# Optimal Status Enforcement in Abstract Argumentation

(including appendix with proofs)

**Andreas Niskanen** and **Johannes P. Wallner** and **Matti Järvisalo**
Helsinki Institute for Information Technology HIIT, Department of Computer Science,
University of Helsinki, Finland

## Abstract

We present complexity results and algorithms for optimal status enforcement in abstract argumentation. Status enforcement is the task of adjusting a given argumentation framework (AF) to support given positive and negative argument statuses, i.e., to accept and reject specific arguments. We study optimal status enforcement as the problem of finding a structurally closest AF supporting given argument statuses. We establish complexity results for optimal status enforcement under several central AF semantics, develop constraint-based algorithms for NP and second-level complete variants of the problem, and empirically evaluate the procedures.

## 1 Introduction

Argumentation is a central topic in modern Artificial Intelligence research [Bench-Capon and Dunne, 2007], motivated by a range of applications in domains such as legal reasoning, multi-agent systems, and decision support [Bench-Capon *et al.*, 2009; McBurney *et al.*, 2012; Amgoud and Prade, 2009]. Argumentation frameworks (AFs) [Dung, 1995] have become the graph-based formal model of choice for many approaches to argumentation in AI, with semantics defining sets of jointly acceptable arguments, i.e., extensions.

Computational approaches with system implementations for reasoning over AFs have recently received notable attention. Two central AF reasoning problems are skeptical and credulous acceptance, i.e., determining if a given argument is supported by a given AF and AF semantics in terms of the argument belonging to all resp. some extensions of the AF. These problems are static (or "non-dynamic"), i.e., defined over a fixed AF. As argumentation is inherently a dynamic process, understanding AF dynamics is an important research problem [Baumann, 2012a; Baumann and Brewka, 2015; Bisquert *et al.*, 2013; Coste-Marquis *et al.*, 2014a; 2014b; Delobelle *et al.*, 2015; Diller *et al.*, 2015]. Central to AF dynamics is the question of how a given AF itself should be adjusted—in analogy with belief change—in light of new knowledge on the arguments the AF should support. Computational approaches to reasoning about AF dynamics are currently at an early stage of development compared to systems for static AF reasoning problems. Extension enforcement [Baumann, 2012b; Bisquert *et al.*, 2013; Coste-Marquis *et al.*, 2015; Wallner *et al.*, 2016]—where, given an AF and a subset of arguments, the task is to find a structurally closest AF that contains the specified subset as (part of) an extension—is one of few AF dynamics problems for which first computational approaches have been recently proposed [Coste-Marquis *et al.*, 2015; Wallner *et al.*, 2016].

In this work we focus on *status enforcement*, a form of AF reasoning that brings together concepts from static credulous/skeptical acceptance and AF dynamics, most closely, extension enforcement. Status enforcement is the task of adjusting a given argumentation framework (AF) to support given positive and negative argument statuses, i.e., adjusting an AF to accept and reject—credulously or skeptically—specific arguments. Intuitively, by enforcing credulously sets of positive and negative argument statuses, any solution AF to the status enforcement problem supports a "point of view" in terms of the positive arguments, at the same time ruling out support for the negative arguments. In the skeptical counterpart, the positive arguments must be supported without any conflicting "points of views". In this work we take on the task of *optimal status enforcement*, i.e., finding a structurally closest AF wrt changes to the attack structure of the AF, supporting given argument statuses. Our main contributions are the following.

(i) For understanding status enforcement, we establish fundamental properties of the problem with connections to extension enforcement and static acceptance problems.

(ii) We establish the computational complexity of optimal status enforcement under central AF semantics (conflict-free, admissible, stable, complete, grounded, and preferred) and parameterizations wrt negative statuses of arguments. Specifically, we identify polynomial-time solvable and NP- and second-level $\Sigma_2^P$-complete variants of the problem.

(iii) We give algorithms for optimal status enforcement, including direct constraint encodings for the NP-complete variants, and counterexample-guided abstraction refinement algorithms based on constrained optimization solvers for variants complete for the second-level of the polynomial hierarchy; and empirically evaluate a prototype implementation of the approaches. Our status enforcement system implementation together with benchmarks used in this paper, as well as full formal proofs of our complexity results, are available via `http://www.cs.helsinki.fi/group/coreo/pakota/`.

## 2 Preliminaries

We recall argumentation frameworks (AFs) [Dung, 1995] and main acceptability AF semantics [Baroni *et al.*, 2011].

**Definition 1.** *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 $b$. An argument $a \in A$ is* defended *(in $F$) by a set $S \subseteq A$ if, for each $b \in A$ such that $(b, a) \in R$, there exists a $c \in S$ such that $(c, b) \in R$.*

**Example 1.** *Let $F = (A, R)$ be an AF with $A = \{a, b, c, d\}$ and $R = \{(a, b), (b, c), (c, d)\}$. The corresponding graph representation is shown in Figure 1a.*

Semantics for AFs are defined through functions $\sigma$ which assign to each AF $F = (A, R)$ a set $\sigma(F) \subseteq 2^A$ of extensions. We consider for $\sigma$ the functions $stb$, $adm$, $com$, $grd$, and $prf$, which stand for stable, admissible, complete, grounded, and preferred, respectively.

**Definition 2.** *Given an AF $F = (A, R)$, the* characteristic function $\mathcal{F}_F : 2^A \to 2^A$ *of $F$ is $\mathcal{F}_F(S) = \{a \in A \mid a \text{ is defended by } S\}$. Moreover, for a set $S \subseteq A$, the* range of $S$ *is $S^+_R = S \cup \{b \mid (a, b) \in R, a \in S\}$.*

**Definition 3.** *Let $F = (A, R)$ be an AF. A set $S \subseteq A$ is* conflict-free (in $F$)*, if there are no $a, b \in S$, such that $(a, b) \in R$. We denote the collection of conflict-free sets of $F$ by $cf(F)$. For a conflict-free set $S \in cf(F)$, it holds that*

- *$S \in stb(F)$ iff $S^+_R = A$;*
- *$S \in adm(F)$ iff $S \subseteq \mathcal{F}_F(S)$;*
- *$S \in com(F)$ iff $S = \mathcal{F}_F(S)$;*
- *$S \in grd(F)$ iff $S$ is the least fixed-point of $\mathcal{F}_F$;*
- *$S \in prf(F)$ iff $S \in adm(F)$ and there is no $T \in adm(F)$ with $S \subset T$.*

For any AF $F$ it holds that $cf(F) \supseteq adm(F) \supseteq com(F) \supseteq prf(F) \supseteq stb(F)$. We use $\sigma$-extension to refer to an extension under a semantics $\sigma$.

