case $F|_{C_i} - D_F(v_{\mathbf{t}}) = F|_{C_i}$ we only write $v|_{C_i} \in cf2(F|_{C_i})$.

- $v|_{C_1} \in cf2(F|_{C_1})$: the sub-framework $F|_{C_1}$ consists of a single SCC, thus we have to check whether $\{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}\} \in nai(F|_{C_1})$, which indeed holds.
- $v|_{C_2 \setminus \{c, e\}} \in cf2(F|_{C_2} - \{c, e\})$: the sub-framework $F|_{C_2} - \{c, e\}$ consists of the empty framework $F_0 = (\emptyset, \emptyset)$, and $v|_{C_2 \setminus \{c, e\}} = \emptyset$, thus the condition $\emptyset \in nai(F_0)$ is fulfilled.
- $v|_{C_3} \in cf2(F|_{C_3})$: the sub-framework $F|_{C_3}$ consists of the single argument d , thus we can see that $v|_{C_3} = \{d \mapsto \mathbf{t}\} \in nai(F|_{C_3})$ holds.
- $v|_{C_4 \setminus \{c, e\}} \in cf2(F|_{C_4} - \{c, e\})$: the sub-framework $F|_{C_4} - \{c, e\}$ consists of the empty framework F_0 and $v|_{C_4 \setminus \{c, e\}} = \emptyset$, thus the condition $\emptyset \in nai(F_0)$ is fulfilled.

The requirement $D_F(v_{\mathbf{t}}) \subseteq v_{\mathbf{f}}$ of Definition 2.6 ensures that the labellings are conflict-free and naive, otherwise we would obtain labellings like $\{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{u}, d \mapsto \mathbf{t}, e \mapsto \mathbf{f}\}$ or $\{a \mapsto \mathbf{f}, b \mapsto \mathbf{t}, c \mapsto \mathbf{t}, d \mapsto \mathbf{t}, e \mapsto \mathbf{u}\}$ which are not naive or even not conflict-free.

Further AF semantics exist; to save space we define them implicitly via ADFs.

2.2 Abstract Dialectical Frameworks

An abstract dialectical framework (ADF) is a directed graph whose nodes represent statements or positions which can be accepted or not. The links represent dependencies: the status of a node s only depends on the status of its parents $par(s)$, that is, the nodes with a direct link to s . Each node s has an associated acceptance condition C_s specifying the exact conditions under which s is accepted. C_s is a function assigning to each subset of $par(s)$ one of the truth values \mathbf{t} , \mathbf{f} . Intuitively, if for some $M \subseteq par(s)$ we have $C_s(M) = \mathbf{t}$, then s will be accepted provided the nodes in M are accepted and those in $par(s) \setminus M$ are not accepted.

Definition 2.8. An *abstract dialectical framework* is a tuple $D = (S, L, C)$ where

- S is a set of statements (positions, nodes),
- $L \subseteq S \times S$ is a set of links,
- $C = \{C_s\}_{s \in S}$ is a collection of total functions $C_s : 2^{par(s)} \rightarrow \{\mathbf{t}, \mathbf{f}\}$, one for each statement s . The function C_s is called *acceptance condition of s* .

A special class of ADFs are *bipolar ADFs* where the links between the statements are attacking, supporting or both.

Definition 2.9. An ADF $D = (S, L, C)$ is *bipolar* iff there are $L^+, L^- \subseteq L$ such that $L = L^+ \cup L^-$, where

- L^+ denotes the set of supporting links, where a link (r, s) is *supporting* in D if and only if for all $R \subseteq par(s)$, we have $C_s(R) = \mathbf{t}$ implies $C_s(R \cup \{r\}) = \mathbf{t}$;
- L^- denotes the set of attacking links, where a link (r, s) is *attacking* in D if and only if for all $R \subseteq par(s)$, we have $C_s(R \cup \{r\}) = \mathbf{t}$ implies $C_s(R) = \mathbf{f}$.⁵

It is often convenient to represent acceptance conditions as propositional formulas; we will do so in this paper. There, each C_s is represented by a propositional formula φ_s over $par(s)$. Then, clearly, for $M \subseteq par(s)$ we have $C_s(M) = \mathbf{t}$ iff $M \models \varphi_s$. In this way, AFs are recast as ADFs as follows: For an AF $F = (A, R)$, the ADF *associated to F* is $D_F = (A, R, C)$ with $C = \{\varphi_a\}_{a \in A}$ and $\varphi_a = \bigwedge_{(b,a) \in R} \neg b$ for $a \in A$. Intuitively, an AF argument is accepted if and only if none of its attackers is accepted.

It may happen that a link $(r, s) \in L$ in an ADF bears no actual significance. Formally, r is *redundant in φ_s* if and only if there is no two-valued interpretation $v : par(s) \setminus \{r\} \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that

$$v\left(\varphi_s^{\{r \mapsto \mathbf{t}\}}\right) \neq v\left(\varphi_s^{\{r \mapsto \mathbf{f}\}}\right)$$

That is, if (r, s) is redundant then r has no influence on the truth value of φ_s whatsoever.⁶

5. Equivalently, to illustrate the symmetry in this definition, a link (r, s) is *attacking* in D if and only if for all $R \subseteq par(s)$, we have $C_s(R) = \mathbf{f}$ implies $C_s(R \cup \{r\}) = \mathbf{f}$.

6. Cast in another way, the redundant links are exactly those that are attacking *and* supporting.

Several semantics can be defined by using three-valued interpretations v to partially evaluate acceptance formulas φ_s . While this style of definition is novel, the resulting semantics have mostly appeared in the literature before (Brewka et al., 2013). Some others are new, but straightforward to define (Strass, 2013); these are the (three-valued) conflict-free, naive and stage semantics.

Definition 2.10. Let $D = (S, L, C)$ be an ADF. A three-valued interpretation v is

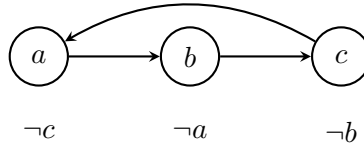
- *admissible* in D , i.e. $v \in \text{adm}(D)$ iff for each $s \in S$ we have:
 - $v(s) = \mathbf{t}$ implies that φ_s^v is irrefutable,
 - $v(s) = \mathbf{f}$ implies that φ_s^v is unsatisfiable;
- *preferred* in D , i.e. $v \in \text{pre}(D)$ iff it is \leq_i -maximal with respect to being admissible;
- *complete* in D , i.e. $v \in \text{com}(D)$ iff for each $s \in S$ we have:
 - $v(s) = \mathbf{t}$ if and only if φ_s^v is irrefutable,
 - $v(s) = \mathbf{f}$ if and only if φ_s^v is unsatisfiable;
- *grounded* in D , i.e. $v \in \text{grd}(D)$ iff v is the \leq_i -least complete interpretation;
- *conflict-free* in D , i.e. $v \in \text{cfl}(D)$ iff for all $s \in S$ we have:
 - $v(s) = \mathbf{t}$ implies that φ_s^v is satisfiable,
 - $v(s) = \mathbf{f}$ implies that φ_s^v is unsatisfiable;
- *naive* in D , i.e. $v \in \text{nai}(D)$ iff it is \leq_i -maximal with respect to being conflict-free;
- *stage* in D , i.e. $v \in \text{stg}(D)$ iff the set $v_{\mathbf{u}}$ is \subseteq -minimal with respect to being conflict-free.

A two-valued interpretation v is a *model* of D , i.e. $v \in \text{mod}(D)$ iff for all $s \in S$ we find $v(s) = v(\varphi_s)$.

Intuitively, an interpretation v is admissible if it can justify the definite stances it takes: for example, whenever v judges a statement s to be true, then this must be justified by the statement's acceptance formula. This justification can take into consideration the definite assignments of v , but must be valid no matter how the undecided statements of v are interpreted. This is elegantly achieved by checking the refutability of the partial evaluation φ_s^v of the acceptance formula of s . Complete interpretations are then the ones whose recommendations are exactly in accordance with the refutability/satisfiability status of v 's assignments. The grounded semantics can consequently be seen as the greatest possible consensus between all acceptable ways of interpreting the ADF at hand. The three-valued notion of conflict-freeness is clearly a weaker version of admissibility, where truth of a statement has to be justified not by irrefutability, but only by satisfiability. (The justification

standard for rejected statements is the same.)⁷ As usual, naive and stage are then those conflict-free interpretations which are information-maximal or undecided-minimal, respectively. A model of an ADF is simply a two-valued complete interpretation. All of these semantics are proper generalisations of the same semantics for AFs (Brewka et al., 2013; Strass, 2013).

Example 2.11. Let $D = (S, L, C)$ be an ADF with $S = \{a, b, c\}$, $L = \{(a, b), (b, c), (c, a)\}$ and the acceptance conditions $\varphi_a = \neg c$, $\varphi_b = \neg a$ and $\varphi_c = \neg b$:



(Note that this is an AF-based ADF with an attack cycle of length three, as in Example 2.3.) Some conflict-free interpretations of D are

$$\begin{array}{ll} v_1 = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}, c \mapsto \mathbf{u}\} & v_2 = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}\} \\ v_3 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}\} & v_4 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{u}, c \mapsto \mathbf{t}\} \end{array}$$

(There are three further conflict-free interpretations, see Example 2.3.) We have a closer look at interpretation v_4 .

- As $v_4(a) = \mathbf{f}$, according to the definition of conflict-free interpretations, $\varphi_a^{v_4}$ needs to be unsatisfiable. Thus we construct the partial valuation of $\varphi_a = \neg c$ by v_4 and obtain $\varphi_a^{v_4} = \neg \top$, which indeed is unsatisfiable.
- As $v_4(c) = \mathbf{t}$, the formula $\varphi_c^{v_4} = \neg b$ needs to be satisfiable, which holds.

On the other hand, consider $v_5 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{u}, c \mapsto \mathbf{f}\}$, which is not conflict-free, as $v_5(c) = \mathbf{f}$ but $\varphi_c^{v_5} = \neg b$ is satisfiable. The naive interpretations of D are v_2, v_3 and v_4 because they are \leq_i -maximal with respect to being conflict-free.

In this paper, we restrict our attention to finite ADFs. This assumption is commonly made, given that finite ADFs are of higher practical relevance and more meaningful when it comes to computational investigations (as complexity problems require finite input sizes).

2.3 Complexity Theory

Assume some fixed finite vocabulary Σ with $|\Sigma| > 1$. A language $L \subseteq \Sigma^*$ is in P iff it can be recognised by a deterministic Turing machine in polynomial time. Complexity class NP contains all problems L that have a polytime-computable witness relation; that is, $L \in \mathsf{NP}$

7. This is in contrast to the definition of (three-valued) conflict-freeness given by Strass and Wallner (2015): consider the AF-based ADF D over $S = \{a, b\}$ with $\varphi_a = \varphi_b = \neg a$, that is, a self-attacking a attacks b . According to Definition 2.10, $\text{cfl}(D) = \{v_0, v_1\}$ with $v_0 = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{u}\}$ and $v_1 = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}\}$. The definition of Strass and Wallner (2015) allows for another conflict-free interpretation $v_2 = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{f}\}$. (Note that both v_0 and v_2 correspond to the conflict-free set \emptyset .) In general, all our conflict-free interpretations are conflict-free according to Strass and Wallner (2015), but not vice versa.

iff there are $W_L \in \mathsf{P}$ and $k \in \mathbb{N}$ such that: $x \in L$ iff there is a y such that $(x, y) \in W_L$ and $|y| \leq |x|^k$. For any class \mathcal{C} of languages, its complement class is $\text{co}\mathcal{C} = \{\bar{L} \mid L \in \mathcal{C}\}$. For example, the class coNP contains all languages L whose complement $\bar{L} = \Sigma^* \setminus L$ is in NP . These two classes give rise to the polynomial hierarchy, that can be defined (using oracle Turing machines) as follows: $\Delta_0^{\mathsf{P}} = \Sigma_0^{\mathsf{P}} = \Pi_0^{\mathsf{P}} = \mathsf{P}$, and for $i \geq 0$, $\Delta_{i+1}^{\mathsf{P}} = \mathsf{P}^{\Sigma_i^{\mathsf{P}}}$, $\Sigma_{i+1}^{\mathsf{P}} = \text{NP}^{\Sigma_i^{\mathsf{P}}}$, $\Pi_{i+1}^{\mathsf{P}} = \text{coNP}^{\Sigma_i^{\mathsf{P}}}$. For any complexity class \mathcal{C} , a Turing machine with access to a \mathcal{C} -oracle can be understood as having a constant-time decision subroutine for problems in \mathcal{C} . For each level i of the polynomial hierarchy, the classes Σ_i^{P} and Π_i^{P} have canonical complete problems, defined via fully quantified Boolean formulae (QBFs) in prenex form. Such formulae have the shape $\mathfrak{D}_1 P_1 \mathfrak{D}_2 P_2 \dots \mathfrak{D}_i P_i \varphi$, where φ is a propositional formula using propositional variables from the pairwise disjoint sets $P_1 \dots P_i$, $\mathfrak{D}_\ell \in \{\forall, \exists\}$ and the quantifiers are alternating, i.e., $\mathfrak{D}_{\ell+1} = \forall$ iff $\mathfrak{D}_\ell = \exists$. Validity of QBFs is determined inductively. Given a (quantifier-free) Boolean formula φ , let $\text{val}(\varphi)$ denote the set of all truth value assignments making φ true. Further, we define:

$$\begin{aligned} \text{val}(\exists P \Psi) &= \{\sigma \mid \text{there are } \pi \in P \rightarrow \{\mathbf{t}, \mathbf{f}\} \text{ and } \sigma' \in \text{val}(\Psi) \text{ with } \sigma' = \pi \uplus \sigma\} \\ \text{val}(\forall P \Psi) &= \{\sigma \mid \text{for all } \pi \in P \rightarrow \{\mathbf{t}, \mathbf{f}\} \text{ there is a } \sigma' \in \text{val}(\Psi) \text{ with } \sigma' = \pi \uplus \sigma\} \end{aligned}$$

A fully quantified QBF $\mathfrak{D}_1 P_1 \mathfrak{D}_2 P_2 \dots \mathfrak{D}_i P_i \varphi$ is *valid* iff $\text{val}(\mathfrak{D}_1 P_1 \mathfrak{D}_2 P_2 \dots \mathfrak{D}_i P_i \varphi)$ contains the empty assignment.

Now, the canonical problem for Σ_i^{P} is as follows: Given a quantified Boolean formula $\Phi = \exists P_1 \forall P_2 \exists P_3 \dots \mathfrak{D}_i P_i \psi$, determine whether Φ is valid. For Π_i^{P} the canonical complete problem is similar, but the formula starts with universal quantification.

Example 2.12. Consider the QBF $\Xi_1 = \exists\{p, q\} \forall\{r\} (\neg r \vee p) \wedge (\neg r \vee \neg q)$ leading to a decision problem belonging to the class Σ_2^{P} . We find that it is valid through the definition of validity provided above. Intuitively, we can confirm the validity as follows: there exists a truth value assignment for p and q (namely $p \mapsto \mathbf{t}$ and $q \mapsto \mathbf{f}$) such that independently of the choice of the assignment for r (be it $r \mapsto \mathbf{t}$ or $r \mapsto \mathbf{f}$), the total assignment will make the formula true.

Now, consider the slightly changed QBF $\Xi_2 = \forall\{p\} \exists\{q\} \forall\{r\} (\neg r \vee p) \wedge (\neg r \vee \neg q)$ leading to a Π_3^{P} problem. We find that it is not valid. Intuitively, this is due to the fact that there is an assignment for p and r (namely $p \mapsto \mathbf{f}$ and $r \mapsto \mathbf{t}$) which makes the formula false independently of what assignment is chosen for q .

While these classes from the polynomial hierarchy are fairly standard, NP and coNP also give rise to the so-called *Boolean hierarchy*. It is rather little-known and defined as follows (Wechsung, 1985). Firstly, for given complexity classes \mathcal{C}_1 and \mathcal{C}_2 define the new classes

$$\begin{aligned} \mathcal{C}_1 \wedge \mathcal{C}_2 &= \{L_1 \cap L_2 \mid L_1 \in \mathcal{C}_1, L_2 \in \mathcal{C}_2\} \\ \mathcal{C}_1 \vee \mathcal{C}_2 &= \{L_1 \cup L_2 \mid L_1 \in \mathcal{C}_1, L_2 \in \mathcal{C}_2\} \end{aligned}$$

Next, set $\mathsf{C}_0^{\text{BH}} = \mathsf{D}_0^{\text{BH}} = \mathsf{P}$ and for $i \geq 0$ define⁸

$$\mathsf{C}_{i+1}^{\text{BH}} = \text{coNP} \wedge \mathsf{D}_i^{\text{BH}} \quad \text{and} \quad \mathsf{D}_{i+1}^{\text{BH}} = \text{NP} \vee \mathsf{C}_i^{\text{BH}}$$

8. This is the Boolean hierarchy between $\Delta_1^{\mathsf{P}} = \mathsf{P}$ and Δ_2^{P} ; there is a Boolean hierarchy between Δ_i^{P} and $\Delta_{i+1}^{\mathsf{P}}$ for all $i \geq 1$ using Σ_i^{P} and Π_i^{P} instead of NP and coNP (Chang & Kadin, 1996).

(Intuitively, C_i^{BH} is for “conjunction” and D_i^{BH} is for “disjunction”.) For example, $D_1^{\text{BH}} = \text{NP}$ and $C_1^{\text{BH}} = \text{coNP}$, while $D_2^{\text{BH}} = \text{NP} \vee C_1^{\text{BH}} = \text{NP} \vee \text{coNP}$ and $C_2^{\text{BH}} = \text{coNP} \wedge D_1^{\text{BH}} = \text{coNP} \wedge \text{NP}$. The class C_2^{BH} is an alternative notation for the class DP, first proposed by Papadimitriou and Yannakakis (1982). Its complement $\text{coDP} = D_2^{\text{BH}}$ contains all languages L for which there are $L_1 \in \text{NP}$ and $L_2 \in \text{coNP}$ with $L = L_1 \cup L_2$. The Boolean hierarchy and the polynomial hierarchy are closely interrelated: Chang and Kadin (1996) showed that the polynomial hierarchy collapses (to the third level) if the Boolean hierarchy collapses. In this paper, when we say that problems are or become “easier” or “harder”, we do this under the assumption that the polynomial hierarchy does not collapse.

3. Decomposing Abstract Dialectical Frameworks

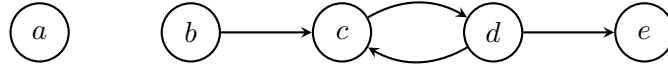
In this section we introduce a decomposition technique for ADFs that is a generalisation of a similar approach for AFs (Baroni et al., 2005). However, the decomposition along the SCCs of an ADF D cannot be performed in the same way as it is done for AFs. If one looks at a set $M \subseteq S$ of statements, the acceptance conditions of the statements in M might still depend on statements that are not contained in M , even if M forms an SCC. To be able to decompose and evaluate an ADF, it is necessary to modify the acceptance conditions of the statements in a way that they only depend on statements also contained in the same component. This modification will be performed depending on the decided truth values of the parents of statements.

We propose a procedure that propagates truth values from independent parts of an ADF to the rest of the ADF. We need to take several facts into account. First, we use three-valued interpretations (labellings) to represent the current acceptance status of statements. As the acceptance conditions of ADFs are defined as functions on two-valued interpretations, we cannot pass on the truth value \mathbf{u} , but we make a statement s forcibly undecided by changing its acceptance formula to $\neg s$. Second, by fixing the truth value of some statements, we might produce *redundancies* in the acceptance conditions of other statements. Eliminating these redundancies from the links and the acceptance formulas is one of the crucial points in the procedure, because by doing so, the dependencies of the statements can change, which has an important influence on the subsequent calls.

Definition 3.1. Let $D = (S, L, C)$ be an ADF and $p, s \in S$. We say that s *u-depends on* p if there is a path from p to s in L but no path from s to p in L (where u stands for unidirectionally). Now let $M \subseteq S$. A statement $s \in S$ is *independent modulo* M iff for each $p \in S$, if s *u-depends on* p then $p \in M$. A set $M \subseteq S$ is *independent* iff there is no $s \in M$ that *u-depends on* a $p \in S \setminus M$. Lastly, define $\text{ind}_D(M) = \{s \in S \mid s \text{ is independent modulo } M \text{ in } D\}$.

Note that dependence here implicitly speaks about strongly connected components (SCCs). Intuitively speaking, statements do not depend on statements in their own SCC, but on all statements in “previous” SCCs. The function ind_D returns the set of all statements that are independent modulo the input set. Note that the topology of the ADF is explicitly given with the set L , thus it is not necessary to consider the acceptance conditions to get the information about independent statements.

Example 3.2. Consider an ADF D with the statements S and links given graphically:



We initially have $ind_D(\emptyset) = \{a, b\} = M_0$. Then $ind_D(M_0) = \{a, b, c, d\} = M_1$ and finally $ind_D(M_1) = \{a, b, c, d, e\} = S$.

Given an independent subset M of statements of an ADF, ignoring all other statements again yields an ADF.

Definition 3.3. Let $D = (S, L, C)$ be an ADF and $M \subseteq S$ be an independent set. The ADF D restricted to M is given by $D|_M = (M, L \cap (M \times M), \{\varphi_s\}_{s \in M})$.

Note that $D|_M$ really is an ADF since its acceptance formulas by presumption do not mention statements not in M .

We next define how to reduce an ADF given a subset M of its statements and an interpretation of this subset. The intuition is that the truth values of statements in M are fixed and can be propagated into the rest of the ADF.⁹ For this definition recall from Definition 2.1 that for a propositional formula φ and a three-valued interpretation v of parts of its signature, φ^v denotes the formula φ where atoms that v maps into $\{\mathbf{t}, \mathbf{f}\}$ have been replaced by their truth values. Through such replacements it may happen that links become redundant. For example, consider the acceptance formula $\varphi_s = a \vee (b \wedge c)$ and the interpretation $v = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}\}$. The reduced formula is $\varphi_s^v = a \vee (\perp \wedge c)$. This formula is equivalent to a and thus c is redundant in φ_s^v . The identification and removal of such redundant parents is an important ingredient of the following definition.

Definition 3.4. Let $D = (S, L, C)$ be an ADF, $M \subseteq S$ and $v : M \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$. The ADF D reduced with v on M is given by $\llbracket D \rrbracket_M^v = (S, \llbracket L \rrbracket_M^v, \{\llbracket \varphi_s \rrbracket_M^v\}_{s \in S})$ with

$$\llbracket \varphi_s \rrbracket_M^v = \begin{cases} \top & \text{if } s \in M \text{ and } v(s) = \mathbf{t} \\ \perp & \text{if } s \in M \text{ and } v(s) = \mathbf{f} \\ \neg s & \text{if } s \in M \text{ and } v(s) = \mathbf{u} \\ \varphi_s^v[r/\top : r \text{ is redundant in } \varphi_s^v] & \text{otherwise} \end{cases}$$

$$\llbracket L \rrbracket_M^v = (L \setminus \{(r, s) \in L \mid r \text{ is redundant in } \llbracket \varphi_s \rrbracket_M^v\}) \cup \{(s, s) \mid v(s) = \mathbf{u}\}.$$

That is, $\llbracket L \rrbracket_M^v$ is L without redundant links and the added self-loops for undecided statements in v . The new acceptance formulas in ADF $\llbracket D \rrbracket_M^v$ fix the truth values of statements in M as v assigns them. Furthermore, the classical ones among these truth values are fixed in acceptance formulas that mention statements in M . Should such replacements make other statements redundant, then these are replaced by a fixed truth value to make the redundancy explicit. In the example above, the partially evaluated formula $a \vee (\perp \wedge c)$ is further transformed into $a \vee (\perp \wedge \top)$, that is, former parent c is replaced by \top . (Since the parent is redundant, it is immaterial which truth value is actually used.) Whenever

9. For the classical truth values \mathbf{t} and \mathbf{f} , the resulting acceptance conditions are clear; for \mathbf{u} we use a self-attack, as it was done for AFs (Baumann, 2011).

$par(s) \cap M = \emptyset$, that is, the parents of s are not affected by v , then $\llbracket \varphi_s \rrbracket_M^v = \varphi_s$, that is, the acceptance formula of s does not change.

Now we present the final ingredient of our decomposition-based scheme, the most important definition of the paper. It describes the actual recursion that is used to assign to a given ADF semantics σ a new semantics σ_2 .

