Controllability of Control Argumentation Frameworks

Andreas Niskanen[1] , **Daniel Neugebauer**[2] and **Matti Järvisalo**[1]
[1]HIIT, Department of Computer Science, University of Helsinki, Finland
[2]Institut für Informatik, Heinrich-Heine Universität Düsseldorf, Germany

Abstract

Control argumentation frameworks (CAFs) allow for modeling uncertainties inherent in various argumentative settings. We establish a complete computational complexity map of the central computational problem of controllability in CAFs for five key semantics. We also develop Boolean satisfiability based counterexample-guided abstraction refinement algorithms and direct encodings of controllability as quantified Boolean formulas, and empirically evaluate their scalability on a range of NP-hard variants of controllability.

1 Introduction

Argumentation is a vibrant area of artificial intelligence research. Various argumentative settings intrinsically involve uncertainties about the existence and relationships between arguments due to different forms of dynamics underlying argumentation [Doutre and Mailly, 2018]. Extending the central formal model of Dung's abstract argumentation frameworks (AFs) [Dung, 1995] to such settings, control argumentation frameworks (CAFs) [Dimopoulos *et al.*, 2018] were recently proposed as a unifying framework to capture uncertainties about the existence of arguments and attacks between arguments. Syntactically, CAFs generalize AFs by allowing for specifying *control*, *uncertain*, and *fixed* arguments, which give rise to uncertainties about the attack relation. CAFs have already proved to be a natural way to approaching argumentation-based negotiation in settings where knowledge of the opponent's profile is incomplete [Dimopoulos *et al.*, 2019] as a suitable formalism for opponent modeling, i.e., modeling uncertainties about an opponent profile.

Lifting the well-studied acceptance problems in AFs, controllability is an analogous central computational problem in CAFs. Reminiscent of particular computational problems in the study of argumentation dynamics [Coste-Marquis *et al.*, 2007; Cayrol *et al.*, 2010; Booth *et al.*, 2013; Coste-Marquis *et al.*, 2014; Doutre *et al.*, 2014; de Saint-Cyr *et al.*, 2016; Diller *et al.*, 2018], in particular non-strict extension enforcement under normal expansions [Baumann and Brewka, 2010; Baumann, 2012; Coste-Marquis *et al.*, 2015], credulous controllability of a set T of arguments refers to being able to use the control arguments of a CAF to ensure that T is contained in some extension of every AF (completion) resulting from instantiating the uncertain part of the CAF. Analogously, skeptical controllability asks if T can be ensured to be contained in all extensions of every completion of the CAF.

Unlike for acceptance [Dvorák and Dunne, 2018] and extension enforcement [Wallner *et al.*, 2017] in AFs, the complexity of variants of controllability in CAFs under central argumentation semantics has not been established, apart from (bounds for) specific semantics [Dimopoulos *et al.*, 2018]. Furthermore, while controllability in CAFs has already been harnessed computationally for argumentation-based negotiation [Dimopoulos *et al.*, 2019], there is significant room for developing computational approaches to controllability and evaluating the empirical hardness of controllability.

In this paper, we establish a complete complexity map of skeptical and credulous controllability under several central argumentation semantics for both general CAFs and so-called simplified CAFs. In particular, we establish tight connections between controllability in CAFs and acceptance in incomplete argumentation frameworks (IAFs) [Baumeister *et al.*, 2018b], and apply these observations to obtain tight complexity results for CAFs via generalizing recent results for acceptance in IAFs [Baumeister *et al.*, 2018a]. On the algorithmic side, we present both direct quantified Boolean satisfiability (QBF) encodings—improving on [Dimopoulos *et al.*, 2018] in terms of compactness of the encodings—and Boolean satisfiability (SAT) based counterexample-guided abstraction refinement procedures for controllability, and provide first results on the empirical hardness of reasoning about controllability with the approaches.

2 Argumentation Frameworks

We recall standard argumentation frameworks [Dung, 1995] and standard AF semantics [Baroni *et al.*, 2018], incomplete AFs [Baumeister *et al.*, 2018b] as useful for our results, and control AFs [Dimopoulos *et al.*, 2018] which we focus on.

Definition 1. An argumentation framework (AF) *is a pair $F = (A, R)$ with a non-empty finite set of arguments A and an attack relation $R \subseteq A \times A$. An argument $a \in A$ is defended by a set $S \subseteq A$ iff for each $(b, a) \in R$ there is a $c \in S$ with $(c, b) \in R$. The characteristic function of F for a set $S \subseteq A$ is $\mathcal{F}_F(S) = \{a \in A \mid a \text{ is defended by } S\}$. The range* *of* S *is* $S_F^+ = S \cup \{a \in A \mid (b,a) \in R, b \in S\}$.

Semantics σ map each AF to a collection of (jointly accepted) subsets of arguments. We consider the admissible (adm), complete (com), preferred (prf), stable (stb), and grounded (grd) semantics.

Definition 2. *Let $F = (A, R)$ be an AF. A set $S \subseteq A$ is conflict-free in F, denoted by $S \in cf(F)$, if there are no $x, y \in S$ with $(x, y) \in R$. For $S \in cf(F)$, it holds that*

- $S \in adm(F)$ *if* $S \subseteq \mathcal{F}_F(S)$*,*
- $S \in com(F)$ *if* $S = \mathcal{F}_F(S)$*,*
- $S \in prf(F)$ *if S is a subset-maximal admissible set,*
- $S \in stb(F)$ *if* $S_F^+ = A$*, and*
- $S \in grd(F)$ *if S is the subset-minimal complete set.*

If $S \in \sigma(F)$ for a semantics σ, we call S a σ-extension. An argument $a \in A$ is skeptically accepted if $a \in \bigcap \sigma(F)$, and credulously accepted if $a \in \bigcup \sigma(F)$.

Incomplete argumentation frameworks [Baumeister *et al.*, 2018b; Baumeister *et al.*, 2018a] allow for representing unquantified uncertainty in an AF via uncertain arguments and attacks. We denote the restriction of an attack relation R to a set of arguments A' by $R|_{A'} = R \cap (A' \times A')$.

