Supplement to KR 2023 paper
Argumentative Reasoning in ASPIC+ under Incomplete Information

Daphne Odekerken[1,2] , **Tuomo Lehtonen**[3] , **AnneMarie Borg**[1] ,
Johannes P. Wallner[4] and **Matti Järvisalo**[3]

[1]Utrecht University, The Netherlands
[2]National Police Lab AI, Netherlands Police, The Netherlands
[3]University of Helsinki, Finland
[4]Graz University of Technology, Austria
{d.odekerken, a.borg}@uu.nl, {tuomo.lehtonen, matti.jarvisalo}@helsinki.fi, wallner@ist.tugraz.at

Proof Details

Proposition 1. *Given an AT $T = (AS, \mathcal{K})$, the corresponding AF $\langle \mathcal{A}, \mathcal{C} \rangle$, and a set of defeasible rules D and an argument $A \in \mathcal{A}$, it holds that at least one $B \in Arg_T(D)$ defeats A if and only if D defeats a rule $r \in \texttt{defrules}(A)$.*

Proof. We prove this in both directions:

- From left to right: suppose that there is some argument B in $Arg_T(D)$ that defeats A on $A' \in \texttt{sub}(A)$. Since A' is defeated, it cannot be observation-based and therefore must be rule-based. Let r be the top rule of A'. Then B is either observation-based or based on some rule $r' \in D$. If B is observation-based then $\texttt{conc}(B) \in \mathcal{K}$ and $\texttt{conc}(B) \in \overline{\texttt{cons}(r)}$. This implies that D defeats r. Otherwise B is based on some rule $r' \in D$. In that case, there must be an argument based on r' in $Arg_T(D) = Arg_T(D \cup \{r'\})$, so r' is applicable by D. In addition, B defeats A', so $\texttt{cons}(r') \in \overline{\texttt{cons}(r)}$ and $r' \nless r$. This implies that D defeats r. Finally, given that r is the top rule of A' and A' is a subargument of A, $r \in \texttt{defrules}(A)$.
- From right to left: assume that $D \subseteq \mathcal{R}$ defeats some $r \in \texttt{defrules}(A)$. Since $r \in \texttt{defrules}(A)$, there must be some subargument of A with top rule r. Let $A' \in \texttt{sub}(A)$ be this subargument. Given that D defeats r, either there is some $l \in \overline{\texttt{cons}(r)}$ in $\mathcal{K}$ or there is an $r' \in D$ whose consequent is a contradictory of r, r' is applicable by D and $r' \nless r$. In the first case, there is an observation-based argument in $Arg_T(D)$ that defeats A on A'. In the second case, given that r' is applicable by D, there must be an argument B in $Arg_T(D)$ with top rule r'. The argument B is not less preferred than A under the last-link principle; therefore B defeats A' and thus also A. □

Proposition 2. *Given an $T = (AS, \mathcal{K})$, the corresponding AF $\langle \mathcal{A}, \mathcal{C} \rangle$, and a set of defeasible rules D, and an argument $A \in \mathcal{A}$, it holds that $Arg_T(D)$ defends A if and only if D defends every rule $r \in \texttt{defrules}(A)$.*

Proof. We prove this in both directions:

- From left to right: suppose that $Arg_T(D)$ defends A and let r be an arbitrary rule in $\texttt{defrules}(A)$. Suppose, towards a contradiction, that r is not defended by D. Then by Definition 18 the set U, consisting of all rules in $\mathcal{R}$ that are not defeated by D, defeats r. Then by Proposition 1 there is some $B \in Arg_T(U)$ that defeats A. Given that $B \in Arg_T(U)$, the argument B can be constructed using $\mathcal{K}$ and U. Due to the way U is constructed (consisting only of rules not defeated by D), by Proposition 1 there is no argument $C \in Arg_T(D)$ that defeats B. Then by Definition 9, $Arg_T(D)$ does not defend A. From this contradiction it follows that r is defended by D.
- From right to left: assume that D defends all $r \in \texttt{defrules}(A)$. Suppose, towards a contradiction, that $Arg_T(D)$ does not defend A. Then by Definition 9 there is some $B \in \mathcal{A}$ that defeats A and no argument in $Arg_T(D)$ defeats B. By Proposition 1, this implies that each rule $r' \in \texttt{defrules}(B)$ is not defeated by D. Therefore it must be that $\texttt{defrules}(B) \subseteq U$ where U is the set of all rules in $\mathcal{R}$ that is not defeated by D. This implies that $B \in Arg_T(U)$. Given that $B \in Arg_T(U)$ defeats A, by Proposition 1 it follows that U defeats a rule $r \in \texttt{defrules}(A)$. But then by Definition 18, r is not defended by D, which contradicts our assumption. Therefore $Arg_T(D)$ must defend A. □

Lemma 3 (Monotonicity of defence)**.** *Let $T = (AS, \mathcal{K})$ be an AT where $AS = (\mathcal{L}, \overline{}, \mathcal{R}, \leq)$. For each $R \subseteq \mathcal{R}$ and $r \in \mathcal{R}$: if R defends r then each R' such that $R \subseteq R' \subseteq \mathcal{R}$ defends r.*

Proof. If R defends r then by Definition 18 the set of rules U in $\mathcal{R}$ that is not defeated by R does not defeat r. Let R' be an arbitrary rule set such that $R \subseteq R' \subseteq \mathcal{R}$ and let U' be the set of all rules in $\mathcal{R}$ that is not defeated by R'. Then $U' \subseteq U$; given that U does not defeat r, U' cannot defeat r either. □

Proposition 3. *Let $T = (AS, \mathcal{K})$ be an AT where $AS = (\mathcal{L}, \overline{}, \mathcal{R}, \leq)$, $R \subseteq \mathcal{R}$ be a set of defeasible rules such that (i) each rule $r \in R$ is applicable by R and (ii) $Arg_T(R)$ is admissible. Let r and r' be rules in $\mathcal{R}$ defended by R. Then*

(1) $Arg_T(R \cup \{r\})$ is admissible and (2) $R \cup \{r\}$ defends r'.

Proof. Suppose that $Arg_T(R)$ is admissible and that r and r' are rules in $\mathcal{R}$ that are defended by R.

1. First we show that $Arg_T(R \cup \{r\})$ is admissible, that is: it defends itself and is conflict-free.
   - By Definition 9 of admissibility, $Arg_T(R)$ defends each argument in $Arg_T(R)$. Then by Proposition 2, R defends each rule in R. Given that r is defended by R as well, R defends each rule in $R \cup \{r\}$. By monotonicity of defence (Lemma 3), $R \cup \{r\}$ defends each rule in $R \cup \{r\}$. Then for each argument A in $Arg_T(R \cup \{r\})$ it holds that each rule in $\texttt{defrules}(A)$ is defended by $R \cup \{r\}$, which by Proposition 2 implies that $Arg_T(R \cup \{r\})$ defends itself.
   - To show admissibility, what remains to be shown is conflict-freeness. Suppose towards a contradiction that $Arg_T(R \cup \{r\})$ is not conflict-free. Then there is some A, B in $Arg_T(R \cup \{r\})$ such that A defeats B. Given that R defends each rule in $R \cup \{r\}$, by Proposition 2 $Arg_T(R)$ defends $Arg_T(R \cup \{r\})$. As $B \in Arg_T(R \cup \{r\})$ is defeated (by A), there must be some argument C in $Arg_T(R)$ that defeats A. However, $Arg_T(R)$ defends itself (by admissibility) so there must be some argument D in $Arg_T(R)$ defeating C. That implies that C and D in $Arg_T(R)$ defeat each other, which contradicts conflict-freeness of the admissible set $Arg_T(R)$. From this contradiction it follows that $Arg_T(R \cup \{r\})$ is conflict-free.