As for enforcement operators [Baumann, 2012b; Coste-Marquis *et al.*, 2015; Wallner *et al.*, 2016], *strict* enforcement requires that the given set $P$ of arguments has to be a $\sigma$-extension, while in *non-strict* enforcement $P$ is required to be part of a $\sigma$-extension. We denote the set of attack structures that strictly enforce $P$ under $\sigma$ for $F$ by $enf^\sigma_s(F, P) = \{R' \mid F' = (A, R'),\ P \in \sigma(F')\}$, and by $enf^\sigma_{ns}(F, P) = \{R' \mid F' = (A, R'),\ \exists E \in \sigma(F') \text{ st } E \supseteq P\}$ the non-strict enforcement. The number of changes of an enforcement is the symmetric difference $|R \Delta R'|$ of two attack structures $R$ and $R'$, i.e., $|R \setminus R'| + |R' \setminus R|$. The optimization problem for extension enforcement is then as follows.

**Extension enforcement** ($x \in \{s, ns\}$)
Input: AF $F = (A, R)$, $P \subseteq A$, and semantics $\sigma$.
Task: Find an AF $F^* = (A, R^*)$ with

$$R^* \in \underset{R' \in enf^\sigma_x(F,P)}{\arg\min} |R \Delta R'|.$$

## 3 Optimal Status Enforcement

In this section we define and give properties of the optimal status enforcement problem.

The operators underlying status enforcement modify the attack structure of a given AF $F$ based on two given sets of arguments, $P$ and $N$, where $P \cap N = \emptyset$. From here on, we will consistently use $P$ and $N$ to denote the sets of arguments that are to be so-called *positively* and *negatively enforced*, respectively. We will consider both credulous and skeptical variants of the status enforcement problem. For the credulous case, the pair $(P, N)$ is said to be enforced in an AF $F'$ if (i) each argument in $P$ is credulously accepted in $F'$; and (ii) each argument in $N$ is *not* credulously accepted in $F'$. In the dual, skeptical case, for $(P, N)$ to be enforced in $F'$ we require that (i) each argument in $P$ is skeptically accepted in $F'$; and (ii) each argument in $N$ is *not* skeptically accepted, in $F'$. In status enforcement, we are given an AF $F$ and the two subsets of its arguments, $P$ and $N$, and the task is to find an AF $F'$ that is structurally close to $F$ and in which $(P, N)$ is enforced.

Formally, we define the modified attack structures for a given AF $F = (A, R)$ for credulous status enforcement as follows. We denote by $cred(F, P, N, \sigma)$ the set

$$\{R' \mid F' = (A, R'), P \subseteq \bigcup \sigma(F'), N \cap \bigcup \sigma(F') = \emptyset\}.$$

In words, in the modified AF $F'$, all arguments in $P$ are credulously accepted (in the union of all $\sigma$-extensions), and each argument in $N$ is not credulously accepted (excluded from the union of $\sigma$-extensions).

For skeptical status enforcement, we denote by $skept(F, P, N, \sigma)$ the set

$$\{R' \mid F' = (A, R'), P \subseteq \bigcap \sigma(F'), N \cap \bigcap \sigma(F') = \emptyset\}.$$

In words, in all modified attack structures each argument in $P$ is contained in all $\sigma$-extensions, while each argument in $N$ is excluded from at least one $\sigma$-extension. Note that, by definition, $A \subseteq \bigcap \sigma(F')$ if $\sigma(F') = \emptyset$. From the considered semantics in this paper, only the stable semantics may admit no extensions for a given AF. This means that if $N = \emptyset$, every positive set $P \subseteq A$ can be skeptically enforced under stable semantics by an AF $F'$ that has no stable extensions. In light of this, we require for skeptical enforcement from here on that $\sigma(F') \neq \emptyset$, i.e., the modified AF admits at least one $\sigma$-extension. In summary, optimal status enforcement is defined as follows.

**Optimal Credulous Status Enforcement**
Input: AF $F = (A, R)$, $P, N \subseteq A$, and semantics $\sigma$.
Task: Find an AF $F^* = (A, R^*)$ with

$$R^* \in \underset{R' \in cred(F,P,N,\sigma)}{\arg\min} |R \Delta R'|.$$

**Optimal Skeptical Status Enforcement**
Input: AF $F = (A, R)$, $P, N \subseteq A$, and semantics $\sigma$.
Task: Find an AF $F^* = (A, R^*)$ with

$$R^* \in \underset{R' \in skept(F,P,N,\sigma)}{\arg\min} |R \Delta R'|.$$

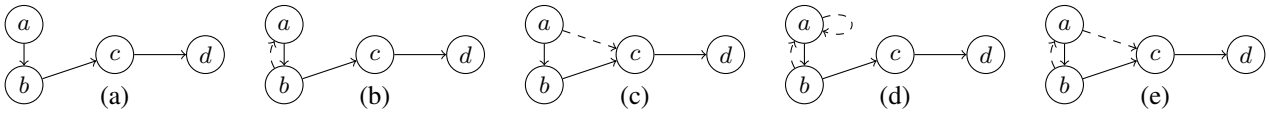


Figure 1: Examples: (a) An AF; (b)–(e) Status enforcement under preferred semantics: enforcing $P = \{d\}$ (b) credulously, (c) skeptically; enforcing $P$ and $N = \{a\}$ (d) credulously, (e) skeptically.

**Example 2.** *In AF $F$ in Figure 1a, the set $\{a, c\}$ is the unique preferred extension. By introducing an attack from $b$ to $a$ (Figure 1b) which yields AF $F'$, we change the preferred extensions to $prf(F') = \{\{a, c\}, \{b, d\}\}$. Thus, under preferred semantics, $d$ is credulously accepted in $F'$. To enforce a positive skeptical status to $d$ under preferred semantics, one can introduce an attack from $a$ to $c$ in $F$ (Figure 1c).*

*For enforcing $P = \{d\}$ and $N = \{a\}$ under preferred semantics, Figure 1d illustrates credulous status enforcement. Here the attack from $b$ to $a$ ensures that there is an admissible set containing $d$, and the self-attack on $a$ enforces that this argument is not contained in any conflict-free set. For skeptical status enforcement with the same sets $P$ and $N$, Figure 1e shows an optimal modification. In this AF we have two preferred extensions $\{a, d\}$ and $\{b, d\}$, which implies that $d$ is skeptically accepted under preferred semantics, while $a$ is not skeptically accepted under preferred semantics, although $a$ is still contained in one preferred extension.*

We now show fundamental properties of the status enforcement operators. We begin with connecting them to extension enforcement. First, in case of so-called unique-status semantics, i.e., semantics $\sigma$ that admit exactly one extension for each AF $F$ ($|\sigma(F)| = 1$), non-strict extension enforcement and enforcing credulous and skeptical statuses coincide when $N = \emptyset$. From the semantics considered in this paper, grounded semantics is a unique-status semantics. Further, strict extension enforcement coincides with enforcing credulous and skeptical statuses when $N = A \setminus P$.

**Proposition 1.** *Let $F = (A, R)$ be an AF, $P \subseteq A$, $N = A \setminus P$, and $\sigma$ a unique-status semantics. It holds that*

- $enf^{\sigma}_{ns}(F, P) = cred(F, P, \emptyset, \sigma) = skept(F, P, \emptyset, \sigma)$;
- $enf^{\sigma}_{s}(F, P) = cred(F, P, N, \sigma) = skept(F, P, N, \sigma)$.