Definition 3. An incomplete argumentation framework (IAF) *is a tuple $\mathcal{I} = (A, A^?, R, R^?)$, where $R, R^? \subseteq (A \cup A^?) \times (A \cup A^?)$ with $A \cap A^? = \emptyset$ and $R \cap R^? = \emptyset$. The set $A^?$ consists of uncertain arguments and the set $R^?$ of uncertain attacks. An IAF is (purely)* argument-incomplete *if $R^? = \emptyset$. A* completion *of $\mathcal{I}$ is an AF $F^* = (A^*, R^*)$ with $A \subseteq A^* \subseteq (A \cup A^?)$ and $R|_{A^*} \subseteq R^* \subseteq (R \cup R^?)|_{A^*}$. An argument $a \in A$ is* possibly *skeptically (credulously) accepted in $\mathcal{I}$ if it is skeptically (credulously) accepted in a completion of $\mathcal{I}$; and* necessarily *skeptically (credulously) accepted in $\mathcal{I}$ if it is skeptically (credulously) accepted in all completions of $\mathcal{I}$.*

A control argumentation framework [Dimopoulos *et al.*, 2018] consists of three parts: the *fixed*, the *uncertain*, and the *control* part. Control AFs allow for modeling the problem of finding a subset of the control part such that no matter what the state of the uncertain part, a certain target is reached.

Definition 4. A control argumentation framework (CAF) *is a triple $\mathcal{C} = (F, C, U)$, where*

- $F = (A_F, R_F)$ *is the* fixed *part with* $R_F \subseteq (A_F \cup A_U) \times (A_F \cup A_U)$*,*
- $U = (A_U, R_U \cup R_U^{\leftrightarrow})$ *is the* uncertain *part with* $R_U, R_U^{\leftrightarrow} \subseteq (A_F \cup A_U) \times (A_F \cup A_U)$*, where $R_U^{\leftrightarrow}$ is symmetric and irreflexive,*
- $C = (A_C, R_C)$ *is the* control *part with* $R_C \subseteq (A_C \times (A_F \cup A_U \cup A_C)) \cup ((A_F \cup A_U \cup A_C) \times A_C)$*,*

A_F, A_U, and A_C are pairwise disjoint sets of arguments, and R_F, R_U, $R_U^{\leftrightarrow}$ and R_C are pairwise disjoint sets of attacks. A subset $A_{conf} \subseteq A_C$ is a control configuration. *A* completion *of CAF $\mathcal{C}$ is an AF $F^* = (A^*, R^*)$ with $A^* = A_F \cup A_C \cup A_U^*$, where $A_U^* \subseteq A_U$ and $(R_F \cup R_C)|_{A^*} \subseteq R^* \subseteq (R_F \cup R_C \cup R_U \cup R_U^{\leftrightarrow})|_{A^*}$ satisfying $(a, b) \in R^*$ or $(b, a) \in R^*$ for all $(a, b) \in R_U^{\leftrightarrow}$.*

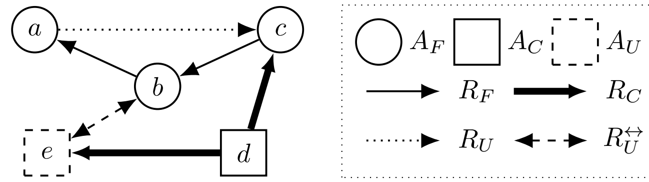

Figure 1: An example CAF (F, C, U) with A_F=$\{a, b, c\}$, A_C=$\{d\}$, A_U=$\{e\}$, R_F=$\{(b, a), (c, b)\}$, R_C=$\{(d, c), (d, e)\}$, R_U=$\{(a, c)\}$, and $R_U^{\leftrightarrow}$=$\{(b, e), (e, b)\}$.

Note that we assume symmetry and irreflexivity to $R_U^{\leftrightarrow}$ without loss of generality.

We focus on the *controllability* problem over control AFs, which asks to decide if there is a *control configuration* such that for all completions of the control AF, a subset of arguments is credulously or skeptically accepted.

Definition 5. *Let $\mathcal{C} = (F, C, U)$, $T \subseteq A_F$, and let σ be an AF semantics. The CAF $\mathcal{C}$ is skeptically (credulously) controllable wrt. T and σ if, for some control configuration $A_{conf} \subseteq A_C$, T is included in every (some) σ-extension of every completion of $\mathcal{C}_{conf} = (F, C_{conf}, U)$, where $C_{conf} = (A_{conf}, R_C|_{A_F \cup A_U \cup A_{conf}})$.*

The CAF in Figure 1 is not credulously controllable w.r.t. $\{a\}$ for any of the five semantics σ we consider, since no matter whether the control argument d is included, there are completions that do not have a in any σ-extension. The CAF is, however, skeptically controllable w.r.t. $\{a\}$ under stb using $A_{conf} = \emptyset$ since each completion either has a unique stable extension containing a or no stable extension at all.

3 Complexity of Controllability

We establish tight complexity results for credulous and skeptical controllability as summarized in Table 1. We assume knowledge of the complexity classes in the polynomial hierarchy (Σ_0^{P}=Π_0^{P}=P, $\Sigma_{i+1}^{\mathsf{P}}$=$\mathsf{NP}^{\Sigma_i^{\mathsf{P}}}$, and Π_{i+1}^{P}=$\mathsf{coNP}^{\Sigma_i^{\mathsf{P}}}$) and the notions of hardness and completeness [Arora and Barak, 2009]. We start from simplified CAFs [Dimopoulos *et al.*, 2018], which are CAFs with no uncertain part, i.e., $A_U = R_U = R_U^{\leftrightarrow} = \emptyset$, denoted by $(F, C, \emptyset)$. First, we establish that the restriction of the controllability problem in simplified CAFs to singleton target sets (called the *conclusion* problem) is equivalent to possible acceptance in argument-incomplete IAFs. Control arguments in such CAFs directly correspond to the IAFs' uncertain arguments, so that both notions of *completion* (and all derived problems) coincide.

Lemma 1. *Given an IAF $\mathcal{I} = (A, A^?, R, \emptyset)$, argument $q \in A$, and semantics σ, q is possibly skeptically (credulously) accepted under σ iff the CAF $\mathcal{C} = (F, C, \emptyset)$, where $A_F = A$, $A_C = A^?$, $R_F = R|_A$ and $R_C = R \setminus (R|_A)$, is skeptically (credulously) controllable with respect to $\{q\}$ and σ.*

This yields hardness of credulous controllability for simplified CAFs via a reduction from possible credulous acceptance in argument-incomplete IAFs. We extend the earlier NP-completeness result for controllability under stb [Dimopoulos *et al.*, 2018] to all five semantics that we consider here.