   Therefore $Arg_T(R \cup \{r\})$ is admissible.
2. From the assumption that R defends r' and Lemma 3 it directly follows that $R \cup \{r\}$ defends r'. □

Proposition 4. *Given an AT $T = (AS, \mathcal{K})$ where $AS = (\mathcal{L}, \overline{}, \mathcal{R}, \leq)$, let C be the least fixpoint of def_T. Then $G(T) = Arg_T(C)$.*

Proof. $def_T(\emptyset) = C_1$ contains all undefeated rules; by Proposition 1, $Arg_T(C_1)$ consists of undefeated arguments only and is therefore admissible. The operator def_T is monotonic and thus has a unique least fixed point. Due to Proposition 3, $Arg_T(def^i_T(\emptyset))$ is admissible for each $i \in \mathbb{N}$. C contains every rule that it defends and thus by Proposition 2 $Arg_T(C)$ contains every argument it defends. Since we assume a finite AT, we reach a fixed-point C at some i. Suppose that $Arg_T(C)$ is not complete. Then there is an argument A defended by $Arg_T(C)$ and not in $Arg_T(C)$. By Proposition 2, we arrive at a contradiction to C being a fixed-point (note that the rules in A are iteratively applicably, bottom-up). Suppose that $Arg_T(C)$ is not grounded. Then the grounded extension $G \subsetneq Arg_T(C)$ is a proper subset. Let j be the first iteration such that this relation holds (first iteration where $Arg_T(C_j)$ is not a subset of the grounded extension). Again via Proposition 2 we arrive at a contradiction: we defend some rule outside the rules of the grounded extension which is applicably by C_j. But then there is an argument defended by $Arg_T(C_j)$, and also defended by the grounded extension. □

Proposition 5. *Given an AT $T = (AS, \mathcal{K})$ where $AS = (\mathcal{L}, \overline{}, \mathcal{R}, \leq)$, the least fixpoint of def_T is reached in at most $|\mathcal{R}|/2$ iterations.*

Proof sketch. Consider the sets $S^i = def^i_T(\emptyset)$ and $D^i = \{r \in \mathcal{R} \mid r \text{ is defeated by } S^i\}$, for $i \in \mathbb{N}$. Note that $S^i \subseteq S^{i+1}$ and $D^i \subseteq D^{i+1}$, and that S^i is disjoint with D^j, for all i, j, because $Arg(S^i)$ is conflict-free by Proposition 3. Moreover, S^i defending a rule that S^{i-1} does not defend requires that S^i defeats some rule that S^{i-1} does not defeat. This implies that if $D^i = D^{i-1}$, then S^i is the least fixed point of $def^i_T(\emptyset)$. Thus, on every iteration either at least one element is added to both S^i and D^i or a least fixed point is reached. Thus, when $i = |\mathcal{R}/2|$, every $r \in \mathcal{R}$ is in either S^i or D^i, and a least fixed point has been reached. □

Proposition 6. *Given an AT $T = (AS, \mathcal{K})$ where $AS = (\mathcal{L}, \overline{}, \mathcal{R}, \leq)$, and C be the least fixpoint of def_T. A literal $l \in \mathcal{L}$ is*

- *unsatisfiable if there is no argument A with $\texttt{conc}(A) = l$,*
- *defended if there is an argument A with $\texttt{conc}(A) = l$ and $\texttt{defrules}(A) \subseteq C$,*
- *out if no argument with conclusion l is based on U, where U is the set of rules not defeated by C, and*
- *blocked otherwise.*

Proof. By Proposition 4 it holds that $Arg_T(C)$ is the grounded extension of the given AT. For any given set of rules D and literal l it holds that there is an argument A based on D concluding l iff one can chain rules iteratively (starting with conclusions of observation-arguments) until we derive l. The statements of the proposition follow directly from definition, and Proposition 1. □

Lemma 1. *Let $T = (AS, \mathcal{K})$ be an AT, let $\mathcal{Q}$ be a set of queryables and let j be a justification status. Given a literal $l \in \mathcal{L}$ and a queryable literal $q \in Q$ where $q \notin \mathcal{K}$ and $\overline{q} \cap \mathcal{K} = \emptyset$, q is j-relevant for l wrt T and $\mathcal{Q}$ iff*

- *there is an AT $T' = (AS, \mathcal{K}')$ with $T \sqsubseteq_Q T'$ such that l is not stable-j wrt T' and*
- *l is stable-j wrt $(AS, \mathcal{K}' \cup \{q\})$.*

Proof. From left to right: if q is j-relevant for l wrt T and $\mathcal{Q}$ then there is some minimal stable-j future theory $(AS, \mathcal{K}^*)$ s.t. $q \in \mathcal{K}^*$; by minimality, l cannot be stable-j in $(AS, \mathcal{K}^* \setminus \{q\})$. From right to left: suppose that an $(AS, \mathcal{K}' \cup \{q\})$ exists s.t. l is stable-j, while l is not stable-j w.r.t. $(AS, \mathcal{K}')$. If $(AS, \mathcal{K}' \cup \{q\})$ is minimal stable-j then we are done; otherwise there is some $\mathcal{K}'' \subset \mathcal{K}'$ such that l is stable-j wrt $(AS, \mathcal{K}'' \cup \{q\})$. Then l cannot be stable-j wrt $(AS, \mathcal{K}'')$, as there is some T'' such that $(AS, \mathcal{K}') \sqsubseteq_Q T''$ in which l was not j and T'' must also be in $(AS, \mathcal{K}'')$. □

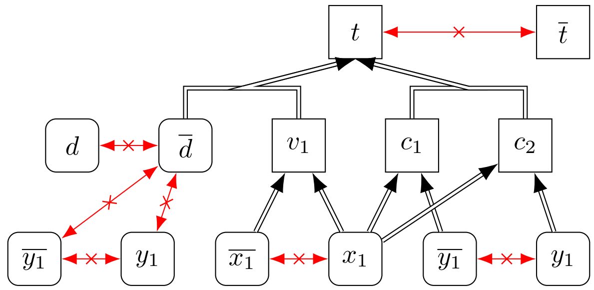

Figure 1: Illustration of the reduction used in Theorem 2 for the formula $\phi = (x_1 \vee y_1) \wedge (x_i \vee \neg y_1)$. The queryables $\overline{y_1}$ and y_1 are displayed twice for readability.