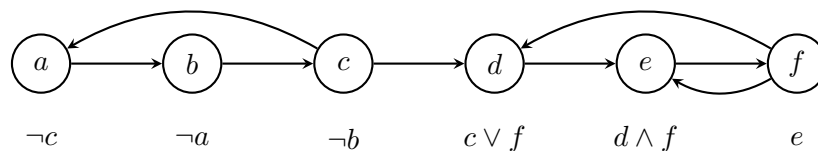
Definition 3.5. Let $D = (S, L, C)$ be an ADF and σ a semantics. Define a set $\sigma_2(D)$ of interpretations as follows:

$$\sigma_2(D) = \sigma_2(ind_D(\emptyset), D), \text{ where for each } M \subseteq S \text{ we define}$$

$$\sigma_2(M, D) = \begin{cases} \sigma(D) & \text{if } M = S \\ \bigcup_{w \in \sigma(D|_M)} \sigma_2(ind_{\llbracket D \rrbracket_M^w}(M), \llbracket D \rrbracket_M^w) & \text{otherwise} \end{cases}$$

The basic underlying intuition of this definition is to recursively decompose a given ADF along its independent statements. We start out with all statements that are independent modulo the empty set, $M_0 = ind_D(\emptyset)$. We now look only at the sub-ADF $D|_{M_0}$ that consists of D restricted to M_0 and consider all its σ -interpretations. For each σ -interpretation w , we use the information it contains (that is, the truth values it assigns) to simplify the rest of the ADF. Simplification means that we propagate the truth values of the interpretation as far as possible and at the same time remove redundant links. We then recursively invoke the definition on the ADF resulting from simplifying D by w . Note that at this point, the statements in M_0 are already dealt with, they have fixed truth values. The main task of the recursive call is to take care of all statements that have newly become independent (modulo M_0). When the sequence of independent statements $M_0 \subseteq M_1 \subseteq \dots$ eventually reaches the fixed-point S , the first case of the definition applies and the recursion stops.¹⁰ An obvious special case are ADFs D with only one strongly connected component. In this case, $ind_D(\emptyset) = S$ and thus $\sigma_2(D) = \sigma(D)$.

Example 3.6. Let the ADF $D = (S, L, C)$ be graphically given as follows:

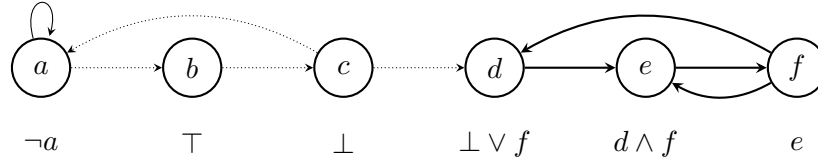


10. To see why, it suffices to see that for each finite ADF D , the set $ind_D(\emptyset)$ is non-empty. This, in turn, is because there cannot be infinite descending chains of SCCs. Then, once we have an interpretation w for the non-empty M_0 , we can simplify D by in essence removing the statements in M_0 from it (yielding D') and are left with obtaining $ind_{D'}(\emptyset)$ again. Note that the restriction to finite ADFs is paramount here, as for example an ADF $D_{\mathbb{N}}^>$ over the natural numbers \mathbb{N} and links $\{(n+1, n) \mid n \in \mathbb{N}\}$ has $ind_{D_{\mathbb{N}}^>}(\emptyset) = \emptyset$, while ADF $D_{\mathbb{N}}^<$ with statements \mathbb{N} and links $\{(n, n+1) \mid n \in \mathbb{N}\}$ has $ind_{D_{\mathbb{N}}^<}(\emptyset) = \{0\}$, but the fixed-point $ind_{D_{\mathbb{N}}^<}(\mathbb{N}) = \mathbb{N}$ cannot be reached in a finite number of steps.

We want to compute $nai_2(D) = nai_2(ind_D(\emptyset), D)$ and thus construct the set $ind_D(\emptyset) = \{a, b, c\} = M_0$. Then we obtain $nai(D|_{M_0}) = \{v_0, v_1, v_2\}$:

$$\begin{aligned} v_0 &= \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}\}, \\ v_1 &= \{a \mapsto \mathbf{f}, b \mapsto \mathbf{u}, c \mapsto \mathbf{t}\}, \\ v_2 &= \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}\}. \end{aligned}$$

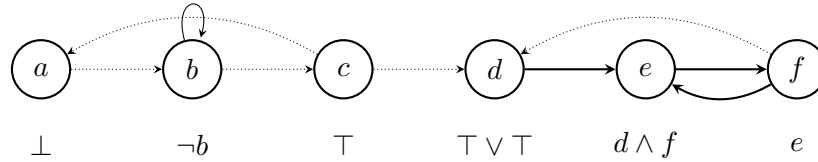
According to Definition 3.5, for each of these interpretations w we construct the respective reduced ADF $\llbracket D \rrbracket_{M_0}^w$ and recursively determine its nai_2 semantics. We begin with $w = v_0 \in nai(D|_{M_0})$ and compute $nai_2(M_1, D_1)$ with $D_1 = \llbracket D \rrbracket_{M_0}^{v_0}$. The ADF D_1 is graphically depicted below; links that have newly become redundant are dotted, links originating in independent statements are thin.



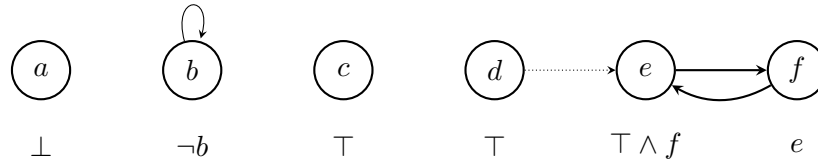
Thus $M_1 = ind_{D_1}(M_0) = S$, and we only need to consider $nai(D_1) = \{v_3, v_4\}$:

$$\begin{aligned} v_3 &= v_0 \cup \{d \mapsto \mathbf{t}, e \mapsto \mathbf{t}, f \mapsto \mathbf{t}\}, \\ v_4 &= v_0 \cup \{d \mapsto \mathbf{f}, e \mapsto \mathbf{f}, f \mapsto \mathbf{f}\}. \end{aligned}$$

We next consider $v_1 \in nai(D|_{M_0})$ and call $nai_2(M_2, D_2)$ with $D_2 = \llbracket D \rrbracket_{M_0}^{v_1}$:



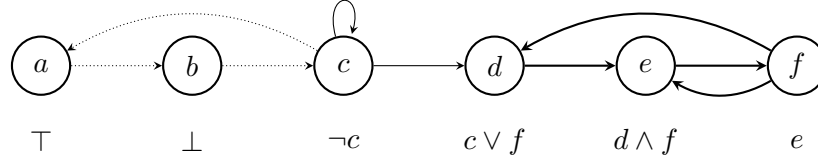
Note that $\varphi_d^{v_1} = \top \vee f$ where f is redundant and thus $\llbracket \varphi_d \rrbracket_{M_0}^{v_1} = \top \vee \top$. For the next step we get $M_2 = ind_{D_2}(M_0) = \{a, b, c, d\}$, and $D_2|_{M_2}$ has the single naive interpretation $v_5 = v_1 \cup \{d \mapsto \mathbf{t}\}$. We compute $nai_2(M_3, D_3)$ with $D_3 = \llbracket D_2 \rrbracket_{M_2}^{v_5}$:



We obtain $M_3 = ind_{D_3}(M_2) = S$ and two naive interpretations for D_3 :

$$\begin{aligned} v_6 &= v_5 \cup \{e \mapsto \mathbf{t}, f \mapsto \mathbf{t}\}, \\ v_7 &= v_5 \cup \{e \mapsto \mathbf{f}, f \mapsto \mathbf{f}\}. \end{aligned}$$

Finally, for $v_2 \in \text{nai}(D|_{M_0})$ the call $\text{nai}_2(M_4, D_4)$ is performed with $D_4 = \llbracket D \rrbracket_{M_0}^{v_2}$:



Thus $M_4 = \text{ind}_{D_4}(M_0) = S$ and the two naive interpretations of D_4 are

$$\begin{aligned} v_8 &= v_2 \cup \{d \mapsto \mathbf{t}, e \mapsto \mathbf{t}, f \mapsto \mathbf{t}\}, \\ v_9 &= v_2 \cup \{d \mapsto \mathbf{t}, e \mapsto \mathbf{f}, f \mapsto \mathbf{f}\}. \end{aligned}$$

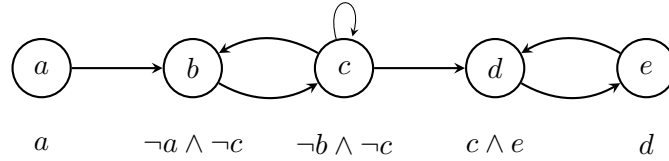
Thus overall, we obtain the set

$$\text{nai}_2(D) = \text{nai}_2(M_1, D_1) \cup \text{nai}_2(M_2, D_2) \cup \text{nai}_2(M_4, D_4) = \{v_3, v_4, v_6, v_7, v_8, v_9\}.$$

In this example the naive and nai_2 interpretations coincide, thus $\text{nai}(D) = \text{nai}_2(D)$.

However, in general we have that $\text{nai} \neq \text{nai}_2$ which is shown in the following example.

Example 3.7. Let the ADF $D = (S, L, C)$ be as given below:



The nai_2 interpretations of this ADF are

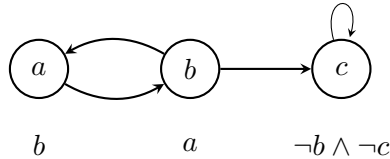
$$\begin{aligned} v_1 &= \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}, d \mapsto \mathbf{t}, e \mapsto \mathbf{t}\}, \\ v_2 &= \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}, d \mapsto \mathbf{f}, e \mapsto \mathbf{f}\}. \end{aligned}$$

But there is the following naive interpretation v_3 of D which is not contained in $\text{nai}_2(D)$

$$v_3 = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, d \mapsto \mathbf{f}, e \mapsto \mathbf{f}\}.$$

Example 3.6 has shown that at least for the case of naive semantics, $\text{nai} \neq \text{nai}_2$. But what about other semantics? Let us look at another example, this time for stage semantics.

Example 3.8. Consider the ADF $D = (S, L, C)$ with $S = \{a, b, c\}$, $L = \{(a, b), (b, a), (b, c), (c, c)\}$ and $\varphi_a = b$, $\varphi_b = a$ and $\varphi_c = \neg b \wedge \neg c$:



We find that D has only one stage interpretation, $stg(D) = \{v_1\}$, with $v_1 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}\}$. On the other hand, for $stg_2(D)$, we start out with $M_1 = ind_D(\emptyset) = \{a, b\}$. There we find $stg(D|_{M_1}) = \{v_2, v_3\}$ with $v_2 = \{a \mapsto \mathbf{f}, b \mapsto \mathbf{f}\}$ and $v_3 = \{a \mapsto \mathbf{t}, b \mapsto \mathbf{t}\}$. The reduced ADF $\llbracket D \rrbracket_{M_1}^{v_2}$ consists only of the self-attacking statement c and $stg(\llbracket D \rrbracket_{M_1}^{v_2}) = \{v_4\}$ with $v_4 = v_2 \circ \{c \mapsto \mathbf{u}\}$. The reduced ADF $\llbracket D \rrbracket_{M_1}^{v_3}$ consists of statement c where φ_c is equivalent to false, thus $stg(\llbracket D \rrbracket_{M_1}^{v_3}) = \{v_5\}$ with $v_5 = v_3 \circ \{c \mapsto \mathbf{f}\}$. Thus $stg_2(D) = \{v_4, v_5\}$; while clearly $v_5 = v_1$, we have $v_4 \in stg_2(D)$ but $v_4 \notin stg(D)$.

Thus for stage semantics, also $stg_2 \neq stg$. The following fundamental result provides the complete picture. For semantics σ, τ , the expression $\sigma \leq \tau$ means that for all ADFs D we have $\sigma(D) \subseteq \tau(D)$.

Theorem 3.9.

1. Let $\sigma \in \{cfi, adm, pre, com, mod\}$. Then $\sigma \leq \sigma_2$.
2. Let $\sigma \in \{nai, stg\}$. Then $\sigma \not\leq \sigma_2$.
3. Let $\sigma \in \{cfi, nai, adm, pre, com, mod\}$. Then $\sigma_2 \leq \sigma$.
4. Let $\sigma \in \{stg\}$. Then $\sigma_2 \not\leq \sigma$.

Proof. 1. Let $D = (S, L, C)$ be an ADF and $v \in \sigma(D)$. Define $M_0 = \emptyset$, $D_1 = D$ and $M_1 = ind_D(\emptyset)$. For $i \geq 1$ define $v_i = v|_{M_i}$, $D_{i+1} = \llbracket D_i \rrbracket_{M_i}^{v_i}$ and $M_{i+1} = ind_{D_{i+1}}(M_i)$. By $\varphi_{s,i}$ we denote the acceptance formula of s in D_i . Let $n \in \mathbb{N}$ be such that $M_n = S$. From Definition 3.5 it follows that $v \in \sigma(D_n)$ if and only if $v \in \sigma_2(D)$, whence we have to show $v \in \sigma(D_n)$. We will show by induction that for all $i \in \mathbb{N}$ we find $v_i \in \sigma(D_i|_{M_i})$. Since $v_n = v|_{M_n} = v|_S = v$ and similarly $D_n|_{M_n} = D_n|_S = D_n$, we infer that $v_n \in \sigma(D_n|_{M_n})$ shows $v \in \sigma(D_n)$.

$i = 1$: Let $s \in M_1 = ind_{D_1}(\emptyset)$. Since s is independent modulo M_1 , the parents of s in D are contained in M_1 . Thus the acceptance function of s in $D_1|_{M_1}$ is exactly that of s in D . Now $v \in \sigma(D)$ and M_1 being independent clearly imply that $v|_{M_1} = v_1 \in \sigma(D_1|_{M_1})$.

$i \rightsquigarrow i + 1$: Let $v_i \in \sigma(D_i|_{M_i})$. Let $s \in S$. If $s \in M_i$, then by definition the acceptance condition of s in D_{i+1} allows only the single truth value $v_i(s)$, therefore assume $s \in M_{i+1} \setminus M_i$. By definition, $M_{i+1} = ind_{D_{i+1}}(M_i)$ whence s is independent modulo M_i in D_{i+1} . Thus $\varphi_{s,i+1}$ mentions only statements in M_{i+1} . For any $r \in M_{i+1}$ we have by definition of $v_{i+1} = v|_{M_{i+1}}$ that $v_{i+1}(r) = v(r)$. If there is any syntactic difference between φ_s and $\varphi_{s,i+1}$, then this difference amounts to replacements of statements by truth values. There are two possible reasons: (a) replacing an $r \in M_{i+1}$ by $v(r) \in \{\mathbf{t}, \mathbf{f}\}$; (b) replacing a redundant $r \in M_{i+1}$ by \top . In both cases, we obtain that $\models \varphi_s^v \equiv \varphi_{s,i+1}^v$: for (a), replacing r by $v(r)$ is obviously compatible with computing the partial evaluation with respect to v ; for (b), replacing a redundant variable in a formula has no influence on the semantics by definition. From this, we can infer that also $\models \varphi_s^v \equiv \varphi_{s,i+1}^{v_{i+1}}$, whence the result follows for $\sigma \in \{cfi, adm, com, mod\}$.

For $\sigma = pre$, assume to the contrary that $v \notin pre_2(D)$. Consider the algorithmic attempt at testing whether $v \in pre_2(D)$, in particular the iteration $i \in \mathbb{N}$

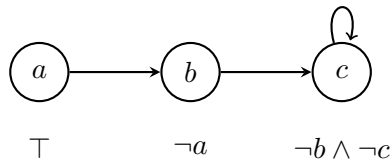
where its negative result becomes obvious, that is, the iteration $i \in \mathbb{N}$ for which $v_i = v|_{M_i} \notin \text{pre}(D_i|_{M_i})$. There are two possible reasons:

- (a) $v_i \notin \text{adm}(D_i|_{M_i})$. Then there is an $s \in M_i$ with $v_i(s) \in \{\mathbf{t}, \mathbf{f}\}$ but where the refutability/satisfiability status of $\varphi_{s,i}$ does not match $v_i(s)$. Since s is independent modulo M_i in D_i , we also have a mismatch between $\varphi_s \equiv \varphi_{s,i}$ and $v_i(s) = v(s)$. But then $v \notin \text{adm}(D) \supseteq \text{pre}(D)$, contradiction.
- (b) there is an interpretation $w_i \in \text{adm}(D_i|_{M_i})$ with $v_i <_i w_i$. We construct an interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ as follows:

$$s \mapsto \begin{cases} w_i(s) & \text{if } s \in M_i \\ v(s) & \text{otherwise} \end{cases}$$

Clearly $v <_i w$ by construction. It remains to show that $w \in \text{adm}(D)$, because then $v \notin \text{pre}(D)$, a contradiction. To show that w is admissible, we have to show that for each $s \in S$, if $w(s) = \mathbf{t}$ then φ_s^w is irrefutable and if $w(s) = \mathbf{f}$ then φ_s^w is unsatisfiable. For $s \in M_i$, this condition holds because $w_i \in \text{adm}(D_i|_{M_i})$. Let $s \in S \setminus M_i$. If s depends only on statements in $S \setminus M_i$, then the condition holds because $v \in \text{pre}(D) \subseteq \text{adm}(D)$. Assume s has parents in M_i . By $v \leq_i w$, the only possible difference between v and w are statements $r \in M_i$ with $v(r) = \mathbf{u}$ and $w(r) \in \{\mathbf{t}, \mathbf{f}\}$. Now if φ_s^v is irrefutable, then φ_s^w is irrefutable as well, the same for unsatisfiability. Thus $w \in \text{adm}(D)$.

2. $\sigma = \text{nai}$: The ADF D from Example 3.6 has $|\text{nai}(D)| = 7$ but $|\text{nai}_2(D)| = 6$.
 $\sigma = \text{stg}$: Consider the following AF-based ADF D with $S = \{a, b, c\}$, $L = \{(a, b), (b, c), (c, c)\}$ and the acceptance conditions $\varphi_a = \top$, $\varphi_b = \neg a$ and $\varphi_c = \neg b \wedge \neg c$:



The interpretation $v = \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}\}$ is not contained in $\text{stg}_2(D)$ because for $D_0 = D$ and $M_0 = \text{ind}_{D_0}(\emptyset) = \{a\}$. Thus, by calling $\text{stg}_2(D_0, M_0)$ we first obtain that $v_0 = \{a \mapsto \mathbf{t}\}$ is the only stage interpretation of the ADF $D_0|_{M_0}$. The next call is $\text{stg}_2(D_1, M_1)$, with $D_1 = \llbracket D_0 \rrbracket_{M_0}^{v_0}$ contains the acceptance conditions $\varphi_a = \top$, $\varphi_b = \neg \top$ and $\varphi_c = \neg b \wedge \neg c$, and the link (a, b) is removed in D_1 . Thus, $M_1 = \text{ind}_{D_1}(M_0) = \{a, b\}$, and $v_1 = v_0 \cup \{b \mapsto \mathbf{f}\} = \text{stg}(D_1|_{M_1})$. Finally, we call $\text{stg}_2(D_2, M_2)$, with $D_2 = \llbracket D_1 \rrbracket_{M_1}^{v_1}$ and the acceptance conditions $\varphi_a = \top$, $\varphi_b = \perp$ and $\varphi_c = \neg \perp \wedge \neg c$, and the redundant link (b, c) was removed in D_2 . Thus, $M_2 = \text{ind}_{D_2}(M_1) = S$ and $\text{stg}(D_2)$ has one interpretation, namely $v_2 = v_1 \cup \{c \mapsto \mathbf{u}\}$.

3. Let $D = (S, L, C)$ be an ADF and $v \in \sigma_2(D)$. Define $D_1 = D$, $M_0 = \emptyset$ and $M_1 = \text{ind}_D(\emptyset)$. By definition of σ_2 , for each $i \geq 1$ there exist an ADF D_i , a set $M_i \subseteq S$

and an interpretation $v_i : M_i \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\} \in \sigma(D_i|_{M_i})$ such that $D_{i+1} = \llbracket D_i \rrbracket_{M_i}^{v_i}$ and $M_{i+1} = \text{ind}_{D_{i+1}}(M_i)$. Assume to the contrary of what we have to show that $v \notin \sigma(D)$. In each of the cases (semantics) below, this assumption will let us conclude the existence of an $s \in S$ with certain properties. To derive the required contradictions, we typically make use of the fact that for *any* $s \in S$, we find some $i \in \mathbb{N}$ with

$$\models \varphi_s^v \equiv \varphi_{s,i}^{v_i}. \quad (\star)$$

To see this, let $i \in \mathbb{N}$ be the least $i \leq |S|$ such that $s \in M_i$. By definition $v_i \in \sigma(D_i|_{M_i})$ and $v_i = v|_{M_i}$, whence $v_i(s) = v(s)$. As s is independent in D_i modulo M_{i-1} , also all parents of s are contained in M_i and thus also independent in D_i (modulo M_{i-1}). Thus, for each parent r of s , we also have $v_i(r) = v(r)$. Equation (\star) follows.

$\sigma = cfi$: There can be the following two cases: (i) there is an $s \in S$ with $v(s) = \mathbf{t}$ but φ_s^v is unsatisfiable: by the above, also $\varphi_{s,i}^{v_i}$ is unsatisfiable and we obtain $v_i \notin cfi(D_i|_{M_i})$, a contradiction. (ii) there is an $s \in S$ with $v(s) = \mathbf{f}$ but φ_s^v is satisfiable: then also $\varphi_{s,i}^{v_i}$ is satisfiable and we obtain $v_i \notin cfi(D_i|_{M_i})$, a contradiction.

$\sigma = nai$: Then there is an interpretation $w \in cfi(D)$ with $v <_i w$, that is, there is a statement $s \in S$ such that $v(s) = \mathbf{u}$ and $w(s) \neq \mathbf{u}$. We now find that $v_i(s) = v(s) = \mathbf{u}$ while $w|_{M_i}(s) = w(s) \neq \mathbf{u}$. Thus $v_i <_i w|_{M_i}$. Since w is conflict-free in D , the interpretation $w|_{M_i}$ is conflict-free in $D_i|_{M_i}$, see (\star) . But then $v_i \notin nai(D_i|_{M_i})$. Contradiction.

$\sigma = adm$: Then there is an $s \in S$ such that either (i) $v(s) = \mathbf{t}$ and φ_s^v is refutable, or (ii) $v(s) = \mathbf{f}$ and φ_s^v is satisfiable. Case (i): Since $v_i \in adm(D_i|_{M_i})$, we have that $\varphi_{s,i}^{v_i}$ is irrefutable. But $\models \varphi_s^v \equiv \varphi_{s,i}^{v_i}$, contradiction. Case (ii) is analogous.

$\sigma = pre$: Then there is an interpretation $w \in adm(D)$ with $v <_i w$, that is, there is a statement $s \in S$ such that $v(s) = \mathbf{u}$ and $w(s) \neq \mathbf{u}$. We now find that $v_i(s) = v(s) = \mathbf{u}$ while $w|_{M_i}(s) = w(s) \neq \mathbf{u}$. Thus $v_i <_i w|_{M_i}$. Since w is admissible in D , the interpretation $w|_{M_i}$ is admissible in $D_i|_{M_i}$, see (\star) . But then $v_i \notin pre(D_i|_{M_i})$. Contradiction.

$\sigma = com$: Then there is an $s \in S$ with $v(s) \in \{\mathbf{t}, \mathbf{f}\}$ and the refutability/satisfiability status of φ_s^v does not match. By (\star) , $v_i \in com(D_i|_{M_i})$ implies that $v(s) = v_i(s)$ matches $\varphi_{s,i}^{v_i} \equiv \varphi_s^v$, a contradiction.

$\sigma = mod$: Then there is a statement $s \in S$ such that $v(s) \neq v(\varphi_s)$. But then again by (\star) , $v_i(s) = v(s) \neq v(\varphi_s) = v_i(\varphi_{s,i})$, and $v_i \notin mod(D_i|_{M_i})$. Contradiction.

4. $\sigma = stg$: A witnessing ADF is presented in Example 3.8. \square

As an easy consequence, we get a number of semantics for which the decomposition-based scheme does not lead to new semantics, but rather new ways to compute the semantics. For the grounded semantics, the equality $grd = grd_2$ follows from the same equality for complete semantics.

Corollary 3.10. *For $\sigma \in \{cfi, adm, pre, com, grd, mod\}$ it holds that $\sigma = \sigma_2$.*

As another result, we can show that for the special case of AFs, our nai_2 semantics coincides with AFs' $cf2$ semantics.

Proposition 3.11. *Let \mathcal{A} be an argumentation framework and $D_{\mathcal{A}}$ its associated ADF. The $cf2$ labellings of \mathcal{A} coincide with the nai_2 interpretations of $D_{\mathcal{A}}$.*

Proof. As labellings of AFs and interpretations of ADFs are basically the same, we will write $v = v_v$ for an interpretation v and its counterpart of a labelling v . We show $nai_2(D_{\mathcal{A}}) \subseteq cf2(\mathcal{A})$.

Towards a contradiction, suppose $v_v \notin cf2(\mathcal{A})$. From above we know that for each $v \in nai_2(D_{\mathcal{A}})$ also $v \in nai(D_{\mathcal{A}})$, thus it also holds that each v_v such that $v = v_v$, $v_v \in nai(\mathcal{A})$. So there exists one base case of the $cf2$ calls, i.e. $cf2(\mathcal{A}')$, with $\mathcal{A}' \subseteq \mathcal{A}$ and $|SCCs(\mathcal{A}')| = 1$, such that $v_v|_{A(\mathcal{A}')} \notin nai(\mathcal{A}')$. As the conflict-free property is preserved by the SCC-recursive schema (Baroni et al., 2005), $v_v|_{A(\mathcal{A}'')}$ is conflict-free in \mathcal{A}' . Thus, there exists another conflict-free labelling $v_w|_{A(\mathcal{A}')} \in cf2(\mathcal{A}')$ which is maximal among the arguments labelled with **t** or **f**, i.e. $v_v|_{A(\mathcal{A}')} <_i v_w|_{A(\mathcal{A}'')}$. This means, there is an argument $s \in A(\mathcal{A}')$ with $v_w|_{A(\mathcal{A}')} (s) \neq \mathbf{u}$ but $v_v|_{A(\mathcal{A}')} (s) = \mathbf{u}$. As v_v is naive in \mathcal{A} we know that s is not attacked by any argument r labelled to **t**, thus we can conclude $v_w|_{A(\mathcal{A}')} (s) = \mathbf{t}$. Furthermore, $v_w|_{A(\mathcal{A}) \setminus A(\mathcal{A}')} = v_v|_{A(\mathcal{A}) \setminus A(\mathcal{A}'')}$ and $v_w \in cf2(\mathcal{A})$.