Theorem 2. *Credulous controllability for simplified CAFs is* NP*-complete under $\sigma \in \{adm, com, prf, stb, grd\}$.*

	General CAFs		Simplified CAFs	
Semantics	Credulous	Skeptical	Credulous	Skeptical
admissible	Σ_3^{P}-c	trivial	NP-c	trivial
complete	Σ_3^{P}-c	Σ_2^{P}-c	NP-c	NP-c
preferred	Σ_3^{P}-c	Σ_3^{P}-c	NP-c	Σ_3^{P}-c
stable	Σ_3^{P}-c	Σ_2^{P}-c	NP-c	Σ_2^{P}-c
grounded	Σ_2^{P}-c	Σ_2^{P}-c	NP-c	NP-c

Table 1: Computational complexity of controllability

Proof. Let $\mathcal{C} = (F, C, \emptyset)$ be a simplified CAF and $T \subseteq A_F$ be a set of target arguments. NP-membership is by nondeterministically guessing a control configuration $A_{conf} \subseteq A_C$ and a σ-extension E of the AF $(A_F \cup A_{conf}, (R_F \cup R_C)|_{A_F \cup A_{conf}})$, and checking in polynomial time if $T \subseteq E$ and E is a σ-extension, with the exception of prf; here, however, it suffices to check admissibility instead of preferredness of E as every admissible set is a subset of some preferred extension. Hardness follows from Lemma 1 by reduction from possible credulous acceptance under σ in argument-incomplete IAFs, which is NP-complete [Baumeister *et al.*, 2018a]. □

For AFs, credulous acceptance under grd coincides with skeptical acceptance under both grd and com. Thus skeptical controllability under grd and com are equivalent to credulous controllability under grd. This yields NP-completeness also for skeptical controllability in simplified CAFs.

Corollary 3. *The skeptical controllability problem for simplified CAFs is* NP*-complete under* $\sigma \in \{com, grd\}$.

For skeptical controllability under stb, we derive Σ_2^{P}-completeness, strengthening the NP-hardness result of [Dimopoulos *et al.*, 2018].

Theorem 4. *Skeptical controllability for simplified CAFs is* Σ_2^{P}*-complete under* stb.

Proof. (Sketch.) As possible skeptical acceptance in argument-incomplete IAFs is Σ_2^{P}-complete under stb [Baumeister *et al.*, 2018a], a reduction by Lemma 1 establishes Σ_2^{P}-hardness for skeptical controllability under stb. For membership, guess a control configuration and use an NP-oracle to check whether every argument in T is skeptically accepted under stb [Dimopoulos and Torres, 1996]. □

Similarly, as skeptical acceptance in AFs under prf is Π_2^{P}-complete [Dunne and Bench-Capon, 2002] and possible skeptical acceptance in argument-incomplete IAFs under prf is Σ_3^{P}-complete [Baumeister *et al.*, 2018a], we have Σ_3^{P}-completeness of skeptical controllability in simplified CAFs under prf via an analogous proof.

Theorem 5. *Skeptical controllability for simplified CAFs is* Σ_3^{P}*-complete under* prf.

Finally, we note that skeptical acceptance for adm is trivial, since the empty set is always admissible, so no argument can ever be in all adm sets.

We turn to skeptical controllability for general CAFs. We establish Σ_2^{P}-completeness of skeptical controllability for general CAFs under stb, strengthening the membership result of [Dimopoulos *et al.*, 2018].

Theorem 6. *Skeptical controllability for general CAFs is* Σ_2^{P}*-complete under* stb.

Proof. Let (F, C, U) be a CAF and $T \subseteq A_F$ a target. Σ_2^{P}-membership follows from the quantifier representation of the problem: Is there a control configuration such that, for all completions $F^* = (A^*, R^*)$ and for all sets of arguments $S \subseteq A^*$ with $T \nsubseteq S$, it holds that $S \notin stb(F^*)$? Each of the quantifiers is polynomial-length-bounded and the condition $S \notin stb(F^*)$ is in P. Note that this representation checks whether every set S *not* containing T is *not* stable, instead of checking whether every stable extension contains T, which is clearly equivalent. Hardness follows from hardness in simplified CAFs. □

Similarly, for prf we establish Σ_3^{P}-completeness.

Theorem 7. *Skeptical controllability for general CAFs is* Σ_3^{P}*-complete under* prf.

Proof. Let (F, C, U) be a CAF and $T \subseteq A_F$ a target. Σ_3^{P}-membership follows from the quantifier representation of the problem: Is there a control configuration such that, for all completions $F^* = (A^*, R^*)$ and for all sets of arguments $S \subseteq A^*$ with $T \nsubseteq S$, it holds that $S \notin adm(F^*)$ or there exists a set $S' \supset S$ with $S' \in adm(F^*)$? Each of the quantifiers is polynomial-length-bounded and the conditions $S \notin adm(F^*)$ and $S' \in adm(F^*)$ are in P. Hardness follows from hardness in simplified CAFs. □

Skeptical controllability under com and grd coincides with credulous controllability under grd. Hence it suffices to consider credulous controllability among the three. We make use of earlier proofs of hardness in the context of IAFs [Baumeister *et al.*, 2018a].

Intuitively, we modify the general reduction from QSAT to IAFs [Baumeister *et al.*, 2018a, Definition 9] by incorporating the control part of CAFs to simulate a partial truth assignment, similarly as uncertain arguments of IAFs are used to simulate an assignment in the reduction to IAFs. This gives a reduction from $\exists X \forall Y \exists Z$-CNF-SAT to CAFs. As in [Baumeister *et al.*, 2018a], we observe a crucial connection between the total assignments over the CNF formula and the σ-extensions of the completions of the corresponding CAF under $\sigma \in \{adm, com, prf, stb\}$: a truth assignment satisfies the formula iff there is a completion with a σ-extension containing a particular query argument. This reduction (see Figure 2) yields Σ_3^{P}-completeness of credulous controllability in general CAFs under $\sigma \in \{adm, com, prf, stb\}$. Thus we close the gap from [Dimopoulos *et al.*, 2018] between Σ_3^{P}-membership and Σ_2^{P}-hardness under stb.

