Logic and Argumentation for New-generation AI

ESSLLI 2026

Liuwen Yu[1]
Leon van der Torre[2]
[1] Luxembourg Institute of Science and Technology
[2] University of Luxembourg

Computational Models of Argument (COMMA)

COMMA:

Conferences:

- **COMMA** (International Conference on Computational Models of Argument), since 2006
- **CLAR** (Conference on Logic and Argumentation), since 2016

Journal: **Argument & Computation**

Knowledge Representation and Reasoning (KRR):

Conferences: KR (International Conference on Principles of Knowledge Representation and Reasoning), since 1989

Journals: KRR is published in Artificial Intelligence Journal, Journal of Artificial Intelligence Research, and specialized journals like Journal of Logic and Computation

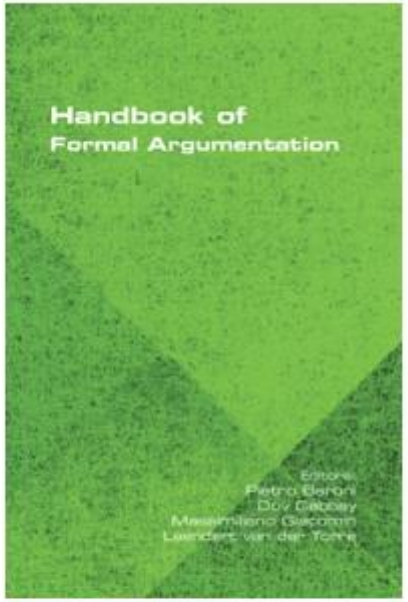
Artificial Intelligence (AI):

Conferences:

- IJCAI (International Joint Conference on Artificial Intelligence), since 1969
- ECAI (European Conference on Artificial Intelligence), since 1974
- AAAI (Association for the Advancement of Artificial Intelligence), since 1979

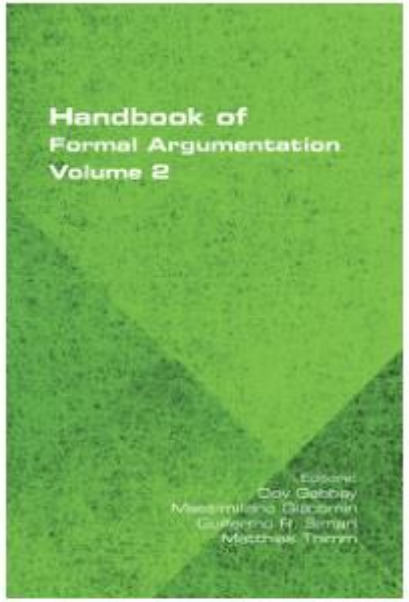
Journals: Artificial Intelligence Journal; Journal of Artificial Intelligence Research

Handbook of Formal Argumentation



28 February 2018

Foundations



August 2021

Extension of Abstract Argumentation

9 December 2024

Applications

College Publications, https://www.collegepublications.co.uk/handbooks/

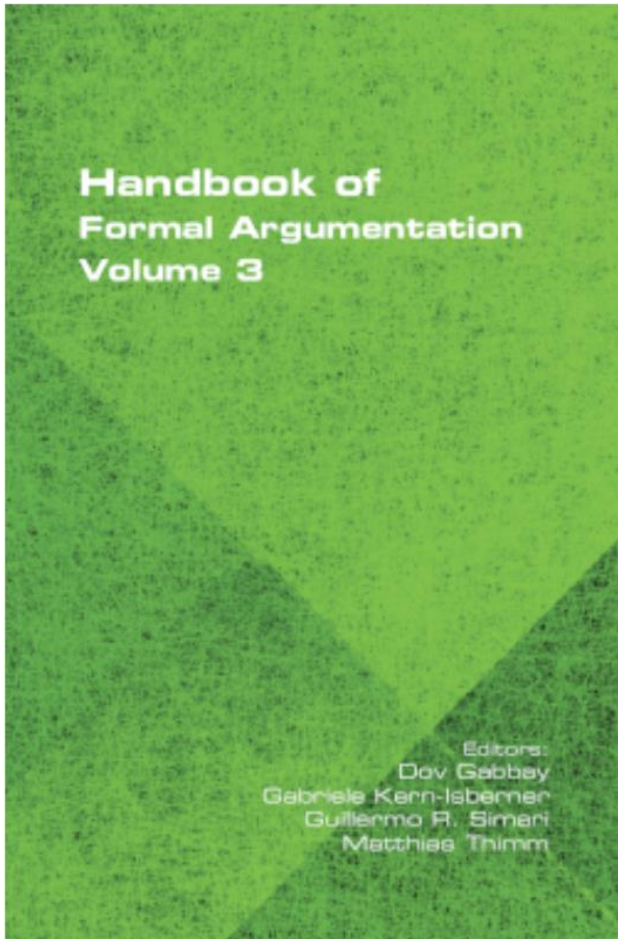
Course Material

1

Thirteen Challenges in Formal and Computational Argumentation

LIUWEN YU, LEENDERT VAN DER TORRE, AND RÉKA MARKOVICH

ABSTRACT. In this chapter, we present thirteen challenges in formal and computational argumentation. They are organized around Dung's attack-defense paradigm shift. First, we describe four challenges pertaining to the diversity of argumentation. Then we discuss five challenges regarding the attack-defense paradigm shift. Finally, we discuss four challenges for computational AI argumentation arising after the paradigm shift. We illustrate these challenges using examples from machine ethics, AI & Law, decision-making, linguistics, philosophy, and other disciplines to illustrate the breadth of argumentation research. We end each challenge by presenting several open questions for further research.



https://yuliuwen.github.io/assets/files/HOFA3_Chapter_Thirteen_Challenges.pdf

Day 1 Dung paradigm shift and reasoning alignment

1.1 AI Logic (JLC corner) is nonmonotonic logic (since 1980)

1.2 Explanation of black boxes in new generation AI: secretary metaphor

1.3 Dung 1995 paradigm shift relates logic and argumentation based explanation

1.4 Reasoning alignment of logic and argumentation based explanation

1.5 Using logic: from inference towards dialogue and decision making, to logic in action

Day 2 Defence first and partial argumentation

2.1 Why unify reasons in philosophy and formal argumentation?

2.2 Rules represent reasons in both logic and argumentation.

2.3 Logic dilemma: prescriptive prioritized default logic or descriptive default logic?

2.4 Argumentation dilemma: weakest link or last link?

2.5 Are the two dilemmas two sides of the same coin?

Day 3 Axiomatic approach and impossibility

3.1 Dung 1995 and 2016/18: unification, representation, axiomatization.

3.2 Representation theorems

3.3 Impossibility: exact representation and natural principles can conflict.

3.4 Escape routes: restrict the domain or allow context dependence.

3.5 Reasons / arguments in progress

Day 4: agentic AI, Fatio and Libra

4.1. New foundations for multiple agents: individual vs collective defense

4.2. New foundations for time: reasons / arguments in progress

4.3. Fatio and Libra

4.4. End-to-end pipelines for traceability and auditability

4.5. Aligning Argumentation as Balancing, Dialogue and Inference (A-BDI)

Day 5 AI&Law and machine ethics

5.1. Key examples: all or nothing, divorce, discretion, …

5.2. The institution vs the individual: decision perspective

5.3. New foundations: bipolar argumentation, qualitative and quantified

5.4. New foundations: contrastive reasons, triples, dynamic scale, attack assignment

5.5. The logic toolbox and combining logics

Classical consequence and operation

Definition (Classical consequence)

Let $\mathcal{L}$ be a propositional language. Classical consequence is the relation $\vdash \subseteq \mathcal{P}(\mathcal{L}) \times \mathcal{L}$ defined by $A \vdash \varphi$ if and only if $\neg\exists v\big[(\forall a \in A)\, v(a) = 1 \wedge v(\varphi) = 0\big]$.

Example

For $A = \{p, p \to q\}$, $A \vdash q$.

Definition (Consequence operation)

The operation corresponding to $\vdash$ is $\mathrm{Cn}\colon \mathcal{P}(\mathcal{L}) \to \mathcal{P}(\mathcal{L})$, defined by $\mathrm{Cn}(A) = \{\varphi \in \mathcal{L} \mid A \vdash \varphi\}$. Hence, $A \vdash \varphi$ if and only if $\varphi \in \mathrm{Cn}(A)$.

Example

For $A = \{p, p \to q\}$, $q \in \mathrm{Cn}(A)$, whereas $r \notin \mathrm{Cn}(A)$.

David Makinson, "Bridges between Classical and Nonmonotonic Logic", Logic Journal of the IGPL 11(1), 2003, pp. 69–96.

Tarskian properties

For all $A, B \subseteq \mathcal{L}$ and $\varphi \in \mathcal{L}$, the consequence relation and its corresponding operation satisfy:

Property	Consequence relation	Consequence operation
Inclusion	If $\varphi \in A$, then $A \vdash \varphi$ *(Reflexivity)*.	$A \subseteq \mathrm{Cn}(A)$.
Monotonicity	If $A \subseteq B$ and $A \vdash \varphi$, then $B \vdash \varphi$.	$A \subseteq B \Longrightarrow \mathrm{Cn}(A) \subseteq \mathrm{Cn}(B)$.
Cut / idempotence	If $A \vdash b$ for every $b \in B$, and $A \cup B \vdash \varphi$, then $A \vdash \varphi$ *(Cut)*.	$\mathrm{Cn}(\mathrm{Cn}(A)) = \mathrm{Cn}(A)$.

David Makinson, “Bridges between Classical and Nonmonotonic Logic”, Logic Journal of the IGPL 11(1), 2003, pp. 69–96.