For better readability we will write D instead of $D_{\mathcal{A}} = (S, L, C)$ to refer to the ADF associated to the argumentation framework \mathcal{A} . Let D_i be the ADF with $M_i \subseteq S$ and the interpretation $v_i : M_i \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\} \in nai(D_i|_{M_i})$ with $v_i = v|_{M_i}$ such that $D_{i+1} = \llbracket D_i \rrbracket_{M_i}^{v_i}$ and $M_{i+1} = ind_{D_{i+1}}(M_i)$ and $s \in M_i$. In particular we have $S(D|_{M_i \setminus M_{i-1}}) = A(\mathcal{A}')$. We have $v_i(s) = v(s) = \mathbf{u}$, but for $w|_{M_i \setminus M_{i-1}} = v_w|_{A(\mathcal{A}'')}$, we have $w|_{M_i}(s) = \mathbf{t}$. Thus, $v_i <_i w|_{M_i}$ and as $v_w \in cf2(\mathcal{A})$ also $w|_{M_i}$ is conflict-free in $D_i|_{M_i}$ (see above). But then $v_i \notin nai(D_i|_{M_i})$, a contradiction.

The other direction is by the same argument. □

Another result says that each stg_2 interpretation is also a nai_2 interpretation.

Proposition 3.12. $stg_2 \leq nai_2$

Proof. Let D be an ADF over S and $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\} \in stg_2(D)$. Define $D_1 = D$ and $M_1 = ind_D(\emptyset)$. By definition of stg_2 , for each $i \geq 1$ there exist an ADF D_i , a set $M_i \subseteq S$ and an interpretation $v_i : M_i \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\} \in stg(D_i|_{M_i})$ with $v_i = v|_{M_i}$ such that $D_{i+1} = \llbracket D_i \rrbracket_{M_i}^{v_i}$ and $M_{i+1} = ind_{D_{i+1}}(M_i)$. We have to show that $v \in nai_2(D)$, and do this by using induction on i to show that there exist $D'_i, M'_i \subseteq S$ and interpretations $v'_i : M'_i \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\} \in nai(D'_i|_{M'_i})$ with $v'_i = v|_{M_i}$ such that $D'_{i+1} = \llbracket D'_i \rrbracket_{M'_i}^{v'_i}$ and $M'_{i+1} = ind_{D'_{i+1}}(M'_i)$. In fact, define $D'_i = D_i$, $M'_i = M_i$ and $v'_i = v_i$.

$i = 1$: We have that $v_1 \in stg(D_1|_{M_1}) \subseteq nai(D_1|_{M_1}) = nai(D'_1|_{M'_1})$ by presumption and since $stg \leq nai$. Furthermore, $D'_2 = D_2 = \llbracket D_1 \rrbracket_{M_1}^{v_1} = \llbracket D'_1 \rrbracket_{M'_1}^{v'_1}$ and $M'_2 = M_2 = ind_{D_2}(M_1) = ind_{D'_2}(M'_1)$.

$i \rightsquigarrow i + 1$: Let $v_i \in stg(D_i|_{M_i})$. Again, $stg(D_i|_{M_i}) \subseteq nai(D_i|_{M_i}) = nai(D'_i|_{M'_i})$ and thus $v'_i = v_i \in nai(D'_i|_{M'_i})$. Furthermore, $D'_{i+1} = D_{i+1} = \llbracket D_i \rrbracket_{M_i}^{v_i} = \llbracket D'_i \rrbracket_{M'_i}^{v'_i}$ and $M'_{i+1} = M_{i+1} = ind_{D_{i+1}}(M_i) = ind_{D'_{i+1}}(M'_i)$. □

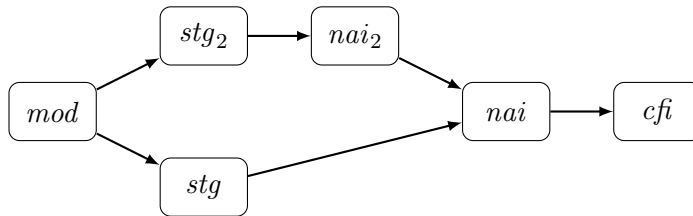


Figure 1: Relation between naive-based ADF semantics in terms of \subseteq -inclusion. An arrow from semantics σ to semantics τ indicates that each σ -interpretation is also a τ -interpretation. If there is no directed path from σ to τ , then one can construct an ADF with a σ -interpretation that is not a τ -interpretation.

Figure 1 gives an overview of the relations between the discussed naive-based ADF semantics. We can see that nai_2 and stg_2 offer further refinements of conflict-freeness that lead to intermediate notions between model (AF stable) and naivety. While one such refinement already existed with stage semantics, we herein present two further refinements that are incomparable to stage semantics, namely nai_2 and stg_2 semantics, with the latter being a further refinement of the former. Instead of taking a global view on an ADF like model, stage, and naive semantics, these two latest additions take a sequence of local views of increasing size, in which statements are successively added as accepted, rejected, or undecided – respectively subject to the notion of conflict-freeness. Arguably, this approach to evaluation of statements could be likened to that of a human reasoner, considering a complex case by starting with the facts at hand, their direct conclusions, and so on, until finally arriving at a verdict taking into account all available information, all the while being subjected to consistency requirements.

Example 3.13. Consider the following situation. There are three persons A, B , and C which support or attack further statements. Person A is against *speed limits* (SL) on German highways and states that B is unreliable. Person B has the goal to make German highways *safer* (S) and states that C is unreliable. Person C wants to save animals by building a *wall* (W) next to highways and states that A is unreliable. In this situation we can identify the following statements a, b, c, sl, s, w with the links and acceptability conditions as shown in Figure 2. Additionally to the attacks in the cycle $\{a, b, c\}$ we identify supports between statements sl and s . When evaluating such a situation humans would first decide which of the persons A, B , and C are reliable, and then depending on the outcome of this evaluation decide on the acceptability of the remaining statements. This leads to the following nai_2 interpretations

$$\begin{aligned}
 v_1 &= \{a \mapsto \mathbf{t}, b \mapsto \mathbf{f}, c \mapsto \mathbf{u}, sl \mapsto \mathbf{f}, s \mapsto \mathbf{f}, w \mapsto \mathbf{t}\}, \\
 v_2 &= \{a \mapsto \mathbf{u}, b \mapsto \mathbf{t}, c \mapsto \mathbf{f}, sl \mapsto \mathbf{t}, s \mapsto \mathbf{t}, w \mapsto \mathbf{f}\}, \\
 v_3 &= \{a \mapsto \mathbf{f}, b \mapsto \mathbf{u}, c \mapsto \mathbf{t}, sl \mapsto \mathbf{t}, s \mapsto \mathbf{t}, w \mapsto \mathbf{t}\}.
 \end{aligned}$$

Thus, when accepting statement a (choosing A to be reliable), this has the influence that statements sl and s will be rejected while statement w still can be accepted. On the other hand, in both cases of accepting either b or c , the statements sl and s will be accepted as

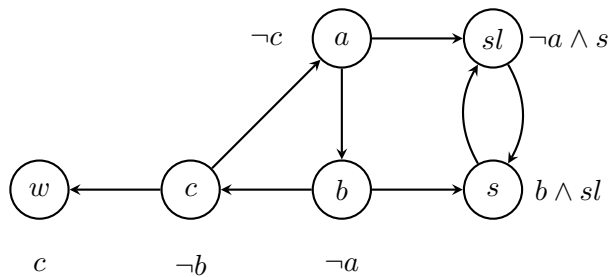


Figure 2: The ADF of Example 3.13

well. In this example one can see the direct influence of the initial cycle (statements a , b and c) on the rest of the ADF.

Such behaviour has been observed in empirical human studies for AFs (Cramer & Guillaume, 2019), which also showed that AF semantics like *cf2* and in later studies the SCF2 semantics (Cramer & van der Torre, 2019) reflect best how humans would evaluate such situations. For ADFs such empirical studies do not exist yet, but it is reasonable to hypothesize that similar behaviour would be observed.

4. Complexity Results

In this section, we analyse the computational complexity of all semantics $\sigma \in \{cfi, nai, stg, nai_2, stg_2\}$. More specifically, we will consider the decision problems of *verification*, (*non-trivial*) *existence*, as well as *credulous* and *sceptical entailment*, respectively defined as follows:

- Ver_σ : Given an ADF D over S and an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$, is $v \in \sigma(D)$?
- $Exists_\sigma$: Given an ADF D over S , does there exist a non-trivial interpretation $v \in \sigma(D)$, that is, one with $v(S) \neq \{\mathbf{u}\}$?
- $Cred_\sigma^{\mathbf{t}} / Cred_\sigma^{\mathbf{f}}$: Given an ADF D over S and an $s \in S$, does there exist an interpretation $v \in \sigma(D)$ with $v(s) = \mathbf{t} / v(s) = \mathbf{f}$?
- $Scep_\sigma^{\mathbf{t}} / Scep_\sigma^{\mathbf{f}}$: Given an ADF D over S and an $s \in S$, is $v(s) = \mathbf{t} / v(s) = \mathbf{f}$ for all $v \in \sigma(D)$?

As we will see later, it is indeed necessary to distinguish the entailment problems $Cred_\sigma^{\mathbf{t}}$ and $Cred_\sigma^{\mathbf{f}}$ (and, respectively, $Scep_\sigma^{\mathbf{t}}$ and $Scep_\sigma^{\mathbf{f}}$), that is, distinguish whether we ask if a statement is sometimes (always) *accepted*, and sometimes (always) *rejected*. For any of the introduced types of decision problems, we say they *coincide* if they yield the same truth values on all inputs.

In several reductions of this paper, we consider quantified Boolean formulas over disjoint vocabularies $P \uplus Q$ with their matrix ψ in either DNF (a disjunction of monomials) or CNF (a conjunction of clauses). They will be used to provide hardness results through reducing checking whether the QBF evaluates to true to some relevant problem at hand.

Sometimes, we cannot use ψ as is, but have to replace atoms from part of its vocabulary, say P , by new literals from a distinct copy of P , the atoms $P' = \{p' \mid p \in P\}$. We will then denote by ψ' the formula obtained from ψ by replacing all positive occurrences of an atom $p \in P$ by the literal $\neg p'$ for the respective $p' \in P'$. For example, for the DNF $\psi = (p_1 \wedge q_1 \wedge \neg p_2) \vee (\neg q_2 \wedge \neg p_1 \wedge p_3)$ we get $\psi' = (\neg p'_1 \wedge q_1 \wedge \neg p_2) \vee (\neg q_2 \wedge \neg p_1 \wedge \neg p'_3)$. The following property of this replacement will be important for us.

Proposition 4.1. *Let ψ be a DNF over $P \uplus Q$. For every interpretation $v : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$, there exists an interpretation $w : P \cup P' \rightarrow \{\mathbf{f}, \mathbf{u}\}$ (with $P' = \{p' \mid p \in P\}$) such that ψ^v is a tautology if and only if ψ'^w is a tautology.*

Proof. For $v : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ define $w : P \cup P' \rightarrow \{\mathbf{f}, \mathbf{u}\}$ thus:

$$w(p) = \begin{cases} \mathbf{f} & \text{if } v(p) = \mathbf{f} \\ \mathbf{u} & \text{otherwise} \end{cases} \quad \text{for } p \in P$$

$$w(p') = \begin{cases} \mathbf{f} & \text{if } v(p) = \mathbf{t} \\ \mathbf{u} & \text{otherwise} \end{cases} \quad \text{for } p \in P$$

It remains to show that ψ^v is a tautology if and only if ψ'^w is a tautology.

if: If ψ^v is refutable, then there is an interpretation $r : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that $r(\psi^v) = (r \circ v)(\psi) = \mathbf{f}$. Define an interpretation $s : P \cup P' \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that $s = w \circ \{p \mapsto \mathbf{t} \mid w(p) = \mathbf{u}\} \circ \{p' \mapsto \mathbf{t} \mid w(p') = \mathbf{u}\}$. Clearly $w \leq_i s$; it furthermore follows that s and v have the same truth values for all literals over P , and that they are compatible with the replacement $p \mapsto \neg p'$:

- $s(p) = \mathbf{t}$ iff $w(p) = \mathbf{u}$ iff $v(p) = \mathbf{t}$;
- $s(p) = \mathbf{f}$ iff $w(p) = \mathbf{f}$ iff $v(p) = \mathbf{f}$;
- $s(p') = \mathbf{t}$ iff $w(p') = \mathbf{u}$ iff $v(p) = \mathbf{f}$;
- $s(p') = \mathbf{f}$ iff $w(p') = \mathbf{f}$ iff $v(p) = \mathbf{t}$; also
- $s(\neg p) = \mathbf{t}$ iff $s(p) = \mathbf{f}$ iff $v(p) = \mathbf{f}$ iff $v(\neg p) = \mathbf{t}$;
- $s(\neg p) = \mathbf{f}$ iff $s(p) = \mathbf{t}$ iff $v(p) = \mathbf{t}$ iff $v(\neg p) = \mathbf{f}$;
- $s(\neg p') = \mathbf{t}$ iff $s(p') = \mathbf{f}$ iff $v(p) = \mathbf{t}$.
- $s(\neg p') = \mathbf{f}$ iff $s(p') = \mathbf{t}$ iff $v(p) = \mathbf{f}$.

Thus $\psi'^s = \psi^v$, consequently $r(\psi'^s) = r(\psi^v) = \mathbf{f}$ and ψ'^w is refutable.

only if: If ψ'^w is refutable, then there is an interpretation $r : P \cup P' \cup Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that $r(\psi'^w) = (r \circ w)(\psi') = \mathbf{f}$. Since ψ' is in DNF, this means that in every monomial of ψ' , there is some literal x' with $(r \circ w)(x') = \mathbf{f}$, that is, either $w(x') = \mathbf{f}$, or $w(x') = \mathbf{u}$ and $r(x') = \mathbf{f}$. We do a case distinction on x' and show that in every monomial ξ of ψ , its corresponding literal x likewise has $(r \circ v)(x) = \mathbf{f}$. Consider an arbitrary, fixed monomial ξ' of $\psi' = \bigvee \xi'$ and its respective false literal x' .

- $x' = p$: Impossible, as positive occurrences of p in ψ have been replaced by $\neg p'$.

- $x' = \neg p$: If $w(\neg p) = \mathbf{f}$ then $w(p) = \mathbf{t}$, which is impossible by construction.
- $x' = p'$: Impossible, as atoms from p' appear only negated in ψ' .
- $x' = \neg p'$: As above, then $w(p') = \mathbf{t}$, which is impossible by definition.
- $x' = q$: Then $r(q) = \mathbf{f}$ and since $x = q$ occurs in ξ we have $(r \circ v)(\xi) = \mathbf{f}$.
- $x' = \neg q$: Then $r(q) = \mathbf{t}$ and since $x = \neg q$ occurs in ξ we get $(r \circ v)(\xi) = \mathbf{f}$.

Since ξ' was an arbitrary monomial, we conclude that $r(\psi^v) = \mathbf{f}$, whence ψ^v is refutable. \square

A similar result holds for satisfiability if ψ is in CNF. We are now ready to present the main complexity results of this paper, tight complexity bounds for all semantics among conflict-free, naive, stage and *nai*₂ for all decision problems introduced above, both for the class of general, unrestricted ADFs, as well as the restricted subclass of bipolar ADFs. The results are grouped together in subsections according to (sub)class and subsections according to decision problems.

4.1 Complexity of General ADFs

We begin with the case of general ADFs, where relationships between statements are unrestricted.

4.1.1 INTERPRETATION VERIFICATION

We start out with verifying if a given interpretation is conflict-free. Roughly, this is done using one satisfiability check and one unsatisfiability check, and the completeness result tells us that we most likely cannot do any better.

Proposition 4.2. *Ver_{cfi} is DP-complete.*

Proof. in DP: Let D be an ADF over S and $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ be an interpretation. To verify that v is conflict-free for D , we have to verify that (1) for all $s \in S$ with $v(s) = \mathbf{t}$, the formula φ_s^v is satisfiable, and (2) for all $s \in S$ with $v(s) = \mathbf{f}$, the formula φ_s^v is unsatisfiable. This can be done in DP by verifying that $\bigwedge_{s \in S, v(s)=\mathbf{t}} \varphi_s^v$ is satisfiable and $\bigvee_{s \in S, v(s)=\mathbf{f}} \varphi_s^v$ is unsatisfiable. Clearly these formulas can be computed in polynomial time.

DP-hard: Let ϕ and ψ be propositional formulas over disjoint vocabularies P_1 and P_2 , respectively. We construct an ADF over statements $S = P_1 \cup P_2 \cup \{x, y\}$ and an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ such that v is conflict-free for D if and only if ϕ is satisfiable and ψ is unsatisfiable. Set $\varphi_p = \neg p$ for all $p \in P_1 \cup P_2$, furthermore set $\varphi_x = \phi$ and $\varphi_y = \psi$. Finally, define v such that $p \mapsto \mathbf{u}$ for $p \in P_1 \cup P_2$, and $x \mapsto \mathbf{t}$ and $y \mapsto \mathbf{f}$. \square

To ease the comprehension of the reductions used in the proofs of our results, we start with one particular reduction that is used to show the Π_2^P -hardness of most other interpretation verification problems.

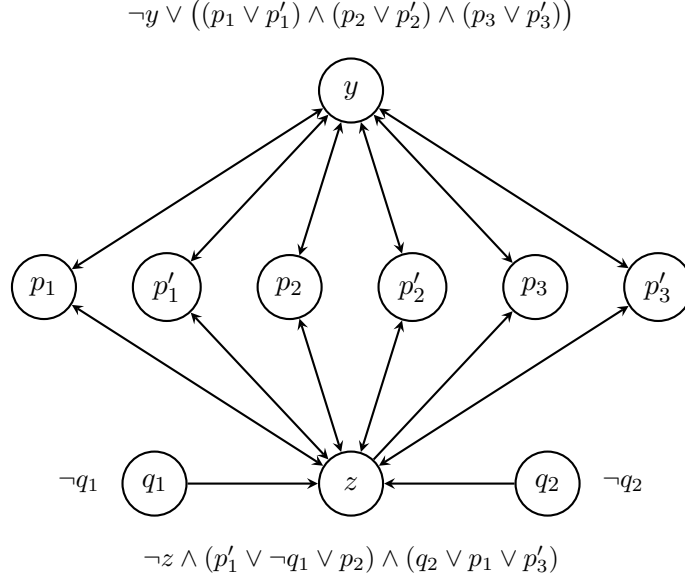


Figure 3: ADF D_Φ for $\Phi = \exists\{p_1, p_2, p_3\}\forall\{q_1, q_2\}(p_1 \wedge q_1 \wedge \neg p_2) \vee (\neg q_2 \wedge \neg p_1 \wedge p_3)$ constructed with Reduction 4.3. To improve the readability of the graph, we omitted the self-loops of all statements and the acceptance formulas $\varphi_p = \neg p \wedge (\neg y \vee z)$ of all $p \in P \cup P'$.

Reduction 4.3. Let $\Phi = \exists P \forall Q \psi$ be a QBF with ψ in DNF. Define an ADF D_Φ over $S = P \cup P' \cup Q \cup \{y, z\}$ with:

$$\begin{aligned} \varphi_p &= \neg p \wedge (\neg y \vee z) && \text{for } p \in P \\ \varphi_{p'} &= \neg p' \wedge (\neg y \vee z) && \text{for } p \in P' \\ \varphi_q &= \neg q && \text{for } q \in Q \\ \varphi_y &= \neg y \vee \bigwedge_{p \in P} (p \vee p') \\ \varphi_z &= \neg z \wedge \neg \psi' \end{aligned}$$

Here ψ' is ψ with all positive occurrences of p replaced by $\neg p'$. Finally, define the interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ such that $v(y) = \mathbf{t}$ and all other statements are mapped to \mathbf{u} .

Intuitively, p and p' serve to guess a valuation for P where setting $p \in S$ to false encodes setting $p \in P$ to false, and setting $p' \in S$ to false encodes setting $p \in P$ to true. All $p, p' \in S$ cannot be set to true, and only be set to false if z is false and y is true; in turn, z can only be set to false if $\neg \psi'$ is unsatisfiable; statement y can only be set to \mathbf{t} or \mathbf{u} . Setting y to true in a conflict-free interpretation v guarantees that for each $p \in P$ at most one of p is false or p' is false in v , but never both. These ideas are reused and (sometimes significantly) elaborated upon in later results. See Figure 3 for the reduction of the QBF $\Phi = \exists\{p_1, p_2, p_3\}\forall\{q_1, q_2\}(p_1 \wedge q_1 \wedge \neg p_2) \vee (\neg q_2 \wedge \neg p_1 \wedge p_3)$ to the ADF D_Φ .

Recall that an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ is naive iff it is conflict-free and \leq_i -maximal with respect to being conflict-free. Thus, to verify that a given interpretation

v is *not* naive, we first check (using an NP oracle) whether v is conflict-free. If v is not conflict-free, we are done; otherwise, we can guess an interpretation v' with $v <_i v'$ and verify in DP (using the NP-oracle again) that v' is conflict-free. Once more, this is the best we can do.

Theorem 4.4. Ver_{nai} is Π_2^P -complete.

Proof. in Π_2^P : Let D be an ADF over S . A given interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ is *not* naive iff it is not conflict-free or if there is a conflict-free interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v <_i w$. Showing that v is not conflict-free can be done in polytime using the NP oracle; otherwise we guess a $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v <_i w$ and use the NP oracle to show that it is conflict-free.

Π_2^P -hard: Let $\Phi = \exists P \forall Q \psi$ be a QBF with ψ in DNF and consider the ADF D_Φ obtained from Reduction 4.3 as well as the interpretation v defined there. We claim that v is naive for D_Φ iff Φ is false, or equivalently, Φ is true iff there is a conflict-free interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v <_i w$.

if: Let $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v <_i w$ be conflict-free. Then there is some $s \in S \setminus \{y\}$ with $w(s) \neq \mathbf{u}$. If $w(s) = \mathbf{t}$ then $\varphi_s^w \equiv \neg \top$ is unsatisfiable in contradiction to w being conflict-free. Thus $w(s) = \mathbf{f}$. We will show that this implies $w(z) = \mathbf{f}$.

- $s = p \in P$: Then $\varphi_p^w = \neg \perp \wedge (\neg \top \vee z^w) \equiv z^w$ is unsatisfiable, that is, $w(z) = \mathbf{f}$.
- $s = p' \in P'$: Analogous.
- $s = q \in Q$: Then $\varphi_q^w = \neg \perp$ is unsatisfiable, contradiction.
- $s = z$: Then $w(z) = \mathbf{f}$.

Since $w(z) = \mathbf{f}$ and w is conflict-free, the formula $\varphi_z^w = \neg \perp \wedge \neg \psi'^w \equiv \neg \psi'^w$ is unsatisfiable, that is, ψ'^w is a tautology. Define an interpretation $w' : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ with $w'(p) = \mathbf{t}$ if $w(p) = \mathbf{f}$ and $w'(p) = \mathbf{f}$ otherwise. Now since ψ'^w is a tautology, Proposition 4.1 shows that $\psi^{w'}$ is a tautology and thus $\exists P \forall Q \psi$ is true.

only if: Let Φ be true. Then there exists an interpretation $w : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that ψ^w is a tautology. Define an interpretation v' as follows:

$$\begin{aligned}
 v'(p) &= \begin{cases} \mathbf{f} & \text{if } w(p) = \mathbf{f} \\ \mathbf{u} & \text{otherwise} \end{cases} && \text{for } p \in P \\
 v'(p') &= \begin{cases} \mathbf{f} & \text{if } w(p) = \mathbf{t} \\ \mathbf{u} & \text{otherwise} \end{cases} && \text{for } p \in P \\
 v'(q) &= \mathbf{u} && \text{for } q \in Q \\
 v'(y) &= \mathbf{t} \\
 v'(z) &= \mathbf{f}
 \end{aligned}$$

Clearly $v <_i v'$ since $v(y) = v'(y)$ and $v(z) = \mathbf{u} <_i \mathbf{f} = v'(z)$. It remains to show that v' is conflict-free for D . This is immediate for p and p' with $v'(p) = \mathbf{f}$ or $v'(p') = \mathbf{f}$, since $v'(y) = \mathbf{t}$. Furthermore, $\varphi_y^{v'}$ is satisfiable since

for each $p \in P$, either $v'(p) = \mathbf{u}$ or $v'(p') = \mathbf{u}$, thus $v'(y) = \mathbf{t}$ is justified. Finally, $\varphi_z^{v'} = \neg \perp \wedge \neg \psi^{v'} \equiv \neg \psi^{v'}$ is unsatisfiable since ψ^w is a tautology (Proposition 4.1). \square

For verifying stage interpretations, membership works in the same way as for naive. For hardness, a close look at Reduction 4.3 reveals that it also works for stage semantics.