Theorem 8. *Credulous controllability for general CAFs is* Σ_3^{P}*-complete under* $\sigma \in \{adm, com, prf, stb\}$.

Proof. (Sketch.) Let (F, C, U) be a CAF and $T \subseteq A_F$ a target. Σ_3^{P}-membership is by the quantifier representation: Is there a control configuration s.t. for all completions $F^* = (A^*, R^*)$ there is a set $T \subseteq E \subseteq A^*$ such that $E \in \sigma(F^*)$? Each of the quantifiers is polynomial-length-bounded and the condition $E \in \sigma(F^*)$ is in P for all considered semantics except prf, which we can again exchange for adm. Hardness follows by a reduction illustrated in Figure 2. □

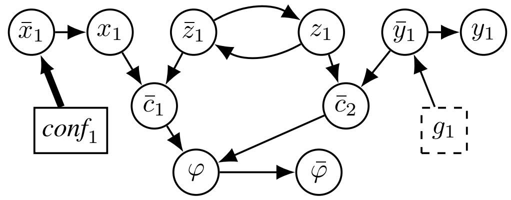

Figure 2: CAF created by the reduction of Theorem 8 for the CNF formula $\varphi = c_1 \wedge c_2$ with clauses $c_1 = x_1 \vee \neg z_1$ and $c_2 = z_1 \vee \neg y_1$. We have $A_U = \{g_1\}$, $A_C = \{conf_1\}$, $R_C = \{(conf_1, \bar{x}_1)\}$, and all other arguments and attacks are in F. There is a control configuration for (F, C, U) such that for all completions, $\{\varphi\}$ is a subset of some σ-extension iff $(\exists \tau_{\{x_1\}})(\forall \tau_{\{y_1\}})(\exists \tau_{\{z_1\}})[c_1 \wedge c_2] = \texttt{true}$. For Theorem 9, we have $Z = \emptyset$, and there is a control configuration for (F, C, U) such that for all completions, $\{\bar{\varphi}\}$ is a subset of the grounded extension iff $(\exists \tau_{\{x_1\}})(\forall \tau_{\{y_1\}})[c_1 \wedge c_2] = \texttt{false}$.

The same reduction for $Z = \emptyset$ gives a reduction from $\exists X \forall Y$-CNF-SAT to credulous controllability under grd for CAFs. As in simplified CAFs, this problem coincides with skeptical controllability under both grd and com. Similarly to [Baumeister *et al.*, 2018a], we prove hardness using an equivalence between completions of the CAF containing a particular query argument in the grounded extension and satisfying total truth assignments to the input formula.

Theorem 9. *For general CAFs, credulous controllability under grd and skeptical controllability under $\sigma \in \{com, grd\}$ are Σ_2^{P}-complete.*

Proof. (Sketch.) Let (F, C, U) be a CAF and $T \subseteq A_F$ a target. Σ_2^{P}-membership is derived from the quantifier representation: Is there a control configuration such that, for all completions $F^* = (A^*, R^*)$, $T \subseteq G$ holds (where $G \in grd(F^*)$ is the grounded extension of F^*)? Each of the quantifiers is polynomial-length-bounded and the condition $T \subseteq G$ with $G \in grd(F^*)$ is in P. Hardness follows by a reduction as illustrated in Figure 2. $\square$

4 Controllability as QBFs

We present QBF encodings of controllability covering the considered problem variants (except for skeptical controllability under prf, the encoding of which is presumably more complicated). Our encodings are more compact than those presented in [Dimopoulos *et al.*, 2018].

Consider controllability of a CAF $\mathcal{C} = (F, C, U)$ w.r.t. a target set $T \subseteq A_F$ under semantics $\sigma \in \{adm, com, stb\}$. Let $\mathcal{A} = A_F \cup A_U \cup A_C$ and $\mathcal{R} = R_F \cup R_U \cup R_U^{\leftrightarrow} \cup R_C$. We declare Boolean variables x_a and y_a for each $a \in \mathcal{A}$, and $r_{a,b}$ for each $(a, b) \in \mathcal{R}$, with the interpretations that for a model τ, $\tau(y_a) = 1$ iff $a \in A^\tau$, $\tau(r_{a,b}) = 1$ iff $(a, b) \in R^\tau$, and $\tau(x_a) = 1$ iff $a \in E \in \sigma(F^\tau)$, where $F^\tau = (A^\tau, R^\tau)$ is a completion of $\mathcal{C}$ under the control configuration $A^\tau_{\mathit{conf}} = \{a \in A_C \mid \tau(y_a) = 1\}$.

As a basis of the QBF encodings (as well as the CEGAR approach described in the next section), we capture semantics by conditioning standard SAT encodings of AF semantics [Besnard and Doutre, 2004] on the uncertain parts of a given CAF, similarly as for IAFs in [Niskanen *et al.*, 2020]. We encode completions with $\varphi_?(\mathcal{C}) = \bigwedge_{a \in A_F} y_a \wedge \bigwedge_{(a,b) \in R_F \cup R_C} r_{a,b} \wedge \bigwedge_{a \in A_U \cup A_C} (\neg y_a \rightarrow (\neg x_a \wedge \bigwedge_{(a,b) \in R_U \cup R_U^{\leftrightarrow}} \neg r_{a,b} \wedge \bigwedge_{(b,a) \in R_U \cup R_U^{\leftrightarrow}} \neg r_{b,a})) \wedge \bigwedge_{(a,b) \in R_U^{\leftrightarrow}} ((y_a \wedge y_b) \rightarrow (r_{a,b} \vee r_{b,a}))$. The formulas $\varphi_{cf}(\mathcal{C}) = \bigwedge_{(a,b) \in \mathcal{R}} (e_{a,b} \rightarrow (\neg x_a \vee \neg x_b))$, $\varphi_{adm}(\mathcal{C}) = \varphi_{cf}(\mathcal{C}) \wedge \bigwedge_{a \in \mathcal{A}} \bigwedge_{(b,a) \in \mathcal{R}} (x_a \rightarrow (e_{b,a} \rightarrow z_b))$, $\varphi_{com}(\mathcal{C}) = \varphi_{adm}(\mathcal{C}) \wedge \bigwedge_{a \in \mathcal{A}} \bigwedge_{(b,a) \in \mathcal{R}} ((e_{b,a} \rightarrow z_b) \rightarrow x_a)$, and $\varphi_{stb}(\mathcal{C}) = \varphi_{cf}(\mathcal{C}) \wedge \bigwedge_{a \in \mathcal{A}} (y_a \rightarrow (x_a \vee z_a))$ capture the semantics $\sigma \in \{adm, com, stb\}$ of the completions, where $z_a = \bigvee_{(b,a) \in \mathcal{R}} (x_b \wedge y_b \wedge r_{b,a})$ for $a \in \mathcal{A}$ and $e_{a,b} = y_a \wedge y_b \wedge r_{a,b}$ for $(a, b) \in \mathcal{R}$. In particular, $\varphi_?(\mathcal{C}) \wedge \varphi_\sigma(\mathcal{C}) \wedge \bigwedge_{t \in T} x_t$ is satisfiable iff there is a control configuration $A_{\mathit{conf}} \subseteq A_C$, a completion of $\mathcal{C}$, and a σ-extension of the completion containing all arguments in T.