Logic for AI is nonmonotonic logic

Non-monotonic Reasoing

- A long-standing area of knowledge representation and thus of AI
- Classical consequence relation is monotonic (e.g. lemmas in proofs)
  - If $S \vdash \varphi$, then $S, T \vdash \varphi$
- But common-sense reasoning is often nonmonotonic
  - additional information may invalidate conclusions drawn earlier

Sources of nonmonotonicity

- Empirical generalisations
  - Adults are usually employed, birds can typically fly,…
- Exceptions to legal rules
  - When a father dies, his son can inherit, except when the son killed the father
- Exceptions to moral principles
  - Normally one should not lie, except when a lie can save lives
- Conflicting information sources
  - Experts who disagree, witnesses who contradict each other, …
- Alternative explanations
  - The grass is wet so it has rained / but the sprinkler was on
- …

Slides from Henry Prakken: http://people.cs.uu.nl/henry/al/chongqing10.html

Day 1 Motivation and reasoning alignment

1.1 AI Logic (JLC corner) is nonmonotonic logic (since 1980)

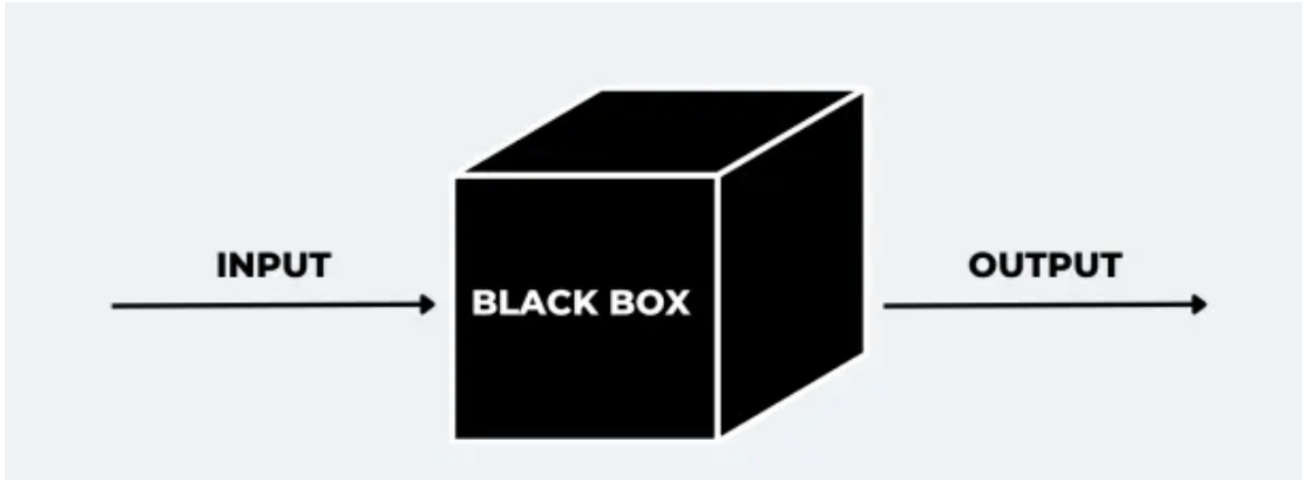
1.2 Explanation of black boxes in new generation AI: secretary metaphor

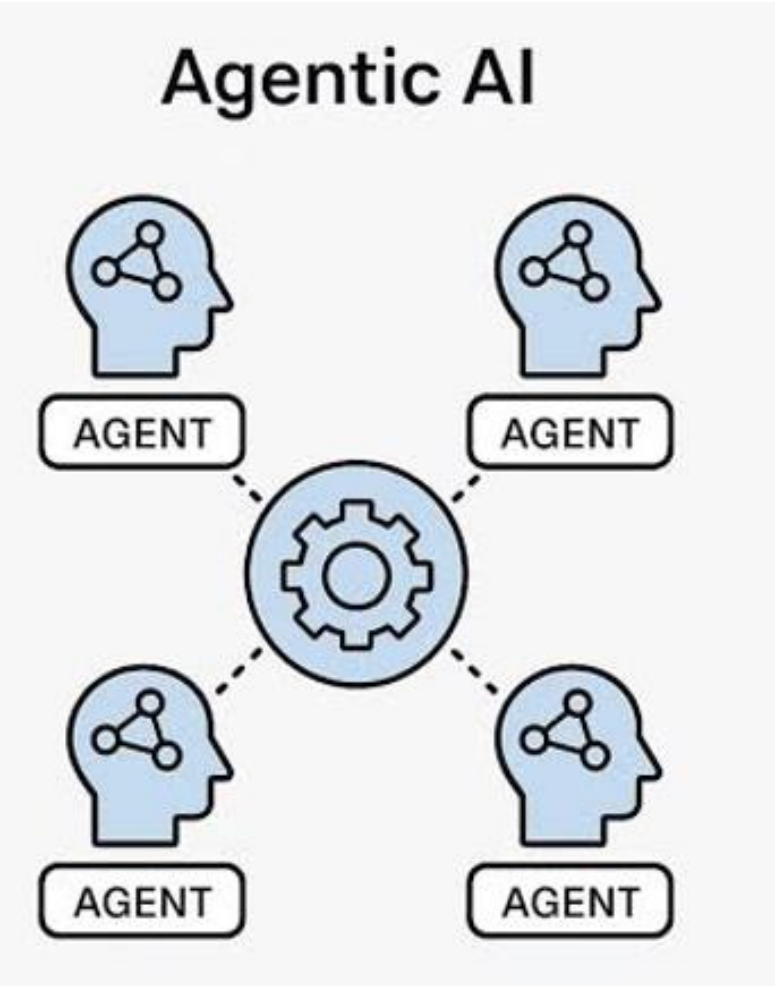
1.3 Dung 1995 paradigm shift relates logic and argumentation based explanation

1.4 Reasoning alignment of logic and argumentation based explanation

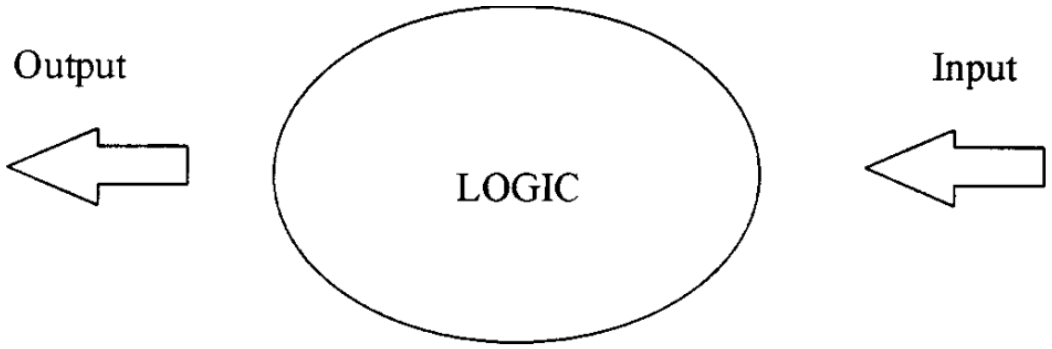
1.5 Using logic: from inference towards dialogue and decision making, to logic in action

New-generation AI: black boxes need explanation

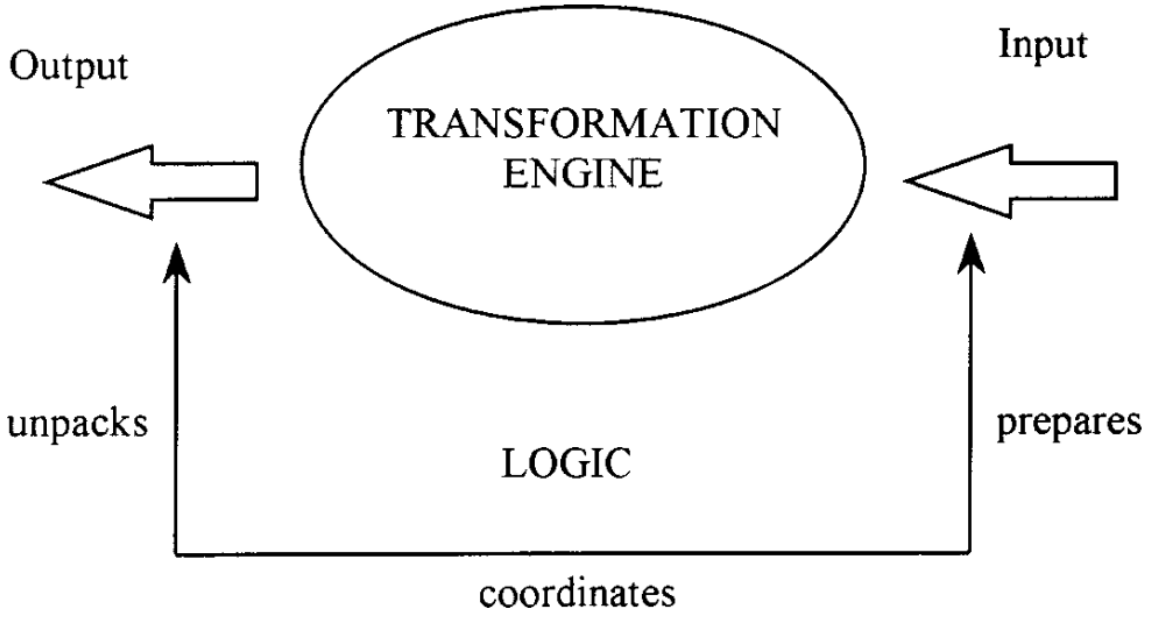




Logic acts as secretary and assistant

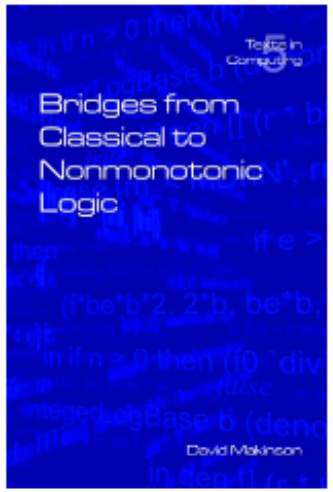


Traditional picture of logic as an inference motor.



Logic assisting a transformation engine

Bridges from Classical to Nonmonotonic Logic



Fixed pivot	Pivotal use	Default use
Assumptions $K \subseteq \mathcal{L}$	add all of K	select compatible parts
Valuations $W \subseteq \mathcal{V}$	restrict to W	select best models
Rules $R \subseteq \mathcal{L}^2$	add all rules in R	vary rules with input

Pivotal: supraclassical and monotonic.
Default: input-dependent and possibly nonmonotonic.

Day 1: assumptions and valuations. **Day 2:** rules.

The pivot must move with the substitution

Main message

Substituting only the schematic positions can destroy consequence.

Example

$\sigma(p) = r$ means: replace every occurrence of p by r.

$K = \{p \to q\},\ A = \{p\},\ \sigma(p) = r,\ \sigma(q) = s.$

$K, \sigma(A) \nvdash \sigma(q)$, whereas $\sigma(K), \sigma(A) \vdash \sigma(q)$.

First bridge: using additional background assumptions

Definition (Pivotal-assumption consequence)

Let $\mathcal{L}$ be a propositional language. Let $K, A \subseteq \mathcal{L}$, where K is the set of background assumptions and A is the current input. For every $\varphi \in \mathcal{L}$,

$$K, A \vdash \varphi \iff \varphi \in P(K, A) \iff K \cup A \vdash \varphi.$$

Thus,

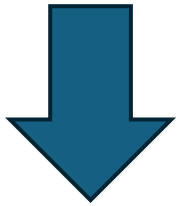
$$P(K, A) = Cn(K \cup A).$$

Example

Let $K = \{p \to q, q \to r\}$ and $A = \{p\}$. Then $q, r \in P(K, A)$.

Default-assumption consequence

"Nonmonotonicity is generated if we allow the background assumption set K to vary with A, in particular, to be diminished when A is inconsistent with K."



maximal subsets L of K that are consistent with A, accept as output ONLY what is common to their separate outputs

Maxiconsistent assumptions

Definition (Consistency with the input)

Let $S \subseteq K$. The set S is *consistent with* A if and only if $S \cup A$ is classically consistent.