Theorem 2. *Deciding whether a queryable is j-relevant for a literal in an AT wrt a set of queryables is Σ_2^P-complete for each $j \in \{$unsatisfiable, defended, out, blocked$\}$. Hardness holds even without preferences.*

Proof of Theorem 2 for unsatisfiable status. Membership in Σ_2^P follows from Lemma 1: a positive instance can be verified with two calls to an oracle for stability, which is coNP-complete by Proposition 7. For Σ_2^P-hardness, we reduce from the Σ_2-ST problem of deciding for a formula ϕ in CNF, quantified over X and Y where $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$ are pairwise disjoint sets, if there is an assignment τ_X to variables in X such that for each assignment τ_Y to variables in Y, $\phi[\tau_X, \tau_Y] =$ False. Construct the following AT T and queryables $\mathcal{Q}$, with $C = c_1 \wedge \ldots \wedge c_p$ the set of clauses in ϕ, and $\overline{X} = \{\overline{x} \mid x \in X\}$, $\overline{Y} = \{\overline{y} \mid y \in Y\}$ and $\overline{C} = \{\overline{c} \mid c \in C\}$. Let $V = \{v_i \mid x_i \in X\}$ and $\overline{V} = \{\overline{v_i} \mid x_i \in X\}$.

$$
\begin{aligned}
\mathcal{Q} &= X \cup \overline{X} \cup Y \cup \overline{Y} \cup \{d, \overline{d}\} \\
\mathcal{L} &= \mathcal{Q} \cup C \cup \overline{C} \cup V \cup \overline{V} \cup \{t, \overline{t}\} \\
{}^{-} &= \{(x, \overline{x}), (\overline{x}, x) \mid x \in X \cup Y \cup V \cup C \cup \{d, t\}\} \cup \\
&\quad \{(y, \overline{d}), (\overline{y}, \overline{d}), (\overline{d}, y), (\overline{d}, \overline{y}) \mid y \in Y\} \\
\mathcal{R} &= \{(\overline{d}, v_1, \ldots, v_n \Rightarrow t)\} \cup \\
&\quad \{(x \Rightarrow c) \mid x \in c\} \cup \{(\overline{x} \Rightarrow c) \mid \neg x \in c\} \cup \\
&\quad \{(y \Rightarrow c) \mid y \in c\} \cup \{(\overline{y} \Rightarrow c) \mid \neg y \in c\} \cup \\
&\quad \{(c_1, \ldots, c_p \Rightarrow t)\} \cup \\
&\quad \{(x_i \Rightarrow v_i), (\overline{x_i} \Rightarrow v_i) \mid x_i \in X\} \\
\mathcal{K} &= \emptyset
\end{aligned}
$$

Then $T = (AS, \mathcal{K})$ where $AS = (\mathcal{L}, {}^{-}, \mathcal{R}, \leq)$ and $\mathcal{Q}$ can be constructed in polynomial time wrt ϕ. We argue that (ϕ, X, Y) is a positive instance of Σ_2-ST iff d is unsatisfiable-relevant for t wrt T.

- From left to right: suppose that (ϕ, X, Y) is a positive instance of Σ_2-ST. Then there is some assignment to variables of X such that for each assignment to variables of Y, $\phi[\tau_X, \tau_Y]$ is False. Let τ_X be this assignment and construct a knowledge base $\mathcal{K}' = \{x \in X \mid \tau_X[x] =$ True$\} \cup \{\overline{x} \in X \mid \tau_X[x] =$ False$\}$. Note that $\mathcal{K}'$ must be consistent, as no $x \in X$ can be assigned both True and False by τ_X. Therefore $T \sqsubseteq_Q (AS, \mathcal{K}')$. Then:
  - t is not stable-unsatisfiable wrt $(AS, \mathcal{K}')$ and $\mathcal{Q}$, because t is not unsatisfiable wrt $(AS, \mathcal{K}' \cup \{\overline{d}\})$: note that none of the contradictories of $\overline{d}$ is in $\mathcal{K}'$; therefore $\mathcal{K}' \cup \{\overline{d}\}$ is a consistent knowledge base. Given that for each $x \in X$ either $x \in \mathcal{K}'$ or $\overline{x} \in \mathcal{K}'$, it must be that for each $x \in X$ either $x \in \mathcal{K}' \cup \{\overline{d}\}$ or $\overline{x} \in \mathcal{K}' \cup \{\overline{d}\}$. This implies that there is an argument for t based on $(\overline{d}, v_1, \ldots, v_n \Rightarrow t)$ in $Arg_{(AS, \mathcal{K}' \cup \{\overline{d}\})}$.
    given that $d \notin \mathcal{K}'$ and for each $y \in Y$ both $y \notin \mathcal{K}'$ and $\overline{y} \notin \mathcal{K}'$, there is an argument for t in $Arg_{(AS, \mathcal{K}')}$.
  - t is stable-unsatisfiable wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$, as we show next. Let $T'' = (AS, \mathcal{K}' \cup \{d\} \cup \mathcal{K}'')$ be an arbitrary AT such that $(AS, \mathcal{K}' \cup \{d\}) \sqsubseteq_Q T''$. Note that $\mathcal{K}'' \subseteq Y \cup \overline{Y}$. Given that there is no assignment τ_Y to variables in Y such that $\phi[\tau_X, \tau_Y]$ is True, there can be no argument for t based on $(c_1, \ldots, c_p \Rightarrow t)$ in $Arg_{T''}$. Furthermore, given that d is in the knowledge base of T'', there can be no argument for t based on $(\overline{d}, v_1, \ldots, v_n \Rightarrow t)$, in $Arg_{T''}$. Since there are no other rules for t and t is not in $\mathcal{Q}$, t must be unsatisfiable wrt T''. As T'' was chosen arbitrarily from all T''' such that $(AS, \mathcal{K}' \cup \{d\}) \sqsubseteq_Q T'''$, we derive that t is stable-unsatisfiable wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$.

  So by Lemma 1, d is unsatisfiable-relevant for t wrt T.
- From right to left: suppose that d is unsatisfiable-relevant for t wrt T. Then by Definition 15, there is some minimal stable-unsatisfiable future theory $T' = (AS, \mathcal{K}' \cup \{d\})$ wrt T and $\mathcal{Q}$. Given that t is stable-unsatisfiable wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$, t must be unsatisfiable wrt $(AS, \mathcal{K}' \cup \{d\})$. Then there is no argument for t in $Arg_{(AS, \mathcal{K}' \cup \{d\})}$.
  In addition, by minimality of $(AS, \mathcal{K}' \cup \{d\})$, t cannot be stable-unsatisfiable wrt $(AS, \mathcal{K}')$ and $\mathcal{Q}$. Then there must be some future argumentation theory of $(AS, \mathcal{K}')$ for which there is some argument for t having the observation-based argument for $\overline{d}$ as a subargument. This must have been the argument based on $(\overline{d}, v_1, \ldots, v_n \Rightarrow t)$. Given that this argument exists, for each $x \in X$, either $x \in \mathcal{K}'$ or $\overline{x} \in \mathcal{K}'$. In addition, for each $y \in Y$: $y \notin \mathcal{K}'$ and $\overline{y} \notin \mathcal{K}'$ (as these are contradictories of $\overline{d}$). Now let τ_X be the assignment to variables in X corresponding to $\mathcal{K}'$: for each $x \in X$, $\tau_X[x] =$ True iff $x \in \mathcal{K}'$ and $\tau_X[x] =$ False iff $\overline{x} \in \mathcal{K}'$.
  Next, we prove that for each assigment τ_Y to variables in Y, $\phi[\tau_X, \tau_Y]$ must be False. Suppose, towards a contradiction, that there is some τ_Y such that $\phi[\tau_X, \tau_Y]$ is True. Let $\mathcal{K}' \cup \{d\} \cup \mathcal{K}''$ be the corresponding knowledge base: $\mathcal{K}'' = \{y \in Y \mid \tau_Y[y] =$ True$\} \cup \{\overline{y} \in Y \mid \tau_Y[y] =$ False$\}$. Then, given that $\phi[\tau_X, \tau_Y]$ is True, there is an argument for t in $Arg_{(AS, \mathcal{K}' \cup \{d\} \cup \mathcal{K}'')}$, which implies that t is not unsatisfiable wrt $(AS, \mathcal{K}' \cup \{d\} \cup \mathcal{K}'')$, but then t was not stable-unsatisfiable wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$; contradiction. Therefore for each assigment τ_Y to variables in Y, $\phi[\tau_X, \tau_Y]$ must be False. In other words: (ϕ, X, Y) is a positive instance of Σ_2-ST. □