Theorem 4.5. *Ver_{stg} is Π_2^P -complete.*

Proof. in Π_2^P : Let D be an ADF over S . A given interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ is *not* stage iff it is not conflict-free or if there is a conflict-free interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $w_{\mathbf{u}} \subsetneq v_{\mathbf{u}}$. Showing that v is not conflict-free can be done in P using the NP oracle; otherwise we guess a $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $w_{\mathbf{u}} \subsetneq v_{\mathbf{u}}$ and use the NP oracle to show that it is conflict-free.

Π_2^P -hard: Let $\Phi = \exists P \forall Q \psi$ be a QBF with ψ in DNF and consider the ADF D_{Φ} obtained from Reduction 4.3. We claim that v is a stage interpretation for D_{Φ} iff Φ is false, or equivalently, Φ is true iff there is a conflict-free interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $w_{\mathbf{u}} \subsetneq v_{\mathbf{u}}$.

if: Let $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $w_{\mathbf{u}} \subsetneq v_{\mathbf{u}}$ be conflict-free. As in the proof of Theorem 4.4 (both results use Reduction 4.3, and in the proof of Theorem 4.4 we only use that there must exist an $s \in S \setminus \{y\}$ with $w(s) \neq \mathbf{u}$, which also applies here), we can argue that in this case $w(z) = \mathbf{f}$ and thus ψ^w is a tautology. (Notice that $w(y) = \mathbf{f}$ is impossible due to conflict-freeness.) Likewise, we can construct an interpretation $w' : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ showing that $\psi^{w'}$ is a tautology and thus $\Phi = \exists P \forall Q \psi$ is true.

only if: Let Φ be true. Again, as in the proof of Theorem 4.4 based on the same reduction, starting from the interpretation $w : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that ψ^w is a tautology, we can construct an interpretation $v' : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ that has $v'_{\mathbf{u}} \subsetneq v_{\mathbf{u}}$ and is conflict-free for D_{Φ} . \square

The same hardness reduction (Reduction 4.3) even works for the nai_2 and stg_2 semantics. It is somewhat harder to show membership in Π_2^P via a reduction to $\text{Ver}_{nai}/\text{Ver}_{stg}$: intuitively, this is done by parallelising the single (independent) verifications of nai/stg interpretations in all SCCs of a given ADF D .

Theorem 4.6. *Ver_{nai_2} and Ver_{stg_2} are Π_2^P -complete.*

Proof. in Π_2^P : We show the reduction of Ver_{nai_2} to Ver_{nai} , since the reduction of Ver_{stg_2} to Ver_{stg} is the same. Let D be an ADF over S and the interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ be given. We recursively compute the unique decomposition of D with respect to v . In the following we denote the independent sets for each recursive call by M_i for $0 \leq i < n$, that is, $M_0 = \text{ind}_D(\emptyset)$ and $M_{i+1} = \text{ind}_{\llbracket D \rrbracket_{M_i}^{v_i}}(M_i)$ with $v_i = v|_{M_i}$. In each recursive call we make a new, distinct copy D_i of the ADF $\llbracket D \rrbracket_{M_i}^{v_i}$ and the respective restricted interpretation $v_i = v|_{M_i}$, that is, for $0 \leq i < n$ define an ADF $D_i = (S_i, L_i, C_i)$ with

statements $S_i = \{s_i \mid s \in M_i\}$, links $L_i = \left\{ (s_i, t_i) \mid s_i, t_i \in S_i, (s, t) \in L(\llbracket D \rrbracket_{M_i}^{v_i}) \right\}$, acceptance formulas $\varphi_{s_i} = \varphi_s[s/s_i : s \in M_i]$ for $\varphi_s \in C(\llbracket D \rrbracket_{M_i}^{v_i})$ and furthermore an interpretation $w_i : S_i \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $w_i(s_i) = v_i(s)$ for all $s \in M_i$. Let $M_k = S$ be the independent set for the last recursive call. (Clearly $k < n$.) Now we have that $v \in \text{nai}_2(D)$ if and only if $w' \in \text{nai}(D')$, where

$$w' = \bigcup_{i=0}^k w_i \quad \text{and} \quad D' = \left(\bigcup_{i=0}^k S_i, \bigcup_{i=0}^k L_i, \bigcup_{i=0}^k C_i \right)$$

The computation of D' can be done in at most n steps (for $|S| = n$) with at most $\frac{n(n+1)}{2}$ statements in D' .

Π_2^P -hard: As in the proof of Theorem 4.4 (Theorem 4.5), that is, via Reduction 4.3: In any nai_2 (stg_2) interpretation, the $q \in Q$ can only be set to \mathbf{u} since $\varphi_q = \neg q$, and the rest of the constructed ADF has only one strongly connected component whence nai_2 and nai (stg_2 and stg) semantics coincide. \square

4.1.2 EXISTENCE OF NON-TRIVIAL INTERPRETATIONS

Deciding whether a given ADF D has at least one non-trivial conflict-free interpretation turns out to be complete for the less well-known complexity class coDP . Intuitively, a coDP -problem allows us to choose whether we “want” to solve an NP- or a coNP -problem, but we have to solve at least one of them correctly. Showing coDP -hardness for $\text{Exists}_{\text{cfi}}$ is comparably easy. The canonical coDP -complete problem is the following, SAT-OR-UNSAT: Given two propositional formulas ϕ and ψ , is ϕ satisfiable or ψ unsatisfiable? Note that the “or” is to be understood logically, that is, it suffices to answer at least one of the questions correctly.

Showing membership of $\text{Exists}_{\text{cfi}}$ for coDP is quite tricky. The first useful observation is that there are essentially only two distinct types of non-trivial conflict-free interpretations:

- (1) those $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v^{-1}(\mathbf{t}) \neq \emptyset$, that is, where some statement is mapped to true;
- (2) those with $v(S) \subseteq \{\mathbf{u}, \mathbf{f}\}$ and $v^{-1}(\mathbf{f}) \neq \emptyset$, that is, all statements are mapped to undefined or false and at least one is mapped to false.

The proof works by showing that existence of non-trivial conflict-free interpretations of type (1) can be decided in NP, and that the existence of those of type (2) can be decided in coNP . In combination, membership for coDP follows.

Showing the first part is straightforward: to decide whether some statement $s \in S$ can be set to true without violating conflict-freeness, we construct the propositional formula $\bigvee_{s \in S} \varphi_s^{\{s \rightarrow \mathbf{t}\}}$ and check if it is satisfiable. If for some $s \in S$ the formula $\varphi_s^{\{s \rightarrow \mathbf{t}\}}$ is satisfiable, then $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(s) = \mathbf{t}$ and $v(s') = \mathbf{u}$ for $s' \in S \setminus \{s\}$ is conflict-free. Otherwise, no $s \in S$ can be set to true in a conflict-free way.

Showing the second part about interpretations $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(S) \subseteq \{\mathbf{u}, \mathbf{f}\}$ (we call them \mathbf{uf} -interpretations) constitutes the main portion of the proof. Roughly,

conflict-free **uf**-interpretations are closed under the least upper bound operator \sqcup_i associated to the information ordering \leq_i on interpretations. That is, whenever v_1 and v_2 are **uf**-interpretations that are conflict-free for D , then the interpretation $v_1 \sqcup_i v_2$ is a **uf**-interpretation that is conflict-free for D as well. Since both $v_1 \leq_i v_1 \sqcup_i v_2$ and $v_2 \leq_i v_1 \sqcup_i v_2$ by definition, there is a unique \leq_i -greatest conflict-free **uf**-interpretation $v_{\max} : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ of D . Our task is to decide whether v_{\max} is non-trivial. We first show how to do this by computing v_{\max} in polynomial time using an NP oracle. The procedure works constructively and begins with the interpretation $v_0 : S \rightarrow \{\mathbf{f}\}$, that is, mapping all statements to false. The computation now stepwise ($j = 0, 1, \dots, n-1$) reassigns $v_{j+1}(s) = \mathbf{u}$ for $s \in v_j^{-1}(\mathbf{f})$ whenever it is the case that assigning $v_j(s) = \mathbf{f}$ is actually not justified because $\varphi_s^{v_j}$ is satisfiable (otherwise, it keeps $v_{j+1}(s) = v_j(s) = \mathbf{f}$). To answer the satisfiability queries, the procedure can use the NP-oracle. The proof finally shows how to combine all the oracle queries into one satisfiability check. This is done by encoding the whole computation into a propositional formula ϕ_{cfi} of polynomial size such that the formula is satisfiable if and only if there is a possible computation that starts with $v_0(S) = \{\mathbf{f}\}$ and ends in the trivial $v_n(S) = \{\mathbf{u}\}$. Since such a computation would show that v_{\max} is trivial, there is a non-trivial conflict-free **uf**-interpretation of D if and only if the formula ϕ_{cfi} is unsatisfiable. This then shows membership in coNP for checking whether there is a non-trivial conflict-free interpretation of type (2), and thus concludes the coDP-membership proof.

Theorem 4.7. *Exists_{cfi} is coDP-complete.*

Proof. in coDP: There are essentially two types of non-trivial conflict-free interpretations $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$:

- (1) those with $v^{-1}(\mathbf{t}) \neq \emptyset$, that is, where some statement is mapped to true;
- (2) those with $v(S) \subseteq \{\mathbf{u}, \mathbf{f}\}$ and $v^{-1}(\mathbf{f}) \neq \emptyset$, that is, all statements are mapped to undefined or false and at least one is mapped to false.

We will argue that the existence of non-trivial conflict-free interpretations of type (1) can be decided in NP, and that the existence of those of type (2) can be decided in coNP. In combination, we get the desired membership for coDP.

Type (1): We test satisfiability of the formula

$$\bigvee_{s \in S} \varphi_s^{\{s \mapsto \mathbf{t}\}}$$

If for some $s \in S$ the formula $\varphi_s^{\{s \mapsto \mathbf{t}\}}$ is satisfiable, then $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(s) = \mathbf{t}$ and $v(s') = \mathbf{u}$ for $s' \in S \setminus \{s\}$ is conflict-free. Otherwise, no $s \in S$ can be set to true without violating conflict-freeness. (If $\varphi_s^{\{s \mapsto \mathbf{t}\}}$ is unsatisfiable, it will stay so no matter if and how we set additional statements in φ_s .)

Type (2): We call an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ a **uf**-interpretation iff $v(S) \subseteq \{\mathbf{u}, \mathbf{f}\}$. These interpretations have the useful property that whenever v_1 and v_2 are

uf-interpretations that are conflict-free for an ADF D , then the interpretation $v_1 \sqcup_i v_2$ with

$$(v_1 \sqcup_i v_2)(s) = \begin{cases} v_1(s) & \text{if } v_1(s) = \mathbf{f} \\ v_2(s) & \text{otherwise} \end{cases}$$

is a **uf**-interpretation that is conflict-free for D as well. Since $v_1, v_2 \leq_i v_1 \sqcup_i v_2$, there is a unique \leq_i -greatest conflict-free **uf**-interpretation $v_{\max} : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ of D . Our task is to decide whether v_{\max} is non-trivial. We start out with showing how to do this by computing v_{\max} in polynomial time using an NP oracle. We will afterwards show how the whole computation can be encoded into a single satisfiability check.

Let D be an ADF over $S = \{s_1, \dots, s_n\}$.

- (a) Set $v_0 = \{s \mapsto \mathbf{f} \mid s \in S\}$.
- (b) For all $j = 0, \dots, n - 1$ do the following:
 - For all $1 \leq i \leq n$ with $v_j(s_i) = \mathbf{f}$, ask the oracle (in multiple queries) if $\varphi_{s_i}^{v_j}$ is unsatisfiable.
 - Set $v_{j+1} = v_j \circ \{s_i \mapsto \mathbf{u} \mid \varphi_{s_i}^{v_j} \text{ is satisfiable}\}$

It is clear that the algorithm loops n times and each iteration asks at most n oracle queries. The procedure computes the \leq_i -greatest conflict-free **uf**-interpretation, so there exists a non-trivial one iff $v_n \neq (S \mapsto \mathbf{u})$.

We encode this computation into a propositional formula ϕ_{cfi} of polynomial size such that ϕ_{cfi} is satisfiable iff the procedure terminates with $v_n = \{s \mapsto \mathbf{u} \mid s \in S\}$ (which is iff D has no non-trivial conflict-free **uf**-interpretation). The main idea is to construct the formula such that each satisfying evaluation corresponds to a run of the procedure that starts with $v_0(S) = \{\mathbf{f}\}$ and ends in $v_n(S) = \{\mathbf{u}\}$. The propositional variables will encode the current interpretation v_j at each step j , and further copies of the ADF's statements (another one for each step j) will provide the variables to encode the necessary oracle calls of each step. The main difficulty of the encoding is to parallelise the oracle calls without sacrificing completeness. We start out with defining the vocabulary P .

$$P = \{s_{i,j}, s_{i,j,k} \mid 1 \leq i, k \leq n, 0 \leq j \leq n\}$$

Intuitively, $s_{i,j}$ is true iff $v_j(s_i) = \mathbf{u}$ (thus $s_{i,j}$ is false iff $v_j(s_i) = \mathbf{f}$). Likewise, $s_{k,i,j}$ is a copy of s_k that is needed to guess whether $\varphi_{s_i}^{v_j}$ is satisfiable. To this end, we denote $\varphi_{i,j} = \varphi_{s_i}[s_k/s_{k,i,j} : 1 \leq k \leq n]$. Now we specify the parts of the formula ϕ_{cfi} .

$$\begin{aligned}
 \phi_0 &= \bigwedge_{1 \leq i \leq n} \neg s_{i,0} \\
 \phi_n &= \bigwedge_{1 \leq i \leq n} s_{i,n} \\
 \phi_{i,j}^? &= (\neg s_{i,j-1} \wedge s_{i,j}) \rightarrow \varphi_{i,j-1} && \text{(for } 1 \leq i, j \leq n) \\
 \phi_{i,j}^{\mathbf{f}} &= \neg s_{i,j-1} \rightarrow \bigwedge_{1 \leq k \leq n} \neg s_{i,k,j-1} && \text{(for } 1 \leq i, j \leq n) \\
 \phi_{cfi} &= \phi_0 \wedge \phi_n \wedge \bigwedge_{1 \leq i, j \leq n} \phi_{i,j}^? \wedge \bigwedge_{1 \leq i, j \leq n} \phi_{i,j}^{\mathbf{f}}
 \end{aligned}$$

First of all, ϕ_0 and ϕ_n encode the intended starting and ending interpretations of the procedure. Formula $\phi_{i,j}^?$ encodes that whenever some statement s_i changes its truth value from $v_{j-1}(s_i) = \mathbf{f}$ to $v_j(s_i) = \mathbf{u}$, then the formula $\varphi_{s_i}^{v_{j-1}}$ is satisfiable. Formula $\phi_{i,j}^{\mathbf{f}}$ expresses that all set truth values $v_{j-1}(s_i) = \mathbf{f}$ must be reflected in the variables used for guessing the satisfiability of $\varphi_{s_i}^{v_{j-1}}$.

To conclude the proof, we will now show that ϕ_{cfi} is satisfiable iff there is a run of the procedure that starts with $v_0(S) = \{\mathbf{f}\}$ and ends in $v_n(S) = \{\mathbf{u}\}$.

if: Let $v_0, v_1, \dots, v_n : S \rightarrow \{\mathbf{u}, \mathbf{f}\}$ be such that $v_0(S) = \{\mathbf{f}\}$ and $v_n(S) = \{\mathbf{u}\}$, and for all $1 \leq i, j \leq n$, we have that $v_{j-1}(s_i) = \mathbf{f}$ and $v_j(s_i) = \mathbf{u}$ imply that the formula $\varphi_{s_i}^{v_{j-1}}$ is satisfiable. We can use the witnesses for satisfiability to define valuations of the $s_{k,i,j}$ and thus build a model of ϕ_{cfi} .

only if: From a model $I \subseteq P$ of ϕ_{cfi} , we first use the truth values of the $s_{i,j}$ for $1 \leq i \leq n$ and $0 \leq j \leq n$ to define a sequence of interpretations $v_0, v_1, \dots, v_n : S \rightarrow \{\mathbf{u}, \mathbf{f}\}$ as follows. Set $v_0(s_i) = \mathbf{f}$ for all $1 \leq i \leq n$; for $1 \leq i, j \leq n$ set

$$v_j(s_i) = \begin{cases} \mathbf{u} & \text{if } I \models s_{i,j} \text{ or } v_{j-1}(s_i) = \mathbf{u} \\ \mathbf{f} & \text{otherwise} \end{cases}$$

This guarantees that $v_n \leq_i v_{n-1} \leq_i \dots \leq_i v_0$ and furthermore each statement s_i changes its truth value at most once, and only from \mathbf{f} to \mathbf{u} . Since $I \models \bigwedge_{1 \leq i, j \leq n} \phi_{i,j}^?$, we know that for all $1 \leq i, j \leq n$ with $v_{j-1}(s_i) = \mathbf{f}$ and $v_j(s_i) = \mathbf{u}$ we have $I \models \varphi_{i,j-1}$. Since also $I \models \bigwedge_{1 \leq i, j \leq n} \phi_{i,j}^{\mathbf{f}}$, we know that $I \not\models s_{i,j-1}$ implies $I \not\models s_{i,k,j-1}$ for all $1 \leq k \leq n$. Now this and the fact that $I \models \varphi_{i,j-1}$ shows that $\varphi_{s_i}^{v_{j-1}}$ is satisfiable, for all $1 \leq i, j \leq n$.

coDP-hard: Given a pair of formulas (ϕ, ψ) over P_1 and P_2 , respectively, we construct the ADF D over $S = P_1 \cup P_2 \cup \{y, z\}$ with

$$\begin{aligned}
 \varphi_p &= \neg p && \text{for } p \in P_1 \cup P_2 \\
 \varphi_y &= \neg y \vee \phi \\
 \varphi_z &= \neg z \wedge \psi
 \end{aligned}$$

It is easy to see that D has a non-trivial conflict-free interpretation v with $v(y) = \mathbf{t}$ iff ϕ is satisfiable, and that D has a non-trivial conflict-free interpretation v with $v(z) = \mathbf{f}$

iff ψ is unsatisfiable. In combination, D has a non-trivial conflict-free interpretation iff ϕ is satisfiable or ψ is unsatisfiable. \square

Fortunately, the amount of work put into this proof “pays off” in that deciding the existence of non-trivial conflict-free interpretations also decides the existence of naive, stage, nai_2 and stg_2 interpretations. The first technical result towards establishing that is the following lemma. It shows how every conflict-free interpretation v gives rise to a naive (or stage) interpretation v' that is “above v ” with respect to some ordering. In case of naive, the ordering is clearly the information ordering \leq_i . In case of stage, the ordering is given by comparing the statements that are assigned the truth value \mathbf{u} by the two interpretations.

Lemma 4.8. *Let D be an ADF over S . For every interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ that is conflict-free for D , there exists:*

1. a naive interpretation $v' : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v \leq_i v'$;
2. a stage interpretation $v'' : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v''_{\mathbf{u}} \subseteq v_{\mathbf{u}}$.

Proof. 1. By induction on the number $n = |v_{\mathbf{u}}|$ of statements that are undefined in v . For $n = 0$, we have that v itself is naive since it is conflict-free and trivially \leq_i -maximal because it is two-valued. Now consider an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $|v_{\mathbf{u}}| = n + 1$ that is conflict-free, but not naive. Then there exists a conflict-free interpretation v' with $v <_i v'$. Then $|v'_{\mathbf{u}}| \leq n$ and by induction hypothesis, there is a naive interpretation v'' with $v <_i v' \leq_i v''$.

2. By induction on the number $n = |v_{\mathbf{u}}|$ of statements that are undefined in v . For $n = 0$, we have that v itself is stage since it is conflict-free and $v_{\mathbf{u}} = \emptyset$ is trivially \subseteq -minimal. Now consider an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $|v_{\mathbf{u}}| = n + 1$ that is conflict-free, but not stage. Then there exists a conflict-free interpretation v' with $v'_{\mathbf{u}} \subsetneq v_{\mathbf{u}}$. Then $|v'_{\mathbf{u}}| \leq n$ and by induction hypothesis, there is an interpretation $v'' : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v''_{\mathbf{u}} \subseteq v'_{\mathbf{u}} \subsetneq v_{\mathbf{u}}$ that is stage for D . \square

The lemma can be used to show that not only do the non-trivial existence problems coincide, but also credulous entailment for conflict-free and naive semantics are equivalent. Intuitively, if an ADF D has a conflict-free interpretation v with, say, $v(s) = \mathbf{t}$, then the lemma above guarantees the existence of a naive w with $v \leq_i w$ and thus $w(s) = v(s) = \mathbf{t}$.

Proposition 4.9. *The following decision problems coincide:*

1. Exists_{cfi} , Exists_{nai} , Exists_{stg} , Exists_{nai_2} , Exists_{stg_2} ;
2. $\text{Cred}_{cfi}^{\mathbf{t}}$ and $\text{Cred}_{nai}^{\mathbf{t}}$;
3. $\text{Cred}_{cfi}^{\mathbf{f}}$ and $\text{Cred}_{nai}^{\mathbf{f}}$.

Proof. Let D be an ADF over S and $s \in S$. We first show the claims about naive semantics. If $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ is conflict-free for D , and $v(s) = \mathbf{t}$ (or $v(s) = \mathbf{f}$, respectively), then by Lemma 4.8 there is a naive interpretation $v' : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v \leq_i v'$, and thus $v'(s) = \mathbf{t}$ (or $v'(s) = \mathbf{f}$, respectively). For the converse direction, if $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ is naive for D , then v is in particular conflict-free. Now for stage semantics. If $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ is conflict-free and non-trivial, then $v_{\mathbf{u}} \subsetneq S$. By Lemma 4.8, there is a stage interpretation v'' with $v''_{\mathbf{u}} \subseteq v_{\mathbf{u}} \subsetneq S$ that is likewise non-trivial. For the converse direction, again every stage interpretation is conflict-free. Now for Exists_{nai_2} . As each nai_2 interpretation is a naive interpretation (Theorem 3.9), if there is a non-trivial nai_2 interpretation, then this interpretation is also a non-trivial naive interpretation. Conversely, if there is no non-trivial nai_2 interpretation, then the trivial interpretation $v(S) = \{\mathbf{u}\}$ is nai_2 . Since it is then also naive, that is, \leq_i -maximal conflict-free, there cannot be another (non-trivial) conflict-free (let alone naive) interpretation. Thus, the existence of the trivial naive interpretation coincides with the existence of the trivial nai_2 interpretation. For Exists_{stg_2} , the argument is the same as $stg_2 \leq nai_2$ by Proposition 3.12. \square

4.1.3 ENTAILMENT

While verification is a quite basic reasoning task, and non-trivial interpretation existence is mostly used to figure out if a given knowledge base is sensible at all, the entailment problem is most likely to be repeatedly used in practice. Recalling that ADFs are intended for modelling argumentation scenarios, entailment queries then allow to answer questions about these scenarios, such as, “Is it the case that there is one possible interpretation of this scenario where statement a is true?” For the conflict-free semantics, this problem is, while infeasible in a conservative sense, still relatively easy.

Theorem 4.10. Cred_{cfi}^t is NP-complete.

Proof. in NP: Given an ADF D over S and an $s \in S$, we consider the interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(s) = \mathbf{t}$ and $v(s') = \mathbf{u}$ for all $s' \in S \setminus \{s\}$. We then compute the formula φ_s^v and verify in NP that it is satisfiable. If φ_s^v is satisfiable, then v is a conflict-free interpretation for D with $v(s) = \mathbf{t}$. If φ_s^v is unsatisfiable, then φ_s^w is unsatisfiable for all $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v \leq_i w$, so there is no conflict-free interpretation where s is true.

NP-hard: We reduce from satisfiability checking. Let ψ be a propositional formula over vocabulary P . Define an ADF D over $S = P \cup \{z\}$ with $\varphi_p = p$ and $\varphi_z = \psi$. It is clear that there is a conflict-free interpretation with $v(z) = \mathbf{t}$ iff ψ is satisfiable. \square

Astonishingly, for similar questions of the form, “Is it the case that there is one possible (conflict-free) interpretation of this scenario where statement a is false?”, giving an answer becomes harder! This asymmetry is quite remarkable, and has its cause in the asymmetry of the definition of a conflict-free interpretation: $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ is conflict-free iff for each $s \in S$ with $v(s) = \mathbf{t}$ the formula φ_s^v is satisfiable, and for each $s \in S$ with $v(s) = \mathbf{f}$ the formula φ_s^v is unsatisfiable. So in one case, there is a satisfiability check, in the other there is an unsatisfiability check. To decide credulous entailment, we clearly have to guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with a desired property (such as $v(s) = \mathbf{t}$ or $v(s) = \mathbf{f}$). And

while the witnesses for verifying $v(s) = \mathbf{t}$ can be guessed alongside v , such is not possible when having to verify $v(s) = \mathbf{f}$. Formally, the hardness part of the result below is proved via a reduction from the problem of deciding whether a quantified Boolean formula $\exists P \forall Q \psi$ is true.

Theorem 4.11. $\text{Cred}_{cft}^{\mathbf{f}}$ is Σ_2^{P} -complete.

Proof. in Σ_2^{P} : Given an ADF D over S and an $s \in S$, we can guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(s) = \mathbf{f}$ and verify in DP that it is conflict-free.