A further observation is that if we enforce $P$ to be contained in (or equal to) a $\sigma$-extension by an AF $F'$, then $F'$ also enforces positive credulous statuses to arguments in $P$. Further, if we enforce a positive skeptical status to a set of arguments $P$ by an AF $F'$, then $F'$ also enforces a positive credulous status to all arguments in $P$. Recall that we require for enforcing positive skeptical statuses that at least one $\sigma$-extension exists in the modified AF.

**Proposition 2.** *The following inclusions hold for any AF $F = (A, R)$, $P \subseteq A$, $N = A \setminus P$, $x \in \{ns, s\}$, and semantics $\sigma$.*

- $enf^{\sigma}_{x}(F, P) \subseteq cred(F, P, \emptyset, \sigma)$;
- $enf^{\sigma}_{s}(F, P) \subseteq skept(F, \emptyset, N, \sigma)$;
- $skept(F, P, \emptyset, \sigma) \subseteq cred(F, P, \emptyset, \sigma)$.

An important question is which pairs of $(P, N)$ can be enforced credulously or skeptically. For all the semantics we consider, there is always an enforcing AF for credulous status enforcement, while for enforcing skeptical statuses, there always exists a solution under complete, grounded, and preferred semantics.

**Proposition 3.** *Let $F = (A, R)$ be an AF, $P, N \subseteq A$ two disjoint sets, $\sigma \in \{cf, adm, com, grd, prf, stb\}$, and $\sigma' \in \{com, grd, prf\}$. It holds that $cred(F, P, N, \sigma) \neq \emptyset$ and $skept(F, P, N, \sigma') \neq \emptyset$.*

*Proof.* (sketch) For enforcing credulous statuses, it holds that for AF $F' = (A, R')$ with $R' = \{(n, n) \mid n \in N\}$ we have $R' \in cred(F, P, N, \sigma)$, except in the case with $\sigma = stb$ and $N \subset A$. In that case, let $x_0 \in (A \setminus N)$ be an arbitrary but fixed argument. It holds that $R'' \in cred(F, P, N, stb)$ for $F'' = (A, R'')$ with $R'' = \{(x_0, n) \mid n \in N\}$.

For enforcing skeptical statuses under complete, grounded, and preferred semantics, note that $A \setminus N$ is the grounded and unique complete and preferred extension of $F'$. □

Enforcing skeptical statuses is trivial under conflict-free, admissible, and other semantics that always admit the empty extension. To see this, note that $\bigcap \sigma(F) = \emptyset$ if $\emptyset \in \sigma(F)$.

**Proposition 4.** *Let $F = (A, R)$ be an AF, $P, N \subseteq A$ two disjoint sets, and $\sigma$ a semantics. Further, let $\mathcal{R} = skept(F, P, N, \sigma)$. If $\sigma$ admits the empty extension for all AFs, i.e. for all AFs $F'$ we have $\emptyset \in \sigma(F')$, then $\mathcal{R} = 2^{A \times A}$ if $P = \emptyset$, and $\mathcal{R} = \emptyset$ otherwise.*

For stable semantics, the possibility of enforcing skeptical statuses depends on whether we have $N = A$ or not. This is because if $E \in stb(F)$, then $E$ contains at least one argument (except for the trivial AF with $A = \emptyset$). Thus, enforcing a negative skeptical status to all arguments in a framework under stable semantics is not possible. Otherwise, if $N \subset A$, one can construct an AF with only attacks originating from an arbitrary argument in $A \setminus N$ to all arguments in $N$.

**Proposition 5.** *Let $F = (A, R)$ be an AF. It holds that $skept(F, \emptyset, A, stb) = \emptyset$. If $P, N \subseteq A$ are two disjoint sets with $N \subset A$, then $skept(F, P, N, stb)$ is non-empty.*

As is the case for credulous and skeptical acceptance in the static, non-dynamic case, enforcing credulous statuses for admissible sets and complete and preferred semantics coincides. Further, enforcing credulous and skeptical statuses under grounded semantics coincides with enforcing skeptical statuses under complete semantics.

**Proposition 6.** *Let $F = (A, R)$ be an AF, $P, N \subseteq A$ two disjoint sets. It holds that*
$cred(F, P, N, adm) = cred(F, P, N, com) = cred(F, P, N, prf)$
*and*
$cred(F, P, N, grd) = skept(F, P, N, grd) = skept(F, P, N, com)$.

## 4 Complexity

Considering the computational complexity of optimal status enforcement, we focus on the following decision problems. Given an AF $F = (A, R)$, two disjoint sets $P, N \subseteq A$, a semantics $\sigma$, and an integer $k \geq 0$, the question is to decide whether there is an $R' \in \mathit{cred}(F, P, N, \sigma)$ (resp. $R' \in \mathit{skept}(F, P, N, \sigma)$) s.t. $F' = (A, R')$ and $|R \Delta R'| \leq k$, i.e., whether there is an enforcing AF with at most $k$ modifications to the attack structure. We distinguish between the general status enforcement problem and the restricted case where $N = \emptyset$, i.e., without negative status to be enforced. Table 1 summarizes our results.

We begin with status enforcement for conflict-free sets, which corresponds simply to addition or removal of self-attacks on the given sets of arguments.

**Proposition 7.** *Optimal credulous status enforcement for conflict-free sets is polynomial-time solvable.*

Skeptical status enforcement for conflict-free and admissible sets is trivial, since the empty set is always conflict-free and admissible (see also Proposition 4).

Credulous and skeptical status enforcement coincides under grounded semantics, which in turn coincides with non-strict extension enforcement under grounded semantics if $N = \emptyset$ (Proposition 1). For complexity of status enforcement under grounded semantics, the following result is a corollary of a previously established NP-completeness result for extension enforcement [Wallner *et al.*, 2016, Theorem 3].

**Corollary 8.** *Credulous and skeptical status enforcement under grounded semantics is* NP*-complete, even if $N = \emptyset$.*

As a further corollary, skeptical status enforcement under complete semantics is NP-complete (see Proposition 6).

**Corollary 9.** *Skeptical status enforcement under complete semantics is* NP*-complete, even if $N = \emptyset$.*

For credulous status enforcement, it turns out that for the remaining semantics the complexities of the general case and the restricted case with $N = \emptyset$ are presumably different. Intuitively, hardness for the restricted case follows from the fact that checking whether an argument is credulously accepted without modifications is NP-hard for these semantics.

**Proposition 10.** *Credulous status enforcement with $N = \emptyset$ is* NP*-complete under admissible, complete, stable, and preferred semantics.*

*Proof.* (sketch) Let $F = (A, R)$ be an AF and $P \subseteq A$. Membership follows from guessing a new AF $F'$, for each argument $p \in P$ a set of arguments $E_p$ with $p \in E_p$, and checking whether each guessed set $E_p$ is a $\sigma$-extension. Verifying whether a set is a $\sigma$-extension can be checked in polynomial time for all considered semantics except preferred semantics, for which it suffices to check whether the set is admissible.

