FAKULTÄT
FÜR !NFORMATIK
Faculty of Informatics

dbai

# Verifiability of Argumentation Semantics

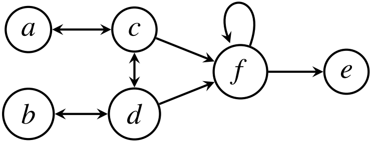
Ringo Baumann, Thomas Linsbichler, Stefan Woltran

6th International Conference on Computational Models of Argument
Potsdam, Germany

September 16, 2016

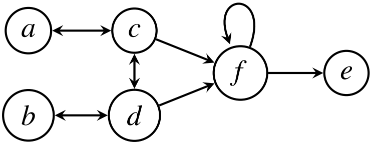
# Introduction

- Abstract Argumentation Framework (AF) [Dung, 1995]:



# Introduction

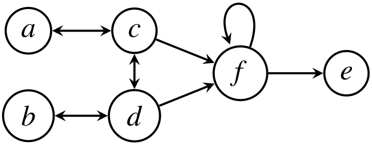
- Abstract Argumentation Framework (AF) [Dung, 1995]:




- Evaluation: argumentation semantics
- Extension: set of jointly acceptable arguments

# Introduction

- Abstract Argumentation Framework (AF) [Dung, 1995]:

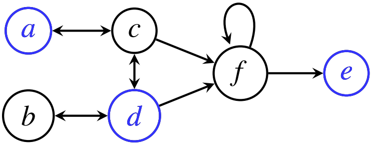



- Evaluation: argumentation semantics
- Extension: set of jointly acceptable arguments

$$\mathit{stb}(\mathcal{F}) =$$

# Introduction

- Abstract Argumentation Framework (AF) [Dung, 1995]:

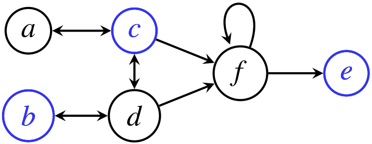



- Evaluation: argumentation semantics
- Extension: set of jointly acceptable arguments

$$\mathit{stb}(\mathcal{F}) = \{\{a, d, e\},$$

# Introduction

- Abstract Argumentation Framework (AF) [Dung, 1995]:

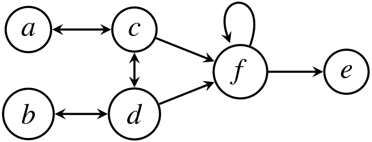



- Evaluation: argumentation semantics
- Extension: set of jointly acceptable arguments

$$\mathit{stb}(\mathcal{F}) = \{\{a, d, e\}, \{b, c, e\}\}$$

# Introduction

- Abstract Argumentation Framework (AF) [Dung, 1995]:




- Evaluation: argumentation semantics
- Extension: set of jointly acceptable arguments

$$\mathit{stb}(\mathcal{F}) = \{\{a, d, e\}, \{b, c, e\}\}$$

- Further semantics: preferred, complete, semi-stable, stage, ...

# Introduction

- Conflict-freeness: basic requirement for argumentation semantics.

# Introduction

- Conflict-freeness: basic requirement for argumentation semantics.

## Example

- Given conflict-free sets $\emptyset$, $\{a\}$, $\{b\}$.

# Introduction

- Conflict-freeness: basic requirement for argumentation semantics.

## Example

- Given conflict-free sets $\emptyset$, $\{a\}$, $\{b\}$.
- Can we compute semantics based on this?

# Introduction

- Conflict-freeness: basic requirement for argumentation semantics.

## Example

- Given conflict-free sets $\emptyset$, $\{a\}$, $\{b\}$.
- Can we compute semantics based on this?
  $\Rightarrow$ Yes, naive semantics (maximal conflict-free sets)

# Introduction

- Conflict-freeness: basic requirement for argumentation semantics.

## Example

- Given conflict-free sets $\emptyset$, $\{a\}$, $\{b\}$.
- Can we compute semantics based on this?
  $\Rightarrow$ Yes, naive semantics (maximal conflict-free sets)

# Introduction

- Conflict-freeness: basic requirement for argumentation semantics.

## Example

- Given conflict-free sets $\emptyset$, $\{a\}$, $\{b\}$.
- Can we compute semantics based on this?
  $\Rightarrow$ Yes, naive semantics (maximal conflict-free sets)

$$\mathcal{F}: a \rightarrow b \qquad \mathcal{G}: a \leftarrow b \qquad \mathcal{H}: a \rightleftarrows b \qquad \cdots$$

$$na(\mathcal{F}) = na(\mathcal{G}) = na(\mathcal{H}) = \cdots = \{\{a\}, \{b\}\}.$$

# Introduction

- Conflict-freeness: basic requirement for argumentation semantics.

## Example

- Given conflict-free sets $\emptyset$, $\{a\}$, $\{b\}$.
- Can we compute semantics based on this?
  $\Rightarrow$ Yes, naive semantics (maximal conflict-free sets)

$$\mathcal{F}: a \rightarrow b \qquad \mathcal{G}: a \leftarrow b \qquad \mathcal{H}: a \leftrightarrows b \qquad \cdots$$

$$\mathit{na}(\mathcal{F}) = \mathit{na}(\mathcal{G}) = \mathit{na}(\mathcal{H}) = \cdots = \{\{a\}, \{b\}\}.$$

$\Rightarrow$ not stage semantics (range-maximal conflict-free sets)

$$\mathit{stg}(\mathcal{F}) = \{\{a\}\}, \mathit{stg}(\mathcal{G}) = \{\{b\}\}, \mathit{stg}(\mathcal{H}) = \{\{a\}, \{b\}\}.$$

# Introduction

- Conflict-freeness: basic requirement for argumentation semantics.