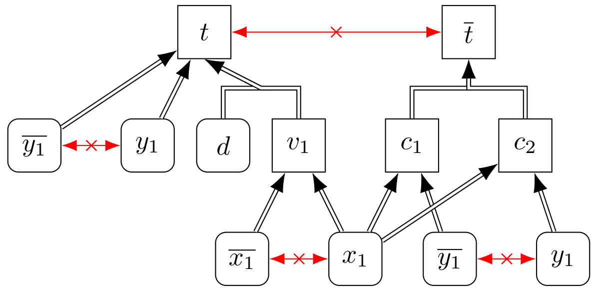

Figure 2: Illustration of the reduction used in Theorem 2 for the defended status for the formula $\phi = (x_1 \vee y_1) \wedge (x_i \vee \neg y_1)$. The queryables $\overline{y_1}$ and y_1 are displayed twice for readability.

Proof of Theorem 2 for defended status. Membership in Σ_2^P follows from Lemma 1: a positive instance can be verified with two calls to an oracle for stability, which is coNP-complete by Proposition 7.

For the Σ_2^P-hardness proof, we reduce from the Σ_2-ST problem of deciding for a formula ϕ in CNF, quantified over X and Y where $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_m\}$ are pairwise disjoint sets, if there exists an assignment τ_X to variables in X such that for each assignment τ_Y to variables in Y, $\phi[\tau_X, \tau_Y] = \text{False}$. Construct the following AT T and queryables $\mathcal{Q}$, with $C = c_1 \wedge \dots \wedge c_p$ the set of clauses in ϕ, and $\overline{X} = \{\overline{x} \mid x \in X\}$, $\overline{Y} = \{\overline{y} \mid y \in Y\}$ and $\overline{C} = \{\overline{c} \mid c \in C\}$. Let $V = \{v_i \mid x_i \in X\}$ and $\overline{V} = \{\overline{v_i} \mid x_i \in X\}$.

$$\begin{aligned}
\mathcal{Q} &= X \cup \overline{X} \cup Y \cup \overline{Y} \cup \{d, \overline{d}\}\\
\mathcal{L} &= \mathcal{Q} \cup C \cup \overline{C} \cup V \cup \overline{V} \cup \{t, \overline{t}\}\\
\overline{} &= \{(x, \overline{x}), (\overline{x}, x) \mid x \in X \cup Y \cup V \cup C \cup \{d, t\}\}\\
\mathcal{R} &= \{(d, v_1, \dots, v_n \Rightarrow t)\} \cup\\
&\quad \{(x \Rightarrow c) \mid x \in c\} \cup \{(\overline{x} \Rightarrow c) \mid \neg x \in c\} \cup\\
&\quad \{(y \Rightarrow c) \mid y \in c\} \cup \{(\overline{y} \Rightarrow c) \mid \neg y \in c\} \cup\\
&\quad \{(c_1, \dots, c_p \Rightarrow \overline{t})\} \cup\\
&\quad \{(x_i \Rightarrow v_i), (\overline{x_i} \Rightarrow v_i) \mid x_i \in X\}\\
&\quad \{(y \Rightarrow t), (\overline{y} \Rightarrow t) \mid y \in Y\}\\
\mathcal{K} &= \emptyset
\end{aligned}$$

Then $T = (AS, \mathcal{K})$ where $AS = (\mathcal{L}, \overline{}, \mathcal{R}, \leq)$ and $\mathcal{Q}$ can be constructed in polynomial time wrt ϕ.

Next, we prove that (ϕ, X, Y) is a positive instance of Σ_2-ST iff d is defended-relevant for t wrt T.

- From left to right: suppose that (ϕ, X, Y) is a positive instance of Σ_2-ST. Then there is some assignment to variables of X such that for each assignment to variables of Y, $\phi[\tau_X, \tau_Y]$ is False. Let τ_X be this assignment and construct a knowledge base $\mathcal{K}' = \{x \in X \mid \tau_X[x] = \text{True}\} \cup \{\overline{x} \in X \mid \tau_X[x] = \text{False}\}$. Note that $\mathcal{K}'$ must be consistent, as no $x \in X$ can be assigned both True and False by τ_X. Therefore $T \sqsubseteq_Q (AS, \mathcal{K}')$. Then:
  - t is not stable-defended wrt $(AS, \mathcal{K}')$ and $\mathcal{Q}$, because t is not defended wrt $(AS, \mathcal{K}')$: given that $d \notin \mathcal{K}'$ and for each $y \in Y$ both $y \notin \mathcal{K}'$ and $\overline{y} \notin \mathcal{K}'$, there is no argument for t in $Arg_{(AS, \mathcal{K}')}$.
  - t is stable-defended wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$, as we show next. Let $T'' = (AS, \mathcal{K}' \cup \{d\} \cup \mathcal{K}'')$ be an arbitrary AT such that $(AS, \mathcal{K}' \cup \{d\}) \sqsubseteq_Q T''$. Note that $\mathcal{K}'' \subseteq Y \cup \overline{Y}$. Given that there is no assignment τ_Y to variables in Y such that $\phi[\tau_X, \tau_Y]$ is True, there can be no argument for $\overline{t}$ in $Arg_{T''}$. Ot the other hand, there is at least one argument for t, based on $(d, v_1, \dots, v_n \Rightarrow t)$, in $Arg_{T''}$. Given that the argument for t is undefeated, t must be defended wrt T''. As T'' was chosen arbitrarily from all T''' such that $(AS, \mathcal{K}' \cup \{d\}) \sqsubseteq_Q T'''$, we derive that t is stable-defended wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$.

  So by Lemma 1, d is defended-relevant for t wrt T.
- From right to left: suppose that d is defended-relevant for t wrt T. Then by Definition 15, there is some minimal stable-defended future theory $T' = (AS, \mathcal{K}' \cup \{d\})$ wrt T and $\mathcal{Q}$. Given that t is stable-defended wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$, t must be defended wrt $(AS, \mathcal{K}' \cup \{d\})$. Then there must have been some argument for t and there is no argument for $\overline{t}$.

  In addition, by minimality of $(AS, \mathcal{K}' \cup \{d\})$, t cannot be stable-defended wrt $(AS, \mathcal{K}')$ and $\mathcal{Q}$. Then there must have been some argument for t having the observation-based argument for d as a subargument. This must have been the argument based on $(d, v_1, \dots, v_n \Rightarrow t)$. Given that this argument exists, for each $x \in X$, either $x \in \mathcal{K}'$ or $\overline{x} \in \mathcal{K}'$. In addition, for each $y \in Y$: $y \notin \mathcal{K}'$ and $\overline{y} \notin \mathcal{K}'$ (by minimality of $(AS, \mathcal{K}' \cup \{d\})$). Now let τ_X be the assignment to variables in X corresponding to $\mathcal{K}'$: for each $x \in X$, $\tau_X[x] = \text{True}$ iff $x \in \mathcal{K}'$ and $\tau_X[x] = \text{False}$ iff $\overline{x} \in \mathcal{K}'$.