Definition (Maximally consistent with A)

Let $K, A \subseteq \mathcal{L}$ and $S \subseteq K$. The set S is *maximally consistent*, or *maxiconsistent*, with A if and only if $S \cup A$ is classically consistent and $T \cup A$ is classically inconsistent for every T such that $S \subsetneq T \subseteq K$.
The family of all such sets is

$$MCS(K, A) = \{S \subseteq K \mid S \text{ is maxiconsistent with } A\}.$$

Example

For $K = \{p \to q, q \to r\}$ and $A = \{p, \neg q\}$, $\mathrm{MCS}(K, A) = \big\{\{q \to r\}\big\}$.

Default-assumption consequence

Definition (Default-assumption consequence)

A formula φ is a *default-assumption consequence of A relative to the default assumptions K* if and only if $S \cup A \vdash \varphi$ for every $S \subseteq K$ maxiconsistent with A. We write $K, A \mid\sim \varphi$. The corresponding operation is

$$C(K, A) = \bigcap \{ Cn(S \cup A) \mid S \subseteq K \text{ and } S \text{ is maxiconsistent with } A \} .$$

Example

For $K = \{p \to q, q \to r\}$ and $A = \{p\}$, $\mathrm{MCS}(K, A) = \{K\}$; hence $K, A \mid\sim q$ and $K, A \mid\sim r$.

The same assumptions, three inputs

Recall

Maxiconsistent sets $\mathrm{MCS}(K, A)$ Output $C(K, A)$ $K, A \mid\sim \varphi \iff \varphi \in C(K, A)$

Example

$K = \{p \to q, q \to r\}$.

Hard input A	$\mathrm{MCS}(K, A)$	Sceptical conclusion
$\{p\}$	$\{K\}$	q, r
$\{p, \neg q\}$	$\{\{q \to r\}\}$	$\neg q$, but not r
$\{p, \neg r\}$	$\{\{p \to q\}, \{q \to r\}\}$	neither q nor $\neg q$

From assumptions to valuations

Syntax $\longrightarrow$ semantics

Maxiconsistent assumption sets $\longleftrightarrow$ best valuations

Models and classical consequence (recap)

Definition (Model set)

Let $\mathcal{V}$ be the set of all classical valuations of $\mathcal{L}$. For $X \subseteq \mathcal{L}$, write $v \models X$ if $v \models \chi$ for every $\chi \in X$, and define

$$\mathrm{Mod}(X) = \{v \in \mathcal{V} \mid v \models X\}.$$

For every $\varphi \in \mathcal{L}$,

$$X \vdash \varphi \quad \Longleftrightarrow \quad \mathrm{Mod}(X) \subseteq \mathrm{Mod}(\{\varphi\}).$$

Example

Every valuation satisfying both p and $p \to q$ also satisfies q. Therefore,

$$\mathrm{Mod}(\{p, p \to q\}) \subseteq \mathrm{Mod}(\{q\}),$$

and hence $\{p, p \to q\} \vdash q$.

Pivotal-valuation consequence

Definition (Pivotal-valuation consequence)

Let $W \subseteq \mathcal{V}$. A formula φ is a *pivotal-valuation consequence of A relative to W* if and only if there is no $v \in W$ such that $v \models A$ and $v \not\models \varphi$. We write $W, A \vdash \varphi$ and define

$$P(W, A) = \{\varphi \in \mathcal{L} \mid W, A \vdash \varphi\}.$$

Example

Let $K = \{p \to q\}$, $W = \mathrm{Mod}(K)$, and $A = \{p\}$. Every valuation in W satisfying A also satisfies q; hence

$$W, A \vdash q.$$

Since $W = \mathrm{Mod}(K)$,

$$P(W, A) = P(K, A) = Cn(K \cup A).$$

Pivotal-valuation consequence

Definition (Pivotal-valuation consequence)

Let $W \subseteq \mathcal{V}$. A formula φ is a *pivotal-valuation consequence of A relative to W* if and only if there is no $v \in W$ such that $v \models A$ and $v \not\models \varphi$. We write $W, A \vdash \varphi$ and define

$$P(W, A) = \{\varphi \in \mathcal{L} \mid W, A \vdash \varphi\}.$$

nonmonotonicity can arise from pivotal valuations if
we allow the restricted set W of valuations to vary with the premise set A

Preferential consequence

Definition (Preferential model and preferential consequence)

A *preferential model* is a pair $(W, <)$, where $W \subseteq \mathcal{V}$ and $< \subseteq W \times W$. For $A \subseteq \mathcal{L}$, define

$$\text{Best}(A) = \{v \in W \mid v \models A \text{ and there is no } w \in W \text{ with } w \models A \text{ and } w < v\}.$$

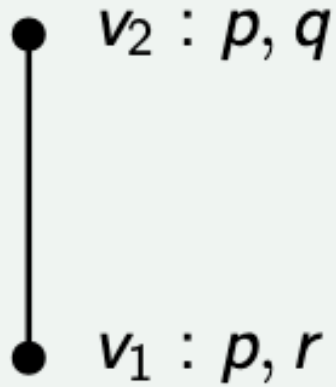
A formula φ is a *preferential consequence of* A if and only if every $v \in \text{Best}(A)$ satisfies φ. We write $A \mid\sim \varphi$.

Example

$W = \{v_1, v_2\}$, with $v_1 < v_2$.

v_1: p, r true; v_2: p, q true.

$p \mid\sim r$, but $p \wedge q \not\mid\sim r$.



Day 1 Dung paradigm shift and reasoning alignment

1.1 AI Logic (JLC corner) is nonmonotonic logic (since 1980)

1.2 Explanation of black boxes in new generation AI: secretary metaphor

1.3 Dung 1995 paradigm shift relates logic and argumentation based explanation

1.4 Reasoning alignment of logic and argumentation based explanation

1.5 Using logic: from inference towards dialogue and decision making, to logic in action

Attack-defense paradigm shift

ELSEVIER

Artificial Intelligence 77 (1995) 321–357

Artificial Intelligence

On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games*

Phan Minh Dung*

Division of Computer Science, Asian Institute of Technology, GPO Box 2754, Bangkok 10501, Thailand

Received June 1993; revised April 1994

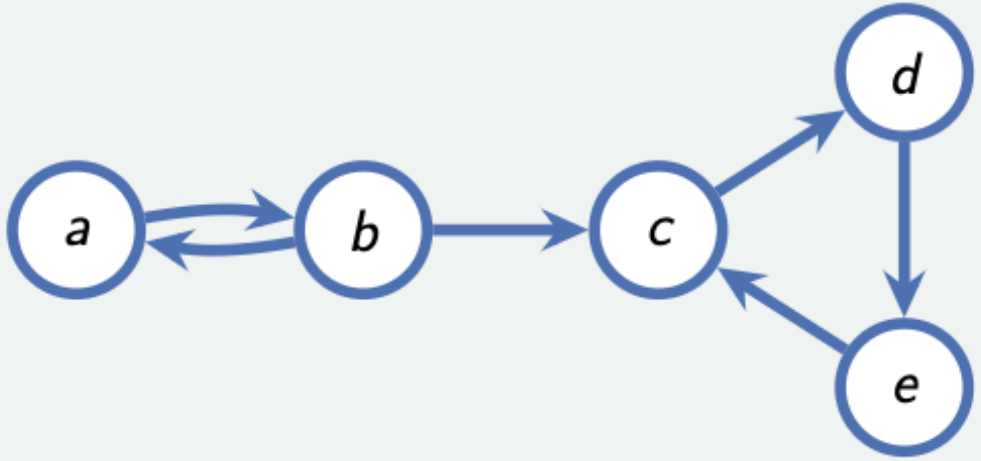
Abstract argumentation framework

Definition (Abstract argumentation framework)

An *abstract argumentation framework* is $AF = \langle AR, attacks \rangle$, where AR is a set of arguments and $attacks \subseteq AR \times AR$ is a binary attack relation.

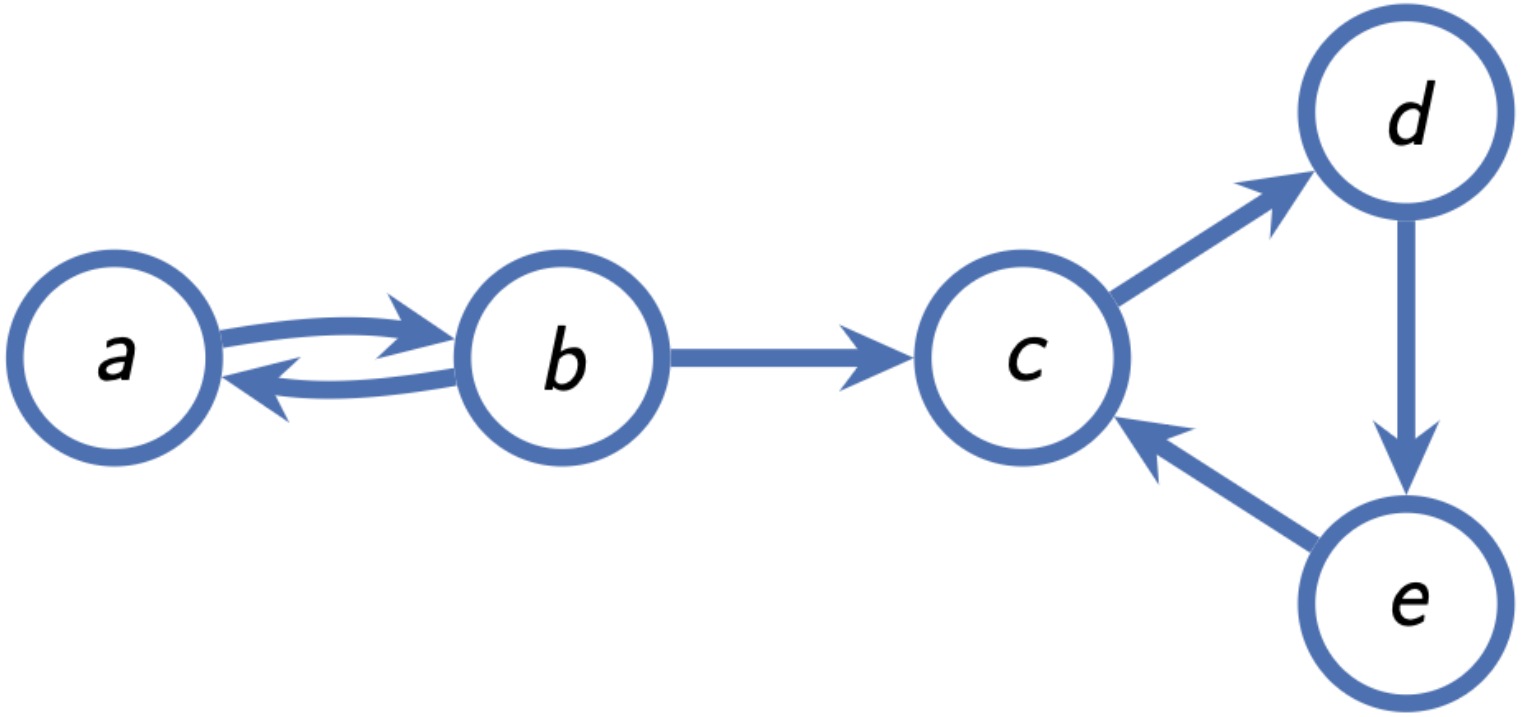
Example

$AR = \{a, b, c, d, e\}$ and $attacks = \{(a, b), (b, a), (b, c), (c, d), (d, e), (e, c)\}$.