## Example (2)

- Given conflict free sets + their range: $(\emptyset, \emptyset)$, $(\{a\}, \{a, b\})$, $(\{b\}, \{b\})$

# Introduction

- Conflict-freeness: basic requirement for argumentation semantics.

## Example (2)

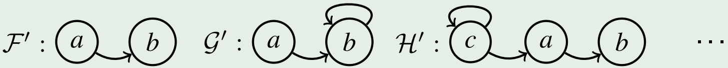
- Given conflict free sets + their range: $(\emptyset, \emptyset)$, $(\{a\}, \{a, b\})$, $(\{b\}, \{b\})$
- Which semantics can we compute based on this?
  $\Rightarrow$ enough to compute stage semantics (range-maximal conflict-free sets)

# Introduction

- Conflict-freeness: basic requirement for argumentation semantics.

## Example (2)

- Given conflict free sets + their range: $(\emptyset, \emptyset)$, $(\{a\}, \{a, b\})$, $(\{b\}, \{b\})$
- Which semantics can we compute based on this?
  $\Rightarrow$ enough to compute stage semantics (range-maximal conflict-free sets)

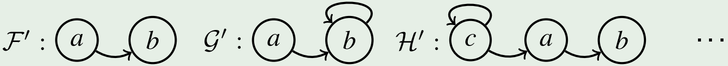


# Introduction

- Conflict-freeness: basic requirement for argumentation semantics.

## Example (2)

- Given conflict free sets + their range: $(\emptyset, \emptyset)$, $(\{a\}, \{a, b\})$, $(\{b\}, \{b\})$
- Which semantics can we compute based on this?
  $\Rightarrow$ enough to compute stage semantics (range-maximal conflict-free sets)



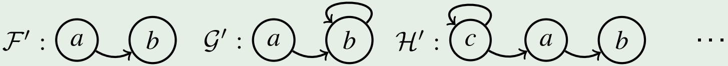

*stg*$(\mathcal{F}') =$ *stg*$(\mathcal{G}') =$ *stg*$(\mathcal{H}') = \cdots = \{\{a\}\}$.

# Introduction

- Conflict-freeness: basic requirement for argumentation semantics.

## Example (2)

- Given conflict free sets + their range: $(\emptyset, \emptyset)$, $(\{a\}, \{a, b\})$, $(\{b\}, \{b\})$
- Which semantics can we compute based on this?
  $\Rightarrow$ enough to compute stage semantics (range-maximal conflict-free sets)




*stg*$(\mathcal{F}') =$ *stg*$(\mathcal{G}') =$ *stg*$(\mathcal{H}') = \cdots = \{\{a\}\}$.

$\Rightarrow$ not preferred semantics (maximal admissible sets)

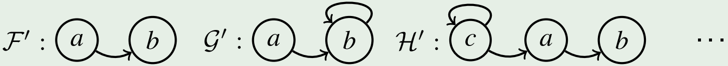
*pr*$(\mathcal{F}') =$ *pr*$(\mathcal{G}') = \{\{a\}\}$, *pr*$(\mathcal{H}') = \{\emptyset\}$.

# Introduction

- Conflict-freeness: basic requirement for argumentation semantics.

## Example (2)

- Given conflict free sets + their range: $(\emptyset, \emptyset)$, $(\{a\}, \{a, b\})$, $(\{b\}, \{b\})$
- Which semantics can we compute based on this?
  $\Rightarrow$ enough to compute stage semantics (range-maximal conflict-free sets)




$$stg(\mathcal{F}') = stg(\mathcal{G}') = stg(\mathcal{H}') = \cdots = \{\{a\}\}.$$

$\Rightarrow$ not preferred semantics (maximal admissible sets)

$$pr(\mathcal{F}') = pr(\mathcal{G}') = \{\{a\}\}, \ pr(\mathcal{H}') = \{\emptyset\}.$$

- Which information on top of conflict-free sets has to be added in order to compute a certain semantics?

# Introduction

- Systematic comparison of argumentation semantics
  - Computational complexity
    [Dunne and Bench-Capon, 2002, Dvořák and Woltran, 2010]
  - Principle-based evaluation [Baroni and Giacomin, 2007]

# Introduction

- Systematic comparison of argumentation semantics
  - Computational complexity
    [Dunne and Bench-Capon, 2002, Dvořák and Woltran, 2010]
  - Principle-based evaluation [Baroni and Giacomin, 2007]

$\Rightarrow$ Hierarchy of verification classes

$\Rightarrow$ Classification of semantics into these classes

$\Rightarrow$ Each "rational" semantics is exactly verifiable by one of these classes

# Introduction

- Systematic comparison of argumentation semantics
  - Computational complexity [Dunne and Bench-Capon, 2002, Dvořák and Woltran, 2010]
  - Principle-based evaluation [Baroni and Giacomin, 2007]

$\Rightarrow$ Hierarchy of verification classes
$\Rightarrow$ Classification of semantics into these classes
$\Rightarrow$ Each “rational” semantics is exactly verifiable by one of these classes

- Strong equivalence
  - Central notion in non-monotonic reasoning [Lifschitz et al., 2001, Turner, 2004, Truszczynski, 2006, Baumann and Strass, 2016]
  - Studied for most argumentation semantics [Oikarinen and Woltran, 2011, Baumann, 2016]

# Introduction

- Systematic comparison of argumentation semantics
  - Computational complexity [Dunne and Bench-Capon, 2002, Dvořák and Woltran, 2010]
  - Principle-based evaluation [Baroni and Giacomin, 2007]

⇒ Hierarchy of verification classes
⇒ Classification of semantics into these classes
⇒ Each “rational” semantics is exactly verifiable by one of these classes

- Strong equivalence
  - Central notion in non-monotonic reasoning [Lifschitz et al., 2001, Turner, 2004, Truszczynski, 2006, Baumann and Strass, 2016]
  - Studied for most argumentation semantics [Oikarinen and Woltran, 2011, Baumann, 2016]

⇒ Missing results for naive and strong admissible semantics
⇒ Characterization theorems for intermediate semantics

# Background

## Definition

An argumentation framework (AF) is a pair $(A, R)$ where

- $A \subseteq \mathcal{U}$ is a finite set of arguments and
- $R \subseteq A \times A$ is the attack relation representing conflicts.