Now, first consider skeptical controllability under $\sigma \in \{com, stb\}$. Let $X = \{y_a \mid a \in A_C\}$ and $Y = \{y_a \mid a \in A_U\} \cup \{r_{a,b} \mid (a, b) \in R_U \cup R_U^{\leftrightarrow}\} \cup \{x_a, z_a \mid a \in A_F \cup A_U \cup A_C\}$. We have that

$$\forall X \exists Y \left(\varphi_\sigma(\mathcal{C}) \wedge \bigvee_{t \in T} \neg x_t \right)$$

is false iff $\mathcal{C}$ is skeptically controllable w.r.t. T under σ. In particular, this is the negation of the QBF in [Dimopoulos *et al.*, 2018], and circumvents the issue of incompatible assignments to universally quantified variables, such as $y_a \wedge y_b \wedge \neg r_{a,b} \wedge \neg r_{b,a}$ for some $(a, b) \in R_U^{\leftrightarrow}$, in the encoding of [Dimopoulos *et al.*, 2018]. Our encoding is also more compact, as we do not have to encode $\neg \varphi_\sigma(\mathcal{C})$ in CNF, which would require as many auxiliary variables as there are clauses in $\varphi_\sigma(\mathcal{C})$ using a standard translation.

For credulous controllability under $\sigma \in \{adm, stb\}$, we use an auxiliary variable ξ interpreted as $\tau(\xi) = 1$ iff τ does not correspond to a valid completion, encoded as $\varphi_\times(\mathcal{C}) = \bigvee_{(a,b) \in R_U \cup R_U^{\leftrightarrow}} (\neg y_a \wedge r_{a,b}) \vee \bigvee_{(a,b) \in R_U^{\leftrightarrow}} (y_a \wedge y_b \wedge \neg r_{a,b} \wedge \neg r_{b,a})$. Let $X = \{y_a \mid a \in A_C\}$, $Y = \{y_a \mid a \in A_U\} \cup \{r_{a,b} \mid (a, b) \in R_U \cup R_U^{\leftrightarrow}\}$, and $Z = \{x_a, z_a \mid a \in A_F \cup A_U \cup A_C\} \cup \{\xi\}$. Then the QBF

$$\exists X \forall Y \exists Z \left(\neg \xi \rightarrow \left(\varphi_\sigma(\mathcal{C}) \wedge \bigwedge_{t \in T} x_t \right) \right) \wedge (\xi \rightarrow \varphi_\times(\mathcal{C}))$$

is true iff $\mathcal{C}$ is credulously controllable w.r.t. T and σ. If ξ is true—quantified existentially after the universal quantifier over the variables representing the completion—the corresponding completion must be invalid to satisfy the formula. If ξ is false and the formula is satisfied, the target is reached.

5 Controllability by SAT-Based CEGAR

As an alternative to QBFs, we develop SAT-based counterexample-guided abstraction refinement (CEGAR) [Clarke *et al.*, 2003; Clarke *et al.*, 2004] procedures for controllability. The algorithms cover all considered problem variants except skeptical controllability under prf, which can be handled similarly as skeptical controllability under com with an additional subset-maximization procedure. CEGAR works on an NP-abstraction, and iteratively queries a SAT solver for a candidate solution and checks for counterexamples to the candidate using the same solver,

Algorithm 1 CEGAR for skeptical controllability

Input: CAF $\mathcal{C} = (F, C, U)$, target $T \subseteq A_F$, $\sigma \in \{com, stb\}$.

1: $\varphi \leftarrow \varphi_\sigma(\mathcal{C})$
2: **while true do**
3: $\quad (result, \tau) \leftarrow \text{SAT}(\varphi \wedge \bigwedge_{t \in T} x_t)$
4: $\quad$ **if** $result = sat$ **then**
5: $\quad\quad (result, \tau) \leftarrow \text{SAT}(\varphi \wedge \psi(A^\tau_{conf}) \wedge \bigvee_{t \in T} \neg x_t)$
6: $\quad\quad$ **if** $result = unsat$ **then return** A^τ_{conf}
7: $\quad\quad \varphi \leftarrow \varphi \wedge \text{RefineControl}(A^\tau_{conf})$
8: $\quad$ **else**
9: $\quad\quad$ **if** $\sigma \neq stb$ **then return** *reject*
10: $\quad\quad$ **while true do**
11: $\quad\quad\quad (result, \tau) \leftarrow \text{SAT}(\varphi)$
12: $\quad\quad\quad$ **if** $result = unsat$ **then break**
13: $\quad\quad\quad \varphi \leftarrow \varphi \wedge \text{RefineControl}(A^\tau_{conf})$
14: $\quad\quad \varphi' \leftarrow \varphi \setminus \varphi_{stb}(\mathcal{C})$
15: $\quad\quad (result, \tau) \leftarrow \text{SAT}(\varphi')$
16: $\quad\quad$ **if** $result = sat$ **then return** A^τ_{conf}
17: $\quad\quad$ **else return** *reject*

Algorithm 2 CEGAR for credulous controllability

Input: CAF $\mathcal{C} = (F, C, U)$, target $T \subseteq A_F$, $\sigma \in \{adm, stb\}$.