Σ_2^{P} -hard: Let $\Phi = \exists P \forall Q \psi$ be a QBF with ψ in DNF. Define an ADF D over $S = P \cup Q \cup \{z\}$ such that:

$$\begin{aligned} \varphi_p &= p && \text{for } p \in P \\ \varphi_q &= \neg q && \text{for } q \in Q \\ \varphi_z &= \neg \psi \end{aligned}$$

We now show that there is a conflict-free interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(z) = \mathbf{f}$ iff the QBF Φ is true.

if: Let $\exists P \forall Q \psi$ be true. Then there exists a $w : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that ψ^w is a tautology, that is, $\varphi_z^w = \neg \psi^w$ is unsatisfiable. We construct a non-trivial interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(z) = \mathbf{f}$ and show that it is conflict-free. Set $v(z) = \mathbf{f}$, $v(q) = \mathbf{u}$ for all $q \in Q$, and $v(p) = w(p)$ for $p \in P$. Clearly v is conflict-free for P and Q , and setting z to false is justified because $\varphi_z^w = \neg \psi^w$ is unsatisfiable.

only if: Let $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ be conflict-free with $v(z) = \mathbf{f}$. It follows that $v(q) = \mathbf{u}$ for all $q \in Q$, and that $\varphi_z^v = \neg \psi^v$ is unsatisfiable, that is, ψ^v is a tautology. Define a two-valued interpretation $w : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ as follows: Set $w(p) = v(p)$ if $v(p) \in \{\mathbf{t}, \mathbf{f}\}$, and $w(p) = \mathbf{t}$ otherwise. Now w extends v on P ($v|_P \leq_i w|_P$), so ψ^w is a tautology and w witnesses that the QBF $\exists P \forall Q \psi$ is true. \square

There is a straightforward way to show that a statement $s \in S$ is *not* sceptically entailed as true by an ADF D over S : guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(s) \neq \mathbf{t}$ and show that v is naive. Since Ver_{nai} is in Π_2^{P} , this straightforward approach yields membership of $\text{Scep}_{nai}^{\mathbf{t}}$ in Π_3^{P} . Fortunately, there is an easier way: we guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(s) = \mathbf{u}$, and verify (using the NP oracle) that v is conflict-free for D , while the augmented interpretation $v \circ \{s \mapsto \mathbf{t}\}$ is *not* conflict-free for D . Intuitively, this identifies statement $s \in S$ as a “troublemaker”, as the one reason that violates conflict-freeness in all interpretations with at least as much information as v . Since among these interpretations at least one must be naive, we have our desired counterexample for sceptical entailment. This yields membership in Π_2^{P} ; as it turns out, that is the best possible bound.

Theorem 4.12. $\text{Scept}_{\text{nai}}^{\mathbf{t}}$ is Π_2^{P} -complete.

Proof. in Π_2^{P} : Let D be an ADF over S and $z \in S$. To show that z is *not* sceptically true in D , we do the following: we guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ and verify (using the NP oracle) that v is conflict-free for D , and furthermore verify that one of (1) $v(z) = \mathbf{f}$, or (2) $v(z) = \mathbf{u}$ and $v \circ \{z \mapsto \mathbf{t}\}$ is not conflict-free for D . We claim that (2) entails that all naive interpretations $v' : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v \leq_i v'$ cannot have $v'(z) = \mathbf{t}$. By Lemma 4.8, it follows that at least one such naive interpretation v' with $v'(z) \neq \mathbf{t}$ exists. Conversely, whenever z is not sceptically true, an interpretation with (1) or (2) must exist. (If z is not sceptically true then there is a naive $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $w(z) = \mathbf{f}$ or $w(z) = \mathbf{u}$, which is the desired witness in both cases, as in the second case w being naive means that w is conflict-free and $w \circ \{z \mapsto \mathbf{t}\}$ is not conflict-free.) It remains to prove the claim. If v is conflict-free and $v \circ \{z \mapsto \mathbf{t}\}$ is not conflict-free, there is a statement $s \in S$ such that (1) $v(s) = \mathbf{t}$, φ_s^v is satisfiable, and $\varphi_s^{v \circ \{z \mapsto \mathbf{t}\}}$ is unsatisfiable; or (2) $v(s) = \mathbf{f}$, φ_s^v is unsatisfiable, and $\varphi_s^{v \circ \{z \mapsto \mathbf{t}\}}$ is satisfiable. Clearly case (2) is impossible, so case (1) holds. But then for any interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v \leq_i w$ and $w(z) = \mathbf{t}$, we find that φ_s^w is unsatisfiable. Consequently, if $v \leq_i w$ and w is conflict-free, then $w(z) \neq \mathbf{t}$.

Π_2^{P} -hard: Let $\Phi = \forall P \exists Q \psi$ be a QBF. We construct an ADF D over $S = P \cup Q \cup \{z\}$ as follows:

$$\begin{aligned} \varphi_p &= p && \text{for } p \in P \\ \varphi_q &= \neg q && \text{for } q \in Q \\ \varphi_z &= \neg z \vee \psi \end{aligned}$$

We now show that z is sceptically true in D iff Φ is true.

if: Let Φ be true and consider a naive interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$. Due to \leq_i -maximality of v it follows that $v(P) \subseteq \{\mathbf{t}, \mathbf{f}\}$, by conflict-freeness we have $v(Q) = \{\mathbf{u}\}$. Since Φ is true, the formula ψ^v is satisfiable whence $v(z) = \mathbf{t}$ by \leq_i -maximality.

only if: We show the contrapositive. Let Φ be false. Then there is an interpretation $v_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that ψ^{v_P} is unsatisfiable. We show that there is a naive interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(z) \neq \mathbf{t}$. Define $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ such that $v(p) = v_P(p)$ for all $p \in P$, $v(q) = \mathbf{u}$ for all $q \in Q$ and $v(z) = \mathbf{u}$. To show that v is naive, it suffices to show that $v(z) \in \{\mathbf{t}, \mathbf{f}\}$ is not consistently possible. First, $\varphi_z^{v \circ \{z \mapsto \mathbf{t}\}} = \neg \top \vee \psi^v \equiv \psi^v = \psi^{v_P}$ is unsatisfiable by assumption. Next, $\varphi_z^{v \circ \{z \mapsto \mathbf{f}\}} = \neg \perp \vee \psi^v \equiv \top$ is satisfiable by construction. \square

The straightforward approach to decide sceptical entailment of truth clearly also works for sceptical entailment of falsity. In this case, however, it turns out that there is no shortcut. For the (quite technical) proof of the result, we adapt and combine proof techniques from Strass and Wallner (2015, Theorem 4.22) and Theorem 4.4.

Theorem 4.13. Scep_{nai}^f is Π_3^P -complete.

Proof. in Π_3^P : Let D be an ADF over S and $s \in S$. To show that s is *not* sceptically false, guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(s) \neq \mathbf{u}$ and verify in Π_2^P that v is naive.

Π_3^P -hard: Let $\Phi = \forall P \exists Q \forall R \psi$ be a QBF with ψ in DNF. We construct an ADF D over the set of statements $S = P \cup Q \cup Q' \cup R \cup \{y, z\}$ with $Q' = \{q' \mid q \in Q\}$ such that z is false in every naive interpretation of D if and only if Φ is true. Define D as follows:

$$\begin{aligned}
 \varphi_p &= p && \text{for } p \in P \\
 \varphi_q &= \neg q \wedge (\neg y \vee z) && \text{for } q \in Q \\
 \varphi_{q'} &= \neg q' \wedge (\neg y \vee z) && \text{for } q \in Q \\
 \varphi_r &= \neg r && \text{for } r \in R \\
 \varphi_y &= \neg y \vee \bigwedge_{q \in Q} (q \vee q') \\
 \varphi_z &= \neg z \wedge \neg \psi'
 \end{aligned}$$

As before ψ' is obtained from ψ by replacing all positive occurrences of $q \in Q$ by $\neg q'$; setting q' to false aims at encoding setting q to true. We start out with some observations and intuitive explanations:

1. The p can be set arbitrarily to true and false; in particular, any naive interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ satisfies $v(P) \subseteq \{\mathbf{t}, \mathbf{f}\}$ due to the \leq_i -maximality condition.
2. The q and q' cannot be set to true, and only to false if z is false and y is true.
3. y cannot be set to false and only be set to true if for each $q \in Q$, at least one of q or q' is undefined; it thereby guarantees that the q, q' encode a two-valued interpretation of Q for each naive pair with y true and z false.
4. z cannot be set to true and only to false if the corresponding partial evaluation of $\neg \psi'$ is unsatisfiable.

We next show that z is false in every naive interpretation of D if and only if Φ is true.

if: Let Φ be true and consider the naive interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$. We have to show that $v(z) = \mathbf{f}$. Assume to the contrary that $v(z) \neq \mathbf{f}$, that is, $v(z) = \mathbf{u}$. It follows that $v(Q) = v(Q') = \{\mathbf{u}\}$ since all of their acceptance conditions are satisfiable. Due to conflict-freeness, $v(y) \neq \mathbf{f}$. We will now construct a conflict-free interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v <_i w$.

As above, $v(P) \subseteq \{\mathbf{t}, \mathbf{f}\}$. As Φ is true, there is an interpretation $v_Q : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that the formula $\psi^{v|_{P \cup v_Q}}$ (over R) is a tautology. Define w as follows:

$$\begin{aligned}
 w(p) &= v(p) && \text{for } p \in P \\
 w(q) &= \begin{cases} \mathbf{f} & \text{if } v_Q(q) = \mathbf{f} \\ \mathbf{u} & \text{otherwise} \end{cases} && \text{for } q \in Q \\
 w(q') &= \begin{cases} \mathbf{f} & \text{if } v_Q(q) = \mathbf{t} \\ \mathbf{u} & \text{otherwise} \end{cases} && \text{for } q \in Q \\
 w(r) &= \mathbf{u} && \text{for } r \in R \\
 w(y) &= \mathbf{t} \\
 w(z) &= \mathbf{f}
 \end{aligned}$$

Clearly $v <_i w$, it remains to show that w is conflict-free. This is clear for $p \in P$ and $q \in Q$ (respectively $q' \in Q'$) and trivial for $r \in R$. It is easy to see that $w(y) = \mathbf{t}$ is justified since for all $q \in Q$, by definition either $w(q) = \mathbf{u}$ or $w(q') = \mathbf{u}$, and thus φ_y^w is satisfiable. Finally, $\varphi_z^w = \neg\perp \wedge \neg\psi'^w \equiv \neg\psi'^w$ is unsatisfiable because $\psi^{v|P \circ v_Q}$ is a tautology.

only if: We show the contrapositive. Let Φ be false. We have to show that there is a naive interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(z) \neq \mathbf{f}$, that is, $v(z) = \mathbf{u}$. Since Φ is false, there is an interpretation $v_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that $\exists Q \forall R \psi^{v_P}$ is false, that is, $\forall Q \exists R \neg\psi^{v_P}$ is true. Then for any interpretation $v_Q : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$, the formula $\neg\psi^{v_P \circ v_Q}$ is satisfiable. Define the interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ as follows:

$$\begin{aligned}
 v(p) &= v_P(p) && \text{for } p \in P \\
 v(q) &= \mathbf{u} && \text{for } q \in Q \\
 v(q') &= \mathbf{u} && \text{for } q \in Q \\
 v(r) &= \mathbf{u} && \text{for } r \in R \\
 v(y) &= \mathbf{t} \\
 v(z) &= \mathbf{u}
 \end{aligned}$$

It is clear that v is conflict-free, it remains to show that v is naive. Assume that there is a conflict-free interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v <_i w$. Then there is some $s \in Q \cup Q' \cup R \cup \{z\}$ with $w(s) \neq \mathbf{u}$. From the definition of the acceptance formulas it follows that $w(s) = \mathbf{f}$.

- $s = q \in Q$: Then $\varphi_q^w = \neg\perp \wedge (\neg\top \vee z)^w$ is unsatisfiable, that is, $w(z) = \mathbf{f}$. It also follows that $w(q') = \mathbf{u}$ since $w(y) = \mathbf{t}$. Then $\varphi_z^w = \neg\perp \wedge \neg\psi'^w \equiv \neg\psi'^w$ is unsatisfiable.

Clearly $w|_Q$ can be interpreted as a (partial) two-valued interpretation of Q . From $v \leq_i w$ we also get that $v_P = v|_P = w|_P$. As concluded above, by assumption we have that for any interpretation $v_Q : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$, the formula $\neg\psi^{v_P \circ v_Q}$ is satisfiable. In particular, $\neg\psi'^{w|_P \circ w|_Q}$ is satisfiable, contradicting unsatisfiability of φ_z^w . The case $s = q' \in Q'$ follows by symmetry.

- $s = r \in R$: Then $\varphi_r^w = \neg\perp$ is unsatisfiable, contradiction.
- $s = z$: Then φ_z^w is unsatisfiable. Contradiction (see case $s = q$).

Thus there exists no such w and v with $v(z) = \mathbf{u}$ is naive. \square

For naive semantics, we have seen (1) an asymmetry in deciding (credulous/sceptical) truth and falsity; and (2) a steady rise in complexity from credulous truth up to sceptical falsity. For stage semantics, surprisingly, these differences vanish: All four decision problems are (more or less) equally hard, namely in the third level of the polynomial hierarchy. For the first problem, this is shown by considering QBFs $\exists P \forall Q \exists R \psi$.

Theorem 4.14. Cred_{stg}^t is Σ_3^P -complete.

Proof. in Σ_3^P : Given an ADF D over S and an $s \in S$, we can guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(s) = \mathbf{t}$ and verify in Π_2^P that it is stage (Theorem 4.5).

Σ_3^P -hard: Let $\Phi = \exists P \forall Q \exists R \psi$ be a QBF with ψ in CNF. Define an ADF D over $S = P \cup P' \cup Q \cup R \cup \{x, y, z\}$ such that:

$$\begin{aligned} \varphi_p &= \neg p \wedge \neg x && \text{for } p \in P \\ \varphi_{p'} &= \neg p' \wedge \neg x && \text{for } p \in P \\ \varphi_q &= q && \text{for } q \in Q \\ \varphi_r &= \neg r && \text{for } r \in R \\ \varphi_x &= \neg x \vee \bigwedge_{p \in P} (p \vee p') \\ \varphi_y &= \neg y \wedge \psi' \\ \varphi_z &= y \end{aligned}$$

We now show that there is a stage interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(z) = \mathbf{t}$ iff the QBF Φ is true.

if: Let $\exists P \forall Q \exists R \psi$ be true. Then there exists a $v_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that for all $v_Q : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$ the formula $\psi^{v_P \circ v_Q}$ is satisfiable. We construct an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ as follows:

$$\begin{aligned} v(p) &= \begin{cases} \mathbf{f} & \text{if } v_P(p) = \mathbf{f} \\ \mathbf{u} & \text{otherwise} \end{cases} && \text{for } p \in P \\ v(p') &= \begin{cases} \mathbf{f} & \text{if } v_P(p) = \mathbf{t} \\ \mathbf{u} & \text{otherwise} \end{cases} && \text{for } p \in P \\ v(q) &\in \{\mathbf{t}, \mathbf{f}\} \text{ arbitrary} && \text{for } q \in Q \\ v(r) &= \mathbf{u} && \text{for } r \in R \\ v(x) &= \mathbf{t} \\ v(y) &= \mathbf{u} \\ v(z) &= \mathbf{t} \end{aligned}$$

Clearly v is conflict-free for $P \cup P' \cup Q \cup R$; setting x to true is justified by definition of v for $P \cup P'$, setting z to true is justified since $v(y) = \mathbf{u}$. It remains to show that v is stage.

Assume to the contrary that v is not stage. Then there is another conflict-free interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $w_{\mathbf{u}} \subsetneq v_{\mathbf{u}}$. First of all, it follows that $w(x) = \mathbf{t}$ (x must be set but cannot be set to false), thus $w(p) = v(p)$ and $w(p') = v(p')$ for $p \in P$. It is also clear that $w(R) = \{\mathbf{u}\}$ and $w(Q) \subseteq \{\mathbf{t}, \mathbf{f}\}$, while not necessarily $w|_Q = v|_Q$. Consequently, $v_{\mathbf{u}} \setminus w_{\mathbf{u}} = \{y\}$. Since w is conflict-free, $w(y) = \mathbf{f}$ and $\varphi_y^w = \neg \perp \wedge \psi'^w \equiv \psi'^w$ is unsatisfiable. However, by the assumption that Φ is true and construction of v , we have $v_P = v|_P = w|_P$ and thus ψ'^w is satisfiable. Contradiction. Thus v is a stage interpretation.

only if: Let $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ be a stage interpretation with $v(z) = \mathbf{t}$. It follows that $v(y) \in \{\mathbf{t}, \mathbf{u}\}$, and since v is conflict-free, $v(y) = \mathbf{u}$. Thus $\varphi_y^v = \neg y \wedge \psi'^v$ is satisfiable, in particular ψ'^v is satisfiable. (Otherwise we could set y to false and v would not be stage.) We have to show that Φ is true. To do this, we can extract a $v_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ as usual from $v|_{P \cup P'}$. We then have to show that for all $v_Q : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$, the formula $\psi^{v_P \circ v_Q}$ is satisfiable.

Assume to the contrary that there is a $v_Q : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that the formula $\psi^{v_P \circ v_Q}$ is unsatisfiable. We can then construct a conflict-free interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $w_{\mathbf{u}} \subsetneq v_{\mathbf{u}}$ as follows:

$$\begin{aligned} w|_{P \cup P'} &= v|_{P \cup P'} \\ w|_Q &= v_Q \\ w(R) &= \{\mathbf{u}\} \\ w(x) &= \mathbf{t} \\ w(y) &= \mathbf{f} \\ w(z) &= \mathbf{f} \end{aligned}$$

The interpretation w is conflict-free, since $\varphi_y^w = \neg \perp \wedge \psi'^w \equiv \psi'^w$ is unsatisfiable by assumption. But then $w_{\mathbf{u}} \subsetneq v_{\mathbf{u}}$ implies that v is not stage. Contradiction. Thus $v_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ is such that for each $v_Q : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$, the formula $\psi^{v_P \circ v_Q}$ is satisfiable. Consequently the QBF Φ is true. \square

For hardness of deciding credulous falsity, we can use a simple extension of the hardness construction used above: basically, the construction relies on a statement y that can be set to \mathbf{u} if the given QBF $\exists P \forall Q \exists R \psi$ is true, and must be set to \mathbf{f} otherwise (due to the inherent \subseteq -minimisation of $v_{\mathbf{u}}$ in stage semantics). The actual reduction now works over a statement z with acceptance formula $\varphi_z = y$; consequently, z can be set to true iff y can be set to \mathbf{u} . In the extended construction below, we now add another statement a with acceptance formula $\varphi_a = \neg z$. Both statements will always be assigned opposite truth values from $\{\mathbf{t}, \mathbf{f}\}$, thus proving the next result.

Proposition 4.15. $\text{Cred}_{stg}^{\mathbf{f}}$ is Σ_3^{P} -complete.

Proof. in Σ_3^{P} : Given an ADF D over S and an $s \in S$, we can guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(s) = \mathbf{f}$ and verify in Π_2^{P} that it is stage (Theorem 4.5).

Σ_3^P -hard: Let $\Phi = \exists P \forall Q \exists R \psi$ be a QBF with ψ in CNF. We extend the construction of Theorem 4.14 with an additional statement $a \in S$ such that $\varphi_a = \neg z$. Now for each stage interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ of D , we have that $v(a) = \mathbf{f}$ iff $v(z) = \mathbf{t}$. \square

To show that a statement s is *not* sceptically entailed as false in an ADF D , we guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(s) \neq \mathbf{f}$ and verify in Π_2^P that v is stage. This approach is optimal, as completeness shows.

Theorem 4.16. $\text{Scep}_{stg}^{\mathbf{f}}$ is Π_3^P -complete.

Proof. in Π_3^P : Let D be an ADF over S and $s \in S$. To show that s is *not* sceptically false, we guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(s) \neq \mathbf{f}$ and verify in Π_2^P that v is stage.

Π_3^P -hard: Let $\Phi = \forall P \exists Q \forall R \psi$ be a QBF with ψ in DNF. We construct an ADF D over the vocabulary $S = P \cup Q \cup Q' \cup R \cup \{x, y, z\}$ with $Q' = \{q' \mid q \in Q\}$ such that z is false in every stage interpretation of D if and only if Φ is true. (We slightly modify the construction of Theorem 4.13.) Define D as follows:

$$\begin{aligned}
 \varphi_p &= \neg p \wedge \neg x && \text{for } p \in P \\
 \varphi_{p'} &= \neg p' \wedge \neg x && \text{for } p \in P \\
 \varphi_q &= \neg q \wedge (\neg y \vee z) && \text{for } q \in Q \\
 \varphi_{q'} &= \neg q' \wedge (\neg y \vee z) && \text{for } q \in Q \\
 \varphi_r &= \neg r && \text{for } r \in R \\
 \varphi_x &= \neg x \vee \bigwedge_{p \in P} (p \vee p') \\
 \varphi_y &= \neg y \vee \bigwedge_{q \in Q} (q \vee q') \\
 \varphi_z &= \neg z
 \end{aligned}$$

Formula ψ' is obtained from ψ by replacing all positive occurrences of $p \in P$ by $\neg p'$ and all positive occurrences of $q \in Q$ by $\neg q'$. We start out with some observations and intuitive explanations:

1. x cannot be set to false and only be set to true if for each $p \in P$, at least one of p or p' is undefined; it thereby guarantees that the p, p' encode a two-valued interpretation of P for each stage interpretation with x true;
2. likewise, y cannot be set to false and only be set to true if for each $q \in Q$, at least one of q or q' is undefined; it thereby guarantees that the q, q' encode a two-valued interpretation of Q for each stage interpretation with y true and z false;
3. the p and p' cannot be set to true, and only to false if x is true; in particular, every stage interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ will set either $v(p) = \mathbf{f}$ or $v(p') = \mathbf{f}$ for $p \in P$, since v is conflict-free and $v_{\mathbf{u}}$ is \subseteq -minimal.

4. The q and q' cannot be set to true, and only to false if z is false and y is true.
5. z is set to false if the corresponding partial evaluation of $\neg\psi'$ is unsatisfiable, otherwise z is set to true.

We now show that z is false in every stage interpretation of D if and only if Φ is true.

if: Let Φ be true and consider the stage interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$. We have to show that $v(z) = \mathbf{f}$. Assume to the contrary that $v(z) \neq \mathbf{f}$, that is, $v(z) \in \{\mathbf{t}, \mathbf{u}\}$. It follows that $v(Q \cup Q') = \{\mathbf{u}\}$ since all of their acceptance conditions are satisfiable. Due to conflict-freeness, $v(x), v(y) \in \{\mathbf{t}, \mathbf{u}\}$. We will now construct a conflict-free interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $w_{\mathbf{u}} \subsetneq v_{\mathbf{u}}$.

As usual, we can extract an interpretation $v_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ from $v(P)$. Since Φ is true, there exists an interpretation $v_Q : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that the formula $\psi^{v_P \circ v_Q}$ (over R) is a tautology. Define w such that:

$$\begin{aligned}
 w(p) &= v(p) && \text{for } p \in P \\
 w(p') &= v(p') && \text{for } p \in P \\
 w(q) &= \begin{cases} \mathbf{f} & \text{if } v_Q(q) = \mathbf{f} \\ \mathbf{u} & \text{otherwise} \end{cases} && \text{for } q \in Q \\
 w(q') &= \begin{cases} \mathbf{f} & \text{if } v_Q(q) = \mathbf{t} \\ \mathbf{u} & \text{otherwise} \end{cases} && \text{for } q \in Q \\
 w(r) &= \mathbf{u} && \text{for } r \in R \\
 w(x) &= \mathbf{t} \\
 w(y) &= \mathbf{t} \\
 w(z) &= \mathbf{f}
 \end{aligned}$$

Clearly $w_{\mathbf{u}} \subsetneq v_{\mathbf{u}}$ since $w(Q \cup Q') \neq \{\mathbf{u}\}$; it remains to show that w is conflict-free. This is clear for $p, p' \in P \cup P'$ and $q, q' \in Q \cup Q'$ and trivial for $r \in R$. It is easy to see that $w(x) = \mathbf{t}$ is justified since for all $p \in P$, by definition either $w(p) = \mathbf{u}$ or $w(p') = \mathbf{u}$, and thus φ_x^w is satisfiable; the same holds for $w(y) = \mathbf{t}$ and $q \in Q$. Finally, $\varphi_z^w = \neg\psi'^w$ is unsatisfiable because $\psi^{v_P \circ v_Q}$ is a tautology.

only if: We show the contrapositive. Let Φ be false. We show that there is a stage interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(z) = \mathbf{t}$. Since Φ is false, there is an interpretation $v_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that $\exists Q \forall R \psi^{v_P}$ is false, that is, $\forall Q \exists R \neg\psi^{v_P}$ is true. Then for any interpretation $v_Q : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$, the formula $\neg\psi^{v_P \circ v_Q}$ is satisfiable.