## Definition

Given an AF $\mathcal{F} = (A, R)$ and $S \subseteq A$,

- $S$ is conflict-free ($S \in cf(\mathcal{F})$) if $\forall a, b \in S : (a, b) \notin R$.
- $a \in A$ is defended by $S$ if $\forall b \in A : (b, a) \in R \Rightarrow \exists c \in S : (c, b) \in R$
- $S^+_\mathcal{F} = S \cup \{a \mid \exists b \in S : (b, a) \in R\}$ (the range of S)
- $S^-_\mathcal{F} = S \cup \{a \mid \exists b \in S : (a, b) \in R\}$ (the anti-range of S)

# Background

## Semantics

Given an AF $\mathcal{F} = (A, R)$, a set $S \subseteq A$ is

- admissible set if $S \in \mathit{cf}(\mathcal{F})$ and each $a \in S$ is defended by $S$,
- complete extension if $S \in \mathit{ad}(\mathcal{F})$ and $a \in S$ if $a$ is defended by $S$,
- naive extension if $S \in \mathit{cf}(\mathcal{F})$ and $\nexists T \in \mathit{cf}(\mathcal{F}) : T \supset S$,
- stable extension if $S \in \mathit{cf}(\mathcal{F})$ and $S^+_{\mathcal{F}} = A$,
- stage extension if $S \in \mathit{cf}(\mathcal{F})$ and $\nexists T \in \mathit{cf}(\mathcal{F}) : T^+_{\mathcal{F}} \supset S^+_{\mathcal{F}}$,
- preferred, grounded, semi-stable, ideal, eager, strongly admissible extensions

# Verifiability

## Definition

We call a function $\mathfrak{r}^x : 2^{\mathcal{U}} \times 2^{\mathcal{U}} \to \left(2^{\mathcal{U}}\right)^n$ which is expressible via basic set operations only[a] neighborhood function.

[a] $\mathfrak{r}^x(A, B)$ is in the language $X ::= A \mid B \mid (X \cup X) \mid (X \cap X) \mid (X \setminus X)$

# Verifiability

## Definition

We call a function $\mathfrak{r}^x : 2^{\mathcal{U}} \times 2^{\mathcal{U}} \to \left(2^{\mathcal{U}}\right)^n$ which is expressible via basic set operations only[a] neighborhood function.

The verification class induced by $\mathfrak{r}^x$ maps each AF $\mathcal{F}$ to

$$\widetilde{\mathcal{F}}^x = \left\{ \left( S, \mathfrak{r}^x(S^+_{\mathcal{F}}, S^-_{\mathcal{F}}) \right) \mid S \in \mathit{cf}(\mathcal{F}) \right\}.$$

[a] $\mathfrak{r}^x(A, B)$ is in the language $X ::= A \mid B \mid (X \cup X) \mid (X \cap X) \mid (X \setminus X)$

# Verifiability

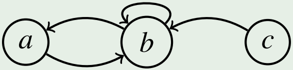
## Definition

We call a function $\mathfrak{r}^x : 2^{\mathcal{U}} \times 2^{\mathcal{U}} \to \left(2^{\mathcal{U}}\right)^n$ which is expressible via basic set operations only[a] neighborhood function.
The verification class induced by $\mathfrak{r}^x$ maps each AF $\mathcal{F}$ to

$$\widetilde{\mathcal{F}}^x = \left\{ \left(S, \mathfrak{r}^x(S^+_{\mathcal{F}}, S^-_{\mathcal{F}})\right) \mid S \in \mathit{cf}(\mathcal{F}) \right\}.$$

[a] $\mathfrak{r}^x(A, B)$ is in the language $X ::= A \mid B \mid (X \cup X) \mid (X \cap X) \mid (X \setminus X)$

## Example

$\mathcal{F}$ :



$\mathfrak{r}^+ : \mathfrak{r}^x(A, B) = A$
$\widetilde{\mathcal{F}}^+ = \{(\emptyset, \emptyset), (\{a\}, \{a, b\}), (\{c\}, \{b, c\}), (\{a, c\}, \{a, b, c\})\}$

# Verifiability

## Definition

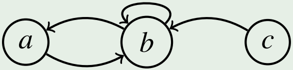
We call a function $\mathfrak{r}^x : 2^{\mathcal{U}} \times 2^{\mathcal{U}} \to \left(2^{\mathcal{U}}\right)^n$ which is expressible via basic set operations only[a] neighborhood function.
The verification class induced by $\mathfrak{r}^x$ maps each AF $\mathcal{F}$ to

$$\widetilde{\mathcal{F}}^x = \left\{ \left(S, \mathfrak{r}^x(S^+_{\mathcal{F}}, S^-_{\mathcal{F}})\right) \mid S \in \mathit{cf}(\mathcal{F}) \right\}.$$

[a] $\mathfrak{r}^x(A, B)$ is in the language $X ::= A \mid B \mid (X \cup X) \mid (X \cap X) \mid (X \setminus X)$