1: $\varphi \leftarrow \varphi_\sigma(\mathcal{C})$
2: **while true do**
3: $\quad (result, \tau) \leftarrow \text{SAT}(\varphi \wedge \bigwedge_{t \in T} x_t)$
4: $\quad$ **if** $result = unsat$ **then return** *reject*
5: $\quad \varphi \leftarrow \varphi \wedge (\psi(A^\tau_{conf}) \rightarrow \text{Refine}(F^\tau))$
6: $\quad$ **while true do**
7: $\quad\quad (result, \tau) \leftarrow \text{SAT}(\varphi \wedge \psi(A^\tau_{conf}) \wedge \bigvee_{t \in T} \neg x_t)$
8: $\quad\quad$ **if** $result = sat$ **then**
9: $\quad\quad\quad (result, \tau) \leftarrow \text{SAT}(\varphi \wedge \chi(F^\tau) \wedge \bigwedge_{t \in T} x_t)$
10: $\quad\quad\quad$ **if** $result = unsat$ **then break**
11: $\quad\quad\quad \varphi \leftarrow \varphi \wedge (\psi(A^\tau_{conf}) \rightarrow \text{Refine}(F^\tau))$
12: $\quad\quad$ **else**
13: $\quad\quad\quad$ **if** $\sigma \neq stb$ **then return** A_{conf}
14: $\quad\quad\quad$ **while true do**
15: $\quad\quad\quad\quad (result, \tau) \leftarrow \text{SAT}(\varphi \wedge \psi(A^\tau_{conf}))$
16: $\quad\quad\quad\quad$ **if** $result = unsat$ **then break**
17: $\quad\quad\quad\quad \varphi \leftarrow \varphi \wedge (\psi(A^\tau_{conf}) \rightarrow \text{Refine}(F^\tau))$
18: $\quad\quad\quad \varphi' \leftarrow \varphi \setminus \varphi_{stb}(\mathcal{C})$
19: $\quad\quad\quad (result, \tau') \leftarrow \text{SAT}(\varphi' \wedge \psi(A^\tau_{conf}))$
20: $\quad\quad\quad$ **if** $result = unsat$ **return** A^τ_{conf}
21: $\quad\quad\quad$ **else** $\text{SAT}(\varphi \wedge \chi(F^\tau))$; **break**
22: $\quad \varphi \leftarrow \varphi \wedge \text{RefineCore}(A^\tau_{conf})$

refining the abstraction until a valid solution is found or all candidates have been ruled out.

First consider second-level complete skeptical controllability under $\sigma \in \{com, stb\}$. As outlined in Algorithm 1, we initialize the abstraction using the SAT encoding $\varphi_\sigma(\mathcal{C})$ (line 1). Then, we iteratively solve the abstraction with the additional clauses $\bigwedge_{t \in T} x_t$ ensuring that T is in a σ-extension of the completion (line 3). If the abstraction is satisfiable (line 4), we obtain the corresponding control configuration A^τ_{conf} (recall Section 4). Given A^τ_{conf}, we check for a counterexample (line 5) as a completion where one of the target arguments is not in a σ-extension. We encode the control configuration using $\psi(A^\tau_{conf}) = \bigwedge_{a \in A^\tau_{conf}} y_a \wedge \bigwedge_{a \in A_C \setminus A^\tau_{conf}} \neg y_a$ and the counterexample query via $\bigvee_{t \in T} \neg x_t$. If there is no counterexample, we have found a valid control configuration and return (line 6). If we obtain a counterexample—a completion F^τ and an extension $E^\tau \in \sigma(F^\tau)$ with $T \not\subseteq E^\tau$—we refine the abstraction by adding the clause $\text{RefineControl}(A^\tau_{conf}) = \bigvee_{a \in \text{AddArg}_\sigma(\tau)} y_a \vee \bigvee_{a \in \text{RemArg}_\sigma(\tau)} \neg y_a$ (line 7), where $\text{AddArg}_{stb}(\tau) = \{a \in A_C \mid \nexists b \in E^\tau, (b, a) \in R_F\}$ and $\text{RemArg}_{stb}(\tau) = \{a \in A_C \mid a \in E^\tau\}$ characterize arguments whose addition or removal can alter the counterexample extension $E^\tau = \{a \in \mathcal{A} \mid \tau(x_a) = 1\}$ under *stb* [Niskanen *et al.*, 2020]. For complete semantics, we have $\text{AddArg}_{com}(\tau) = \text{AddArg}_{stb}(\tau)$, and $\text{RemArg}_{com}(\tau) = \{a \in A_C \mid a \in E^\tau \cup A \setminus (E^\tau)^+_{F^\tau}\}$, that is, for removal we need to also consider removing arguments that are not attacked by E^τ.

If there is no solution to the abstraction (line 8), if $\sigma \neq stb$ we reject (line 9) since $\mathcal{C}$ is then not controllable. The case $\sigma = stb$ is more delicate as there may be a control configuration for which no completion has a stable extension, making $\mathcal{C}$ controllable. Thus we filter out all control configurations admitting a completion with a stable extension (lines 10–13). Finally, we check if there still are control configurations left (lines 14–15). If there are, we return the corresponding control configuration (line 16), and otherwise reject (line 17).

Algorithm 2 for credulous controllability under $\sigma \in \{adm, stb\}$—a third-level complete problem—solves the abstraction $\varphi_\sigma(\mathcal{C})$ (line 1) iteratively under the constraint $\bigwedge_{t \in T} x_t$ (line 3). If unsatisfiable, there is no valid control configuration (line 4). Otherwise, we begin the search for a counterexample, i.e., a completion for the current control configuration that has no extension containing the target. The current completion F^τ is not such a counterexample, so we refine the abstraction via $\psi(A^\tau_{conf}) \rightarrow \text{Refine}(F^\tau)$ (line 5), where $\text{Refine}(F^\tau)$ is exactly the so-called *strong refinement* earlier proposed for IAFs [Niskanen *et al.*, 2020]. We iteratively search for a completion F^τ with $E^\tau \in \sigma(F^\tau)$ and $T \not\subseteq E^\tau$ (line 7). If such a candidate counterexample is found (line 8), we check if F^τ is a counterexample by checking if there is a further extension containing T (line 9), encoding the completion via $\chi(F^\tau) = \bigwedge_{a \in A^\tau} y_a \wedge \bigwedge_{a \in \mathcal{A} \setminus A^\tau} \neg y_a \wedge \bigwedge_{(a,b) \in R^\tau} r_{a,b} \wedge \bigwedge_{(a,b) \in \mathcal{R} \setminus R^\tau} \neg r_{a,b}$. If there is no such extension, we have a counterexample to A^τ_{conf} (line 10). Otherwise, we refine the abstraction (line 11).