Define the interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ as follows:

$$\begin{aligned}
 v(p) &= \begin{cases} \mathbf{f} & \text{if } v_P(p) = \mathbf{f} \\ \mathbf{u} & \text{otherwise} \end{cases} && \text{for } p \in P \\
 v(p') &= \begin{cases} \mathbf{f} & \text{if } v_P(p) = \mathbf{t} \\ \mathbf{u} & \text{otherwise} \end{cases} && \text{for } p \in P \\
 v(q) &= \mathbf{u} && \text{for } q \in Q \\
 v(q') &= \mathbf{u} && \text{for } q \in Q \\
 v(r) &= \mathbf{u} && \text{for } r \in R \\
 v(x) &= \mathbf{t} \\
 v(y) &= \mathbf{t} \\
 v(z) &= \mathbf{t}
 \end{aligned}$$

It is clear that v is conflict-free for $P \cup P' \cup Q \cup Q' \cup R$. For x and y it follows by definition, and finally $v(z) = \mathbf{t}$ is justified since $\neg\psi^{v|_{P \cup P' \cup Q}} = \neg\psi^{v|_{P \cup P'}}$ is satisfiable by the argument above. It remains to show that v is stage. Assume that there is a conflict-free interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $w_{\mathbf{u}} \subsetneq v_{\mathbf{u}}$. Then there is some $s \in P \cup P' \cup Q \cup Q' \cup R$ with $s \in v_{\mathbf{u}} \setminus w_{\mathbf{u}}$. From the definition of the acceptance formulas it follows that $w(s) = \mathbf{f}$.

- $s = p \in P$: From $v(p) = \mathbf{u}$ we get $v(p') = \mathbf{f}$ by definition. By $w_{\mathbf{u}} \subseteq v_{\mathbf{u}}$ we get $w(p') = \mathbf{f}$. Along with $w(p) = \mathbf{f}$ this means that φ_x^w is unsatisfiable, in contradiction to $w(x) = \mathbf{t}$ and w being conflict-free.
- $s = p' \in P'$: Symmetric.
- $s = q \in Q$: Then $\varphi_q^w = \neg\perp \wedge (\neg\top \vee z)^w$ is unsatisfiable, that is, $w(z) = \mathbf{f}$.
- $s = q' \in Q'$: Symmetric.
- $s = r \in R$: Then $\varphi_r^w = \neg\perp$ is unsatisfiable, contradiction.

In any case, $w(z) = \mathbf{f}$. Then $\varphi_z^w = \neg\psi^{w|_Q}$ is unsatisfiable. Clearly $w|_Q$ can be interpreted as a (partial) two-valued interpretation of Q . From $w_{\mathbf{u}} \subseteq v_{\mathbf{u}}$ and w being conflict-free we furthermore get that $v|_{P \cup P'} = w|_{P \cup P'}$. So again we can extract an interpretation $w_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ from $w|_{P \cup P'}$, namely $w_P = v_P$. As concluded above, by assumption we have that for any interpretation $w_Q : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$, the formula $\neg\psi^{v_P \circ w_Q} = \neg\psi^{w_P \circ w_Q}$ is satisfiable. In particular, $\neg\psi^{w|_{P \cup P'} \circ w|_{Q \cup Q'}}$ is satisfiable, in contradiction to $\varphi_z^w = \neg\psi^{w|_Q}$ being unsatisfiable. Thus there exists no such w and v with $v(z) = \mathbf{t}$ is stage. \square

In the step from $\text{Scep}_{stg}^{\mathbf{f}}$ to $\text{Scep}_{stg}^{\mathbf{t}}$ we can use the same construction extension as in the step from $\text{Cred}_{stg}^{\mathbf{t}}$ to $\text{Cred}_{stg}^{\mathbf{f}}$.

Proposition 4.17. $\text{Scep}_{stg}^{\mathbf{t}}$ is Π_3^{P} -complete.

Proof. in Π_3^{P} : Let D be an ADF over S and $s \in S$. To show that s is *not* sceptically true, we guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(s) \neq \mathbf{t}$ and verify in Π_2^{P} that v is stage.

Π_3^P -hard: Let $\Phi = \forall P \exists Q \forall R \psi$ be a QBF with ψ in DNF. We adapt the construction of ADF D from Theorem 4.16 by adding a statement $a \in S$ and setting $\varphi_a = \neg z$. As in the proof of Theorem 4.16, Φ is true iff $v(z) = \mathbf{f}$ for all stage interpretations $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$. Furthermore, in each stage interpretation v , we have $v(z), v(a) \in \{\mathbf{t}, \mathbf{f}\}$ and $v(z) = \mathbf{f}$ iff $v(a) = \mathbf{t}$. Consequently, the QBF Φ is true iff $v(a) = \mathbf{t}$ in each stage interpretation v of D . \square

For the *nai*₂ and *stg*₂ semantics, we can directly use that the relevant entailment decision problems (or their complements, respectively) are polynomially interreducible.

Proposition 4.18. *Let $\sigma \in \{\text{nai}_2, \text{stg}_2\}$. The following problems can be polynomially reduced to each other:*

- $\text{Cred}_\sigma^{\mathbf{t}}$ and $\text{Cred}_\sigma^{\mathbf{f}}$,
- $\text{Scep}_\sigma^{\mathbf{t}}$ and $\text{Scep}_\sigma^{\mathbf{f}}$,
- $\text{co-Scep}_\sigma^{\mathbf{f}}$ and $\text{Cred}_\sigma^{\mathbf{t}}$.

Proof. Let D be an ADF over S and $s \in S$. Let s_1, s_2, s_3 be fresh argument names not occurring in S and let ψ be a propositional formula over $\{s_1, s_2\}$. Let now $D[s, \psi]$ be the ADF obtained from D by adding the arguments s_1, s_2, s_3 and setting $\varphi_{s_1} = s$, $\varphi_{s_2} = \neg s$ and $\varphi_{s_3} = \psi$. Then we obtain that $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ is a σ interpretation of D iff v' is a σ interpretation of $D[s, \psi]$, where v' coincides with v on S , and

- $v'(s_1) = \mathbf{t}$ and $v'(s_2) = \mathbf{f}$ in case $v(s) = \mathbf{t}$,
- $v'(s_1) = \mathbf{f}$ and $v'(s_2) = \mathbf{t}$ in case $v(s) = \mathbf{f}$,
- $v'(s_1) = \mathbf{t}$ and $v'(s_2) = \mathbf{t}$ in case $v(s) = \mathbf{u}$,

and $v'(s_3)$ is obtained as by evaluating ψ for $s_1 \mapsto v'(s_1)$ and $s_2 \mapsto v'(s_2)$. That this is the case follows from the definition of σ taking into account that s_1 and s_2 depend on s and that s_3 depends on s_1 and s_2 .

This observation can now be leveraged for establishing the above results:

- s is credulously (sceptically) true in D iff s_3 is credulously (sceptically) false in $D[s, \neg s_1 \vee s_2]$;
- s is credulously (sceptically) false in D iff s_3 is credulously (sceptically) true in $D[s, \neg s_1 \wedge s_2]$.
- The following are equivalent:
 - s is not sceptically false in D
 - not every σ interpretation v of D satisfies $v(s) = \mathbf{f}$
 - there is some σ interpretation v of D with $v(s) = \mathbf{t}$ or $v(s) = \mathbf{u}$
 - there is some σ interpretation v' of $D[s, s_1]$ with $v'(s_3) = \mathbf{t}$
 - s_3 is credulously true in $D[s, s_1]$.

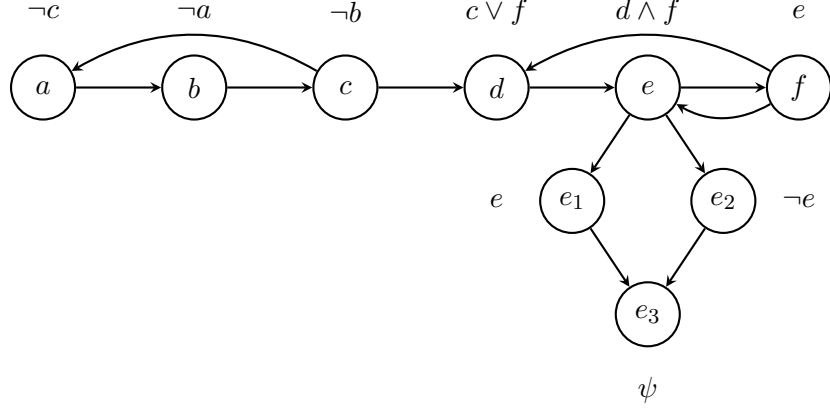


Figure 4: The ADF $D[e, \psi]$ from the proof of Proposition 4.18 applied to ADF D from Example 3.6. The important aspect of the construction is the “diamond” subgraph below (and including) statement e : We can see that whenever there is a σ -interpretation (where $\sigma \in \{nai_2, stg_2\}$) setting e to \mathbf{t} , then immediately e_1 must be set to \mathbf{t} and e_2 must be set to \mathbf{f} . For $\psi = \neg e_1 \vee e_2$, this leads to the reduced acceptance formula $\neg \top \vee \perp$, which is unsatisfiable, whence only \mathbf{f} can be assigned to e_3 . Similarly, whenever e can be set to \mathbf{f} and we have $\psi = \neg e_1 \wedge e_2$, then the reduced $\neg \perp \wedge \top$ is irrefutable and e_3 must be set to \mathbf{t} . For any interpretation with e set to \mathbf{u} , the connective used in ψ makes the difference. Both the reduced versions of φ_{e_1} and φ_{e_2} are satisfiable and so could be set to \mathbf{t} or \mathbf{u} . However, the only way to justify setting e_3 to \mathbf{f} is to have the reduced version of $\neg e_1 \vee e_2$ to be unsatisfiable, which is only the case for $e_1 \mapsto \mathbf{t}$ and $e_2 \mapsto \mathbf{f}$; the only way to justify setting e_3 to \mathbf{t} is to have the reduced version of $\neg e_1 \wedge e_2$ to be satisfiable, which is only the case for $e_1 \mapsto \mathbf{f}$ and $e_2 \mapsto \mathbf{t}$. For $\psi = e_1 \wedge \neg e_2$ as in the third item of Proposition 4.18, whenever e is set to \mathbf{t} , the reduced version of φ_{e_3} is still satisfiable and \mathbf{f} is ruled out as possible truth value for e_3 .

• The following are equivalent:

- s is credulously true in D
- there is some σ interpretation v of D with $v(s) = \mathbf{t}$
- not every σ interpretation v of D satisfies $v(s) = \mathbf{f}$ or $v(s) = \mathbf{u}$
- not every σ interpretation v' of $D[s, s_1 \wedge \neg s_2]$ satisfies $v'(s_3) = \mathbf{f}$
- s_3 is not sceptically false in $D[s, s_1 \wedge \neg s_2]$. □

A pictorial illustration of $D[e, \psi]$ obtained from the ADF D from Example 3.6 and the statement $e \in S$ is in Figure 4.

Together with the observation that the hardness reduction of Theorem 4.13 (for $\text{Scep}_{nai}^{\mathbf{f}}$) works for the nai_2 and stg_2 semantics as well, the proposition leads to the following results.

σ	<i>cfi</i>	<i>nai</i>	<i>stg</i>	<i>nai₂</i>	<i>stg₂</i>
Ver_σ	DP-c (Proposition 4.2)	Π_2^P -c (Theorem 4.4)	Π_2^P -c (Theorem 4.5)	Π_2^P -c (Theorem 4.6)	Π_2^P -c (Theorem 4.6)
Exists_σ	coDP-c (Theorem 4.7)	coDP-c (Proposition 4.9)	coDP-c (Proposition 4.9)	coDP-c (Proposition 4.9)	coDP-c (Proposition 4.9)
Cred_σ^t	NP-c (Theorem 4.10)	NP-c (Proposition 4.9)	Σ_3^P -c (Theorem 4.14)	Σ_3^P -c (Theorem 4.19)	Σ_3^P -c (Theorem 4.19)
Cred_σ^f	Σ_2^P -c (Theorem 4.11)	Σ_2^P -c (Proposition 4.9)	Σ_3^P -c (Proposition 4.15)	Σ_3^P -c (Theorem 4.19)	Σ_3^P -c (Theorem 4.19)
Scep_σ^t	trivial	Π_2^P -c (Theorem 4.12)	Π_3^P -c (Proposition 4.17)	Π_3^P -c (Theorem 4.19)	Π_3^P -c (Theorem 4.19)
Scep_σ^f	trivial	Π_3^P -c (Theorem 4.13)	Π_3^P -c (Theorem 4.16)	Π_3^P -c (Theorem 4.19)	Π_3^P -c (Theorem 4.19)

Table 1: Complexity results for naive-based semantics of abstract dialectical frameworks; \mathcal{C} -c stands for \mathcal{C} -complete.

Theorem 4.19. Let $\sigma \in \{nai_2, stg_2\}$.

- Cred_σ^t and Cred_σ^f are Σ_3^P -complete.
- Scep_σ^t and Scep_σ^f are Π_3^P -complete.

Proof. in Σ_3^P / Π_3^P : Let D be an ADF over S and $s \in S$. For Cred_σ^t and Cred_σ^f , we can guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(s) = \mathbf{t}$ (or $v(s) = \mathbf{f}$, respectively) and verify in Π_2^P that it is σ (Theorem 4.6). For Scep_σ^t and Scep_σ^f , to show that s is *not* sceptically true, we guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(s) \neq \mathbf{t}$ (or $v(s) \neq \mathbf{f}$, respectively) and verify in Π_2^P that v is σ (Theorem 4.6).

Σ_3^P / Π_3^P -hardness: We first note that the reduction from QBF of shape $\forall P \exists Q \forall R \psi$ presented in Theorem 4.13 works for σ as well, thus establishing Π_3^P -hardness of Scep_σ^f . Then, the second part of Theorem 4.18 implies Π_3^P -hardness of Scep_σ^t and the third part of the same theorem yields Σ_3^P -hardness of Cred_σ^t . Consequently, applying the first part of Theorem 4.18 we obtain Σ_3^P -hardness also of Cred_σ^f . \square

4.2 Complexity of Bipolar ADFs

In line with previous complexity results of general and bipolar ADFs, also in our case the decision problems for bipolar ADFs are potentially easier. The reason is that while for general ADFs we have to solve one satisfiability and one irrefutability check and thus get DP-hardness, for bipolar ADFs the respective checks are easier because we can use the additional information about link polarities.

Here is how it works: Suppose we want to check if a statement $s \in S$ of a bipolar ADF $D = (S, L^+ \cup L^-, C)$ has been correctly set to true in an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$. As usual, we consult the partially evaluated acceptance formula φ_s^v . Since D is bipolar, we have the additional information that all parents of s are supporting or attacking, and we even know which. To find out whether φ_s^v is satisfiable, we create a canonical two-valued interpretation $v^+ : S \rightarrow \{\mathbf{t}, \mathbf{f}\}$, the best possible scenario for the statements that are set to \mathbf{u} in v :

$$v^+(r) = \begin{cases} v(r) & \text{if } v(r) \in \{\mathbf{t}, \mathbf{f}\} \\ \mathbf{t} & \text{if } v(r) = \mathbf{u} \text{ and } (r, s) \in L^+ \\ \mathbf{f} & \text{otherwise} \end{cases}$$

That is, we set all supporters of s to true and all attackers of s to false. If v^+ is no model for φ_s^v , then the formula is unsatisfiable. (Assume to the contrary that φ_s^v had a model, then changing the truth value of supporters to true and attackers to false would preserve the model property by the definition of bipolarity, contradiction.) Note that if v is two-valued, then $v^+ = v$ and we just have to check that φ_s^v evaluates to \mathbf{t} . A symmetric construction (setting supporters to false and attackers to true) is possible when checking for refutability. Thus in any case we only have to evaluate a given formula with a two-valued interpretation that can basically be read off the given pair of ADF and three-valued interpretation.

While membership of this decision problem in \mathbf{P} formally follows from earlier work by Strass and Wallner (2015, proof of Proposition 5.1), we can even improve that upper bound to membership in \mathbf{L} , that is, we can show that the problem can be decided using only logarithmic space. For technical reasons, we apply the notion of bipolarity to propositional formulas, where a formula is bipolar iff in its associated characteristic Boolean function, each atom is supporting or attacking. We assume that polarities are represented along with the formula, for example by prepending a list of tuples of the form $(p, \{+\})$ that indicate if an atom p is supporting, attacking or both.¹¹

Lemma 4.20. *Let φ be a bipolar propositional formula over P and $v : P \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ an interpretation. Deciding satisfiability of φ^v can be done in \mathbf{L} .*

Proof. Evaluating the truth value of a propositional formula can be done in \mathbf{L} (Papadimitriou, 1995, Theorem 16.1). The only additional work we have to do is reconstructing the truth value $v^+(p)$ of any $p \in P$ with $v(p) = \mathbf{u}$ we encounter during evaluation. However, this is easy since we have the list of polarities given. \square

As we shall see, this in essence yields a decrease in complexity by one level of the polynomial hierarchy, across the board.

4.2.1 INTERPRETATION VERIFICATION

As a first result, Lemma 4.20 more or less directly leads to verification of conflict-free interpretations being decidable in polynomial time. In contrast to the general case, where

11. There is also a representation result for bipolar Boolean functions (Strass, 2015, Theorem 1). Roughly, atom polarities can be represented implicitly in the formula, by letting each atom occur only in at most one polarity, that is, only after an even or after an odd number of negations, where these negations are counted from the root of the formula tree.

the problem is DP-hard, here we only have to solve the “easy”, bipolar versions of the associated (un)satisfiability problems.

Proposition 4.21. Ver_{cf}^\pm is in L.

Proof. Let D be a bipolar ADF over S and $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ be an interpretation. By Lemma 4.20, for each $s \in S$ with $v(s) = \mathbf{t}$ we can decide in L whether φ_s^v is satisfiable, and for each $s \in S$ with $v(s) = \mathbf{f}$ we can decide in L whether φ_s^v is unsatisfiable. There are at most $|S|$ such checks, so we only need one additional counter that keeps track of which statement’s truth value we are currently verifying. \square

In accordance with the general case, verifying naive interpretations is one level harder than verifying conflict-free interpretations. Interestingly, the problem is still harder than for AFs (where it can be decided in logarithmic space). The hardness reduction below will serve as a blueprint for further reductions. The main difference to the previously developed reduction techniques is that now we have to construct bipolar ADFs. In the case below, we achieve this by duplicating the vocabulary P of a given formula ψ and in the formula replace all negative occurrences of any $p \in P$ by a new atom p' encoding that p is false. The resulting formula is strictly positive and therefore bipolar. Clearly, we have to make sure that all pairs p and p' are only interpreted in a way that corresponds to a two-valued interpretation of P on the original formula ψ .

Proposition 4.22. Ver_{nai}^\pm is coNP-complete.

Proof. in coNP: Let D be an ADF over S and $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ be an interpretation. To verify that v is *not* naive, we can either verify in P that v is not conflict-free, or guess an interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v <_i w$ and verify in P that w is conflict-free (Proposition 4.21).

coNP-hard: We reduce from tautology checking. Let $\psi = \psi_1 \vee \dots \vee \psi_n$ be a propositional formula in disjunctive normal form over vocabulary P . We duplicate this vocabulary into $P' = \{p' \mid p \in P\}$ and replace all negative occurrences of p by p' :

$$\psi' = \psi[\neg p/p' : p \in P]$$

Note that ψ' is a strictly positive formula, that is, contains no negation. Now define the ADF $D_\psi = (S, L, C)$ over $S = P \cup P' \cup \{x, y, z\}$ given by

$$\begin{aligned} \varphi_p &= \neg p \wedge (\neg x \vee y \vee z) && \text{for } p \in P \\ \varphi_{p'} &= \neg p' \wedge (\neg x \vee y \vee z) && \text{for } p \in P \\ \varphi_x &= \neg x \vee \bigwedge_{p \in P} (p \vee p') \\ \varphi_y &= \neg y \wedge \bigvee_{p \in P} (p \wedge p') \\ \varphi_z &= \neg z \wedge \psi' \end{aligned}$$

Observe that the defined ADF is bipolar. Furthermore, define an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ by $v(x) = \mathbf{t}$, $v(y) = \mathbf{f}$ and $v(s) = \mathbf{u}$ for all $s \in S \setminus \{x, y\}$. To show

that v is naive for D_ψ iff ψ is a tautology, we now show that v is not naive for D_ψ iff ψ is refutable.

if: Let ψ be refutable. Then there is a valuation $v_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that $v_P(\psi) = \mathbf{f}$.

Define an interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}\}$ as follows:

$$\begin{aligned} w(p) &= \begin{cases} \mathbf{f} & v_P(p) = \mathbf{f} \\ \mathbf{u} & \text{otherwise} \end{cases} && \text{for } p \in P \\ w(p') &= \begin{cases} \mathbf{f} & v_P(p) = \mathbf{t} \\ \mathbf{u} & \text{otherwise} \end{cases} && \text{for } p \in P \\ w(x) &= \mathbf{t} \\ w(y) &= \mathbf{f} \\ w(z) &= \mathbf{f} \end{aligned}$$

Clearly $v <_i w$, it remains to show that w is conflict-free for D_ψ .

For each $p \in P$, either (a) $w(p) = \mathbf{u}$ and $w(p') = \mathbf{f}$, or (b) $w(p) = \mathbf{f}$ and $w(p') = \mathbf{u}$; thus φ_x^w is satisfiable (so $w(x) = \mathbf{t}$ is justified) and φ_y^w is unsatisfiable (so $w(y) = \mathbf{f}$ is justified). For each $p \in P$, the formulas φ_p^w and $\varphi_{p'}^w$ are unsatisfiable because $w(x) = \mathbf{t}$, $w(y) = \mathbf{f}$ and $w(z) = \mathbf{f}$. Finally, $\varphi_z^w = \neg\perp \wedge \psi'^w \equiv \psi'^w$, so it remains to show that ψ'^w is unsatisfiable. We know that $v_P(\psi) = \mathbf{f}$, and since $\psi = \psi_1 \vee \dots \vee \psi_n$ this means that for all $1 \leq i \leq n$, we have $v_P(\psi_i) = \mathbf{f}$. Let $i \in \{1, \dots, n\}$ be arbitrary; we show that $\psi_i'^w \equiv \perp$. Monomial ψ_i is a conjunction of literals from P , so $v_P(\psi_i) = \mathbf{f}$ can have only two reasons:

- ψ_i contains an atom $p \in P$ with $v_P(p) = \mathbf{f}$. Then by definition $w(p) = \mathbf{f}$ and ψ_i' contains the same atom, whence $\psi_i'^w \equiv \perp$.
- ψ_i contains a negative literal $\neg p$ for a $p \in P$ with $v_P(p) = \mathbf{t}$. Then by definition ψ_i' contains the atom p' , furthermore $w(p') = \mathbf{f}$ and thus $\psi_i'^w \equiv \perp$.

Hence, w is conflict-free for D_ψ with $v <_i w$, thus v is not naive for D_ψ .

only if: Let $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v <_i w$ be conflict-free for D_ψ . Since $v <_i w$, there is some $s \in S \setminus \{x, y\}$ with $w(s) \neq \mathbf{u}$.

- $s \in P \cup P'$: Then $w(s) = \mathbf{f}$ by definition of φ_s ; furthermore φ_s^w is unsatisfiable since w is conflict-free, thus $w(z) = \mathbf{f}$.
- $s = z$: Then $w(z) = \mathbf{f}$ by definition of φ_z .

In any case, $w(z) = \mathbf{f}$ whence $\varphi_z^w = \neg\perp \wedge \psi'^w \equiv \psi'^w$ is unsatisfiable. By definition, $s \in P \cup P'$ implies $v(s) \in \{\mathbf{f}, \mathbf{u}\}$. Since φ_y^v is unsatisfiable, for each $p \in P$ we have $v(p) = \mathbf{f}$ or $v(p') = \mathbf{f}$. Since φ_x^v is satisfiable, for each $p \in P$ we have $v(p) \neq v(p')$. In combination, for each $p \in P$, either $w(p) = \mathbf{f}$ or $w(p') = \mathbf{f}$ (but not both or neither). This lets us construct a valuation $v_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ by

$$v_P(p) = \begin{cases} \mathbf{t} & \text{if } w(p') = \mathbf{f} \\ \mathbf{f} & \text{otherwise} \end{cases}$$

We now show that $v_P(\psi) = \mathbf{f}$ by showing that for all $1 \leq i \leq n$, we have $v_P(\psi_i) = \mathbf{f}$. Since ψ'^w is unsatisfiable and ψ' is a strictly positive formula, each ψ'_i for $1 \leq i \leq n$ must be unsatisfiable already. Let $i \in \{1, \dots, n\}$. By the above reasoning, ψ'_i contains an atom $s \in P \cup P'$ that is replaced by \mathbf{f} in $\psi'_i{}^w$, that is, $w(s) = \mathbf{f}$.

- If $s = p \in P$, then p occurs positively in ψ_i and $v_P(p) = \mathbf{f}$ by definition, whence $v_P(\psi_i) = \mathbf{f}$.
- If $s = p' \in P'$, then p occurs negatively in ψ_i and $v_P(p) = \mathbf{t}$ by definition, whence $v_P(\psi_i) = \mathbf{f}$.

Altogether it follows that $v_P(\psi) = \mathbf{f}$, whence ψ is refutable. \square

For the stage semantics, we can borrow membership from naive semantics, and hardness from Dung's argumentation frameworks.

Proposition 4.23. Ver_{stg}^\pm is coNP-complete.