## Example

$\mathcal{F}:$




$\mathfrak{r}^+ : \mathfrak{r}^x(A, B) = A$
$\widetilde{\mathcal{F}}^+ = \{(\emptyset, \emptyset), (\{a\}, \{a, b\}), (\{c\}, \{b, c\}), (\{a, c\}, \{a, b, c\})\}$

$\mathfrak{r}^{-\pm} : \mathfrak{r}^x(A, B) = (B, A \setminus B)$
$\widetilde{\mathcal{F}}^{-\pm} = \{(\emptyset, \emptyset, \emptyset), (\{a\}, \{a, b\}, \emptyset), (\{c\}, \{c\}, \{b\}), (\{a, c\}, \{a, b, c\}, \emptyset)\}$

# Verifiability

- Neighborhood functions for $n = 1$:

$$\begin{aligned}
\mathfrak{r}^{\epsilon}(A,B) &= \emptyset \\
\mathfrak{r}^{+}(A,B) &= A \\
\mathfrak{r}^{-}(A,B) &= B \\
\mathfrak{r}^{\mp}(A,B) &= B \setminus A \\
\mathfrak{r}^{\pm}(A,B) &= A \setminus B \\
\mathfrak{r}^{\cap}(A,B) &= A \cap B \\
\mathfrak{r}^{\cup}(A,B) &= A \cup B \\
\mathfrak{r}^{\Delta}(A,B) &= (A \cup B) \setminus (A \cap B)
\end{aligned}$$

- $2^7 + 1$ syntactically different neighborhood functions
- $r^{x_1,\dots,x_n}(A,B) ::= (r^{x_1}(A,B),\dots,r^{x_n}(A,B))$

# Verifiability

## Definition

$\mathfrak{r}^x$ is more informative than $\mathfrak{r}^y$ ($\mathfrak{r}^x \succeq \mathfrak{r}^y$): there is a function $\delta : \left(2^{\mathcal{U}}\right)^n \to \left(2^{\mathcal{U}}\right)^m$ such that $\delta\left(\mathfrak{r}^x(A,B)\right) = \mathfrak{r}^y\left(A,B\right)$ for any $A, B \subseteq \mathcal{U}$.

In case $\mathfrak{r}^x \approx \mathfrak{r}^y$ ($\mathfrak{r}^x \succeq \mathfrak{r}^y$ and $\mathfrak{r}^y \succeq \mathfrak{r}^x$), we say that $\mathfrak{r}^x$ represents $\mathfrak{r}^y$.

# Verifiability

## Definition

$\mathfrak{r}^x$ is more informative than $\mathfrak{r}^y$ ($\mathfrak{r}^x \succeq \mathfrak{r}^y$): there is a function $\delta : \left(2^{\mathcal{U}}\right)^n \to \left(2^{\mathcal{U}}\right)^m$ such that $\delta\left(\mathfrak{r}^x(A,B)\right) = \mathfrak{r}^y\left(A,B\right)$ for any $A, B \subseteq \mathcal{U}$.

In case $\mathfrak{r}^x \approx \mathfrak{r}^y$ ($\mathfrak{r}^x \succeq \mathfrak{r}^y$ and $\mathfrak{r}^y \succeq \mathfrak{r}^x$), we say that $\mathfrak{r}^x$ represents $\mathfrak{r}^y$.

## Example

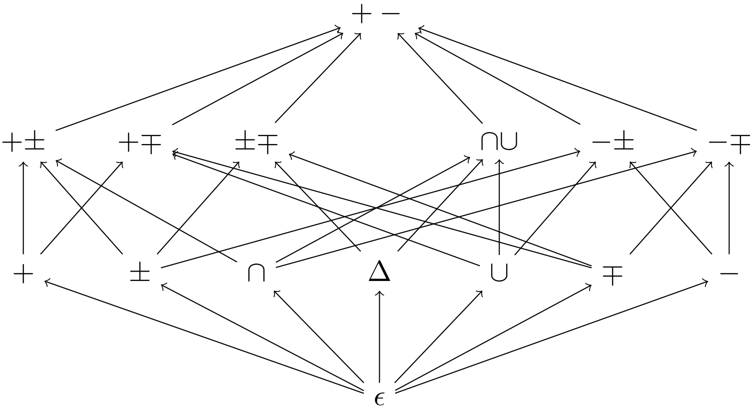
- $\delta_1(\mathfrak{r}^{+\pm}(A,B)) = \delta_1(A, A \setminus B) =_{def} (A, A \setminus (A \setminus B)) = (A, A \cap B) = \mathfrak{r}^{+\cap}(A,B)$;
- $\delta_2(\mathfrak{r}^{+\cap}(A,B)) = \delta_2(A, A \cap B) =_{def} (A \setminus (A \cap B), A \cap B) = (A \setminus B, A \cap B) = \mathfrak{r}^{\pm\cap}(A,B)$;
- $\delta_3(\mathfrak{r}^{\pm\cap}(A,B)) = \delta_3(A \setminus B, A \cap B) =_{def} ((A \setminus B) \cup (A \cap B), A \setminus B) = (A, A \setminus B) = \mathfrak{r}^{+\pm}(A,B)$.

$\Rightarrow \mathfrak{r}^{+\pm} \approx \mathfrak{r}^{+\cap} \approx \mathfrak{r}^{\pm\cap}$

# Verifiability

**Lemma**

*All neighborhood functions are represented by the ones depicted below and the $\prec$-relation represented by arcs holds.*



# Verifiability

## Definition

A semantics $\sigma$ is verifiable by the verification class induced by the neighborhood function $\mathfrak{r}^x$ ($x$-verifiable) iff there is a function $\gamma_\sigma : \left(2^{\mathcal{U}}\right)^n \to 2^{2^{\mathcal{U}}}$ s.t.

$$\forall \mathcal{F} : \quad \gamma_\sigma\left(\widetilde{\mathcal{F}}^x\right) = \sigma(\mathcal{F}).$$

Moreover, $\sigma$ is exactly $x$-verifiable iff $\sigma$ is $x$-verifiable and there is no $\mathfrak{r}^y$ with $\mathfrak{r}^y \prec \mathfrak{r}^x$ such that $\sigma$ is $y$-verifiable.