  Next, we prove that for each assigment τ_Y to variables in Y, $\phi[\tau_X, \tau_Y]$ must be False. Suppose, towards a contradiction, that there is some τ_Y such that $\phi[\tau_X, \tau_Y]$ is True. Let $\mathcal{K}' \cup \{d\} \cup \mathcal{K}''$ be the corresponding knowledge base: $\mathcal{K}'' = \{y \in Y \mid \tau_Y[y] = \text{True}\} \cup \{\overline{y} \in Y \mid \tau_Y[y] = \text{False}\}$. Then, given that $\phi[\tau_X, \tau_Y]$ is True, there is an argument for $\overline{t}$ in $Arg_{(AS, \mathcal{K}' \cup \{d\} \cup \mathcal{K}'')}$, which implies that t is not defended wrt $(AS, \mathcal{K}' \cup \{d\} \cup \mathcal{K}'')$, but then t was not stable-defended wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$; contradiction. Therefore for each assigment τ_Y to variables in Y, $\phi[\tau_X, \tau_Y]$ must be False. In other words: (ϕ, X, Y) is a positive instance of Σ_2-ST. □

Proof of Theorem 2 for out status. The Σ_2^P-membership proof is similar to the proof for defended-relevance.

For the Σ_2^P-hardness proof, we reduce from the Σ_2-ST problem of deciding for a formula ϕ in CNF, quantified over X and Y where $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_m\}$ are pairwise disjoint sets, if there exists an assignment τ_X to variables in X such that for each assignment τ_Y to variables in Y, $\phi[\tau_X, \tau_Y] = \text{False}$. Construct the following AT T and queryables $\mathcal{Q}$, with $C = c_1 \wedge \dots \wedge c_p$ the set of clauses in ϕ, and $\overline{X} = \{\overline{x} \mid x \in X\}$, $\overline{Y} = \{\overline{y} \mid y \in Y\}$ and $\overline{C} = \{\overline{c} \mid$

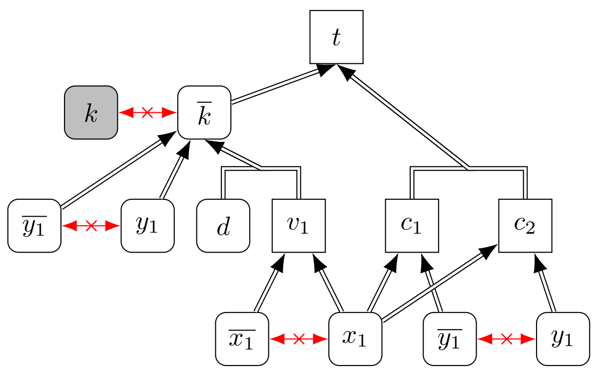

Figure 3: Illustration of the reduction used in Theorem 2 for the out status for the formula $\phi = (x_1 \vee y_1) \wedge (x_i \vee \neg y_1)$. The queryables $\overline{y_1}$ and y_1 are displayed twice for readability.

$c \in C\}$. Let $V = \{v_i \mid x_i \in X\}$ and $\overline{V} = \{\overline{v_i} \mid x_i \in X\}$.

$$\mathcal{Q} = X \cup \overline{X} \cup Y \cup \overline{Y} \cup \{d, \overline{d}, k, \overline{k}\}$$
$$\mathcal{L} = \mathcal{Q} \cup C \cup \overline{C} \cup V \cup \overline{V} \cup \{t, \overline{t}\}$$
$$\overline{} = \{(x, \overline{x}), (\overline{x}, x) \mid x \in X \cup Y \cup V \cup C \cup \{d, t, k\}\}$$
$$\begin{aligned}\mathcal{R} = \{&(d, v_1, \ldots, v_n \Rightarrow \overline{k})\} \cup \\ &\{(x \Rightarrow c) \mid x \in c\} \cup \{(\overline{x} \Rightarrow c) \mid \neg x \in c\} \cup \\ &\{(y \Rightarrow c) \mid y \in c\} \cup \{(\overline{y} \Rightarrow c) \mid \neg y \in c\} \cup \\ &\{(c_1, \ldots, c_p \Rightarrow t)\} \cup \\ &\{(x_i \Rightarrow v_i), (\overline{x_i} \Rightarrow v_i) \mid x_i \in X\} \\ &\{(y \Rightarrow \overline{k}), (\overline{y} \Rightarrow \overline{k}) \mid y \in Y\} \\ &\{(\overline{k} \Rightarrow t)\}\end{aligned}$$
$$\mathcal{K} = \{k\}$$

Then $T = (AS, \mathcal{K})$ where $AS = (\mathcal{L}, \overline{}, \mathcal{R}, \leq)$ and $\mathcal{Q}$ can be constructed in polynomial time wrt ϕ.

Next, we prove that (ϕ, X, Y) is a positive instance of Σ_2-ST iff d is out-relevant for t wrt T.

- From left to right: suppose that (ϕ, X, Y) is a positive instance of Σ_2-ST. Then there is some assignment to variables of X such that for each assignment to variables of Y, $\phi[\tau_X, \tau_Y]$ is False. Let τ_X be this assignment and construct a knowledge base $\mathcal{K}' = \{k\} \cup \{x \in X \mid \tau_X[x] = \text{True}\} \cup \{\overline{x} \in X \mid \tau_X[x] = \text{False}\}$. Note that $\mathcal{K} \subseteq \mathcal{K}'$ and that $\mathcal{K}'$ must be consistent, as no $x \in X$ can be assigned both True and False by τ_X. Therefore $T \sqsubseteq_Q (AS, \mathcal{K}')$. Then:
  - t is not stable-out wrt $(AS, \mathcal{K}')$ and $\mathcal{Q}$, because t is not out wrt $(AS, \mathcal{K}')$: given that $d \notin \mathcal{K}'$ and for each $y \in Y$ both $y \notin \mathcal{K}'$ and $\overline{y} \notin \mathcal{K}'$, there is no argument for t in $Arg_{(AS, \mathcal{K}')}$.
  - t is stable-out wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$, as we show next. Let $T'' = (AS, \mathcal{K}' \cup \{d\} \cup \mathcal{K}'')$ be an arbitrary AT such that $(AS, \mathcal{K}' \cup \{d\}) \sqsubseteq_Q T''$. Note that $\mathcal{K}'' \subseteq Y \cup \overline{Y}$. Given that there is no assignment τ_Y to variables in Y such that $\phi[\tau_X, \tau_Y]$ is True, there can be no argument for t based on $(c_1, \ldots, c_p \Rightarrow t)$ in $Arg_{T''}$. Ot the other hand, there is at least one argument for t, based on $(\overline{k} \Rightarrow t)$, in $Arg_{T''}$. In fact, every argument for t in $Arg_{T''}$ must be based on $(\overline{k} \Rightarrow t)$ and is therefore defeated by the observation-based (undefeated) argument k, which must be in $G(T'')$. Given that there is an argument for t in Arg_T but every argument for t in Arg_T is defeated by an argument in $G(T'')$, t must be out wrt T''. As T'' was chosen arbitrarily from all T''' such that $(AS, \mathcal{K}' \cup \{d\}) \sqsubseteq_Q T'''$, we derive that t is stable-out wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$.