Hardness follows in all cases from a straightforward reduction from the static credulous acceptance problem for an argument $a$ (an NP-complete problem for all considered semantics [Dimopoulos and Torres, 1996]), and constructing an instance for credulous status enforcement with $P = \{a\}$, and allowing zero modifications ($k = 0$). □

Table 1: Complexity results for status enforcement.

| | $N = \emptyset$ | | $N$ unrestricted | |
|---|---|---|---|---|
| $\sigma$ | credulous | skeptical | credulous | skeptical |
| Conflict-free | in P | trivial | in P | trivial |
| Admissible | NP-c | trivial | $\Sigma_2^P$-c | trivial |
| Stable | NP-c | $\Sigma_2^P$-c | $\Sigma_2^P$-c | $\Sigma_2^P$-c |
| Complete | NP-c | NP-c | $\Sigma_2^P$-c | NP-c |
| Grounded | NP-c | NP-c | NP-c | NP-c |
| Preferred | NP-c | in $\Sigma_3^P$ | $\Sigma_2^P$-c | in $\Sigma_3^P$ |

In contrast, credulous status enforcement under stable, admissible, complete, and preferred semantics is $\Sigma_2^P$-complete if $N \neq \emptyset$. Intuitively, the jump in complexity is due to coNP-completeness of verifying that an argument is not credulously accepted in a given AF. Thus the problem can be decided by a non-deterministic guess of a new attack structure and verifying that all negative statuses are credulously enforced.

**Theorem 11.** *Credulous status enforcement under stable, admissible, complete, and preferred semantics is $\Sigma_2^P$-complete.*

Complexity of skeptical status enforcement under stable semantics is established similarly as for credulous status enforcement under that semantics. Here second-level hardness comes from the fact that verifying skeptical acceptance in a fixed AF is coNP-complete under stable semantics.

**Corollary 12.** *Skeptical status enforcement under stable semantics is $\Sigma_2^P$-complete, even if $N = \emptyset$.*

For skeptical status enforcement under preferred semantics we show membership in $\Sigma_3^P$, which is due to the fact that checking skeptical acceptance in a fixed AF under preferred semantics is $\Pi_2^P$-complete [Dunne and Bench-Capon, 2002].

**Proposition 13.** *Enforcing skeptical acceptance under preferred semantics is in $\Sigma_3^P$.*

## 5 Algorithms

We present declarative encodings of optimal status enforcement for NP variants of the problem, and, based on the encodings, develop counterexample-guided abstraction refinement (CEGAR) [Clarke *et al.*, 2003] algorithms based on maximum satisfiability (MaxSAT) and SAT solvers for optimally solving $\Sigma_2^P$–complete variants of status enforcement. In detail, we provide MaxSAT encodings for $N = \emptyset$ under admissible and stable semantics, and CEGAR for $\Sigma_2^P$ credulous status enforcement for arbitrary $N$ under admissible and stable, as well as skeptical status enforcement under stable semantics. This covers all the non-trivial problem variants considered (except for grounded) by Proposition 6.

For background on MaxSAT, recall that for a Boolean variable $x$, there are two literals, $x$ and $\neg x$. A clause is a disjunction ($\vee$) of literals. A truth assignment $\tau$ is a function from variables to true (1) and false (0). Satisfaction is defined as usual. A Partial MaxSAT (or simply MaxSAT) instance consists of hard clauses $\varphi_h$ and soft clauses $\varphi_s$. An assignment $\tau$ is a solution to a MaxSAT instance $(\varphi_h, \varphi_s)$ if $\tau$ satisfies $\varphi_h$. The cost of $\tau$, $c(\tau)$, is the number of clauses in $\varphi_s$ not satisfied by $\tau$. A solution $\tau$ to a MaxSAT instance $\varphi$ is optimal if $c(\tau) \leq c(\tau')$ for any solution $\tau'$ to $\varphi$.

Let $F = (A, R)$ be an AF, and $P, N \subseteq A$ disjoint sets of arguments whose statuses are to be enforced under a semantics $\sigma$. To encode the credulous status enforcement problem in MaxSAT, we define variables $x_a^p$ for each $a \in A$ and $p \in P$ and $r_{a,b}$ for each $a, b \in A$. Now $\tau(x_a^p) = 1$ corresponds to $a \in E_p$, where $E_p$ is any $\sigma$-extension containing the enforced argument $p$. Likewise, $\tau(r_{a,b}) = 1$ iff $(a, b) \in R'$, where $R'$ is a solution attack structure. For skeptical status enforcement, instead of variables $x_a^p$, we define variables $x_a^n$ for each $a \in A$ and $n \in N$ as indicators for $a \in E_n$, where $E_n$ is any $\sigma$-extension that does not include the argument $n$.

For both credulous and skeptical status enforcement, the soft clauses encode modifications to the attack structure by $\varphi_s = \bigwedge_{a,b \in A} \alpha_{a,b}$, where

$$\alpha_{a,b} = \begin{cases} r_{a,b} & \text{if } (a, b) \in R, \\ \neg r_{a,b} & \text{if } (a, b) \notin R. \end{cases}$$

For credulous status enforcement, the hard clauses are

$$\psi(cred, F, P, N, \sigma) = \bigwedge_{p \in P} \Big( \varphi_\sigma^p(F) \wedge x_p^p \wedge \bigwedge_{n \in N} \neg x_n^p \Big),$$

where $\varphi_\sigma^p(F)$ encodes semantics $\sigma$ so that the $x_a^p$ variables correspond to $E_p \in \sigma(F')$ with $F' = (A, R')$ and $R'$ defined via the attack variables $r_{a,b}$. For conflict-freeness, we have $\varphi_{cf}^p(F) = \bigwedge_{a,b \in A} \Big( \neg r_{a,b} \vee \neg x_a^p \vee \neg x_b^p \Big)$, for admissible sets we use formula $\varphi_{adm}^p(F)$ defined as

$$\varphi_{cf}^p(F) \wedge \bigwedge_{a,b \in A} \Big( (x_a^p \wedge r_{b,a}) \rightarrow \bigvee_{c \in A} (x_c^p \wedge r_{c,b}) \Big),$$

and for stable semantics

$$\varphi_{stb}^p(F) = \varphi_{cf}^p(F) \wedge \bigwedge_{a \in A} \Big( \neg x_a^p \rightarrow \bigvee_{b \in A} (x_b^p \wedge r_{b,a}) \Big).$$

If $N = \emptyset$, each satisfying assignment to $\psi(cred, F, P, \emptyset, \sigma)$ corresponds to an $R' \in cred(F, P, \emptyset, \sigma)$ and vice versa, for $\sigma \in \{adm, com, prf, stb\}$.

Note that the encodings allow for capturing several refinements of the problem. For example, refinements of the optimality criterion, e.g., more elaborate cost models for expressing relative "strength" of, or "confidence" in, attacks can be accounted for by using non-unit weights on the soft clauses; similarly, hard constraints on changes to the attack structure can be enforced by making the corresponding soft clauses hard. Also, enforcing the existence of $\sigma$-extensions attacking certain arguments is possible. Furthermore, e.g., a bounded number of additional arguments can also be allowed.