Proof. Containment works as for naive semantics, hardness carries over from AFs (Dvořák & Woltran, 2011). \square

Concerning nai_2 semantics, we can reuse existing reductions in both directions. For membership in the general case, we reduced Ver_{nai_2} to Ver_{nai} (Theorem 4.6). That worked by simply duplicating the ADF into several renamed copies. Since this clearly preserves bipolarity, the same reduction works in the bipolar case. For hardness, we observe that the hardness reduction given for naive semantics above constructs BADFs with only one strongly connected component. In that case, nai and nai_2 semantics coincide by definition, whence verifying nai_2 interpretations for those BADFs is as hard as verifying nai interpretations.

Proposition 4.24. $\text{Ver}_{nai_2}^\pm$ is coNP-complete.

Proof. in coNP: It suffices to observe that the reduction from Ver_{nai_2} to Ver_{nai} given in Theorem 4.6 preserves bipolarity.

coNP-hard: The hardness reduction given in Proposition 4.22 constructs a BADF D_ψ with $\text{ind}_{D_\psi}(\emptyset) = S$. That is, there is only one strongly connected component and therefore $nai_2(D_\psi) = nai(D_\psi)$ by definition. Thus the same hardness reduction also works for nai_2 semantics: for the interpretation v constructed there, we have $v \in nai_2(D_\psi)$ iff $v \in nai(D_\psi)$ iff ψ is a tautology. \square

For stg_2 , we again can conclude hardness from existing results about AFs, membership works as for nai_2 semantics.

Proposition 4.25. $\text{Ver}_{stg_2}^\pm$ is coNP-complete.

Proof. For membership, again the reduction from Ver_{stg_2} to Ver_{stg} given in Theorem 4.6 preserves bipolarity. Hardness carries over from AFs (Dvořák & Gaggl, 2016). \square

4.2.2 EXISTENCE OF NON-TRIVIAL INTERPRETATIONS

For general ADFs, deciding the existence of non-trivial conflict-free interpretations turned out to have one of the most complicated membership proofs (Theorem 4.7). For the special case of bipolar ADFs, we can reuse the main idea of that proof; with the distinction that deciding the bipolar (un)satisfiability problems can be done in polynomial time.

Proposition 4.26. *Exists $_{cfi}^{\pm}$ is P-complete.*

Proof. in P: Let D be a bipolar ADF over S . In accordance with the developments in the proof of Theorem 4.7, there are only two types of non-trivial conflict-free interpretations:

1. Those interpretations $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v^{-1}(\mathbf{t}) \neq \emptyset$. To check whether there is such a $v \in cfi(D)$, for each $s \in S$ we test whether the formula $\varphi_s^{\{\mathbf{s} \rightarrow \mathbf{t}\}}$ is satisfiable. This can be done in logarithmic space (and thus polynomial time) for each single $s \in S$ by Lemma 4.20, and there are $|S|$ such checks. Otherwise, no $s \in S$ can be set to true without violating conflict-freeness.
2. Those interpretations that set all statements to undefined or false, and at least one of them to false, that is, those $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(S) \subseteq \{\mathbf{u}, \mathbf{f}\}$ and $v^{-1}(\mathbf{f}) \neq \emptyset$. To decide whether there is such a $v \in cfi(D)$, we can use the procedure from the proof of Theorem 4.7, where we computed the unique \leq_i -greatest conflict-free \mathbf{uf} -interpretation v_{\max} using an NP oracle. In this case we do not need oracle calls, but can test satisfiability in polytime using Lemma 4.20.

P-hard: We present a logspace reduction from Horn satisfiability. Let T be a set of Horn clauses over vocabulary P , that is, a set of clauses with at most one positive literal each. We denote $T = T^+ \cup T^-$ where T^+ is the set of all definite Horn clauses of T (those with exactly one positive literal) and T^- is the set of all strictly negative clauses of T . We define a new, definite Horn theory T' over $P' = P \cup \{q\}$ with $q \notin P$ as follows:

$$T' = T^+ \cup \{\psi \vee p \mid \psi \in T^-, p \in P'\}$$

We denote by $Th(T') = \{p \in P' \mid T' \models p\}$ the *theory* of T' . From our construction it follows that T is unsatisfiable iff $Th(T') = P'$: The only way to conclude all atoms from P' , in particular q , is the falsity of one $\psi \in T^-$, since q does not even occur in T .

Now we use T' to define an ADF $D_{T'}$ over P' given by

$$\varphi_p = \neg p \wedge \bigvee_{\psi \vee p \in T'} \neg \psi \quad \text{for } p \in P'$$

It is clear that $D_{T'}$ is bipolar since each $p \in P'$ attacks itself and all other statements that occur in φ_p are supporting. Also note that no conflict-free interpretation for $D_{T'}$ can set any statement to true, that is, $v \in cfi(D_{T'})$ implies $v(S) \subseteq \{\mathbf{f}, \mathbf{u}\}$.

The construction can be computed in logarithmic space as follows. We assume to use a Turing machine with one read-only input tape containing P and T in the form of

numbered atoms and clauses, one working tape with logarithmic space restriction and one write-only output tape for producing $D_{T'}$. On the working tape, we need one counter i that expresses that we are currently constructing the acceptance formula for $p_i \in P'$; another counter j that expresses that we currently look at clause $\psi_j \in T$. For each $p_i \in P$, we go through all clauses $\psi_j \in T$ and check whether ψ_j is of the form $\psi \vee p_i$ or ψ_j is purely negative. If so, we copy the negation of ψ (or ψ_j in the purely negative case) to the initially empty disjunction in φ_{p_i} on the output tape. After all original $p_i \in P$ have been considered, we construct φ_q likewise, clearly using all and only purely negative clauses.

To finalise the proof by showing that $D_{T'}$ has a non-trivial conflict-free interpretation iff T is satisfiable, we show next (by contraposition) that $\text{cfi}(D_{T'}) = \{(P' \mapsto \mathbf{u})\}$ iff $\text{Th}(T') = P'$.

if: Let $v \in \text{cfi}(D_{T'})$ with $v(r) \neq \mathbf{u}$ for some $r \in P'$. We will show that $r \notin \text{Th}(T')$.

Define the sets $F = \{p \in P' \mid v(p) \neq \mathbf{u}\}$ and $U = P' \setminus F$, and an interpretation $w : P' \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that $w = (F \mapsto \mathbf{f}) \circ (U \mapsto \mathbf{t})$. We will show that w is a model of T' , whence by $r \in F$ and thus $w(r) = \mathbf{f}$ it follows that $T' \not\models r$. Let $\psi \vee p \in T'$ be a clause.

- $p \in F$. Then by definition of φ_p , we have $v(p) = \mathbf{f}$, whence $\varphi_p^v \equiv \bigvee_{\xi \vee p \in T'} \neg \xi^v$ is unsatisfiable. Thus in particular we find that $\neg \psi^v$ is unsatisfiable, that is, the purely negative clause ψ^v is true. Since w maps the atoms in F just as v does, also ψ^w is true, whence $w(\psi \vee p) = \mathbf{t}$.
- $p \in U$. The fact that $w(p) = \mathbf{t}$ by definition implies that $w(\psi \vee p) = \mathbf{t}$.

Thus w is a model for T' with $w(r) = \mathbf{f}$ and $\text{Th}(T') \subsetneq P'$.

only if: Let $\text{Th}(T') \subsetneq P'$. Define an interpretation $v : P' \rightarrow \{\mathbf{u}, \mathbf{f}\}$ such that

$$v(p) = \begin{cases} \mathbf{u} & \text{if } p \in \text{Th}(T') \\ \mathbf{f} & \text{otherwise} \end{cases} \quad \text{for } p \in P'$$

Clearly v is non-trivial by assumption. It remains to show that v is conflict-free, where it obviously suffices to consider only $P' \setminus \text{Th}(T')$. Let $p \in P' \setminus \text{Th}(T')$ be arbitrary; we show that φ_p^v is unsatisfiable. By assumption, $T' \not\models p$, that is, there is a model $w : P' \rightarrow \{\mathbf{t}, \mathbf{f}\}$ of T' with $w(p) = \mathbf{f}$. Since w is a model of T' , for all clauses $\psi \vee p \in T'$ we find $w(\psi \vee p) = \mathbf{t}$. Since furthermore $w(p) = \mathbf{f}$, it follows that $w(\psi) = \mathbf{t}$ for all $\psi \vee p \in T'$. In particular, each such ψ is non-empty, and since it is a negative clause, it must contain a negative literal $\neg r_\psi$ for an atom $r_\psi \in P' \setminus \text{Th}(T')$, that is, with $v(r_\psi) = \mathbf{f}$. It follows thus that v is a model for the conjunction $\bigwedge_{\psi \vee p \in T'} \psi$, whence

$$\top \equiv \left(\bigwedge_{\psi \vee p \in T'} \psi \right)^v \equiv \left(\neg \bigvee_{\psi \vee p \in T'} \neg \psi \right)^v \equiv \left(p \vee \neg \bigvee_{\psi \vee p \in T'} \neg \psi \right)^v \equiv (\neg \varphi_p)^v$$

Therefore $\varphi_p^v \equiv \perp$ is unsatisfiable. Since $p \in P' \setminus \text{Th}(T')$ was arbitrary, v is conflict-free. \square

The proof of Proposition 4.9 entails that a similar result holds for bipolar ADFs.

Corollary 4.27. *The following decision problems coincide:*

1. $\text{Exists}_{cfi}^{\pm}$, $\text{Exists}_{nai}^{\pm}$, $\text{Exists}_{stg}^{\pm}$, $\text{Exists}_{nai_2}^{\pm}$, $\text{Exists}_{stg_2}^{\pm}$;
2. $\text{Cred}_{cfi}^{\mathbf{t},\pm}$ and $\text{Cred}_{nai}^{\mathbf{t},\pm}$;
3. $\text{Cred}_{cfi}^{\mathbf{f},\pm}$ and $\text{Cred}_{nai}^{\mathbf{f},\pm}$.

4.2.3 ENTAILMENT

Credulous reasoning over conflict-free interpretations works as in the general case, only that we do not need an oracle and the problem thus becomes easier.

Proposition 4.28. $\text{Cred}_{cfi}^{\mathbf{t},\pm}$ is in L.

Proof. Given a BADF D over S and an $z \in S$, to check whether z is credulously true we check in L whether the interpretation $v = (S \mapsto \mathbf{u}) \circ \{z \mapsto \mathbf{t}\}$ is conflict-free. (The interpretation need not be constructed explicitly, we can “keep in mind” its easy structure.) If so, we have found a witness; if not, then φ_s^v is unsatisfiable and will be so for any $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v \leq_i w$. \square

In line with the complexities for the general case, it is harder to decide whether a particular statement can be set to false in a conflict-free manner.

Proposition 4.29. $\text{Cred}_{cfi}^{\mathbf{f},\pm}$ is NP-complete.

Proof. in NP: Given a BADF D over S and an $z \in S$, we can guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(z) = \mathbf{f}$ and verify in L that it is conflict-free (Proposition 4.21).

NP-hard: We reduce from refutability checking. Let $\psi = \psi_1 \vee \dots \vee \psi_n$ be a propositional formula in DNF over a vocabulary P . We assume w.l.o.g. that in each monomial ψ_i , each atom occurs only once (negated or unnegated), as additional occurrences would be redundant (or make the monomial redundant). Now we define the (bipolar) ADF D_ψ over $S = P \cup \{d_1, \dots, d_n, z\}$ by

$$\begin{aligned} \varphi_p &= p && \text{for } p \in P \\ \varphi_{d_i} &= \psi_i && \text{for } 1 \leq i \leq n \\ \varphi_z &= d_1 \vee \dots \vee d_n \end{aligned}$$

We show that there is a conflict-free interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(z) = \mathbf{f}$ iff ψ is refutable.

if: Let ψ be refutable. Then there is a valuation $v_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ with $v_P(\psi) = \mathbf{f}$, that is, $v_P(\psi_i) = \mathbf{f}$ for all $1 \leq i \leq n$. Define an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ as follows:

$$\begin{aligned} v(p) &= v_P(p) && \text{for } p \in P \\ v(d_i) &= \mathbf{f} && \text{for } 1 \leq i \leq n \\ v(z) &= \mathbf{f} \end{aligned}$$

We have to show that v is conflict-free for D_ψ . This is clear for $p \in P$ and z ; for the d_i , it follows since $v_P(\psi_i) = \mathbf{f}$ implies $\psi_i^v \equiv \perp$ by definition of v .

only if: Let $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ be conflict-free with $v(z) = \mathbf{f}$. Then $\varphi_z^v = (d_1 \vee \dots \vee d_n)^v$ is unsatisfiable, that is, $v(d_1) = \dots = v(d_n) = \mathbf{f}$. Consequently, all of $\varphi_{d_1}^v, \dots, \varphi_{d_n}^v$ are unsatisfiable. Define a valuation $v_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ by $v_P = v|_P$. By the above, $v_P(\psi_i) = \mathbf{f}$ for all $1 \leq i \leq n$, whence $v_P(\psi) = \mathbf{f}$. Thus ψ is refutable. \square

The proof of the membership part of the next result works as in the general case, only that this time we need no oracle to verify conflict-freeness of interpretations. In the hardness part, we can reuse the reduction from above and employ the fact that for the BADFs obtained via that reduction there are no naive interpretations where the statement in question remains undecided. Therefore, for that BADF class, sceptical truth is exactly the co-problem of credulous falsity and coNP-hardness of one follows from NP-hardness of the other.

Proposition 4.30. $\text{Scep}_{\text{nai}}^{\mathbf{t}, \pm}$ is coNP-complete.

Proof. in coNP: Let D be a BADF over S and $z \in S$. To show that z is *not* sceptically true for D , we guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ and show in \mathbf{L} that v is conflict-free, and furthermore show that one of (1) $v(z) = \mathbf{f}$ or (2) $v(z) = \mathbf{u}$ and $v \circ \{z \mapsto \mathbf{t}\}$ is not conflict-free.

coNP-hard: We can use the reduction from the hardness proof of Proposition 4.29. In the BADF constructed there, all naive interpretations v are two-valued with respect to z (φ_z^v is either satisfiable or unsatisfiable, and z does not occur in any acceptance formula). Thus, there is a naive interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ where $v(z) \neq \mathbf{t}$ iff there is a naive interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ where $v(z) = \mathbf{f}$. \square

The hardness part of the proof of the next result is the most complicated one when considering all decision problems for bipolar ADFs so far. It reduces QBF-truth for formulas of the form $\forall P \exists Q \psi$ to naive sceptical falsity for bipolar ADFs. Here, bipolarity of the constructed ADF is achieved by splitting the given conjunctive normal form ψ into clauses and creating a statement for each clause. However, this bars us from using conflict-free semantics' built-in (un)satisfiability check for solving the (un)satisfiability problem associated to $\forall P \exists Q \psi$. Alas, we can regain this ability by moving the inner satisfiability check $\exists Q \psi$ from the level of a single statement to the level of an interpretation, and dealing with the outer quantification $\forall P$ on the level of all interpretations.

Theorem 4.31. $\text{Scep}_{\text{nai}}^{\mathbf{f}, \pm}$ is Π_2^{P} -complete.

Proof. in Π_2^{P} : Let D be a BADF over S and $z \in S$. To show that z is not sceptically false, we guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(z) \neq \mathbf{f}$ and verify in coNP that it is naive (Proposition 4.22).

Π_2^P -hard: Let $\Phi = \forall P \exists Q \psi$ be a QBF with $\psi = \psi_1 \wedge \dots \wedge \psi_n$ in CNF. Define ψ' as ψ with all positive occurrences of $q \in Q$ replaced by $\neg q'$. Further define a BADF D_Φ over S as follows:

$$\begin{aligned}
 S &= P \cup Q \cup Q' \cup \{c_1, \dots, c_n, x, y, z\} \\
 \varphi_p &= p && \text{for } p \in P \\
 \varphi_q &= \neg q \wedge (\neg x \vee y \vee z) && \text{for } q \in Q \\
 \varphi_{q'} &= \neg q' \wedge (\neg x \vee y \vee z) && \text{for } q \in Q \\
 \varphi_{c_i} &= \neg c_i \wedge \neg \psi'_i && \text{for } 1 \leq i \leq n \\
 \varphi_x &= \neg x \vee \bigwedge_{q \in Q} (q \vee q') \\
 \varphi_y &= \neg y \wedge \bigvee_{q \in Q} (q \wedge q') \\
 \varphi_z &= \neg z \wedge (c_1 \vee \dots \vee c_n)
 \end{aligned}$$

We show that z is sceptically false in D_Φ iff Φ is true.

if: Let Φ be true. Then for each $v_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$, there is a $v_Q : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$ such that $v_Q(v_P(\psi)) = \mathbf{t}$, that is, for all $1 \leq i \leq n$, we have $v_Q(v_P(\psi_i)) = \mathbf{t}$. We show that every conflict-free interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(z) = \mathbf{u}$ can be extended to a conflict-free interpretation $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v \leq_i w$ and $w(z) = \mathbf{f}$. Clearly, interpretations with $v(z) = \mathbf{t}$ are not conflict-free, so this establishes the result.

Let $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ be conflict-free with $v(z) = \mathbf{u}$. It follows that for all $q \in Q$, the formulas φ_q^v and $\varphi_{q'}^v$ are satisfiable, whence $v(Q \cup Q') = \{\mathbf{u}\}$. This in turn implies that both φ_x and φ_y are satisfiable, whence $v(x) \neq \mathbf{f}$ and $v(y) = \mathbf{u}$. We now define $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ as follows: For $p \in P$, we set $w(p) = v(p)$ if $v(p) \in \{\mathbf{t}, \mathbf{f}\}$ and $w(p) \in \{\mathbf{t}, \mathbf{f}\}$ arbitrary otherwise. This defines a valuation $w|_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$, thus by assumption there exists a $v_Q : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$ with $v_Q(w|_P(\psi)) = \mathbf{t}$. We let

$$\begin{aligned}
 w(q) &= \begin{cases} \mathbf{f} & \text{if } v_Q(q) = \mathbf{f} \\ \mathbf{u} & \text{otherwise} \end{cases} && \text{for } q \in Q \\
 w(q') &= \begin{cases} \mathbf{f} & \text{if } v_Q(q) = \mathbf{t} \\ \mathbf{u} & \text{otherwise} \end{cases} && \text{for } q \in Q \\
 w(c_i) &= \mathbf{f} && \text{for } 1 \leq i \leq n \\
 w(x) &= \mathbf{t} \\
 w(y) &= \mathbf{f} \\
 w(z) &= \mathbf{f}
 \end{aligned}$$

We have to show that w is conflict-free. This is easy to see for z and the $q, q' \in Q \cup Q'$; for x and y it holds since for each $q \in Q$ we have either $w(q) = \mathbf{f}$ or $w(q') = \mathbf{f}$ by definition. It remains to show that for each $i \in \{1, \dots, n\}$, the formula $\varphi_{c_i}^w = \neg \perp \wedge \neg \psi_i'^w$ is unsatisfiable. Let $i \in \{1, \dots, n\}$. By construction,

$\neg\psi_i^{w|P}$ is equivalent to a conjunction of atoms from $Q \cup Q'$. By assumption, we have that $\psi_i^{w|P \circ v_Q} \equiv \top$.

- ψ_i contains a literal p (or $\neg p$) for some $p \in P$ with $w(p) = \mathbf{t}$ ($w(p) = \mathbf{f}$). Then the conjunction for $\neg\psi_i^{w|P}$ contains $\neg\top$ (or $\neg\neg\perp$) and is unsatisfiable.
- ψ_i contains an atom $q \in Q$ with $v_Q(q) = \mathbf{t}$. Then by definition $w(q') = \mathbf{f}$, furthermore ψ_i' contains $\neg q'$ and the conjunction for $\varphi_{c_i}^w = \neg\psi_i'^w$ contains $\neg\neg\perp$ whence it is unsatisfiable.
- ψ_i contains a literal $\neg q$ for a $q \in Q$ with $v_Q(q) = \mathbf{f}$. Then by definition $w(q) = \mathbf{f}$, furthermore ψ_i' contains $\neg q$ and the conjunction for $\varphi_{c_i}^w = \neg\psi_i'^w$ contains $\neg\neg\perp$ whence it is unsatisfiable.

Thus for each conflict-free interpretation v for D_Φ there is a conflict-free w with $v \leq_i w$ and $w(z) = \mathbf{f}$ whence z is sceptically false in D_Φ .

only if: Let Φ be false, that is, let $v_P : P \rightarrow \{\mathbf{t}, \mathbf{f}\}$ be such that ψ^{v_P} is unsatisfiable. Then for each $v_Q : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$, we have $\psi^{v_P \circ v_Q} \equiv \perp$, that is, for each $v_Q : Q \rightarrow \{\mathbf{t}, \mathbf{f}\}$ there is a $j \in \{1, \dots, n\}$ such that $\psi_j^{v_P \circ v_Q} \equiv \perp$. We construct an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ as follows:

$$\begin{aligned}
 v(p) &= v_P(p) && \text{for } p \in P \\
 v(q) &= \mathbf{u} && \text{for } q \in Q \\
 v(q') &= \mathbf{u} && \text{for } q' \in Q' \\
 v(c_i) &= \begin{cases} \mathbf{u} & \text{if } \varphi_{c_i}^{v_P} \text{ is satisfiable} \\ \mathbf{f} & \text{otherwise} \end{cases} && \text{for } 1 \leq i \leq n \\
 v(x) &= \mathbf{t} \\
 v(y) &= \mathbf{u} \\
 v(z) &= \mathbf{u}
 \end{aligned}$$

To conclude the proof, we show that v is naive. It is immediate that v is conflict-free. Assume to the contrary that there is a $w : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v <_i w$. Then $w^{-1}(\mathbf{f}) \setminus v^{-1}(\mathbf{f}) \neq \emptyset$. If $w(y) = \mathbf{f}$ then for each $q \in Q$, either $w(q) = \mathbf{f}$ or $w(q') = \mathbf{f}$. If $w(c_i) = \mathbf{f} \neq v(c_i)$ for some $i \in \{1, \dots, n\}$, then $w(s) = \mathbf{f}$ for some $s \in Q \cup Q'$ (since $\varphi_{c_i}^w$ is unsatisfiable with $v_P \leq_i w$ although $\varphi_{c_i}^{v_P}$ with $v_P(P) \subseteq \{\mathbf{t}, \mathbf{f}\}$ is satisfiable). If $w(s) = \mathbf{f}$ for some $s \in Q \cup Q'$ then also $w(z) = \mathbf{f}$. Thus φ_z^w is unsatisfiable, that is, $\varphi_{c_i}^w = \neg\psi_i'^w \equiv \perp$ for all $1 \leq i \leq n$, whence $\psi_i'^w \equiv \top$ for all $1 \leq i \leq n$. But then $v_P \leq_i v <_i w$ implies that ψ^{v_P} is satisfiable, contradiction. Thus v with $v(z) = \mathbf{u}$ is naive. \square

Figure 5 shows the BADF D_Φ for the QBF $\Phi = \forall\{p\}\exists\{q_1, q_2\}(p \vee \neg q_1) \wedge (\neg p \vee q_2) \wedge (q_1 \vee \neg q_2)$.

Once this result is obtained, the remaining problems for stage and *nai*₂ semantics in the bipolar case can be tackled using techniques that are either standard or known from the general case. We first show how credulous/sceptical truth for stage semantics can be reduced to credulous/sceptical falsity.

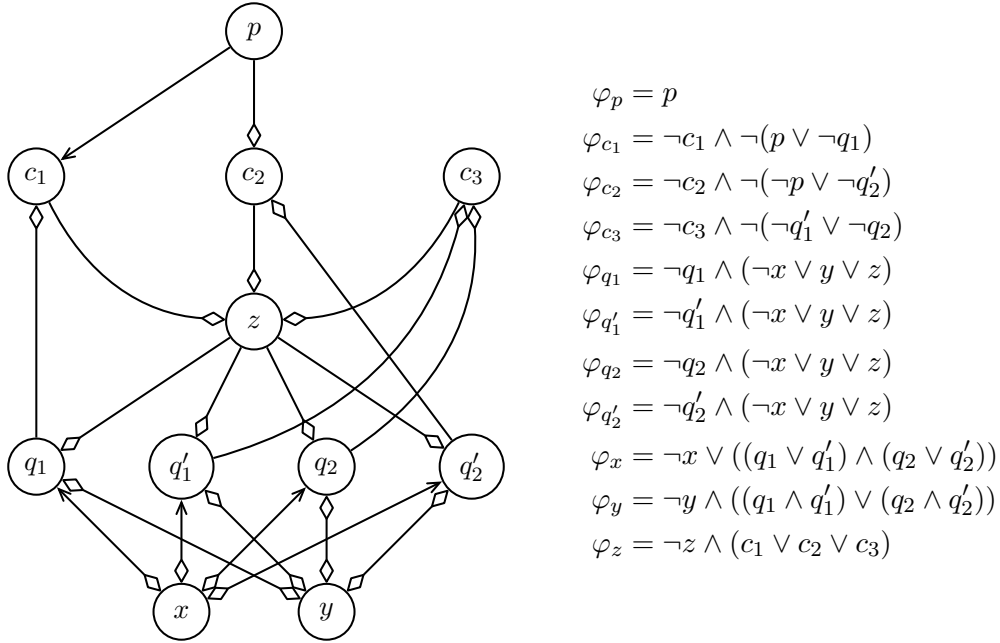


Figure 5: BADF D_Φ (Theorem 4.31) for $\Phi = \forall\{p\}\exists\{q_1, q_2\}(p \vee \neg q_1) \wedge (\neg p \vee q_2) \wedge (q_1 \vee \neg q_2)$. Attacking links are shown with arrow tips, supporting links have diamond tips. All nodes are self-looping; the loops have been omitted from the presentation for clarity. The QBF Φ is true, since for $p \mapsto \mathbf{t}$ we find $q_2 \mapsto \mathbf{t}$ and $q_1 \mapsto \mathbf{t}$; for $p \mapsto \mathbf{f}$ we find $q_1 \mapsto \mathbf{f}$ and $q_2 \mapsto \mathbf{f}$. Consequently, the constructed BADF D_Φ has two naive interpretations: $v_1 = \{p \mapsto \mathbf{t}, c_1 \mapsto \mathbf{f}, c_2 \mapsto \mathbf{f}, c_3 \mapsto \mathbf{f}, q_1 \mapsto \mathbf{u}, q_1' \mapsto \mathbf{f}, q_2 \mapsto \mathbf{u}, q_2' \mapsto \mathbf{f}, x \mapsto \mathbf{t}, y \mapsto \mathbf{f}, z \mapsto \mathbf{f}\}$ and $v_2 = \{p \mapsto \mathbf{f}, c_1 \mapsto \mathbf{f}, c_2 \mapsto \mathbf{f}, c_3 \mapsto \mathbf{f}, q_1 \mapsto \mathbf{f}, q_1' \mapsto \mathbf{u}, q_2 \mapsto \mathbf{f}, q_2' \mapsto \mathbf{u}, x \mapsto \mathbf{t}, y \mapsto \mathbf{f}, z \mapsto \mathbf{f}\}$. As expected (since Φ is true), z is false in both of them.