  So by Lemma 1, d is out-relevant for t wrt T.
- From right to left: suppose that d is out-relevant for t wrt T. Then by Definition 15, there is some minimal stable-out future theory $T' = (AS, \mathcal{K}' \cup \{d\})$ wrt T and $\mathcal{Q}$. Given that t is stable-out wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$, t must be out wrt $(AS, \mathcal{K}' \cup \{d\})$. Then there must have been some argument for t in $Arg_{T'}$ and each argument for t in $Arg_{T'}$ is defeated by an argument in $G(T')$. This implies that there is no argument based on $(c_1, \ldots, c_p \Rightarrow t)$, as this argument for t would have been undefeated.
  In addition, by minimality of $(AS, \mathcal{K}' \cup \{d\})$, t cannot be stable-out wrt $(AS, \mathcal{K}')$ and $\mathcal{Q}$. So there is an AT $(AS, \mathcal{K}'')$ such that $(AS, \mathcal{K}') \sqsubseteq_Q (AS, \mathcal{K}'')$ and t is not out wrt $(AS, \mathcal{K}'')$. Note that there can be no argument based on $(c_1, \ldots, c_p \Rightarrow t)$ in $Arg_{(AS, \mathcal{K}'')}$, as that would imply that there would be an argument based on $(c_1, \ldots, c_p \Rightarrow t)$ in $Arg_{(AS, \mathcal{K}'' \cup \{d\})}$, while $(AS, \mathcal{K}' \cup \{d\}) \sqsubseteq_Q (AS, \mathcal{K}'' \cup \{d\})$ and t is supposed to be stable-out wrt $(AS, \mathcal{K}' \cup \{d\})$. Therefore t is not defended wrt $(AS, \mathcal{K}'')$. t cannot be blocked wrt $(AS, \mathcal{K}'')$ either, as there is no "equally strong" argument defeating any argument for t. This implies that t must be unsatisfiable wrt $(AS, \mathcal{K}'')$, which means that t was unsatisfiable wrt $(AS, \mathcal{K}')$ as well. Then there is no argument for t, so for each $y \in Y$: $y \notin \mathcal{K}'$ and $\overline{y} \notin \mathcal{K}'$.
  Given that there is an argument for t in $Arg_{T'}$ and $T' = (AS, \mathcal{K}' \cup \{d\})$, the argument for t in $Arg_{(AS, \mathcal{K}' \cup \{d\})}$ must have been based on $(d, v_1, \ldots, v_n \Rightarrow \overline{k})$ and $(\overline{k} \Rightarrow t)$. This implies that for each $x \in X$, either $x \in \mathcal{K}'$ or $\overline{x} \in \mathcal{K}'$. Now let τ_X be the assignment to variables in X corresponding to $\mathcal{K}'$: for each $x \in X$, $\tau_X[x] = \text{True}$ iff $x \in \mathcal{K}'$ and $\tau_X[x] = \text{False}$ iff $\overline{x} \in \mathcal{K}'$.
  Next, we prove that for each assigment τ_Y to variables in Y, $\phi[\tau_X, \tau_Y]$ must be False. Suppose, towards a contradiction, that there is some τ_Y such that $\phi[\tau_X, \tau_Y]$ is True. Let $\mathcal{K}' \cup \{d\} \cup \mathcal{K}^*$ be the corresponding knowledge base: $\mathcal{K}^* = \{y \in Y \mid \tau_Y[y] = \text{True}\} \cup \{\overline{y} \in Y \mid \tau_Y[y] = \text{False}\}$. Then, given that $\phi[\tau_X, \tau_Y]$ is True, there is an argument for t in $Arg_{(AS, \mathcal{K}' \cup \{d\} \cup \mathcal{K}^*)}$, which implies that t is defended wrt $(AS, \mathcal{K}' \cup \{d\} \cup \mathcal{K}^*)$, but then t was not stable-out wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$; contradiction. Therefore for each assigment τ_Y to variables in Y, $\phi[\tau_X, \tau_Y]$ must be False. In other words: (ϕ, X, Y) is a positive instance of Σ_2-ST. □

Proof of Theorem 2 for blocked status. The Σ_2^P-membership proof is similar to the proof for defended-

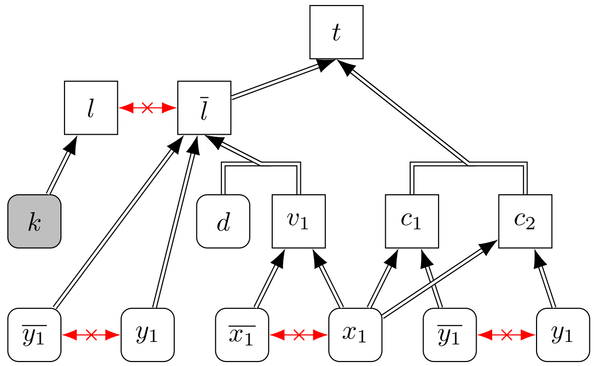

Figure 4: Illustration of the reduction used in Theorem 2 for the blocked status for the formula $\phi = (x_1 \vee y_1) \wedge (x_i \vee \neg y_1)$. The queryables $\overline{y_1}$ and y_1 are displayed twice for readability.

relevance.

For the Σ_2^P-hardness proof, we reduce from the Σ_2-ST problem of deciding for a formula ϕ in CNF, quantified over X and Y where $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_m\}$ are pairwise disjoint sets, if there exists an assignment τ_X to variables in X such that for each assignment τ_Y to variables in Y, $\phi[\tau_X, \tau_Y] = \text{False}$. Construct the following AT T and queryables $\mathcal{Q}$, with $C = c_1 \wedge \ldots \wedge c_p$ the set of clauses in ϕ, and $\overline{X} = \{\overline{x} \mid x \in X\}$, $\overline{Y} = \{\overline{y} \mid y \in Y\}$ and $\overline{C} = \{\overline{c} \mid c \in C\}$. Let $V = \{v_i \mid x_i \in X\}$ and $\overline{V} = \{\overline{v_i} \mid x_i \in X\}$.

$$\mathcal{Q} = X \cup \overline{X} \cup Y \cup \overline{Y} \cup \{d, \overline{d}, k, \overline{k}\}$$
$$\mathcal{L} = \mathcal{Q} \cup C \cup \overline{C} \cup V \cup \overline{V} \cup \{t, \overline{t}, l, \overline{l}\}$$
$$\overline{} = \{(x, \overline{x}), (\overline{x}, x) \mid x \in X \cup Y \cup V \cup C \cup \{d, t, k, l\}\}$$
$$\begin{aligned}\mathcal{R} = \{&(d, v_1, \ldots, v_n \Rightarrow \overline{l})\} \cup \\ &\{(x \Rightarrow c) \mid x \in c\} \cup \{(\overline{x} \Rightarrow c) \mid \neg x \in c\} \cup \\ &\{(y \Rightarrow c) \mid y \in c\} \cup \{(\overline{y} \Rightarrow c) \mid \neg y \in c\} \cup \\ &\{(c_1, \ldots, c_p \Rightarrow t)\} \cup \\ &\{(x_i \Rightarrow v_i), (\overline{x_i} \Rightarrow v_i) \mid x_i \in X\} \\ &\{(y \Rightarrow \overline{l}), (\overline{y} \Rightarrow \overline{l}) \mid y \in Y\} \\ &\{(k \Rightarrow l), (\overline{l} \Rightarrow t)\}\end{aligned}$$
$$\mathcal{K} = \{k\}$$

Then $T = (AS, \mathcal{K})$ where $AS = (\mathcal{L}, \overline{}, \mathcal{R}, \leq)$ and $\mathcal{Q}$ can be constructed in polynomial time wrt ϕ.