For $N \neq \emptyset$, due to second-level hardness, we propose a CEGAR approach described as Algorithm 1 which relies on iterative (Max)SAT calls to solve status enforcement optimally. We first apply MaxSAT to $\psi(cred, F, P, N, \sigma)$ to generate a candidate solution (Line 3), which optimally solves the subproblem of enforcing each argument in $P$ to be accepted credulously, at the same time enforcing that each generated witness extension does not include arguments in $N$. We then check whether this candidate is also a solution for the status enforcement problem by asking whether in the modified AF there exists a $\sigma$-extension containing some $n \in N$ via a SAT-check in Line 4. If no such $\sigma$-extension exists, $\tau$ represents an optimal solution to the credulous status enforcement instance. Otherwise we refine the initial formula by excluding the current candidate attack structure and ask for another modification to the AF.

**Algorithm 1** CEGAR-based status enforcement for AF $F = (A, R)$, $P, N \subseteq A$, $\sigma \in \{adm, stb\}$, $M \in \{cred, skept\}$

1: $\chi \leftarrow \psi(M, F, P, N, \sigma)$
2: **while** true **do**
3: $\quad (c, \tau) \leftarrow \text{MAXSAT}(\chi, \varphi_s)$
4: $\quad result \leftarrow \text{SAT}(\text{CHECK}(M, A, \tau, P, N, \sigma))$
5: $\quad$ **if** $result = unsatisfiable$ **then** return $(c, \tau)$
6: $\quad$ **else** $\chi \leftarrow \chi \wedge \text{REFINE}(\tau)$

For checking whether there is a $\sigma$-extension containing an $n \in N$ in the AF $F' = (A, R')$, with $R'$ defined via truth assignment $\tau$, we use formulas $\text{CHECK}(cred, A, \tau, P, N, \sigma) = \phi_\sigma(A, \tau) \wedge \bigvee_{n \in N} x_n$. Formula $\phi_\sigma(A, \tau)$ encodes credulous acceptance in the static case with $\phi_{cf}(A, \tau) = \bigwedge_{\tau(r_{a,b})=1} (\neg x_a \vee \neg x_b)$ for conflict-free sets. Here, variables $x_a$ for $a \in A$ encode that $a$ is in the $\sigma$-extension. For admissible sets and stable extensions we define

$$\phi_{adm}(A, \tau) = \phi_{cf}(A, \tau) \wedge \bigwedge_{\tau(r_{b,a})=1} \Big( x_a \rightarrow \bigvee_{\tau(r_{c,b})=1} x_c \Big);$$

$$\phi_{stb}(A, \tau) = \phi_{cf}(A, \tau) \wedge \bigwedge_{a \in A} \Big( \neg x_a \rightarrow \bigvee_{\tau(r_{b,a})=1} x_b \Big).$$

If a candidate is not successfully verified, we refine formula $\chi$ of Algorithm 1 with

$$\text{REFINE}(\tau) = \neg \Big( \bigwedge_{\tau(r_{a,b})=1} r_{a,b,} \wedge \bigwedge_{\tau(r_{a,b})=0} \neg r_{a,b} \Big).$$

For skeptical status enforcement under stable semantics we slightly adapt Algorithm 1 by using

$$\psi(skept, F, P, N, \sigma) = \bigwedge_{n \in N} \Big( \varphi_\sigma^n(F) \wedge \neg x_n^n \wedge \bigwedge_{p \in P} x_p^n \Big),$$

$\text{CHECK}(skept, A, \tau, P, N, stb) = \phi_{stb}(A, \tau) \wedge \bigvee_{p \in P} \neg x_p$. For the special case $N = \emptyset$, we enforce a stable extension containing $P$ via $\psi$.

## 6 Experiments

We have implemented the MaxSAT encodings and the CEGAR-procedures, obtaining the first system for optimal status enforcement. Here we present an overview of an empirical evaluation of the system.

We generated benchmark instances following essentially a standard model for random directed graphs.[1] For each $|A| \in \{20, 40, \ldots, 200\}$ and $p \in \{0.05, 0.1, \ldots, 0.35\}$[2], we

[1] Based on an initial evaluation, the ICCMA'15 argumentation system competition [Thimm *et al.*, 2016] instances are currently too large in terms of the number of arguments to be suitable as basis for status enforcement benchmarks.

[2] Non-trivial instances arose mainly with $p \leq 0.35$.

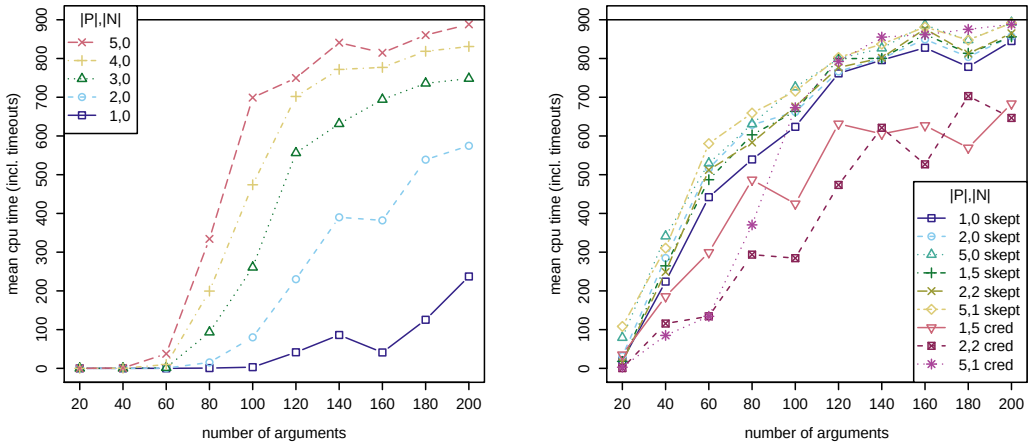


Figure 2: Credulous admissible ($N = \emptyset$) (left), CEGAR on credulous ($N \neq \emptyset$) and skeptical stable (right).

generated ten random AFs with $|A|$ arguments by including individual attacks with probability $p$. For each AF, we randomly picked 5 arguments, of which we enforced $|P| \in \{1, 2, \ldots, 5\}$ positively, and finally picked $|N| \in \{0, 1, 2, 5\}$ arguments from the set $A \setminus P$ to be enforced negatively. We used OpenWBO [Martins *et al.*, 2014] as the MaxSAT solver, and ran the experiments on 2.83-GHz Intel Xeon E5440 4-core nodes with 32-GB RAM and Debian GNU/Linux 8 under 900-second per-instance timeout.

We provide results for two central AF semantics, admissible and stable, for both credulous and skeptical variants of optimal status enforcement. Mean runtimes with timeouts included as 900s are shown in Figure 2 for the NP problems of credulous status enforcement with $|N|$ under admissible semantics (left) and for the $\Sigma_2^P$ skeptical and credulous status enforcement problems under stable semantics (right). In summary, the procedures generally scale up to at least 100 arguments. As expected, increasing the size of $P$ makes the problem harder (left); with $|P| = 2$, the approach still scales to 200 arguments and beyond. For the harder case $|P| = 5$, most (65/70) instances are solved at $|A| = 80$, after which timeouts start increasing linearly, with 68/70 timeouts at $|A| = 200$. For the CEGAR approach (right), credulous status enforcement is easier than skeptical under stable semantics. Interestingly, the empirical hardness of skeptical status enforcement under stable semantics is not significantly affected by different choices for size of $P$ and $N$.