If no counterexample candidate F^τ is found (line 12), if $\sigma \neq stb$ we return the current A^τ_{conf} (line 13). The case $\sigma = stb$ is more complex as we need to check if some completion has no stable extension, in which case the target would not be reached. This is checked similarly as in Algorithm 1 (lines 14–21), with the distinctions that SAT calls are performed under the current configuration via $\psi(A_{conf})$ (lines 15 and 19) and that the refinement is on completions (line 17).

When we exit the inner loop, we have a counterexample F^τ with no $E \in \sigma(F^\tau)$, $T \subseteq E$. On line 10 this was deduced via

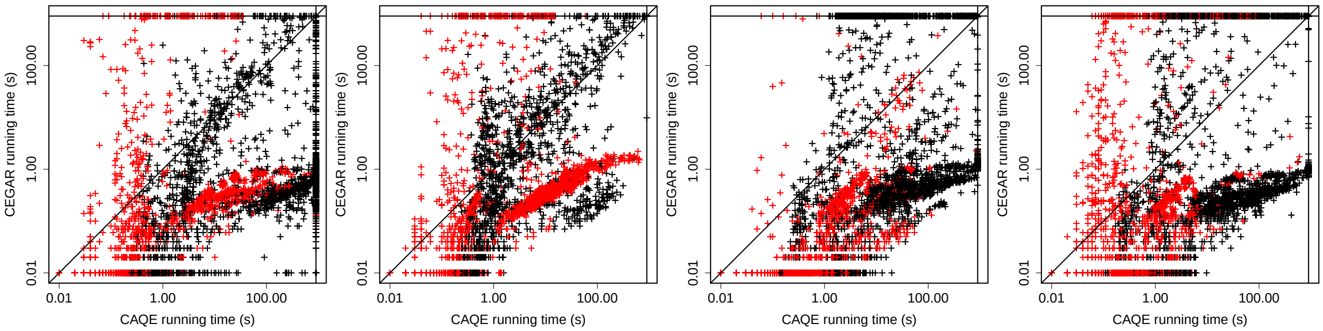

Figure 3: QBF vs CEGAR: skeptical controllability under com (left), stb (left center) and credulous under adm (right center), stb (right).

the SAT solver reporting unsatisfiability, while on line 21 we make an additional SAT solver call which is guaranteed to be unsatisfiable. Therefore in both cases the SAT solver provides an *unsatisfiable core* $U \subseteq \chi(F^\tau)$ containing exactly those units in $\chi(F^\tau)$ used in the unsatisfiability proof. Thus, we refine the control configuration via $\text{REFINECORE}(A^\tau_{conf}) = \bigvee_{a \in (A_C \setminus A^\tau_{conf}), \neg y_a \in U} y_a \vee \bigvee_{a \in A^\tau_{conf}, y_a \in U} \neg y_a$.

6 Empirical Evaluation

We overview results from an evaluation of the QBF and CEGAR approaches to controllability. We used the QBF solver CAQE 4.0.0 [Tentrup, 2019] with the Bloqqer [Heule *et al.*, 2015] preprocessor and the flag `--qdo` to obtain assignments corresponding to control configurations. For CEGAR we used the Glucose 4.1 SAT solver [Audemard and Simon, 2018]. The experiments were run on Intel Xeon E5-2680 v4 2.4-GHz, 256-GB nodes with CentOS 7 under a per-instance 900-s time and 64-GB memory limit. The implementation, benchmarks, and runtime data are available online.

We generated CAFs from the 2019 ICCMA competition AFs (http://argumentationcompetition.org/2019/iccma-instances.tar.gz) as follows. For each AF, the query argument from ICCMA 2019 is the singleton target. For each $p_C \in \{0.05, 0.1, 0.15, 0.2\}$, each non-query argument is a control argument with probability p_C. For each $p_U \in \{0, 0.05, 0.1, 0.15, 0.2\}$, each argument (apart from control and query arguments) is uncertain with probability p_U. Each attack is uncertain with probability $p_U/2$ unless the source or the target is a control argument, and has uncertain direction with probability $p_U/2$ unless the reverse direction is already a fixed or an uncertain attack. The rest of the arguments and attacks remain fixed. This yielded a total of 6520 CAFs, out of which 1304 are simplified CAFs with no uncertain part.

We consider skeptical controllability under com and stb and credulous controllability under adm and stb. The relative efficiency of the approaches depends on the instance (see Figure 3). In general, runtimes do not correlate well with instance sizes, but rather with the benchmark domain of the original AF. Interestingly Bloqqer is able to solve a noticeable number of the QBF instances (i.e., without calling CAQE), those are shown in red. The QBF approach exhibits fewer timeouts, most noticeably on skeptical controllability under stb. The relative mean runtimes (Figure 4) depend on the problem variant. For skeptical under com, increasing both p_C and p_U tends to increase empirical hardness, with CEGAR exhibiting lower mean runtimes. The effect of p_C is more modest on credulous under stb, where the mean runtimes of the QBF approach are considerably lower than for CEGAR.

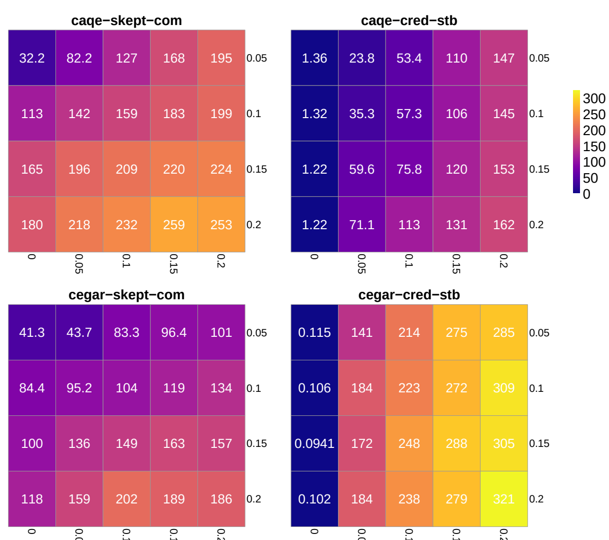

Figure 4: Mean runtimes: QBF (top) and CEGAR (bottom), skeptical under com (left) and credulous under stb (right) with different values of p_C (rows) and p_U (columns).