# Verifiability

**Proposition**

Complete semantics is exactly $+-$-verifiable.

# Verifiability

## Proposition

Complete semantics is exactly $+-$-verifiable.

## Proof

- Verifiability:

$$\gamma_{co}(\widetilde{\mathcal{F}}^{+-}) = \{S \mid (S, S^+, S^-) \in \widetilde{\mathcal{F}}^{+-}, (S^- \setminus S^+) = \emptyset,$$
$$\forall (\bar{S}, \bar{S}^+, \bar{S}^-) \in \widetilde{\mathcal{F}}^{+-} : \bar{S} \supset S \Rightarrow (\bar{S}^- \setminus S^+) \neq \emptyset)\}$$

# Verifiability

## Proposition

Complete semantics is exactly $+-$-verifiable.

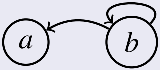
## Proof

- Verifiability:

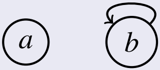
$$\gamma_{co}(\widetilde{\mathcal{F}}^{+-}) = \{S \mid (S, S^+, S^-) \in \widetilde{\mathcal{F}}^{+-}, (S^- \setminus S^+) = \emptyset,$$
$$\forall(\bar{S}, \bar{S}^+, \bar{S}^-) \in \widetilde{\mathcal{F}}^{+-} : \bar{S} \supset S \Rightarrow (\bar{S}^- \setminus S^+) \neq \emptyset)\}$$

- Exactness:

$+\pm$ : $\mathcal{F}_1$ :




$\mathcal{F}_1'$ :




  - $\widetilde{\mathcal{F}_1}^{+\pm} = \{(\emptyset, \emptyset, \emptyset), (\{a\}, \{a\}, \emptyset)\} = \widetilde{\mathcal{F}_1'}^{+\pm}$
  - $co(\mathcal{F}_1) = \{\emptyset\} \neq \{\{a\}\} = co(\mathcal{F}_1')$

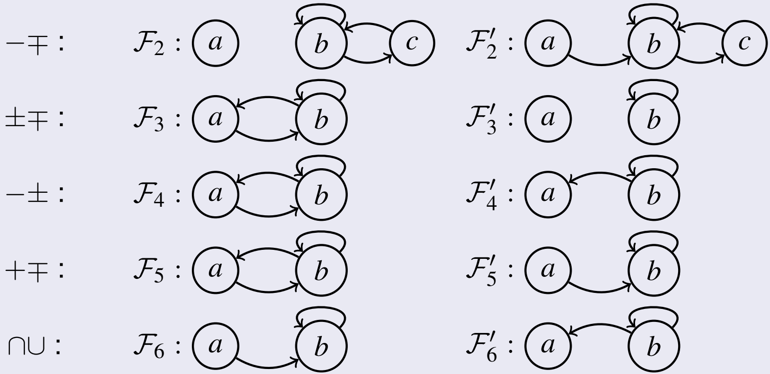
$\Rightarrow$ *co* is not $+\pm$-verifiable

# Verifiability

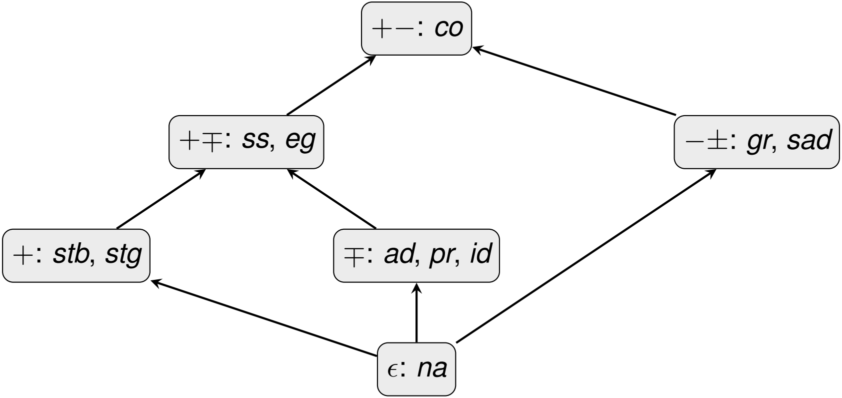
## Proposition

Complete semantics is exactly $+-$-verifiable.

## Proof (ctd.)



# Verifiability



# Verifiability

## Definition

We call a semantics $\sigma$ rational if self-loop-chains are irrelevant.

That is, for every AF $\mathcal{F}$ it holds that $\sigma(\mathcal{F}) = \sigma(\mathcal{F}^l)$, where
$\mathcal{F}^l = (A_\mathcal{F}, R_\mathcal{F} \setminus \{(a,b) \in R_\mathcal{F} \mid (a,a),(b,b) \in R_\mathcal{F}, a \neq b\})$.