## 7 Related Work

A majority of argumentation system implementations for AFs [Cerutti *et al.*, 2014; Dvořák *et al.*, 2014; Egly *et al.*, 2010; Nofal *et al.*, 2014] focus on the static problems of skeptical and credulous acceptance under different semantics; the ICCMA'15 argumentation system competition [Thimm *et al.*, 2016] also focused on these problems. Status enforcement, as focused on in this work, adopts the notions of skeptical and credulous acceptance into a dynamic setting.

There is recent work focusing on different revision operators for AFs [Baumann, 2012a; Baumann and Brewka, 2015; Bisquert *et al.*, 2013; Booth *et al.*, 2013; Coste-Marquis *et al.*, 2014a; 2014b; Delobelle *et al.*, 2015; Diller *et al.*, 2015; Liao *et al.*, 2011]. Operators giving rise to computational problems concerning dynamics of AFs can be categorized into ones based on *semantical* [Booth *et al.*, 2013; Coste-Marquis *et al.*, 2014a; 2014b; Diller *et al.*, 2015] and *structural* [Baumann, 2012b; Delobelle *et al.*, 2015; Coste-Marquis *et al.*, 2015; Wallner *et al.*, 2016] notions of distance between AFs. Status enforcement falls into the structural distance category. As for the problem statement of optimal status enforcement, [Doutre *et al.*, 2014; Kontarinis *et al.*, 2013] suggest a similar problem setting (though in the latter in terms of subset-minimal instead of optimal structural changes). However, no algorithms for optimal status enforcement are proposed.

Only few systems exist for enforcement problems; for extension enforcement, two have been recently proposed [Coste-Marquis *et al.*, 2015; Wallner *et al.*, 2016]. The closest to this work is [Wallner *et al.*, 2016], with CEGAR-style algorithms for second-level extension enforcement problems.

## 8 Conclusions

We presented properties, complexity analysis, and algorithms for optimal status enforcement as a form of AF dynamics in abstract argumentation. Complexity of optimal status enforcement ranges from polytime-solvable to (at least) completeness for the second level of the polynomial hierarchy. We also proposed and evaluated a first prototype system for optimal status enforcement via employing MaxSAT solvers.

## Acknowledgments

This work has been funded by Academy of Finland under grants 251170 COIN, 276412, and 284591.

## References


[Amgoud and Prade, 2009] L. Amgoud and H. Prade. Using arguments for making and explaining decisions. *Artificial Intelligence*, 173(3-4):413–436, 2009.

[Baroni *et al.*, 2011] P. Baroni, M. Caminada, and M. Giacomin. An introduction to argumentation semantics. *Knowledge Engineering Review*, 26(4):365–410, 2011.

[Baumann and Brewka, 2015] R. Baumann and G. Brewka. AGM meets abstract argumentation: Expansion and revision for Dung frameworks. In *Proc. IJCAI*, pages 2734–2740. AAAI Press, 2015.

[Baumann, 2012a] R. Baumann. Normal and strong expansion equivalence for argumentation frameworks. *Artificial Intelligence*, 193:18–44, 2012.

[Baumann, 2012b] R. Baumann. What does it take to enforce an argument? Minimal change in abstract argumentation. In *Proc. ECAI*, volume 242 of *FAIA*, pages 127–132, 2012.

[Bench-Capon and Dunne, 2007] T.J.M. Bench-Capon and P.E. Dunne. Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10-15):619–641, 2007.

[Bench-Capon *et al.*, 2009] T.J.M. Bench-Capon, H. Prakken, and G. Sartor. Argumentation in legal reasoning. In *Argumentation in Artificial Intelligence*, pages 363–382. Springer, 2009.

[Bisquert *et al.*, 2013] P. Bisquert, C. Cayrol, F. Dupin de Saint-Cyr, and M. Lagasquie-Schiex. Enforcement in argumentation is a kind of update. In *Proc. SUM*, volume 8078 of *LNCS*, pages 30–43. Springer, 2013.

[Booth *et al.*, 2013] R. Booth, S. Kaci, T. Rienstra, and L. W. N. van der Torre. A logical theory about dynamics in abstract argumentation. In *Proc. SUM*, volume 8078 of *LNCS*, pages 148–161. Springer, 2013.

[Cerutti *et al.*, 2014] F. Cerutti, M. Giacomin, M. Vallati, and M. Zanella. An SCC recursive meta-algorithm for computing preferred labellings in abstract argumentation. In *Proc. KR*, pages 42–51. AAAI Press, 2014.

[Clarke *et al.*, 2003] E.M. Clarke, O. Grumberg, S. Jha, Y. Lu, and H. Veith. Counterexample-guided abstraction refinement for symbolic model checking. *Journal of the ACM*, 50(5):752–794, 2003.

[Coste-Marquis *et al.*, 2014a] S. Coste-Marquis, S. Konieczny, J. Mailly, and P. Marquis. On the revision of argumentation systems: Minimal change of arguments statuses. In *Proc. KR*, pages 52–61. AAAI Press, 2014.

[Coste-Marquis *et al.*, 2014b] S. Coste-Marquis, S. Konieczny, J. Mailly, and P. Marquis. A translation-based approach for revision of argumentation frameworks. In *Proc. JELIA*, volume 8761 of *LNCS*, pages 397–411. Springer, 2014.

[Coste-Marquis *et al.*, 2015] S. Coste-Marquis, S. Konieczny, J. Mailly, and P. Marquis. Extension enforcement in abstract argumentation as an optimization problem. In *Proc. IJCAI*, pages 2876–2882. AAAI Press, 2015.

[Delobelle *et al.*, 2015] J. Delobelle, S. Konieczny, and S. Vesic. On the aggregation of argumentation frameworks. In *Proc. IJCAI*, pages 2911–2917. AAAI Press, 2015.

[Diller *et al.*, 2015] M. Diller, A. Haret, T. Linsbichler, S. Rümmele, and S. Woltran. An extension-based approach to belief revision in abstract argumentation. In *Proc. IJCAI*, pages 2926–2932. AAAI Press, 2015.

[Dimopoulos and Torres, 1996] Y. Dimopoulos and A. Torres. Graph theoretical structures in logic programs and default theories. *Theoretical Computer Science*, 170(1-2):209–244, 1996.

[Doutre *et al.*, 2014] S. Doutre, A. Herzig, and L. Perrussel. A dynamic logic framework for abstract argumentation. In *Proc. KR*, pages 62–71. AAAI Press, 2014.

[Dung, 1995] P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–358, 1995.

[Dunne and Bench-Capon, 2002] P.E. Dunne and T.J.M. Bench-Capon. Coherence in finite argument systems. *Artificial Intelligence*, 141(1/2):187–203, 2002.