7 Conclusions

We provided complexity results for credulous and skeptical controllability in CAFs under five central semantics. Almost all cases are complete for the second or third level of the polynomial hierarchy. Due to this, we presented two declarative approaches, based on direct QBF encodings and SAT-based CEGAR, to controllability, and first results on the empirical hardness of controllability. Covering further semantics, developing further CEGAR refinement strategies, and evaluating different QBF solvers are directions for further work.

Acknowledgments

This work was financially supported by Academy of Finland grants 276412, 312662, 322869 and Univ. Helsinki DoCS doctoral programme. Computational resources were provided by FGCI (urn:nbn:fi:research-infras-2016072533).

References

[Arora and Barak, 2009] S. Arora and B. Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, 2009.

[Audemard and Simon, 2018] G. Audemard and L. Simon. On the Glucose SAT solver. *Int. J. Artif. Intell. T.*, 27(1):1840001:1–1840001:25, 2018.

[Baroni *et al.*, 2018] P. Baroni, M. Caminada, and M. Giacomin. Abstract argumentation frameworks and their semantics. In *Handbook of Formal Argumentation*, chapter 4, pages 159–236. College Publications, 2018.

[Baumann and Brewka, 2010] R. Baumann and G. Brewka. Expanding argumentation frameworks: Enforcing and monotonicity results. In *COMMA*, volume 216 of *FAIA*, pages 75–86. IOS Press, 2010.

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

[Baumeister *et al.*, 2018a] D. Baumeister, D. Neugebauer, and J. Rothe. Credulous and skeptical acceptance in incomplete argumentation frameworks. In *COMMA*, volume 305 of *FAIA*, pages 181–192. IOS Press, 2018.

[Baumeister *et al.*, 2018b] D. Baumeister, D. Neugebauer, J. Rothe, and H. Schadrack. Verification in incomplete argumentation frameworks. *Artif. Intell.*, 264:1–26, 2018.

[Besnard and Doutre, 2004] P. Besnard and S. Doutre. Checking the acceptability of a set of arguments. In *NMR*, pages 59–64, 2004.

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

[Cayrol *et al.*, 2010] C. Cayrol, F. Dupin de Saint-Cyr, and M. Lagasquie-Schiex. Change in abstract argumentation frameworks: Adding an argument. *J. Artif. Intell. Res.*, 38:49–84, 2010.

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

[Clarke *et al.*, 2004] E.M. Clarke, A. Gupta, and O. Strichman. SAT-based counterexample-guided abstraction refinement. *IEEE T-CAD*, 23(7):1113–1123, 2004.

[Coste-Marquis *et al.*, 2007] S. Coste-Marquis, C. Devred, S. Konieczny, M. Lagasquie-Schiex, and P. Marquis. On the merging of Dung's argumentation systems. *Artif. Intell.*, 171(10-15):730–753, 2007.

[Coste-Marquis *et al.*, 2014] S. Coste-Marquis, S. Konieczny, J. Mailly, and P. Marquis. On the revision of argumentation systems: Minimal change of arguments statuses. In *KR*, pages 52–61. AAAI Press, 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 *IJCAI*, pages 2876–2882. AAAI Press, 2015.

[de Saint-Cyr *et al.*, 2016] F. Dupin de Saint-Cyr, P. Bisquert, C. Cayrol, and M. Lagasquie-Schiex. Argumentation update in YALLA (Yet Another Logic Language for Argumentation). *Int. J. Approx. Reasoning*, 75:57–92, 2016.

[Diller *et al.*, 2018] M. Diller, A. Haret, T. Linsbichler, S. Rümmele, and S. Woltran. An extension-based approach to belief revision in abstract argumentation. *Int. J. Approx. Reasoning*, 93:395–423, 2018.

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

[Dimopoulos *et al.*, 2018] Y. Dimopoulos, J. Mailly, and P. Moraitis. Control argumentation frameworks. In *AAAI*, pages 4678–4685. AAAI Press, 2018.

[Dimopoulos *et al.*, 2019] Y. Dimopoulos, J. Mailly, and P. Moraitis. Argumentation-based negotiation with incomplete opponent profiles. In *AAMAS*, pages 1252–1260. IFAAMAS, 2019.

[Doutre and Mailly, 2018] S. Doutre and J. Mailly. Constraints and changes: A survey of abstract argumentation dynamics. *Argument & Computation*, 9(3):223–248, 2018.

[Doutre *et al.*, 2014] S. Doutre, A. Herzig, and L. Perrussel. A dynamic logic framework for abstract argumentation. In *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. *Artif. Intell.*, 77(2):321–358, 1995.

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

[Dvorák and Dunne, 2018] W. Dvorák and P.E. Dunne. Computational problems in formal argumentation and their complexity. In *Handbook of Formal Argumentation*, chapter 13, pages 631–687. College Publications, 2018.

[Heule *et al.*, 2015] M. Heule, M. Järvisalo, F. Lonsing, M. Seidl, and A. Biere. Clause elimination for SAT and QSAT. *J. Artif. Intell. Res.*, 53:127–168, 2015.

[Niskanen *et al.*, 2020] A. Niskanen, D. Neugebauer, M. Järvisalo, and J. Rothe. Deciding acceptance in incomplete argumentation frameworks. In *AAAI*. AAAI Press, 2020.

[Tentrup, 2019] L. Tentrup. CAQE and QuAbS: Abstraction based QBF solvers. *Journal on Satisfiability, Boolean Modeling and Computation*, 11(1):155–210, 2019.

[Wallner *et al.*, 2017] J.P. Wallner, A. Niskanen, and M. Järvisalo. Complexity results and algorithms for extension enforcement in abstract argumentation. *J. Artif. Intell. Res.*, 60:1–40, 2017.