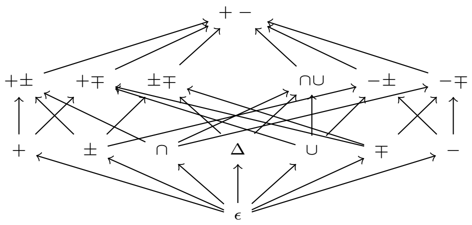
# Verifiability

## Definition

We call a semantics $\sigma$ rational if self-loop-chains are irrelevant.
That is, for every AF $\mathcal{F}$ it holds that $\sigma(\mathcal{F}) = \sigma(\mathcal{F}^l)$, where
$\mathcal{F}^l = (A_\mathcal{F}, R_\mathcal{F} \setminus \{(a,b) \in R_\mathcal{F} \mid (a,a),(b,b) \in R_\mathcal{F}, a \neq b\})$.

## Theorem

Every semantics which is rational is exactly verifiable by a verification class induced by one of the neighborhood functions below.



# Strong Equivalence

**Definition**

Two AFs $\mathcal{F}$ and $\mathcal{G}$ are strongly equivalent w.r.t. semantics$\sigma$ ($\mathcal{F} \equiv^{\sigma}_{E} \mathcal{G}$) iff for all AFs $\mathcal{H}$: $\sigma(\mathcal{F} \cup \mathcal{H}) = \sigma(\mathcal{G} \cup \mathcal{H})$

# Strong Equivalence

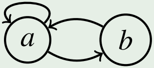
## Definition

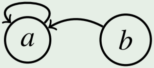
Two AFs $\mathcal{F}$ and $\mathcal{G}$ are strongly equivalent w.r.t. semantics$\sigma$ ($\mathcal{F} \equiv^{\sigma}_{E} \mathcal{G}$) iff for all AFs $\mathcal{H}$: $\sigma(\mathcal{F} \cup \mathcal{H}) = \sigma(\mathcal{G} \cup \mathcal{H})$

$\Rightarrow$ syntactical criteria exist

## Example (stable semantics)

- $stb$-kernel: $\mathcal{F}^{k(stb)} = (A, R \setminus \{(a,b) \mid a \neq b, (a,a) \in R\})$.
- Theorem: $\mathcal{F}^{k(stb)} = \mathcal{G}^{k(stb)} \Leftrightarrow \mathcal{F}$ and $\mathcal{G}$ are strongly equivalent.

$\mathcal{F}$ : 

$\mathcal{G}$ : 

We have $\mathcal{F}^{k(stb)} = \mathcal{G}^{k(stb)} = \mathcal{G}$. Thus, $\mathcal{F}$ and $\mathcal{G}$ are strong equivalent.

# Strong Equivalence

**Definition ($\sigma$-kernel)**

Let $\mathcal{F} = (A, R)$. We define $\sigma$-kernels $\mathcal{F}^{k(\sigma)} = \left(A, R^{k(\sigma)}\right)$ whereby

1. $R^{k(stb)} = R \setminus \{(a,b) \,|a \neq b, (a,a) \in R\}$,
2. $R^{k(ad)} = R \setminus \{(a,b) \,|a \neq b, (a,a) \in R, \{(b,a),(b,b)\} \cap R \neq \emptyset\}$,
3. $R^{k(gr)} = R \setminus \{(a,b) \,|a \neq b, (b,b) \in R, \{(a,a),(b,a)\} \cap R \neq \emptyset\}$,
4. $R^{k(co)} = R \setminus \{(a,b) \,|a \neq b, (a,a),(b,b) \in R\}$.
5. $R^{k(na)} = R \cup \{(a,b) \mid a \neq b, \{(a,a),(b,a),(b,b)\} \cap R \neq \emptyset\}$.

# Strong Equivalence

**Theorem**

*Strong equivalence is characterizable through kernels (see below).*

$$\mathcal{F} \equiv^{\sigma}_{E} \mathcal{G} \Leftrightarrow \mathcal{F}^{k} = \mathcal{G}^{k}$$

| $stg$ | $stb$ | $ss$ | $eg$ | $ad$ | $pr$ | $id$ | $gr$ | $sad$ | $co$ | $na$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $k(stb)$ | $k(stb)$ | $k(ad)$ | $k(ad)$ | $k(ad)$ | $k(ad)$ | $k(ad)$ | $k(gr)$ | $k(gr)$ | $k(co)$ | $k(na)$ |

# Intermediate Semantics

- *stb* and *stg* are both characterizable through $k(stb)$.
- Does this also hold for arbitrary semantics $\sigma$ with *stb*$(\mathcal{F}) \subseteq \sigma(\mathcal{F}) \subseteq$ *stg*$(\mathcal{F})$ for each AF $\mathcal{F}$?
  (e.g. when obtained from SESAME [Besnard et al., 2016])

# Intermediate Semantics

- *stb* and *stg* are both characterizable through $k(stb)$.
- Does this also hold for arbitrary semantics $\sigma$ with *stb*$(\mathcal{F}) \subseteq \sigma(\mathcal{F}) \subseteq$ *stg*$(\mathcal{F})$ for each AF $\mathcal{F}$?
  (e.g. when obtained from SESAME [Besnard et al., 2016])

## Example

- "Stagle semantics":
  $S \in$ *sta*$(\mathcal{F}) \Leftrightarrow S \in$ *cf*$(\mathcal{F})$, $S^+_\mathcal{F} \cup S^-_\mathcal{F} = A_\mathcal{F}$ and $\forall T \in cf(\mathcal{F}) : S^+_\mathcal{F} \not\subset T^+_\mathcal{F}$

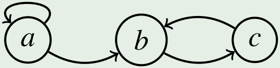
# Intermediate Semantics

- *stb* and *stg* are both characterizable through $k(stb)$.
- Does this also hold for arbitrary semantics $\sigma$ with *stb*$(\mathcal{F}) \subseteq \sigma(\mathcal{F}) \subseteq$ *stg*$(\mathcal{F})$ for each AF $\mathcal{F}$?
  (e.g. when obtained from SESAME [Besnard et al., 2016])

## Example

- "Stagle semantics":
  $S \in \textit{sta}(\mathcal{F}) \Leftrightarrow S \in \textit{cf}(\mathcal{F}), S^+_{\mathcal{F}} \cup S^-_{\mathcal{F}} = A_{\mathcal{F}}$ and $\forall T \in cf(\mathcal{F}) : S^+_{\mathcal{F}} \not\subset T^+_{\mathcal{F}}$

$\mathcal{F}$ :



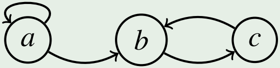

- *stb*$(\mathcal{F}) = \emptyset \subset$ *sta*$(\mathcal{F}) = \{\{b\}\} \subset$ *stg*$(\mathcal{F}) = \{\{b\}, \{c\}\}$.