[Dvořák *et al.*, 2014] W. Dvořák, M. Järvisalo, J.P. Wallner, and S. Woltran. Complexity-sensitive decision procedures for abstract argumentation. *Artificial Intelligence*, 206:53–78, 2014.

[Egly *et al.*, 2010] U. Egly, S.A. Gaggl, and S. Woltran. Answer-set programming encodings for argumentation frameworks. *Argument & Computation*, 1(2):147–177, 2010.

[Kontarinis *et al.*, 2013] D. Kontarinis, E. Bonzon, N. Maudet, A. Perotti, L. van der Torre, and S. Villata. Rewriting rules for the computation of goal-oriented changes in an argumentation system. In *Proc. CLIMA*, volume 8143 of *LNCS*, pages 51–68. Springer, 2013.

[Liao *et al.*, 2011] B.S. Liao, L. Jin, and R. C. Koons. Dynamics of argumentation systems: a division-based method. *Artificial Intelligence*, 175(11):1790–1814, 2011.

[Martins *et al.*, 2014] R. Martins, V.M. Manquinho, and I. Lynce. Open-WBO: A modular MaxSAT solver. In *Proc. SAT*, volume 8561 of *LNCS*, pages 438–445, 2014.

[McBurney *et al.*, 2012] P. McBurney, S. Parsons, and I. Rahwan, editors. *ArgMAS 2011 Revised Selected Papers*, volume 7543 of *LNCS*. Springer, 2012.

[Nofal *et al.*, 2014] S. Nofal, K. Atkinson, and P. E. Dunne. Algorithms for decision problems in argument systems under preferred semantics. *Artificial Intelligence*, 207:23–51, 2014.

[Thimm *et al.*, 2016] M. Thimm, S. Villata, F. Cerutti, N. Oren, H. Strass, and M. Vallati. Summary report of the first international competition on computational models of argumentation. *AI Magazine*, 37(1):102–104, 2016.

[Wallner *et al.*, 2016] J.P. Wallner, A. Niskanen, and M. Järvisalo. Complexity results and algorithms for extension enforcement in abstract argumentation. In *Proc. AAAI*. AAAI Press, 2016.

## Appendix

*Proof of Theorem 11 and Corollary 12.* For membership, recall that verifying credulous acceptance of an argument under admissible, complete, preferred, and stable semantics is in NP. This means that verifying whether an argument is not credulously accepted is in coNP. Verifying whether an argument is skeptically accepted in an AF under stable semantics is in coNP. This implies that verifying whether an argument is not skeptically accepted under stable semantics is in NP. To see membership for all problems mentioned in the theorem and corollary, consider a non-deterministic guess of a modified attack structure with at most $k$ changes. Further, for each subproblem in NP (i.e., credulous acceptance or non skeptical acceptance), we guess corresponding sets of arguments of the new framework and verify whether they are $\sigma$-extensions. Finally, we verify the coNP subproblems (non credulous acceptance and skeptical acceptance) with a coNP oracle. This concludes membership in $\Sigma_2^P$.

For hardness, we use a reduction from the $\Sigma_2^P$-complete problem of deciding whether a given quantified Boolean formula $\phi = \exists X \forall Y \psi$ in prenex normal form is valid. W.l.o.g. we assume that $\psi$ is in disjunctive normal form. Let $C$ be the set of conjunctions in $\psi$. Further, we define $\overline{Z} = \{\overline{z} \mid z \in Z\}$ as a renaming of elements in a set. Also, let $n = |X|$, and $D = \{d_i^x \mid x \in X, 1 \leq i \leq n+1\}$. We construct an AF $F = (A, R)$ with

$$
\begin{aligned}
A =& X \cup \overline{X} \cup Y \cup \overline{Y} \cup C \cup D \cup \{q, q', \overline{q}\} \\
R =& \{(x, \overline{x}), (\overline{x}, x), (x, x), (\overline{x}, \overline{x}) \mid x \in X\} \cup \\
& \{(y, \overline{y}), (\overline{y}, y) \mid y \in Y\} \cup \\
& \{(z, c) \mid z \in X \cup Y, c \in C, \neg z \in c\} \cup \\
& \{(\overline{z}, c) \mid z \in X \cup Y, c \in C, z \in c\} \cup \\
& \{(c, \overline{q}) \mid c \in C\} \cup \\
& \{(x, d_i^x), (\overline{x}, d_i^x), (d_i^x, q') \mid x \in X, 1 \leq i \leq n+1\} \cup \\
& \{(\overline{q}, q)\}.
\end{aligned}
$$

W.l.o.g. we assume that $X$, $\overline{X}$, $Y$, $\overline{Y}$, $C$, $D$, and $\{q, q', \overline{q}\}$ are disjoint sets. An illustration of the reduction is shown in Figure 3 with conjunctions $c = \neg x \wedge \neg y$ and $c' = x \wedge y$.

Let $\hat{F} = \{F' \mid F' = (A, R'), |R \Delta R'| \leq n\}$. We show that the following statements are equivalent.

1. $\exists F' \in \hat{F}$ s.t. $\exists E \in adm(F')$ with $q, q' \in E$ and $\forall E' \in adm(F')$ we have $\overline{q} \notin E'$;
2. $\exists F' \in \hat{F}$ s.t. $\exists E \in stb(F')$ with $q, q' \in E$ and $\forall E' \in stb(F')$ we have $\overline{q} \notin E'$;

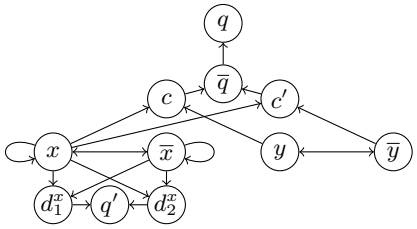


Figure 3: Illustration of hardness proof of Theorem 11.

3. $\exists F' \in \hat{F}$ s.t. $stb(F') \neq \emptyset$ and $\forall E' \in stb(F')$ we have $q, q' \in E'$;
4. $\phi = \exists X \forall Y \psi(X, Y)$ is valid.

We begin with showing that the fourth item implies each of the other three. Assume $\phi$ is valid. Then there exists a truth assignment $\tau$ on $X$ s.t. for all truth assignments $\tau'$ that assign the same value as $\tau$ to variables in $X$ we have $\tau' \models \psi$. Let $\tau$ and $\tau'$ be such a truth assignments. Let

$$
\begin{aligned}
X' &= \{x \in X \mid \tau'(x) = 1\}, \\
\overline{X}' &= \{\overline{x} \in \overline{X} \mid \tau'(x) = 0\}, \\
Y' &= \{y \in Y \mid \tau'(y) = 1\},
\end{aligned}
$$

and

$$
\overline{Y}' = \{\overline{y} \in \overline{Y} \mid \tau'(y) = 0\}.
$$

Further, let $R' = R \setminus \{(z, z) \mid z \in X' \cup \overline{X}'\}$, i.e., we remove self-attacks from arguments in $X'$ resp. from $\overline{x}$ where $x \notin X'$. It follows that for $F' = (A, R')$ we have $F' \in \hat{F}$. We now construct a stable extension $E \in stb(F')$ s.t. $q, q' \in E$. This implies the first condition of the first three items in the list above.