We prove that (ϕ, X, Y) is a positive instance of Σ_2-ST iff d is blocked-relevant for t wrt T.

- From left to right: suppose that (ϕ, X, Y) is a positive instance of Σ_2-ST. Then there is some assignment to variables of X such that for each assignment to variables of Y, $\phi[\tau_X, \tau_Y]$ is False. Let τ_X be this assignment and construct a knowledge base $\mathcal{K}' = \{k\} \cup \{x \in X \mid \tau_X[x] = \text{True}\} \cup \{\overline{x} \in X \mid \tau_X[x] = \text{False}\}$. Note that $\mathcal{K} \subseteq \mathcal{K}'$ and that $\mathcal{K}'$ must be consistent, as no $x \in X$ can be assigned both True and False by τ_X. Therefore $T \sqsubseteq_Q (AS, \mathcal{K}')$. Then:
  - t is not stable-blocked wrt $(AS, \mathcal{K}')$ and $\mathcal{Q}$, because t is not blocked wrt $(AS, \mathcal{K}')$: given that $d \notin \mathcal{K}'$ and for each $y \in Y$ both $y \notin \mathcal{K}'$ and $\overline{y} \notin \mathcal{K}'$, there is no argument for t in $Arg_{(AS, \mathcal{K}')}$.
  - t is stable-blocked wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$, as we show next. Let $T'' = (AS, \mathcal{K}' \cup \{d\} \cup \mathcal{K}'')$ be an arbitrary AT such that $(AS, \mathcal{K}' \cup \{d\}) \sqsubseteq_Q T''$. Note that $\mathcal{K}'' \subseteq Y \cup \overline{Y}$. Given that there is no assignment τ_Y to variables in Y such that $\phi[\tau_X, \tau_Y]$ is True, there can be no argument for t based on $(c_1, \ldots, c_p \Rightarrow t)$ in $Arg_{T''}$. Ot the other hand, there is at least one argument for t, based on $(\overline{l} \Rightarrow t)$, in $Arg_{T''}$. In fact, every argument for t in $Arg_{T''}$ must be based on $(\overline{l} \Rightarrow t)$ and is therefore defeated by the argument based on $(k \Rightarrow l)$, which is defeated by all arguments for $\overline{l}$. Given that there is an argument for t in Arg_T but every argument for t in Arg_T is defeated by an argument in $Arg_{T''}$ that is not in or defeated by any argument in $G(T'')$, t must be blocked wrt T''. As T'' was chosen arbitrarily from all T''' such that $(AS, \mathcal{K}' \cup \{d\}) \sqsubseteq_Q T'''$, we derive that t is stable-blocked wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$.

  So by Lemma 1, d is blocked-relevant for t wrt T.
- From right to left: suppose that d is blocked-relevant for t wrt T. Then by Definition 15, there is some minimal stable-blocked future theory $T' = (AS, \mathcal{K}' \cup \{d\})$ wrt T and $\mathcal{Q}$. Given that t is stable-blocked wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$, t must be blocked wrt $(AS, \mathcal{K}' \cup \{d\})$. Then there must have been some argument for t in $Arg_{T'}$ and each argument for t in $Arg_{T'}$ is defeated by an argument in $Arg_{T'}$. This implies that there is no argument based on $(c_1, \ldots, c_p \Rightarrow t)$, as this argument for t would have been undefeated.
  In addition, by minimality of $(AS, \mathcal{K}' \cup \{d\})$, t cannot be stable-blocked wrt $(AS, \mathcal{K}')$ and $\mathcal{Q}$. So there is an AT $(AS, \mathcal{K}'')$ such that $(AS, \mathcal{K}') \sqsubseteq_Q (AS, \mathcal{K}'')$ and t is not blocked wrt $(AS, \mathcal{K}'')$. Note that there can be no argument based on $(c_1, \ldots, c_p \Rightarrow t)$ in $Arg_{(AS, \mathcal{K}'')}$, as that would imply that there would be an argument based on $(c_1, \ldots, c_p \Rightarrow t)$ in $Arg_{(AS, \mathcal{K}'' \cup \{d\})}$, while $(AS, \mathcal{K}' \cup \{d\}) \sqsubseteq_Q (AS, \mathcal{K}'' \cup \{d\})$ and t is supposed to be stable-blocked wrt $(AS, \mathcal{K}' \cup \{d\})$. Therefore t is not defended wrt $(AS, \mathcal{K}'')$. t cannot be out wrt $(AS, \mathcal{K}'')$ either, as there is no "stronger" argument defeating any argument for t. This implies that t must be unsatisfiable wrt $(AS, \mathcal{K}'')$, which means that t was unsatisfiable wrt $(AS, \mathcal{K}')$ as well. Then there is no argument for t, so for each $y \in Y$: $y \notin \mathcal{K}'$ and $\overline{y} \notin \mathcal{K}'$.
  Given that there is an argument for t in $Arg_{T'}$ and $T' = (AS, \mathcal{K}' \cup \{d\})$, the argument for t in $Arg_{(AS, \mathcal{K}' \cup \{d\})}$ must have been based on $(d, v_1, \ldots, v_n \Rightarrow \overline{l})$ and $(\overline{l} \Rightarrow t)$. This implies that for each $x \in X$, either $x \in \mathcal{K}'$ or $\overline{x} \in \mathcal{K}'$. Now let τ_X be the assignment to variables in X corresponding to $\mathcal{K}'$: for each $x \in X$, $\tau_X[x] = \text{True}$ iff $x \in \mathcal{K}'$ and $\tau_X[x] = \text{False}$ iff $\overline{x} \in \mathcal{K}'$.
  Next, we prove that for each assigment τ_Y to variables in Y, $\phi[\tau_X, \tau_Y]$ must be False. Suppose, towards a contradiction, that there is some τ_Y such that $\phi[\tau_X, \tau_Y]$ is True.

Let $\mathcal{K}' \cup \{d\} \cup \mathcal{K}^*$ be the corresponding knowledge base: $\mathcal{K}^* = \{y \in Y \mid \tau_Y[y] = \text{True}\} \cup \{\overline{y} \in Y \mid \tau_Y[y] = \text{False}\}$. Then, given that $\phi[\tau_X, \tau_Y]$ is True, there is an argument for t in $Arg_{(AS, \mathcal{K}' \cup \{d\} \cup \mathcal{K}^*)}$, which implies that t is defended wrt $(AS, \mathcal{K}' \cup \{d\} \cup \mathcal{K}^*)$, but then t was not stable-blocked wrt $(AS, \mathcal{K}' \cup \{d\})$ and $\mathcal{Q}$; contradiction. Therefore for each assigment τ_Y to variables in Y, $\phi[\tau_X, \tau_Y]$ must be False. In other words: (ϕ, X, Y) is a positive instance of Σ_2-ST. □

Proposition 9. *Let $T = (AS, \mathcal{K})$ be an AT, $\mathcal{Q}$ a set of queryables and j a justification status. Given $l \in \mathcal{L}$ and $q \in Q$ where $q \notin \mathcal{K}$ and $\overline{q} \cap \mathcal{K} = \emptyset$*

- *if $T' = (AS, \mathcal{K}') \sqsupseteq_Q T$ such that l is not stable-j wrt T', then for each $\mathcal{K}'' \subseteq \mathcal{K}'$, l is not stable-j wrt $(AS, \mathcal{K}'')$.*
- *if $T' = (AS, \mathcal{K}') \sqsupseteq_Q T$ such that l is stable-j wrt T' and $q \notin \mathcal{K}'$, then for each consistent $\mathcal{K}'' \supseteq \mathcal{K}'$, l is stable-j wrt $(AS, \mathcal{K}'' \setminus \{q\})$.*

Proof. For the firts item, suppose that $T' = (AS, \mathcal{K}')$ is a future AT with $T \sqsubseteq_Q T'$ such that l is not stable-j wrt T'. Then there is some T^* with $(AS, \mathcal{K}') \sqsubseteq_Q T^*$ such that l is not j wrt T*. For each $\mathcal{K}'' \subseteq \mathcal{K}'$ it must be that $(AS, \mathcal{K}'') \sqsubseteq_Q T^*$ as well; therefore l is not stable-j wrt $(AS, \mathcal{K}'')$ and $\mathcal{Q}$.