# Intermediate Semantics

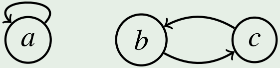
- *stb* and *stg* are both characterizable through $k(stb)$.
- Does this also hold for arbitrary semantics $\sigma$ with $\mathit{stb}(\mathcal{F}) \subseteq \sigma(\mathcal{F}) \subseteq \mathit{stg}(\mathcal{F})$ for each AF $\mathcal{F}$?
  (e.g. when obtained from SESAME [Besnard et al., 2016])

## Example

- "Stagle semantics":
  $S \in \mathit{sta}(\mathcal{F}) \Leftrightarrow S \in \mathit{cf}(\mathcal{F}), S^{+}_{\mathcal{F}} \cup S^{-}_{\mathcal{F}} = A_{\mathcal{F}}$ and $\forall T \in \mathit{cf}(\mathcal{F}) : S^{+}_{\mathcal{F}} \not\subset T^{+}_{\mathcal{F}}$

$\mathcal{F}$ :



- $\mathit{stb}(\mathcal{F}) = \emptyset \subset \mathit{sta}(\mathcal{F}) = \{\{b\}\} \subset \mathit{stg}(\mathcal{F}) = \{\{b\}, \{c\}\}$.

$\mathcal{F}^{k(\mathit{stb})}$ :



- $\mathit{sta}\left(\mathcal{F}^{k(\mathit{stb})}\right) = \{\{b\}, \{c\}\} \Rightarrow \mathcal{F} \not\equiv^{\mathit{sta}}_{E} \mathcal{F}^{k(\mathit{stb})}, \mathcal{F}^{k(\mathit{stb})} = \left(\mathcal{F}^{k(\mathit{stb})}\right)^{k(\mathit{stb})}$

$\Rightarrow$ Stagle semantics is not compatible with the stable kernel.

# Intermediate Semantics

## Theorem

For each semantics $\sigma$ which is $+$-verifiable and *stb*-*stg*-intermediate, it holds that

$$\mathcal{F}^{k(\mathit{stb})} = \mathcal{G}^{k(\mathit{stb})} \Leftrightarrow \mathcal{F} \equiv_E^\sigma \mathcal{G}.$$

# Intermediate Semantics

## Theorem

For each semantics $\sigma$ which is $+$-verifiable and *stb*-*stg*-intermediate, it holds that

$$\mathcal{F}^{k(\mathit{stb})} = \mathcal{G}^{k(\mathit{stb})} \Leftrightarrow \mathcal{F} \equiv_E^\sigma \mathcal{G}.$$

## Theorem

For each semantics $\sigma$ which is $+\mp$-verifiable and $\rho$-*ad*-intermediate with $\rho \in \{\mathit{ss}, \mathit{id}, \mathit{eg}\}$, it holds that

$$\mathcal{F}^{k(\mathit{ad})} = \mathcal{G}^{k(\mathit{ad})} \Leftrightarrow \mathcal{F} \equiv_E^\sigma \mathcal{G}.$$

## Theorem

For each semantics $\sigma$ which is $-\pm$-verifiable and *gr*-*sad*-intermediate, it holds that

$$\mathcal{F}^{k(\mathit{gr})} = \mathcal{G}^{k(\mathit{gr})} \Leftrightarrow \mathcal{F} \equiv_E^\sigma \mathcal{G}.$$

# Conclusion

Summary:

- Hierarchy of verification classes
- Each “rational” semantics is exactly verifiable by a certain class
- Characterization of strong equivalence for intermediate semantics

Future work:

- Semantics not captured by the approach, e.g. *cf2* semantics [Baroni et al., 2005]
- Investigating labelling-based semantics [Caminada and Gabbay, 2009]
- Use classification as distance measure [Doutre and Mailly, 2016]

# References I


Baroni, P. and Giacomin, M. (2007).
On principle-based evaluation of extension-based argumentation semantics.
Artif. Intell., 171(10-15):675–700.

Baroni, P., Giacomin, M., and Guida, G. (2005).
SCC-Recursiveness: A general schema for argumentation semantics.
Artif. Intell., 168(1-2):162–210.

Baumann, R. (2016).
Characterizing equivalence notions for labelling-based semantics.
In Principles of Knowledge Representation and Reasoning: Proceedings of the 15th International Conference, pages 22–32.

Baumann, R. and Strass, H. (2016).
An abstract logical approach to characterizing strong equivalence in logic-based knowledge representation formalisms.
In Principles of Knowledge Representation and Reasoning: Proceedings of the 15th International Conference, pages 525–528.

Besnard, P., Doutre, S., Ho, V. H., and Longin, D. (2016).
SESAME - A system for specifying semantics in abstract argumentation.
In Proc. SAFA@COMMA, pages 40–51.