Lemma 4.32.

1. $\text{Cred}_{stg}^{\mathbf{t}, \pm}$ can be polynomially reduced to $\text{Cred}_{stg}^{\mathbf{f}, \pm}$;
2. $\text{Scep}_{stg}^{\mathbf{t}, \pm}$ can be polynomially reduced to $\text{Scep}_{stg}^{\mathbf{f}, \pm}$.

Proof. Let D be a BADF over S and $z \in S$. Construct D' over $S' = S \cup \{z'\}$ as an extension of D with $\varphi_{z'} = \neg z$. (Clearly D' is bipolar since D is bipolar and z is attacking in $\varphi_{z'}$.) By construction, $stg(D)$ and $stg(D')$ correspond one-to-one, and in each $v \in stg(D')$ we find $v(z') \in \{\mathbf{t}, \mathbf{f}\}$: Either $v(z) \in \{\mathbf{f}, \mathbf{u}\}$, then $v(z') = \mathbf{t}$; or $v(z) = \mathbf{t}$, then $v(z') = \mathbf{f}$. We now show that we can decide credulous (sceptical) truth of z in D by deciding credulous (sceptical) falsity of z' in D' .

1. The following are equivalent:
 - There is a $v \in stg(D')$ with $v(z') = \mathbf{f}$,
 - there is a $v \in stg(D')$ with $\varphi_{z'}^v$ unsatisfiable,
 - there is a $v \in stg(D')$ with $v(z) = \mathbf{t}$,
 - there is a $w \in stg(D)$ with $w(z) = \mathbf{t}$.

2. The following are equivalent:

- There is a $v \in stg(D')$ with $v(z') \neq \mathbf{f}$,
- there is a $v \in stg(D')$ with $v(z') = \mathbf{t}$,
- there is a $v \in stg(D')$ with $v(z) \in \{\mathbf{f}, \mathbf{u}\}$,
- there is a $w \in stg(D)$ with $w(z) \in \{\mathbf{f}, \mathbf{u}\}$,
- there is a $w \in stg(D)$ with $w(z) \neq \mathbf{t}$. □

Together with the fact that hardness for credulous and sceptical truth carries over from AF results shown by Dvořák and Woltran (2010), this leads to tight bounds for all reasoning problems over stage semantics.

Theorem 4.33. $\text{Cred}_{stg}^{\mathbf{t},\pm}$ and $\text{Cred}_{stg}^{\mathbf{f},\pm}$ are Σ_2^P -complete; $\text{Scep}_{stg}^{\mathbf{t},\pm}$ and $\text{Scep}_{stg}^{\mathbf{f},\pm}$ are Π_2^P -complete.

Proof. Containment of $\text{Cred}_{stg}^{\mathbf{t},\pm}$ and $\text{Cred}_{stg}^{\mathbf{f},\pm}$ in Σ_2^P is straightforward: given a BADF D over S and a $z \in S$, we can guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(z) = \mathbf{t}$ (or $v(z) = \mathbf{f}$) and verify in coNP that v is stage (Proposition 4.23). Likewise we can obtain membership of $\text{Scep}_{stg}^{\mathbf{t},\pm}$ and $\text{Scep}_{stg}^{\mathbf{f},\pm}$ in Π_2^P : given a BADF D over S and a $z \in S$, to show that z is *not* sceptically true (false), we can guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(z) \neq \mathbf{t}$ (or $v(z) \neq \mathbf{f}$) and verify in coNP that v is stage (Proposition 4.23).

Hardness of $\text{Cred}_{stg}^{\mathbf{t},\pm}$ for Σ_2^P and that of $\text{Scep}_{stg}^{\mathbf{t},\pm}$ for Π_2^P carries over from AFs (Dvořák & Woltran, 2010). By Lemma 4.32, this yields the hardness results also for $\text{Cred}_{stg}^{\mathbf{f},\pm}$ and $\text{Scep}_{stg}^{\mathbf{f},\pm}$. □

For nai_2 and stg_2 semantics, we can make use of the techniques established for the general case (in particular Proposition 4.18).

Proposition 4.34. Let $\sigma \in \{nai_2, stg_2\}$. The following problems can be polynomially reduced to each other:

- $\text{Cred}_{\sigma}^{\mathbf{t},\pm}$ and $\text{Cred}_{\sigma}^{\mathbf{f},\pm}$,
- $\text{Scep}_{\sigma}^{\mathbf{t},\pm}$ and $\text{Scep}_{\sigma}^{\mathbf{f},\pm}$,
- $\text{co-Scep}_{\sigma}^{\mathbf{f},\pm}$ and $\text{Cred}_{\sigma}^{\mathbf{t},\pm}$.

Proof. It suffices to observe that all constructions in the proof of Proposition 4.18 preserve bipolarity. □

σ	<i>cfi</i>	<i>nai</i>	<i>stg</i>	<i>nai</i> ₂	<i>stg</i> ₂
Ver_σ^\pm	in L (Proposition 4.21)	coNP-c (Proposition 4.22)	coNP-c (Proposition 4.23)	coNP-c (Proposition 4.24)	coNP-c (Proposition 4.25)
Exists_σ^\pm	P-c (Proposition 4.26)	P-c (Corollary 4.27)	P-c (Corollary 4.27)	P-c (Corollary 4.27)	P-c (Corollary 4.27)
$\text{Cred}_\sigma^{\mathbf{t},\pm}$	in L (Proposition 4.28)	in L (Corollary 4.27)	Σ_2^{P} -c (Theorem 4.33)	Σ_2^{P} -c (Theorem 4.35)	Σ_2^{P} -c (Theorem 4.35)
$\text{Cred}_\sigma^{\mathbf{f},\pm}$	NP-c (Proposition 4.29)	NP-c (Corollary 4.27)	Σ_2^{P} -c (Theorem 4.33)	Σ_2^{P} -c (Theorem 4.35)	Σ_2^{P} -c (Theorem 4.35)
$\text{Scep}_\sigma^{\mathbf{t},\pm}$	trivial	coNP-c (Proposition 4.30)	Π_2^{P} -c (Theorem 4.33)	Π_2^{P} -c (Theorem 4.35)	Π_2^{P} -c (Theorem 4.35)
$\text{Scep}_\sigma^{\mathbf{f},\pm}$	trivial	Π_2^{P} -c (Theorem 4.31)	Π_2^{P} -c (Theorem 4.33)	Π_2^{P} -c (Theorem 4.35)	Π_2^{P} -c (Theorem 4.35)

Table 2: Complexity results for naive-based semantics of bipolar abstract dialectical frameworks; \mathcal{C} -c stands for \mathcal{C} -complete.

Together with the observation that the hardness reduction of Theorem 4.31 (for $\text{Cred}_{nai}^{\mathbf{f},\pm}$) works for the *nai*₂ and *stg*₂ semantics as well, the proposition leads to the following results.

Theorem 4.35. *Let $\sigma \in \{nai_2, stg_2\}$.*

- $\text{Cred}_\sigma^{\mathbf{t},\pm}$ and $\text{Cred}_\sigma^{\mathbf{f},\pm}$ are Σ_2^{P} -complete.
- $\text{Scep}_\sigma^{\mathbf{t},\pm}$ and $\text{Scep}_\sigma^{\mathbf{f},\pm}$ are Π_2^{P} -complete.

Proof. membership: Let D be a BADF over S and $z \in S$. For $\text{Cred}_\sigma^{\mathbf{t},\pm}$ and $\text{Cred}_\sigma^{\mathbf{f},\pm}$, we can guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(z) = \mathbf{t}$ (or $v(z) = \mathbf{f}$, respectively) and verify in coNP that it is σ (Proposition 4.24, Proposition 4.25). For $\text{Scep}_\sigma^{\mathbf{t},\pm}$ and $\text{Scep}_\sigma^{\mathbf{f},\pm}$, to show that z is *not* sceptically true, we guess an interpretation $v : S \rightarrow \{\mathbf{t}, \mathbf{f}, \mathbf{u}\}$ with $v(z) \neq \mathbf{t}$ (or $v(z) \neq \mathbf{f}$, respectively) and verify in coNP that v is σ (Proposition 4.24, Proposition 4.25).

hardness: We first note that the reduction from Theorem 4.31 works for *nai*₂ and *stg*₂ as well, since the BADF D_Φ constructed there satisfies $\text{ind}_{D_\Phi}(\emptyset) = P$ and $\text{ind}_{D_\Phi}(P) = S$, that is, there are only two “layers” of strongly connected components, where the first is only used for guessing. This establishes Π_2^{P} -hardness of $\text{Scep}_\sigma^{\mathbf{f},\pm}$. Then, the second part of Proposition 4.34 implies Π_2^{P} -hardness of $\text{Scep}_\sigma^{\mathbf{t},\pm}$ and the third part of the same result yields Σ_2^{P} -hardness of $\text{Cred}_\sigma^{\mathbf{t},\pm}$. Consequently, applying the first part of Proposition 4.34 we obtain Σ_2^{P} -hardness also of $\text{Cred}_\sigma^{\mathbf{f},\pm}$. \square

5. Discussion

We introduced and studied a scheme to decompose abstract dialectical frameworks along their strongly connected components. For several semantics, our scheme leads to a new way to compute interpretations, among them admissible, complete, preferred, grounded and two-valued model semantics. Thereby, these semantics turn out to be SCC-recursive, for ADFs just as for AFs. This is strong evidence that SCC-recursiveness is a useful notion for the study not only of AF, but also for ADF semantics.

For the semantics that do not possess the SCC-recursiveness property, our scheme leads to new semantics which arguably remedy some of the original semantics’ shortcomings, such as naive and stage semantics. For these semantics, we studied their relationship to the base semantics and computational properties in exhaustive detail.

Due to the generality of ADFs, this paper – as a byproduct – defines the *nai₂* and *stg₂* semantics also for logic programs. That is, when a normal logic program fails to have (stable) models due to odd-length negative cycles, our decomposition-based scheme can straightforwardly be applied to the logic program’s associated ADF (Brewka & Woltran, 2010; Strass, 2015) to compute *nai₂* and *stg₂* interpretations. Computationally, this is quite economic since the increase in size from logic program to ADF is at most linear.

We also presented numerous novel results on the computational complexity of naive-based semantics for abstract dialectical frameworks. An overview can be found in Tables 1 and 2. The main lesson learned is that naive-based semantics for ADFs are – computationally speaking – not at all “naive”, not even in the case of bipolar ADFs. However, we also observed a difference in complexity between bipolar and general ADFs (one level of the polynomial hierarchy), as previously observed for other semantics (Strass & Wallner, 2015).

Our analysis paves the way for implementing naive-based ADF semantics, for example by adding adequate ASP encodings for the verification and existence problem to the DIAMOND system (Ellmauthaler & Strass, 2014) and its more recent incarnation, GODIAMOND (Strass & Ellmauthaler, 2017). For the sceptical and credulous entailment problems in the third level of the polynomial hierarchy, encodings based on QBFs seem possible (Diller, Wallner, & Woltran, 2014). In terms of possible native ADF algorithms, similar SCC-decomposition-based computation approaches for AFs seem to be promising starting points (Baumann et al., 2011; Liao, 2013). In general, the fact that we have succeeded in applying SCC-recursive decomposition in a semantics-neutral way to graphs with arbitrary argument relations (instead of just attacks) suggests indeed that this SCC-based decomposition procedure can potentially be used to improve the efficiency of ADF-related implementations, something that opens up avenues for future research.

For future work, we plan to consider further semantics. For example, in this paper we have not considered the ADF stable model semantics for clarity, as it also uses notions of reduct and partial evaluation that are subtly different from the ones employed in this paper. Naturally, we also want to implement our decomposition-based scheme to verify whether there is a performance gain in comparison to conventional evaluation methods. Finally, this paper can serve as a starting point for considering arbitrary splittings of ADFs, an important next step as also recognised by Linsbichler (2014).

Acknowledgements This research was partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 389792660 – TRR 248 (see <https://perspicuous-computing.science>), by the Bundesministerium für Bildung und Forschung (BMBF) Förderkennzeichen 01|S20056_NAVAS, and by the European Research Council through the ERC Consolidator Grant 771779 (DeciGUT). We thank several anonymous reviewers for their helpful feedback.

References

- Al-Abdulkarim, L., Atkinson, K., & Bench-Capon, T. J. M. (2014). Abstract dialectical frameworks for legal reasoning. In Hoekstra, R. (Ed.), *Proceedings of the Twenty-Seventh Annual Conference on Legal Knowledge and Information Systems (JURIX)*, Vol. 271 of *FAIA*, pp. 61–70. IOS Press.
- Al-Abdulkarim, L., Atkinson, K., & Bench-Capon, T. J. M. (2015). Evaluating the use of abstract dialectical frameworks to represent case law. In Sichelman, T., & Atkinson, K. (Eds.), *Proceedings of the 15th International Conference on Artificial Intelligence and Law (ICAAIL)*, pp. 156–160. ACM.
- Al-Abdulkarim, L., Atkinson, K., & Bench-Capon, T. J. M. (2016). A methodology for designing systems to reason with legal cases using abstract dialectical frameworks. *Artificial Intelligence and Law*, 24(1), 1–49.
- Al-Abdulkarim, L., Atkinson, K., Bench-Capon, T. J. M., Whittle, S., Williams, R., & Wolfenden, C. (2017). Noise induced hearing loss: An application of the angelic methodology. In Wyner, A. Z., & Casini, G. (Eds.), *Proceedings of the Thirtieth International Conference on Legal Knowledge and Information Systems (JURIX)*, Vol. 302 of *FAIA*, pp. 79–88. IOS Press.
- Amgoud, L., & Cayrol, C. (2002). A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3), 197–215.
- Amgoud, L., Cayrol, C., Lagasquie-Schiex, M., & Livet, P. (2008). On bipolarity in argumentation frameworks. *International Journal of Intelligent Systems*, 23, 1–32.
- Amgoud, L., & Vesic, S. (2010). Handling inconsistency with preference-based argumentation. In Deshpande, A., & Hunter, A. (Eds.), *Scalable Uncertainty Management – 4th International Conference (SUM)*, Vol. 6379 of *LNCS*, pp. 56–69. Springer.
- Baroni, P., Caminada, M., & Giacomin, M. (2011). An introduction to argumentation semantics. *Knowledge Engineering Review*, 26(4), 365–410.
- Baroni, P., Giacomin, M., & Guida, G. (2005). SCC-recursiveness: A general schema for argumentation semantics. *Artificial Intelligence*, 168(1–2), 162–210.
- Baumann, R. (2011). Splitting an argumentation framework. In Delgrande, J. P., & Faber, W. (Eds.), *Logic Programming and Nonmonotonic Reasoning (LPNMR)*, Vol. 6645 of *LNCS*, pp. 40–53. Springer.
- Baumann, R., Brewka, G., & Wong, R. (2011). Splitting argumentation frameworks: An empirical evaluation. In Modgil, S., Oren, N., & Toni, F. (Eds.), *Theory and Appli-*

- cations of Formal Argumentation – First International Workshop (TAFE)*, Vol. 7132 of *LNCS*, pp. 17–31. Springer.
- Bench-Capon, T. J. M. (2003). Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, *13*(3), 429–448.
- Brewka, G., & Gordon, T. F. (2010). Carneades and abstract dialectical frameworks: A reconstruction. In Baroni, P., Cerutti, F., Giacomin, M., & Simari, G. R. (Eds.), *Proceedings of the Third International Conference on Computational Models of Argument (COMMA)*, Vol. 216 of *FAIA*, pp. 3–12. IOS Press.
- Brewka, G., Strass, H., Ellmauthaler, S., Wallner, J. P., & Woltran, S. (2013). Abstract Dialectical Frameworks Revisited. In Rossi, F. (Ed.), *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 803–809. IJCAI/AAAI.
- Brewka, G., & Woltran, S. (2010). Abstract Dialectical Frameworks. In Lin, F., Sattler, U., & Truszczyński, M. (Eds.), *Proceedings of the Twelfth International Conference on Principles of Knowledge Representation and Reasoning (KR)*, pp. 102–111. AAAI Press.
- Chang, R., & Kadin, J. (1996). The Boolean hierarchy and the polynomial hierarchy: A closer connection. *SIAM Journal on Computing*, *25*(2), 340–354.
- Cramer, M., & Guillaume, M. (2019). Empirical study on human evaluation of complex argumentation frameworks. In Calimeri, F., Leone, N., & Manna, M. (Eds.), *Logics in Artificial Intelligence – 16th European Conference (JELIA)*, Vol. 11468 of *LNCS*, pp. 102–115. Springer.
- Cramer, M., & van der Torre, L. (2019). SCF2 – an argumentation semantics for rational human judgments on argument acceptability. In Beierle, C., Ragni, M., Stolzenburg, F., & Thimm, M. (Eds.), *Proceedings of the 8th Workshop on Dynamics of Knowledge and Belief (DKB-2019) and the 7th Workshop KI & Kognition (KIK-2019) co-located with 44nd German Conference on Artificial Intelligence (KI 2019)*, Vol. 2445 of *CEUR Workshop Proceedings*, pp. 24–35. CEUR-WS.org.
- Diller, M., Wallner, J. P., & Woltran, S. (2014). Reasoning in abstract dialectical frameworks using Quantified Boolean Formulas. In Parsons, S., Oren, N., & Reed, C. (Eds.), *Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA)*, Vol. 266 of *FAIA*, pp. 241–252. IOS Press.
- Dung, P. M. (1995). On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n -Person Games. *Artificial Intelligence*, *77*(2), 321–358.
- Dunne, P. E., Hunter, A., McBurney, P., Parsons, S., & Wooldridge, M. (2011). Weighted argument systems: Basic definitions, algorithms, and complexity results. *Artificial Intelligence*, *175*(2), 457–486.
- Dvořák, W., & Gaggl, S. A. (2016). Stage semantics and the SCC-recursive schema for argumentation semantics. *Journal of Logic and Computation*, *26*(4), 1149–1202.
- Dvořák, W., & Woltran, S. (2010). Complexity of semi-stable and stage semantics in argumentation frameworks. *Information Processing Letters*, *110*(11), 425–430.

- Dvořák, W., & Woltran, S. (2011). On the intertranslatability of argumentation semantics. *Journal of Artificial Intelligence Research*, 41, 445–475.
- Ellmauthaler, S., & Strass, H. (2014). The DIAMOND system for computing with abstract dialectical frameworks. In Parsons, S., Oren, N., & Reed, C. (Eds.), *Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA)*, Vol. 266 of *FAIA*, pp. 233–240. IOS Press.
- Gaggl, S. A., Rudolph, S., & Strass, H. (2015). On the computational complexity of naive-based semantics for abstract dialectical frameworks. In Yang, Q., & Wooldridge, M. (Eds.), *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2985–2991. IJCAI/AAAI.
- Gaggl, S. A., & Strass, H. (2014). Decomposing abstract dialectical frameworks. In Parsons, S., Oren, N., & Reed, C. (Eds.), *Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA)*, Vol. 266 of *FAIA*, pp. 281–292. IOS Press.
- Gaggl, S. A., & Woltran, S. (2013). The cf2 argumentation semantics revisited. *Journal of Logic and Computation*, 23(5), 925–949.
- Li, H., Oren, N., & Norman, T. J. (2011). Probabilistic argumentation frameworks. In Modgil, S., Oren, N., & Toni, F. (Eds.), *Revised Selected Papers of the First International Workshop on Theory and Applications of Formal Argumentation (TAFAs)*, Vol. 7132 of *LNCS*, pp. 1–16. Springer.
- Liao, B. (2013). Toward incremental computation of argumentation semantics: A decomposition-based approach. *Annals of Mathematics and Artificial Intelligence*, 67(3–4), 319–358.
- Liao, B. (2014). *Efficient Computation of Argumentation Semantics*. Intelligent systems series. Academic Press.
- Linsbichler, T. (2014). Splitting abstract dialectical frameworks. In Parsons, S., Oren, N., & Reed, C. (Eds.), *Proceedings of the Fifth International Conference on Computational Models of Argument (COMMA)*, Vol. 266 of *FAIA*, pp. 357–368. IOS Press.
- Nouioua, F., & Risch, V. (2011). Argumentation frameworks with necessities. In Benferhat, S., & Grant, J. (Eds.), *Proceedings of the Fifth International Conference on Scalable Uncertainty Management (SUM)*, Vol. 6929 of *LNCS*, pp. 163–176. Springer.
- Papadimitriou, C. H. (1995). *Computational Complexity*. Addison-Wesley.
- Papadimitriou, C. H., & Yannakakis, M. (1982). The complexity of facets (and some facets of complexity). In Lewis, H. R., Simons, B. B., Burkhard, W. A., & Landweber, L. H. (Eds.), *Proceedings of the Fourteenth Annual ACM Symposium on Theory of Computing (STOC)*. ACM.
- Polberg, S., & Doder, D. (2014). Probabilistic abstract dialectical frameworks. In Fermé, E., & Leite, J. (Eds.), *Proceedings of the Fourteenth European Conference on Logics in Artificial Intelligence (JELIA)*, Vol. 8761 of *LNCS*, pp. 591–599. Springer.
- Rahwan, I., Simari, G. R., & van Benthem, J. (2009). *Argumentation in Artificial Intelligence*, Vol. 47. Springer.

- Spanring, C. (2012). Intertranslatability results for abstract argumentation semantics. Master's thesis, Vienna University of Technology.
- Strass, H. (2013). Approximating operators and semantics for abstract dialectical frameworks. *Artificial Intelligence*, 205, 39–70.
- Strass, H. (2015). Expressiveness of two-valued semantics for abstract dialectical frameworks. *Journal of Artificial Intelligence Research*, 54, 193–231.
- Strass, H. (2017). *Abstract Dialectical Frameworks. An Analysis of their Properties and Role in Knowledge Representation and Reasoning*. Habilitationsschrift. Leipzig University. <http://nbn-resolving.de/urn:nbn:de:bsz:15-qucosa2-167205>.
- Strass, H. (2018). Instantiating rule-based defeasible theories in abstract dialectical frameworks and beyond. *Journal of Logic and Computation*, 28(3), 605–627.
- Strass, H., & Ellmauthaler, S. (2017). goDIAMOND 0.6.6 system description. In Gaggl, S. A., Linsbichler, T., Maratea, M., & Woltran, S. (Eds.), *Second International Competition on Computational Models of Argumentation*. <https://argumentationcompetition.org/2017/goDIAMOND.pdf>.
- Strass, H., & Wallner, J. P. (2015). Analyzing the Computational Complexity of Abstract Dialectical Frameworks via Approximation Fixpoint Theory. *Artificial Intelligence*, 226, 34–74.
- Toni, F., & Sergot, M. (2011). Argumentation and answer set programming. In Balduccini, M., & Son, T. (Eds.), *Logic Programming, Knowledge Representation, and Nonmonotonic Reasoning*, Vol. 6565 of *LNCS*, pp. 164–180. Springer.
- Verheij, B. (1996). Two approaches to dialectical argumentation: Admissible sets and argumentation stages. In Meyer, J.-J. C., & van der Gaag, L. C. (Eds.), *Proceedings of the Eighth Dutch Conference on Artificial Intelligence (NAIC)*, pp. 357–368.
- Wechsung, G. (1985). On the Boolean closure of NP. In Budach, L. (Ed.), *Proceedings of the International Conference on Fundamentals of Computation Theory (FCT)*, Vol. 199 of *LNCS*, pp. 485–493. Springer.
- Zhang, X., & Lin, Z. (2009). An argumentation-based approach to handling inconsistencies in DL-Lite. In Mertsching, B., Hund, M., & Aziz, Z. (Eds.), *KI 2009: Advances in Artificial Intelligence*, Vol. 5803 of *LNCS*, pp. 615–622. Springer.