For the second item, suppose that $T' = (AS, \mathcal{K}')$ is a future AT with $T \sqsubseteq_Q T'$ such that l is stable-j wrt T' and $q \notin \mathcal{K}'$. Then for each $\mathcal{K}''$ such that $\mathcal{K}' \subseteq \mathcal{K}''$, l is stable-j wrt $(AS, \mathcal{K}'')$, and since $q \notin \mathcal{K}'$, $\mathcal{K}'' \setminus \{q\} \subseteq \mathcal{K}'$. Thus, l is stable-j wrt $(AS, \mathcal{K}'' \setminus \{q\})$.

□

Benchmark Details

As benchmarks, we consider both real-world and synthetic data. *For real-world benchmarks*, we generated instances for the stability and relevance problems based on the argumentation system $AS = (\mathcal{L}, \overline{\ \ }, \mathcal{R}, \leq)$ and set of queryables $\mathcal{Q}$ used in an inquiry system for the intake of online trade fraud at the Netherlands Police. In this setting, $|\mathcal{L}| = 60$, $|\mathcal{Q}| = 30$ and $|\mathcal{R}| = 43$. All literals in $\mathcal{L} \setminus \mathcal{Q}$ have a single contradictory: their negation. Considering the queryables in $\mathcal{Q}$, 19 queryables have a single contradictory; three literals have two contradictories; seven literals have three contradictories and one literal has four contradictories. All rules are equally preferred: $\leq = \emptyset$. Most rules have one (13) or two (14) antecedents; four rules have three antecedents; eight rules have four and the remaining four rules have five antecedents. The rules are defined in such a way that they form a tree-like structure, without (support) cycles. Thanks to this structure, each literal can be assigned a finite layer, which informally is the largest number of rule applications to reach a queryable. Out of the 60 literals, 40 have layer 0 (this includes all 30 queryables); 6 have layer 1; 5 have layer 2; 6 have layer 3 and 3 have layer 4. Six of the literals are considered as "topics": these are the literals for which the stability status is needed and are typically situated in high layers (2, 3 or 4). To generate stability instances, we obtained knowledge bases by randomly sampling 25 consistent subsets of each size between 1 and 15 from $\mathcal{Q}$, as well as the empty knowledge base. For each of the six topics, we test all four stability statuses. Similarly, instances for relevance were created for each combination of stability instances and each queryable in $\mathcal{Q}$.

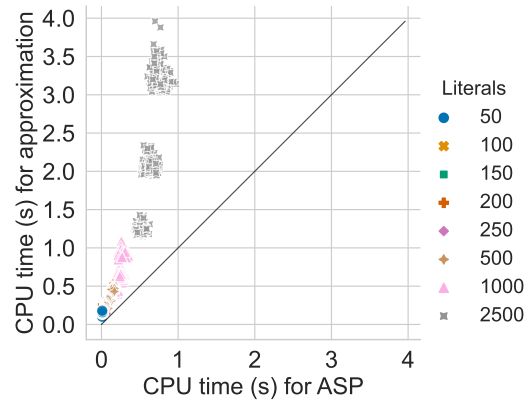

Figure 5: Runtime comparison of our ASP approach and the approximation algorithm for finding the stability status of all literals.

For a further scalability study, we also consider *synthetic data*. For this, we generated argumentation theories and queryable sets that are parametrised by the size of the language $|\mathcal{L}|$ and rule set size $|\mathcal{R}|$. We generated instances with the following settings:

- Language size $|\mathcal{L}|$: flexible, in $[50, 100, 150, 200, 250, 500, 1000, 2500]$
- Rule set size $|\mathcal{R}|$: flexible, in $[0.5 \cdot |\mathcal{L}|, |\mathcal{L}|, 1.5 \cdot |\mathcal{L}|]$
- Rule antecedent distribution: $\{1 : |\mathcal{R}|/3, 2 : |\mathcal{R}|/3, 3 : |\mathcal{R}|/9, 4 : |\mathcal{R}|/9, 5 : |\mathcal{R}|/9\}$
- Literal layer distribution: $\{0 : \frac{2}{3} \cdot |\mathcal{L}|, 1 : |\mathcal{L}|/10, 2 : |\mathcal{L}|/10, 3 : |\mathcal{L}|/10, 4 : \text{rest}\}$
- Queryable/literal ratio $|\mathcal{Q}|/|\mathcal{L}|$: 0.5
- Axiom/queryable ratio $|\mathcal{K}|/|\mathcal{Q}|$: 0.5

Similarly as for the fraud data set, all queryables have layer 0; that is: there are no rules for queryable literals. We obtained a partial ordering for the rule preferences by considering all rules with contradictory consequents and sampling half of them into $\leq$. All literals with layers 3 or 4 are considered as topics.

Additional Results

Figure 5 shows that our exact ASP-based approach is able to detect the stability status of all literals faster than the existing approximation algorithm for all instances in the synthetic dataset.

ASP in Brief

An ASP program π consists of rules r of the form $b_0 \leftarrow b_1, \ldots, b_k, \mathit{not}\, b_{k+1}, \ldots, \ \mathit{not}\, b_m$, where each b_i is an atom. A rule is positive if $k = m$ and a fact if $m = 0$. A literal is an atom b_i or $\mathit{not}\, b_i$. A rule without head b_0 is a constraint and a shorthand for $a \leftarrow b_1, \ldots, b_k, \mathit{not}\, b_{k+1}, \ldots, \ \mathit{not}\, b_m, \mathit{not}\, a$ for a fresh a. An atom b_i is $p(t_1, \ldots, t_n)$ with each t_j either a constant or a variable. An answer set program is ground if it is free of variables. For a non-ground program, GP is the set of rules obtained by applying all possible substitutions from the variables to the set of constants appearing in the program. An interpretation I, i.e., a subset of all the ground atoms, satisfies a positive rule $r = h \leftarrow b_1, \ldots, b_k$ iff all positive body elements $b_1, \ldots, b_k$ being in I implies that the head atom is in I. For a program π consisting only of positive rules, let $Cl(\pi)$ be the uniquely determined interpretation I that satisfies all rules in π and no subset of I satisfies all rules in π. Interpretation I is an answer set of a ground program π if $I = Cl(\pi^I)$ where $\pi^I = \{(h \leftarrow b_1, \ldots, b_k) \mid (h \leftarrow b_1, \ldots, b_k, not\, b_{k+1}, \ldots, not\, b_m) \in \pi, \{b_{k+1}, \ldots, b_m\} \cap I = \emptyset\}$ is the reduct; and of a non-ground program π if I is an answer set of GP of π.