Consider the arguments in $C$ that are defended by $X' \cup \overline{X}' \cup Y' \cup \overline{Y}' = E'$, i.e., $\mathcal{F}_{F'}(E') \cap C = C'$. From the assumption that $\tau' \models \psi$ we can conclude that $C' \neq \emptyset$, since

$$
\begin{aligned}
& c \in C' && (1)\\
\text{iff } & c \in (\mathcal{F}_{F'}(E') \cap C) && (2)\\
\text{iff } & \forall (b, c) \in R' \exists a \in E' \text{ s.t. } (a, b) \in R' && (3)\\
\text{iff } & z \in c \text{ implies } z \in E' \text{ and} && (4)\\
& \neg z \in c \text{ implies } \overline{z} \in E' && (5)\\
\text{iff } & z \in c \text{ implies } \tau'(z) = 1 \text{ and} && (6)\\
& \neg z \in c \text{ implies } \tau'(z) = 0 && (7)\\
\text{iff } & \tau' \models c. && (8)
\end{aligned}
$$

By assumption $\exists c \in C$ s.t. $\tau' \models c$, and thus $C' \neq \emptyset$. It immediately follows from construction of $F'$ and previous observations that $\{q, q'\} \cup E' \cup C'$ is stable in $F'$.

Next, we show that $\forall E \in adm(F')$ we have $\overline{q} \notin E$. This, together with previous results, implies that the fourth item in the list implies all other three. Suppose $\exists E \in adm(F')$ s.t. $\overline{q} \in E$. Then $\overline{q} \in \mathcal{F}_{F'}(E)$ and $C \cap E = \emptyset$.

$$
\begin{aligned}
& \exists E \in adm(F') \text{ s.t. } \overline{q} \in E' && (9)\\
\text{only if } & \forall c \in C \exists b \in E \text{ s.t. } (b, c) \in R' && (10)\\
\text{iff } & \forall c \in C \exists z \in X \cup Y \text{ s.t.} && (11)\\
& z \in c \text{ implies } \overline{z} \in E && (12)\\
& \neg z \in c \text{ implies } z \in E && (13)\\
\text{only if } & \exists \tau'' \text{ s.t. } \forall x \in X && (14)\\
& \tau''(x) = 1 \text{ implies } x \in E, && (15)\\
& \tau''(x) = 0 \text{ implies } \overline{x} \in E \text{ and} && (16)\\
& \forall c \in C, \tau'' \not\models c. && (17)
\end{aligned}
$$

Notice that $\tau''$ is compatible with $\tau$ in the sense that $\forall x \in X$ we have $\tau''(x) = 1$ implies $\tau(x) = 1$ and $\tau''(x) = 0$. Thus

the partial assignment $\tau''$ can be completed to one that assigns the same variables to $X$ as $\tau$ and does not satisfy $\psi$. But this implies $\phi$ is not valid, which is a contradiction.

The previous proof directly implies that the fourth item implies the first and second. For the third one, it is immediate that there exists a stable extension in $F'$. Suppose there is a stable extension $T$ in $F'$ s.t. $q$ is not contained in it. It is immediate that $T$ is not stable, since no admissible set in $F'$ contains the only attacker of $q$, namely $\overline{q}$. Finally, to see that there is no stable extension $T'$ with $q' \notin T'$, by previous proof we know that each $d_i^x$ is attacked by each stable extension. Thus, if $q'$ is not contained in $T'$, then no argument in $T'$ attacks $q'$, and therefore $T'$ is not stable in $F'$.

We now proceed to the other direction of the hardness proof. We show that the first three items individually imply that the fourth item in the list holds. Assume that $F' \in \hat{F}$ with $F' = (A, R')$. First note that all three items imply condition (i) $\exists E \in adm(F')$ s.t. $q, q' \in E$. Consider $R_i = \{(a, d_i^x) \mid a \in A, x \in X\} \cup \{(d_i^x, q') \mid x \in X\}$, i.e., the set of attacks that originate from arguments in $A$ to arguments $d_i^x$ and from $d_i^x$ to $q'$ for a fixed $i$. It follows that $\exists i$ s.t. $|(R \Delta R') \cap R_i| = 0$, i.e., there exists an index $i$ s.t. the attack relation $R_i$ is unchanged in $R'$ compared to $R$. This holds since there are $n+1$ attack structures $R_i$ but only at most $n$ modifications in $R'$ compared to $R$.

Since $q' \in E$ by assumption (i), it follows that for each $x \in X$ we have $x \in E$ or $\overline{x} \in E$, since these are the only attackers of $d_i^x$ which attacks $q'$ (admissibility of $E$). Since we have at most $n$ changes in $R \Delta R'$, it immediately follows that for each $x \in X$ one of the following statements holds:

- $(x, x) \notin R'$ and $(\overline{x}, \overline{x}) \in R'$, or
- $(x, x) \in R'$ and $(\overline{x}, \overline{x}) \notin R'$.

This also implies that $(R \Delta R') \subseteq \{(x, x), (\overline{x}, \overline{x} \mid x \in X\}$. Summarizing, it holds that for each $x \in X$ we have removed exactly one self-attack from either $x$ or $\overline{x}$ in $R'$ compared to $R$ and these are the only changes.

From this and assuming one of the first three items in the list it follows that $\forall E' \in adm(F')$ we have $\overline{q} \notin E'$ (referred to as condition (ii)), since $q \in E'$ and $(\overline{q}, q) \in R'$. Let $\tau$ be a truth assignment s.t. $\tau(x) = 1$ iff $(x, x) \notin R'$. Now suppose that $\tau \not\models \psi$ (i.e., $\phi$ is not valid). Let $E'' = (E \cap (X \cup \overline{X})) \cup \{y \in Y \mid \tau(y) = 1\} \cup \{\overline{y} \in \overline{Y} \mid \tau(y) = 0\} \cup \{\overline{q}, q'\}$. It follows that $E'' \in cf(F')$ (all self-attacks on the corresponding $x$ and $\overline{x}$ are not present in $F'$). It is immediate that $E''$ attacks all arguments in $A \setminus E''$ in $F'$ except for arguments $C$.

$$
\begin{aligned}
& E'' \text{ attacks all arguments in } C && (18)\\
\text{iff } & \forall c \in C \exists a \in E'' \text{ s.t. } (a, c) \in R' && (19)\\
\text{iff } & \forall c \in C \exists z \in X \cup Y && (20)\\
& z \in c \text{ implies } \overline{z} \in E'' \text{ and} && (21)\\
& \neg z \in c \text{ implies } z \in E'' && (22)\\
\text{iff } & \forall c \in C \exists z \in X \cup Y && (23)\\
& z \in C \text{ implies } \tau(z) = 0 && (24)\\
& \neg z \in C \text{ implies } \tau(z) = 1 && (25)\\
\text{iff } & \tau \not\models \psi. && (26)
\end{aligned}
$$

This implies that $E'' \in stb(F')$ which is a contradiction (see condition (ii)). Therefore $\tau \models \psi$ and thus $\phi$ is valid. □