# References II


- Caminada, M. and Amgoud, L. (2007).
  On the evaluation of argumentation formalisms.
  Artif. Intell., pages 286–310.
- Caminada, M. and Gabbay, D. M. (2009).
  A logical account of formal argumentation.
  Studia Logica, 93(2):109–145.
- Doutre, S. and Mailly, J.-G. (2016).
  Quantifying the difference between argumentation semantics.
  In Computational Models of Argument - Proceedings of COMMA 2016, volume 287 of Frontiers in Artificial Intelligence and Applications, pages 255–262. IOS Press.
- Dung, P. M. (1995).
  On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games.
  Artif. Intell., 77(2):321–357.
- Dunne, P. E. and Bench-Capon, T. J. M. (2002).
  Coherence in finite argument systems.
  Artif. Intell., 141(1/2):187–203.
- Dvořák, W. and Woltran, S. (2010).
  Complexity of semi-stable and stage semantics in argumentation frameworks.
  Inf. Process. Lett., 110(11):425–430.

# References III


Lifschitz, V., Pearce, D., and Valverde, A. (2001).
Strongly equivalent logic programs.
ACM Transactions on Computational Logic, 2(4):526–541.

Oikarinen, E. and Woltran, S. (2011).
Characterizing strong equivalence for argumentation frameworks.
Artif. Intell., 175(14-15):1985–2009.

Truszczynski, M. (2006).
Strong and uniform equivalence of nonmonotonic theories - an algebraic approach.
Annals of Mathematics and Artificial Intelligence, 48(3-4):245–265.

Turner, H. (2004).
Strong equivalence for causal theories.
In 7th International Conference on Logic Programming and Nonmonotonic Reasoning, Proceedings, volume 2923 of Lecture Notes in Computer Science, pages 289–301. Springer.

# Verifiability

$$\gamma_{\mathit{na}}(\widetilde{\mathcal{F}}_A^{\epsilon}) = \{S \mid S \in \widetilde{\mathcal{F}}, S \text{ is } \subseteq\text{-maximal in } \widetilde{\mathcal{F}}\};$$
$$\gamma_{\mathit{stg}}(\widetilde{\mathcal{F}}_A^{+}) = \{S \mid (S, S^+) \in \widetilde{\mathcal{F}}^+, S^+ \text{ is } \subseteq\text{-maximal in } \{C^+ \mid (C, C^+) \in \widetilde{\mathcal{F}}^+\}\};$$
$$\gamma_{\mathit{stb}}(\widetilde{\mathcal{F}}_A^{+}) = \{S \mid (S, S^+) \in \widetilde{\mathcal{F}}^+, S^+ = A\};$$
$$\gamma_{\mathit{ad}}(\widetilde{\mathcal{F}}_A^{\mp}) = \{S \mid (S, S^{\mp}) \in \widetilde{\mathcal{F}}^{\mp}, S^{\mp} = \emptyset\};$$
$$\gamma_{\mathit{pr}}(\widetilde{\mathcal{F}}_A^{\mp}) = \{S \mid S \in \gamma_{\mathit{ad}}(\widetilde{\mathcal{F}}_A^{\mp}), S \text{ is } \subseteq\text{-maximal in } \gamma_{\mathit{ad}}(\widetilde{\mathcal{F}}_A^{\mp})\};$$
$$\gamma_{\mathit{ss}}(\widetilde{\mathcal{F}}_A^{+\mp}) = \{S \mid S \in \gamma_{\mathit{ad}}(\widetilde{\mathcal{F}}_A^{\mp}), S^+ \text{ is } \subseteq\text{-maximal in } \{C^+ \mid (C, C^+, C^{\mp}) \in \widetilde{\mathcal{F}}^{+\mp}, C \in \gamma_{\mathit{ad}}(\widetilde{\mathcal{F}}_A^{\mp})\}\};$$
$$\gamma_{\mathit{id}}(\widetilde{\mathcal{F}}_A^{\mp}) = \{S \mid S \text{ is } \subseteq\text{-maximal in } \{C \mid C \in \gamma_{\mathit{ad}}(\widetilde{\mathcal{F}}_A^{\mp}), C \subseteq \bigcap \gamma_{\mathit{pr}}(\widetilde{\mathcal{F}}_A^{\mp})\}\};$$
$$\gamma_{\mathit{eg}}(\widetilde{\mathcal{F}}_A^{+\mp}) = \{S \mid S \text{ is } \subseteq\text{-maximal in } \{C \mid C \in \gamma_{\mathit{ad}}(\widetilde{\mathcal{F}}_A^{\mp}), C \subseteq \bigcap \gamma_{\mathit{ss}}(\widetilde{\mathcal{F}}_A^{+\mp})\}\};$$
$$\gamma_{\mathit{sad}}(\widetilde{\mathcal{F}}_A^{-\pm}) = \{S \mid (S, S^-, S^{\pm}) \in \widetilde{\mathcal{F}}^{-\pm}, \exists (S_0, S_0^-, S_0^{\pm}), \ldots, (S_n, S_n^-, S_n^{\pm}) \in \widetilde{\mathcal{F}}^{-\pm} :$$
$$(\emptyset = S_0 \subset \cdots \subset S_n = S \wedge \forall i \in \{1, \ldots, n\} : S_i^- \subseteq S_{i-1}^{\pm})\};$$
$$\gamma_{\mathit{gr}}(\widetilde{\mathcal{F}}_A^{-\pm}) = \{S \mid S \in \gamma_{\mathit{sad}}(\widetilde{\mathcal{F}}_A^{-\pm}), \forall (\bar{S}, \bar{S}^-, \bar{S}^{\pm}) \in \widetilde{\mathcal{F}}^{-\pm} : \bar{S} \supset S \Rightarrow (\bar{S}^- \setminus S^{\pm}) \neq \emptyset)\}.$$