Phan Minh Dung, "On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games", Artificial Intelligence 77(2), 1995, pp. 321–357.

Abstract argumentation semantics



Which arguments should be accepted?

Abstract argumentation semantics: gunfight rules

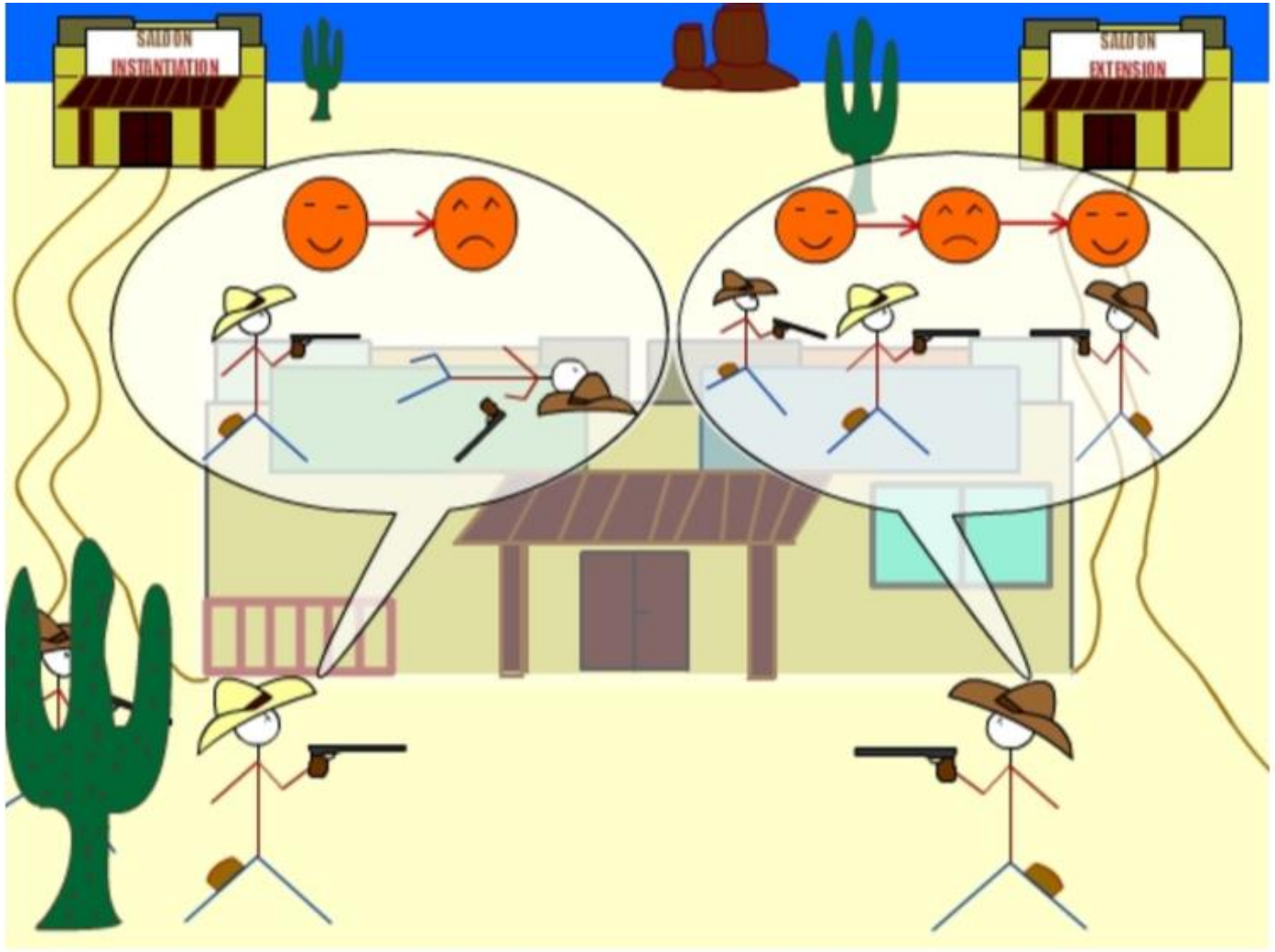


Figure from Serena Villata. Meta-argumentation for MAS: Coalition Formation, Merging Views, Subsumption Relations and Dependence Network

Abstract argumentation semantics: gunfight rules

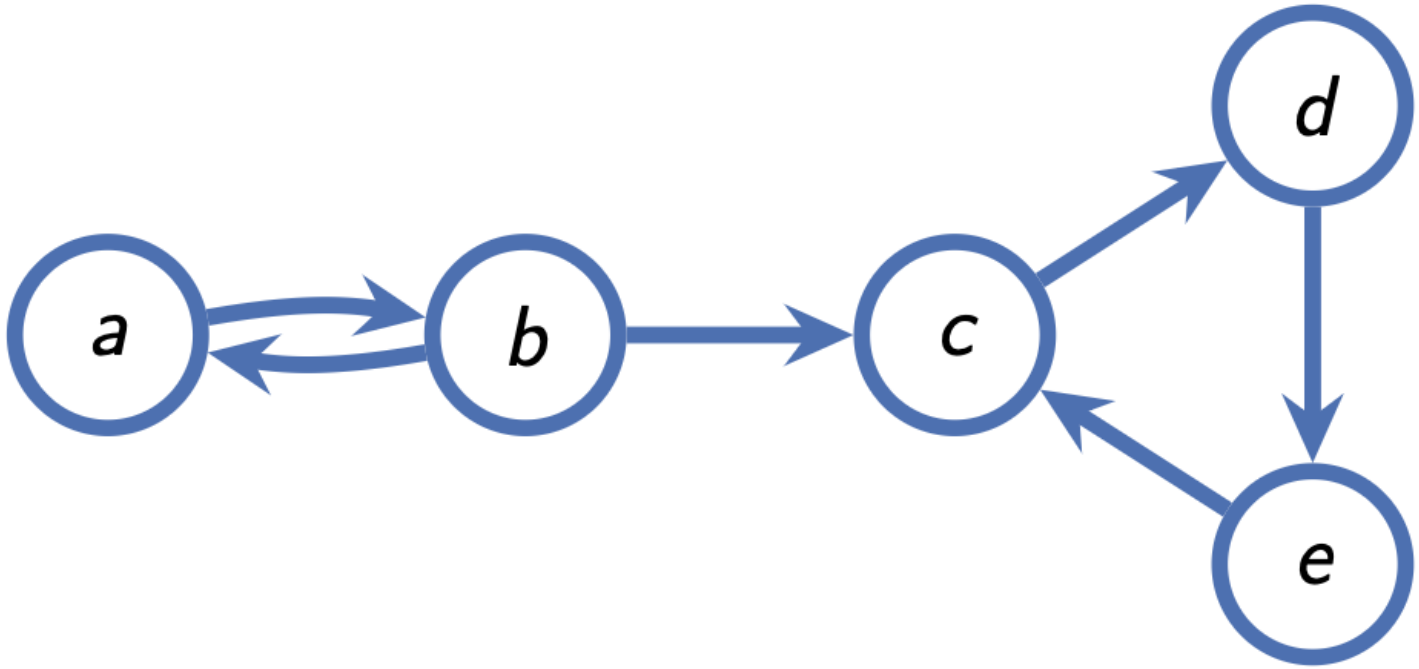
Argument

Attack

Gunfight rules:

iff all attackers are

iff there is an attacker



Abstract argumentation semantics: gunfight rules

Argument

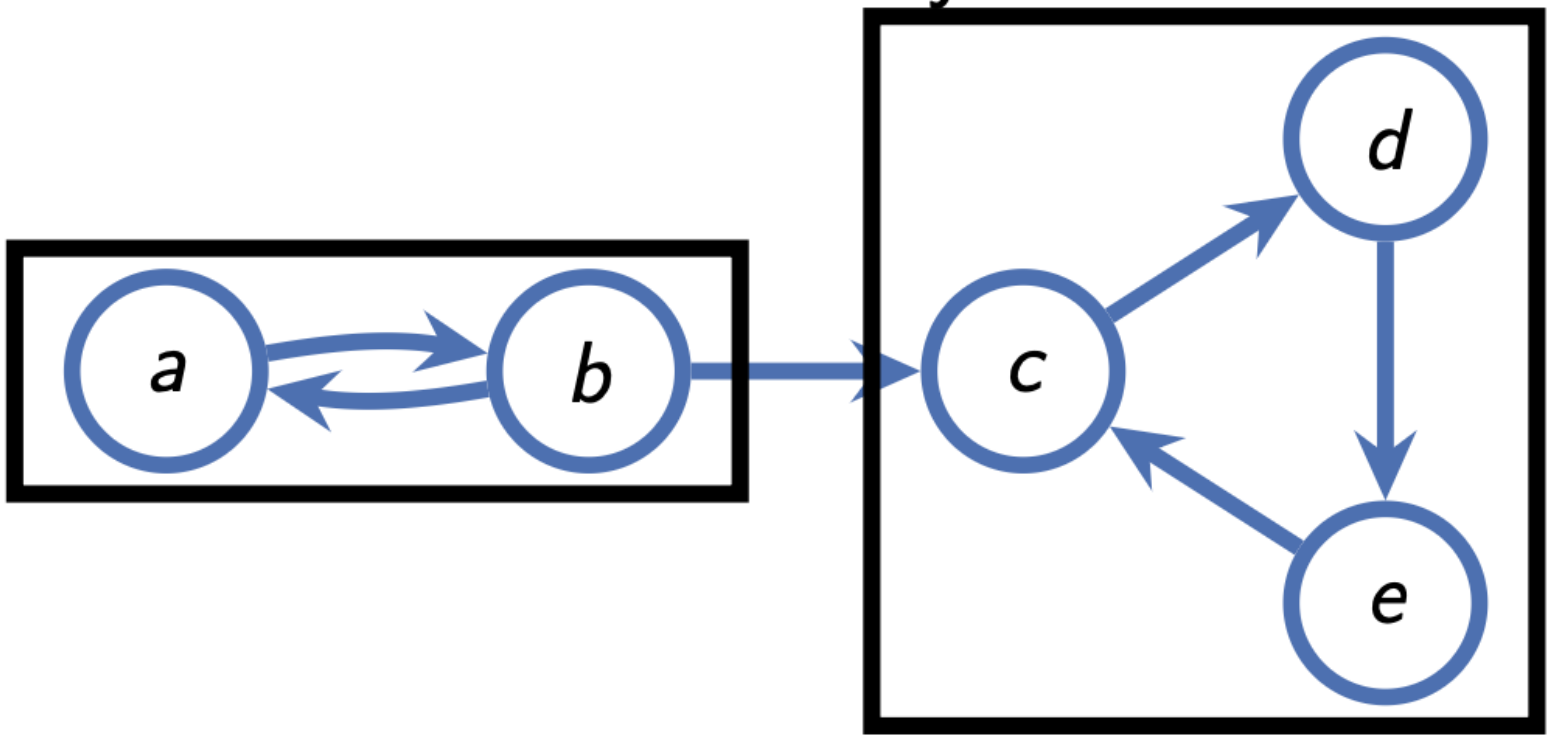
Attack

Gunfight rules:

iff all attackers are

iff there is an attacker

odd vs. even cycles



Abstract argumentation semantics: gunfight rules

Argument

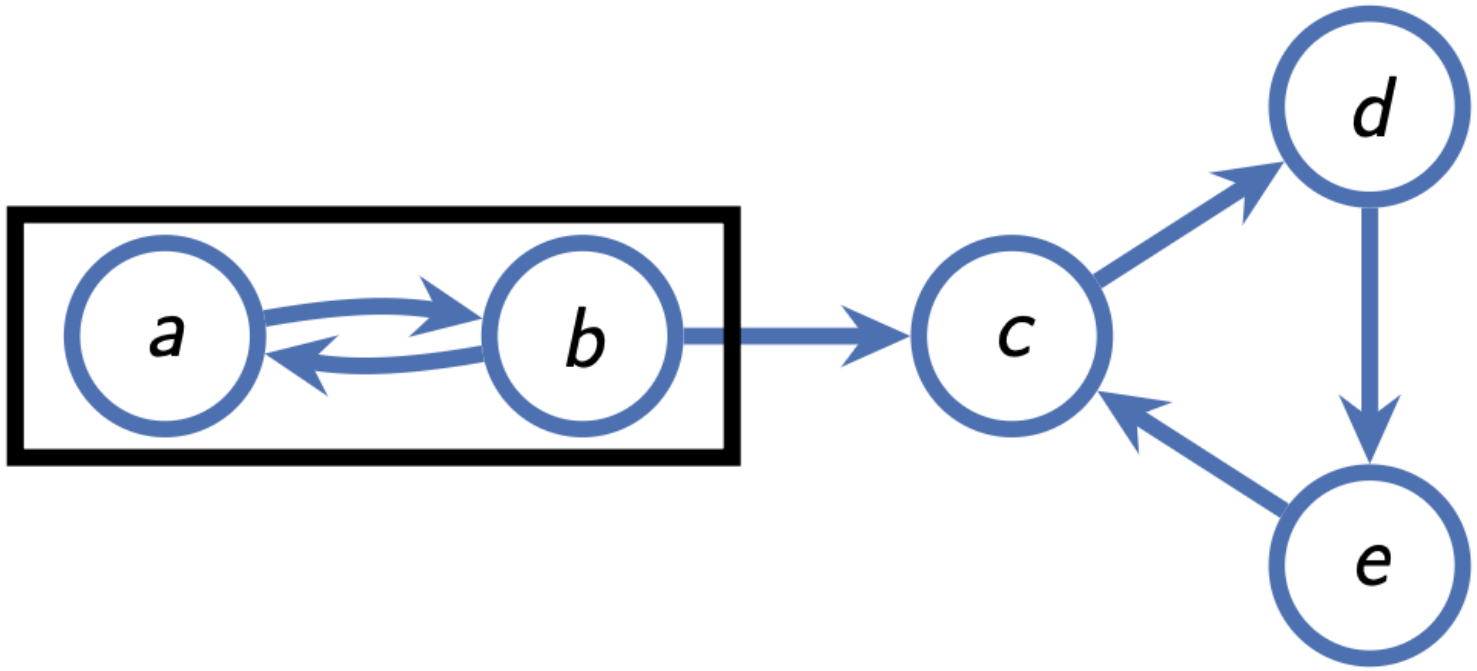
Attack

Gunfight rules:

iff all attackers are

iff there is an attacker

We start from the left



Abstract argumentation semantics: gunfight rules

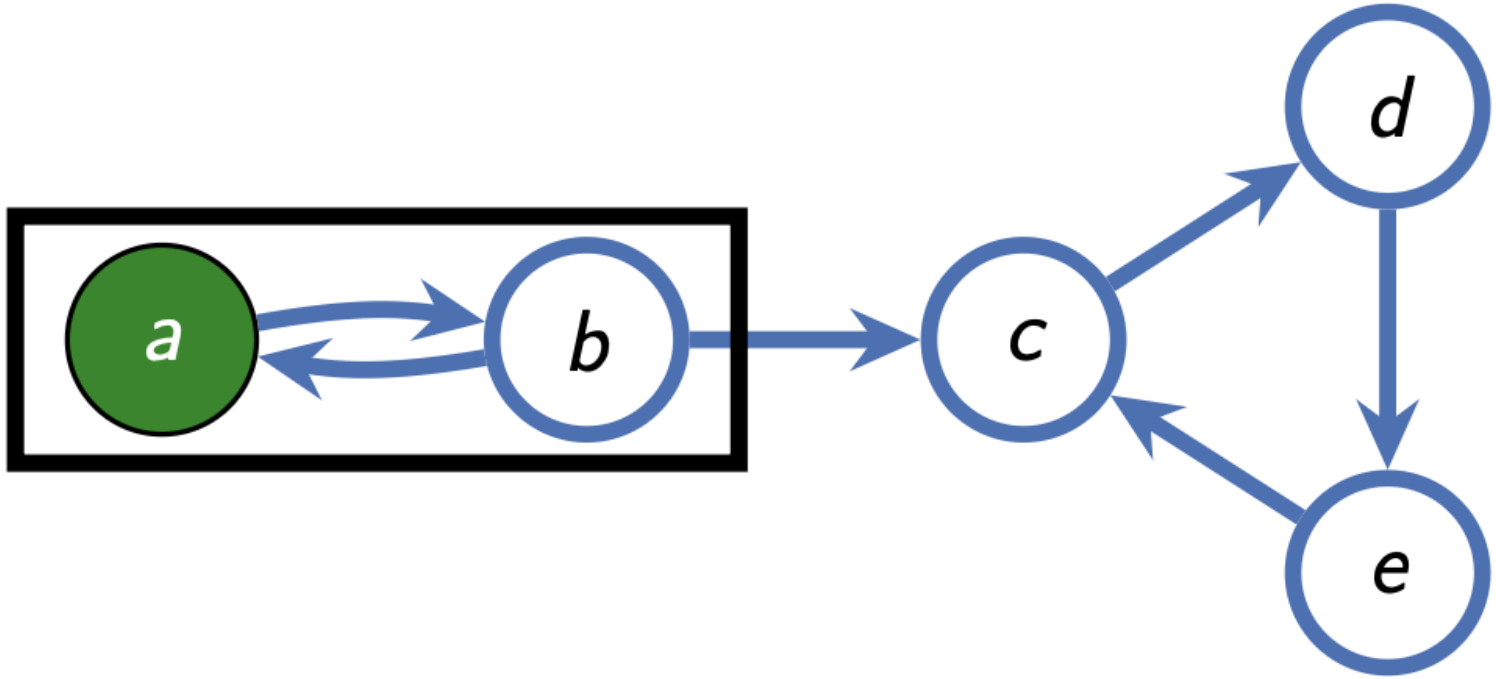
Argument

Attack

Gunfight rules:

iff all attackers are

iff there is an attacker



Abstract argumentation semantics: gunfight rules

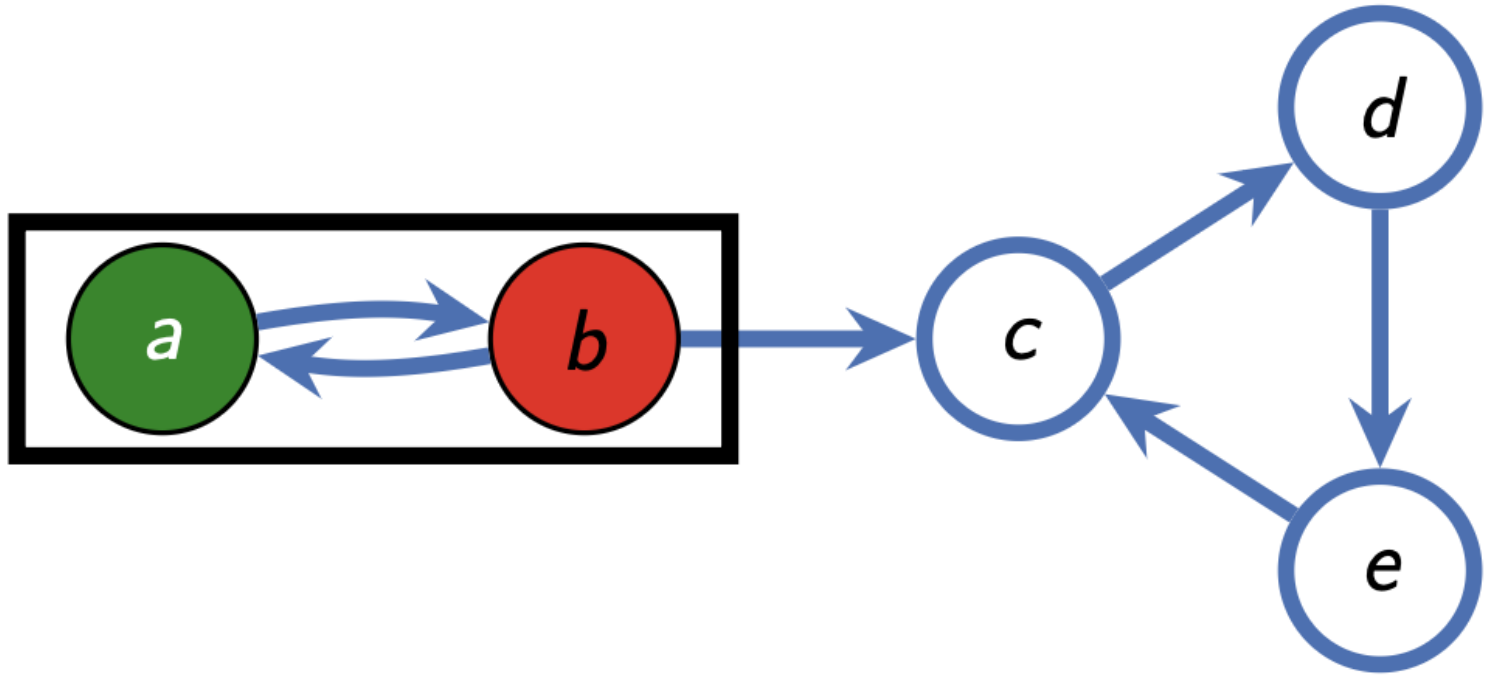
Argument

Attack

Gunfight rules:

iff all attackers are

iff there is an attacker



Abstract argumentation semantics: gunfight rules

Argument

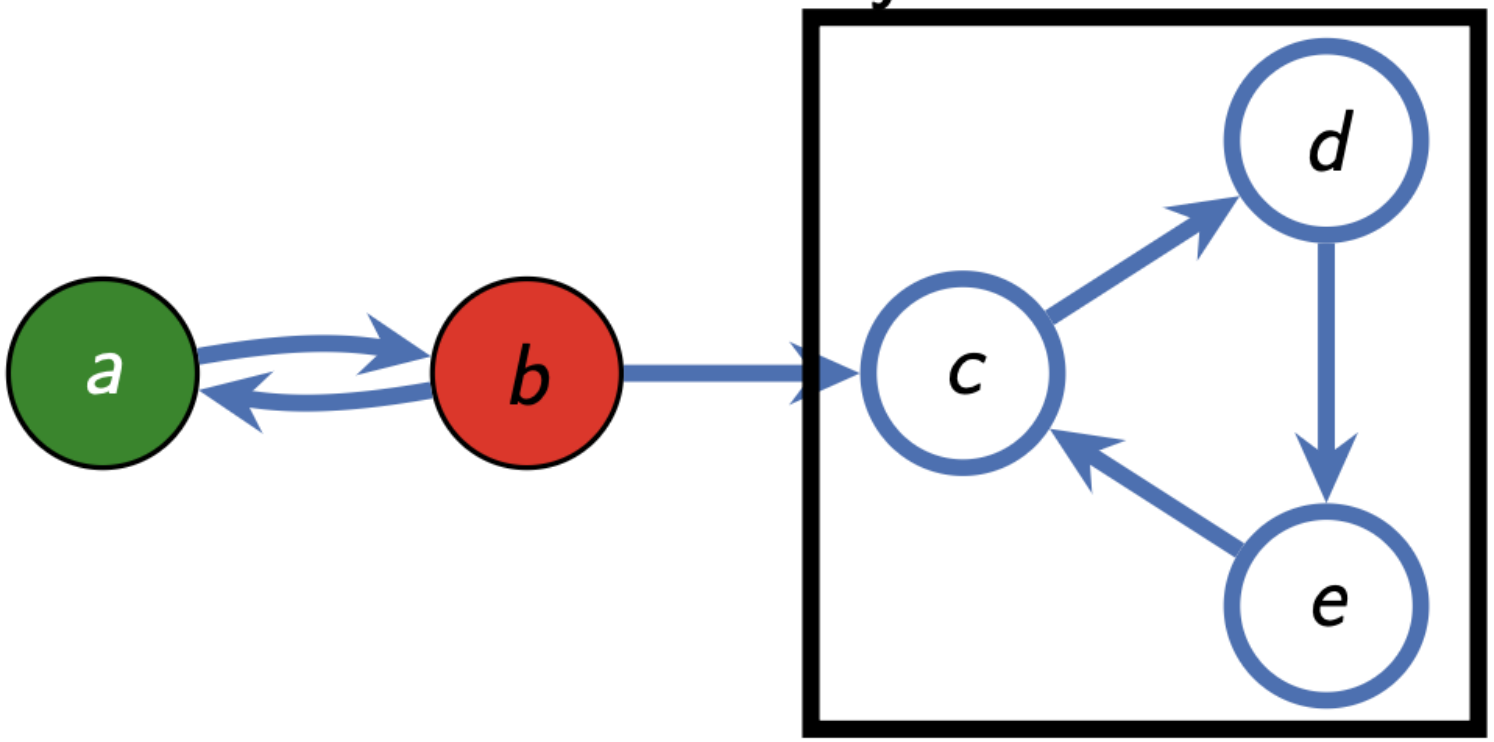
Attack

Gunfight rules:

iff all attackers are

iff there is an attacker

odd vs. even cycles



Abstract argumentation semantics: gunfight rules

Argument

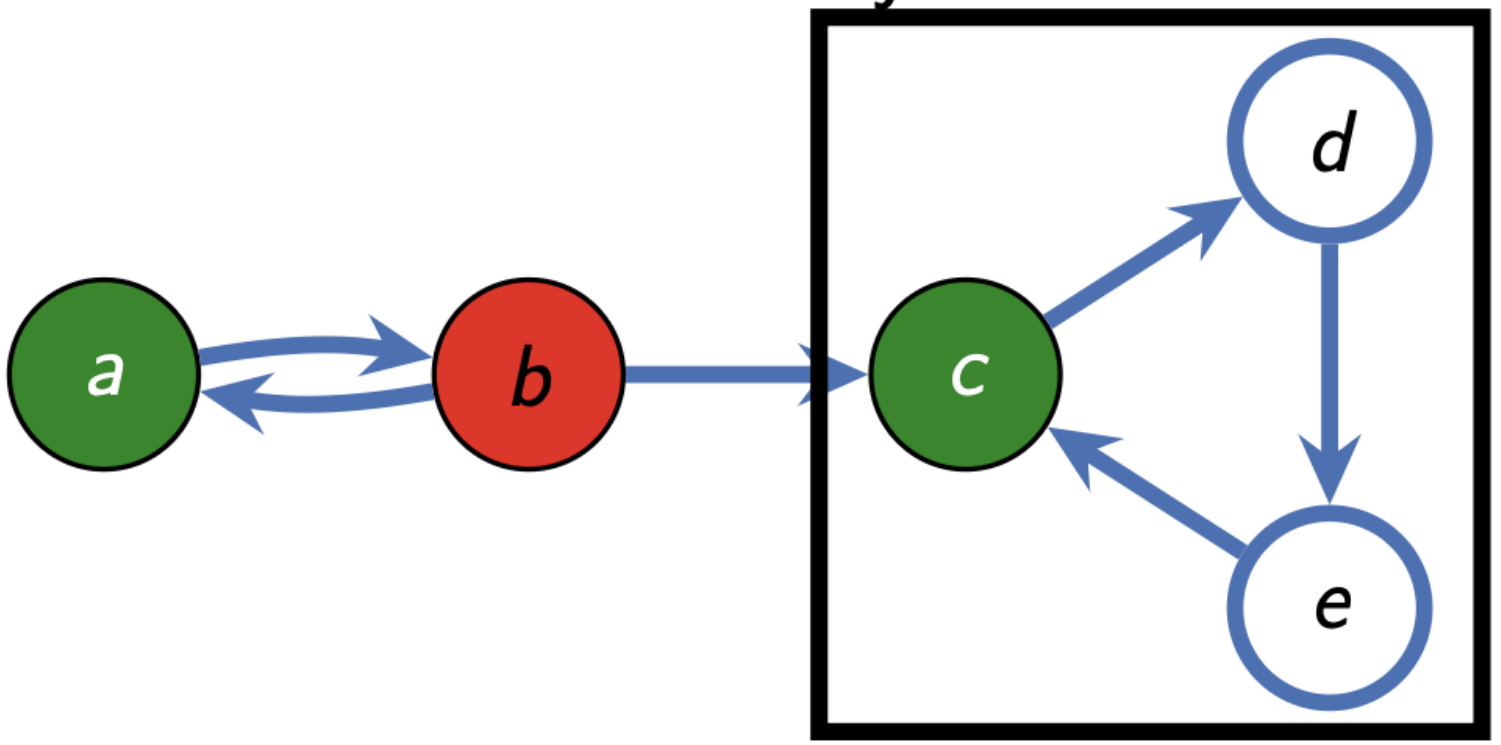
Attack

Gunfight rules:

iff all attackers are

iff there is an attacker

odd vs. even cycles



Abstract argumentation semantics: gunfight rules

Argument

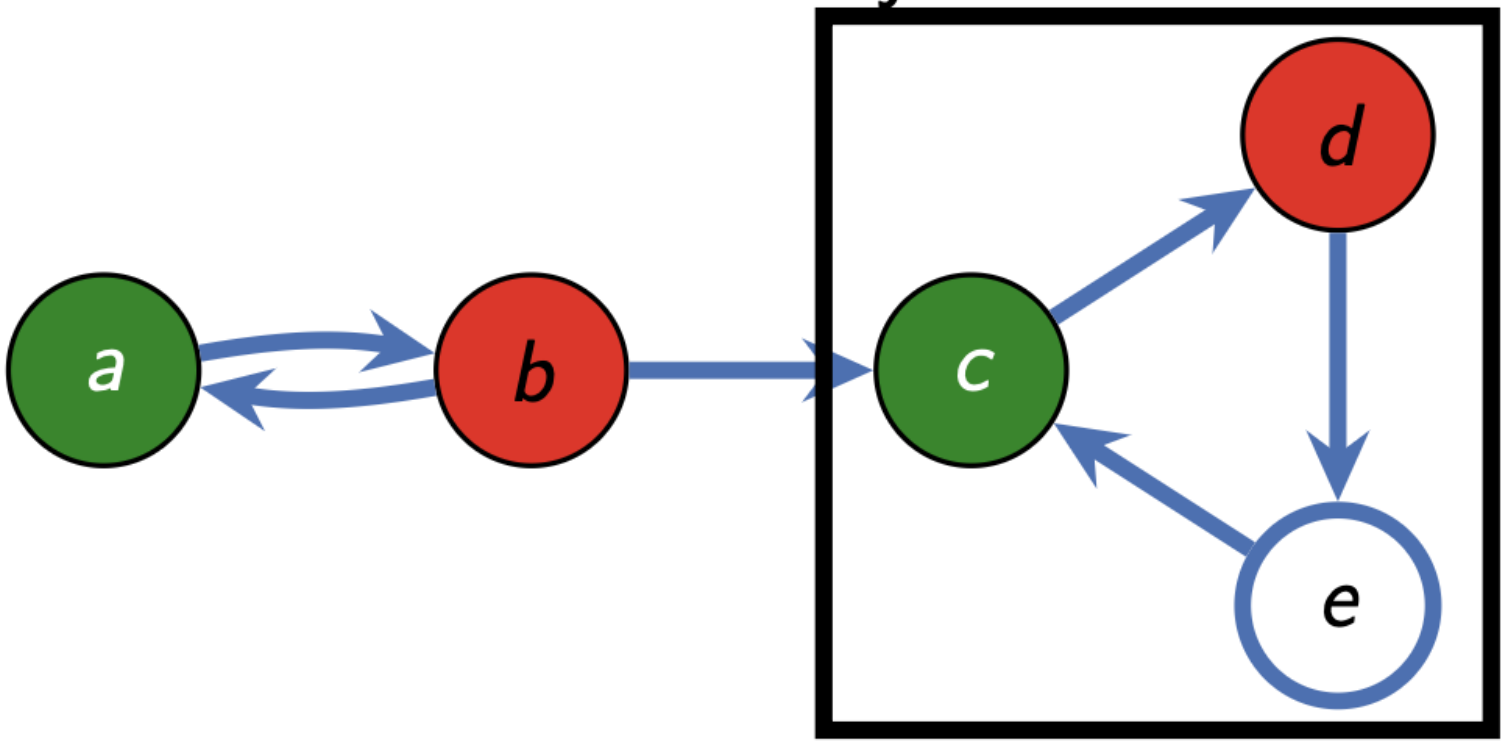
Attack

Gunfight rules:

iff all attackers are

iff there is an attacker

odd vs. even cycles



Abstract argumentation semantics: gunfight rules

Argument

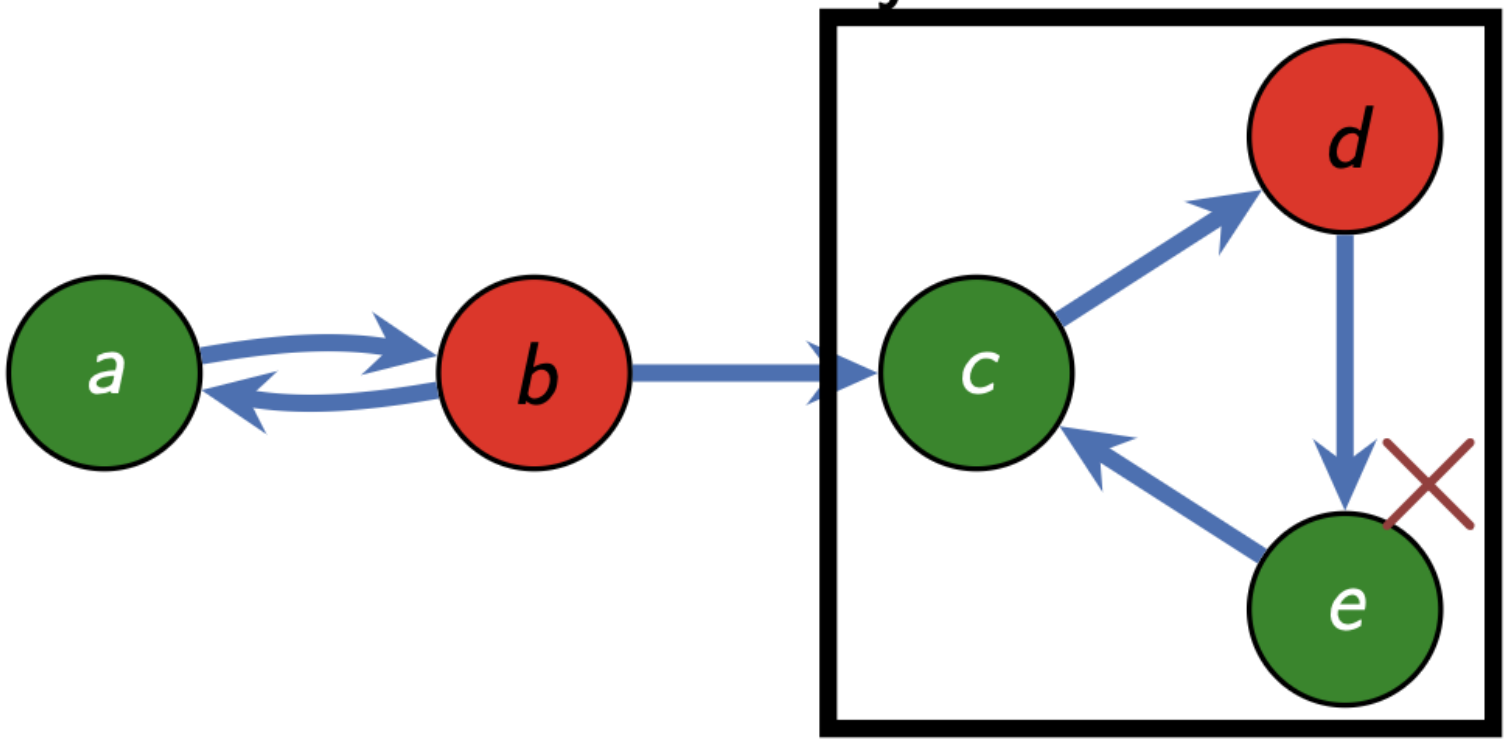
Attack

Gunfight rules:

iff all attackers are

iff there is an attacker

odd vs. even cycles



Abstract argumentation semantics: gunfight rules

Argument

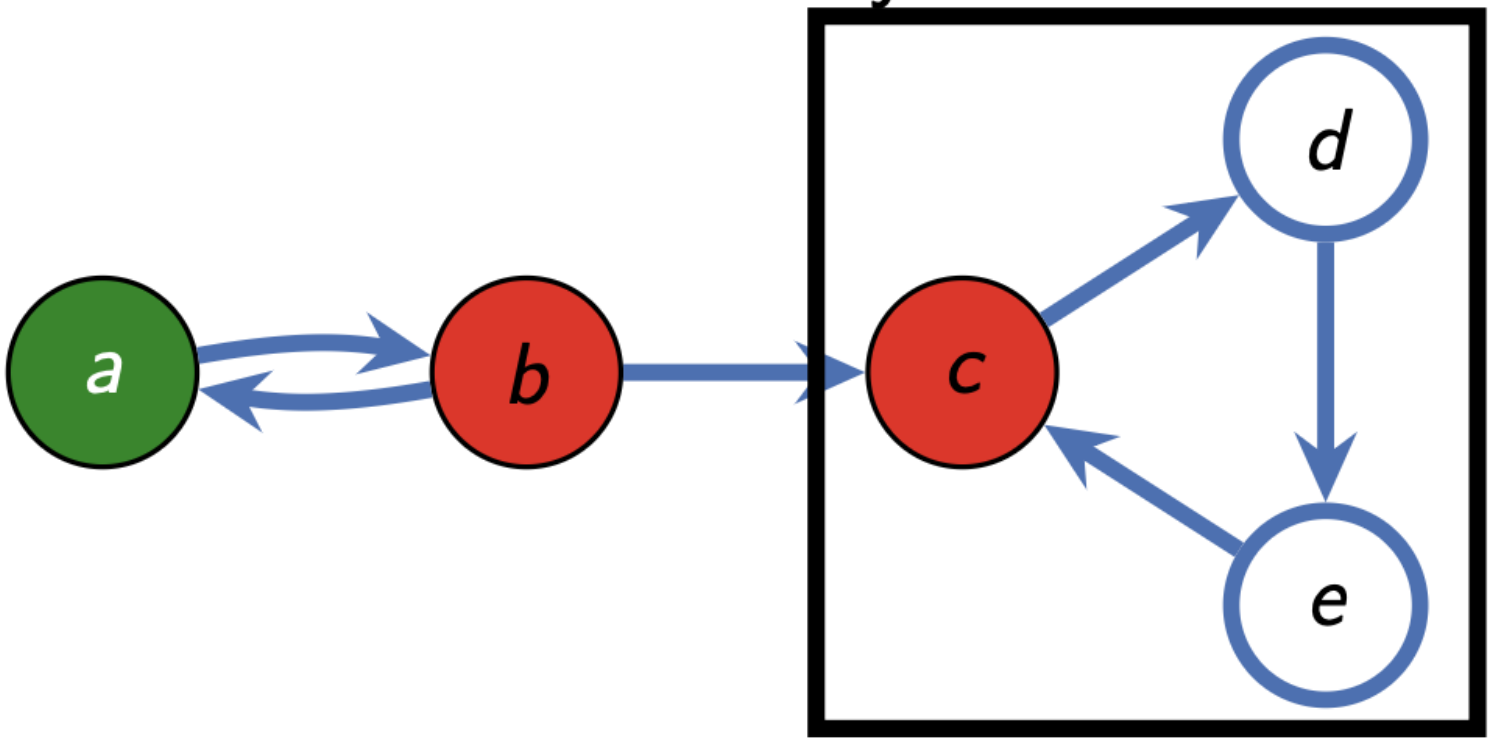
Attack

Gunfight rules:

iff all attackers are

iff there is an attacker

odd vs. even cycles



Abstract argumentation semantics: gunfight rules

Argument

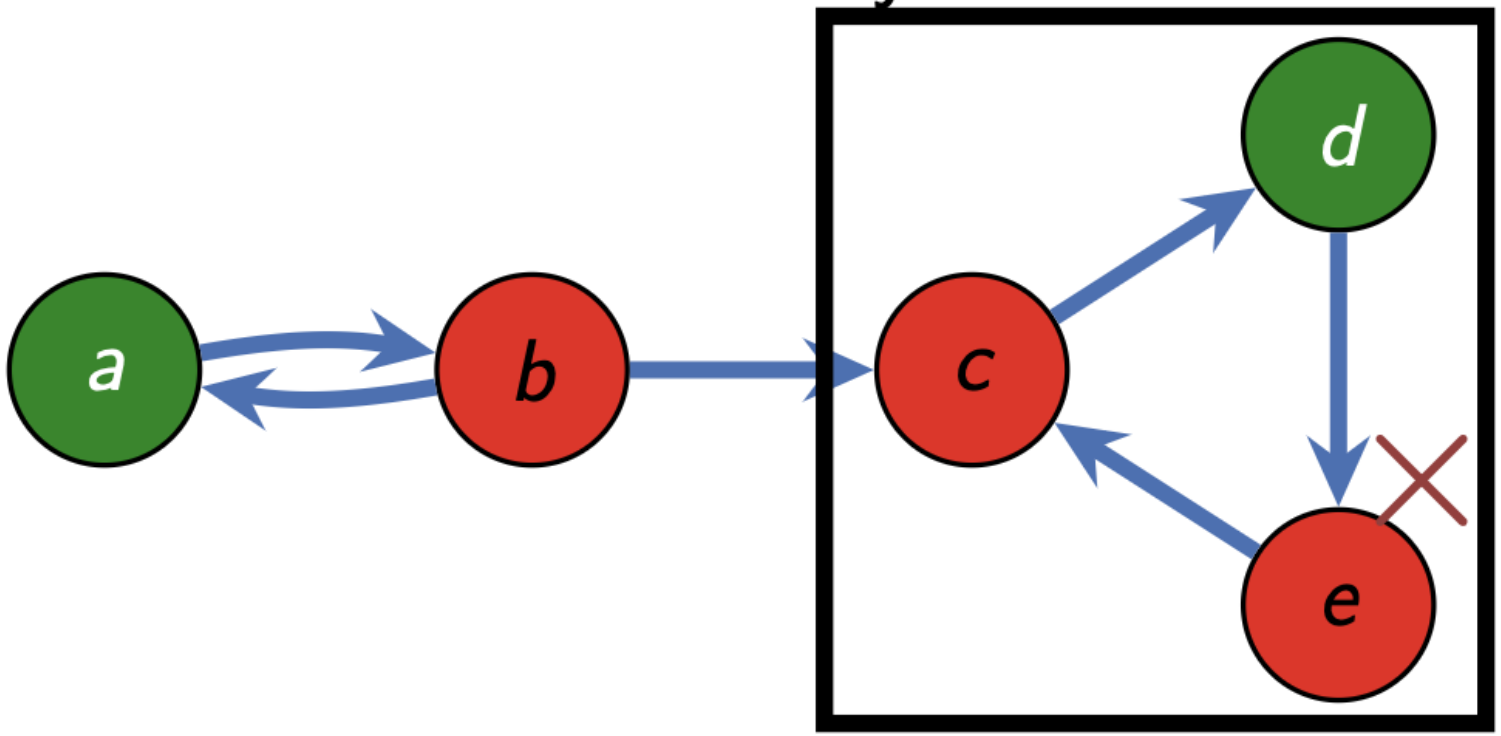
Attack

Gunfight rules:

iff all attackers are

iff there is an attacker

odd vs. even cycles



Abstract argumentation semantics: gunfight rules

Argument

Attack

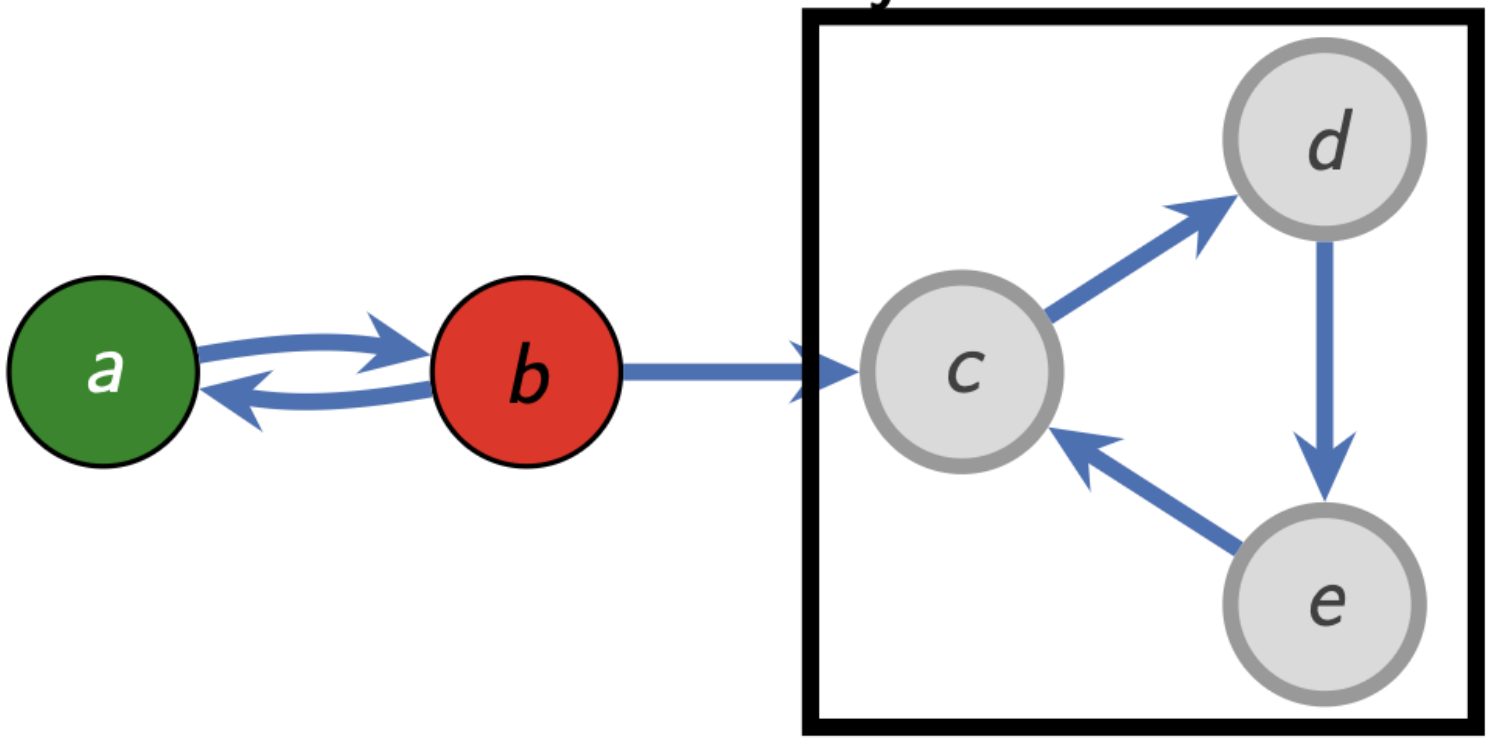
Gunfight rules:

iff all attackers are

iff there is an attacker

iff not all attackers are and no attacker is

odd vs. even cycles



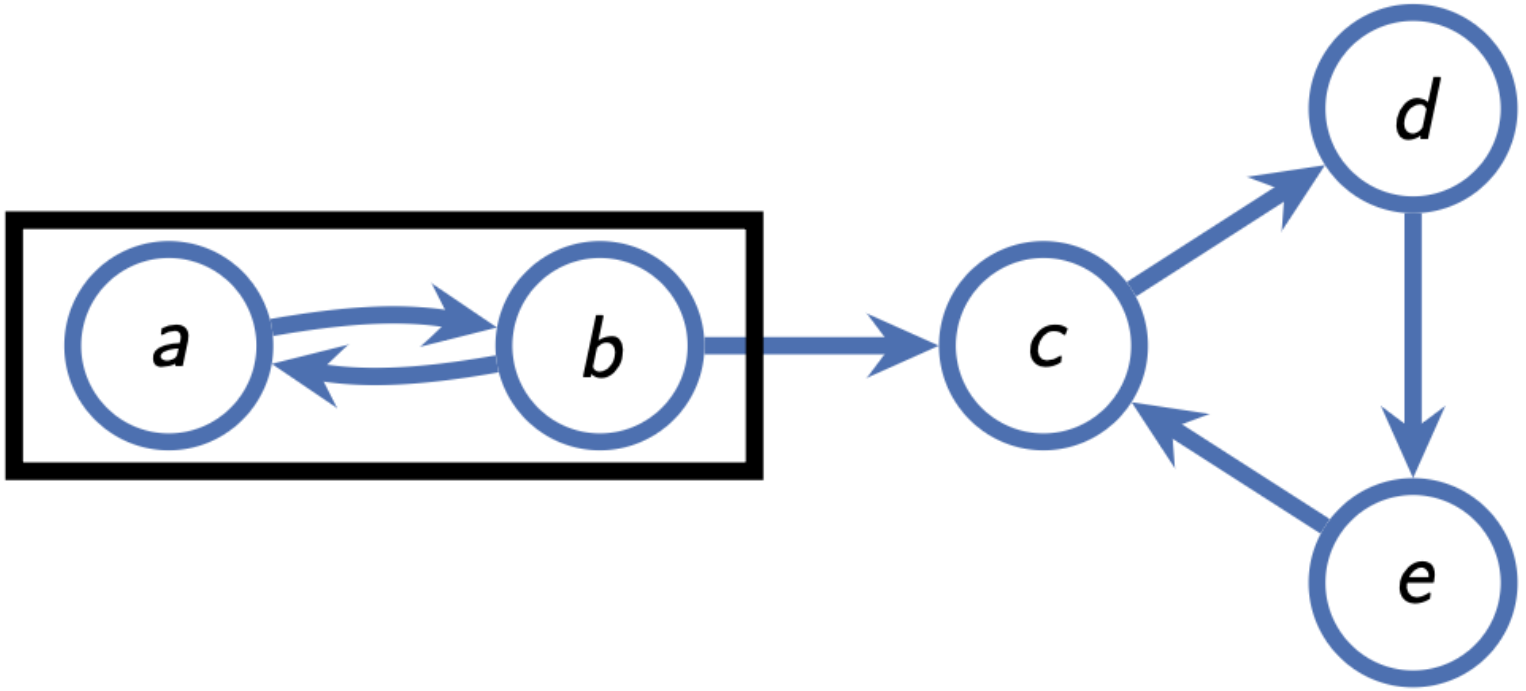
Abstract argumentation semantics: gunfight rules

Argument

Attack

Gunfight rules:

- iff all attackers are
- iff there is an attacker
- iff not all attackers are and no attacker is



Abstract argumentation semantics: gunfight rules

Argument

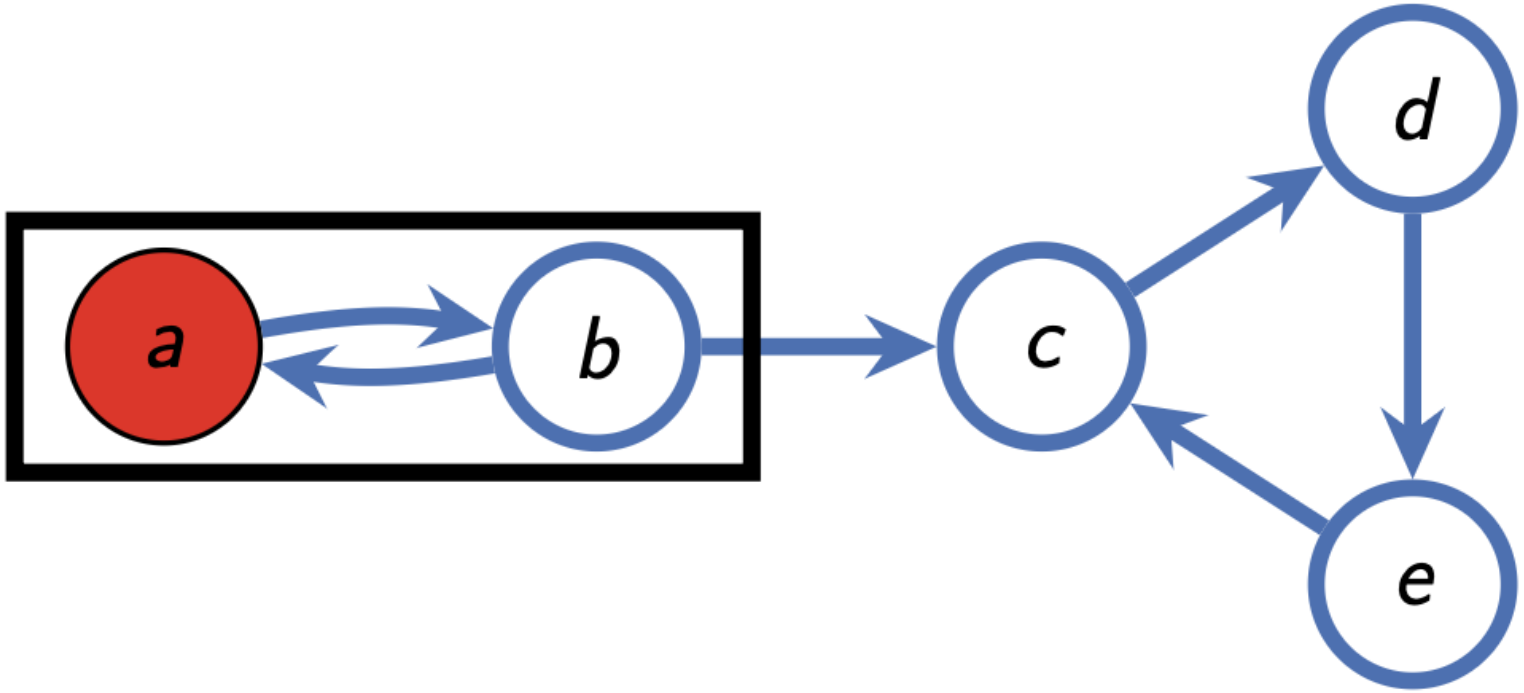
Attack

Gunfight rules:

iff all attackers are

iff there is an attacker

iff not all attackers are and no attacker is



Abstract argumentation semantics: gunfight rules

Argument

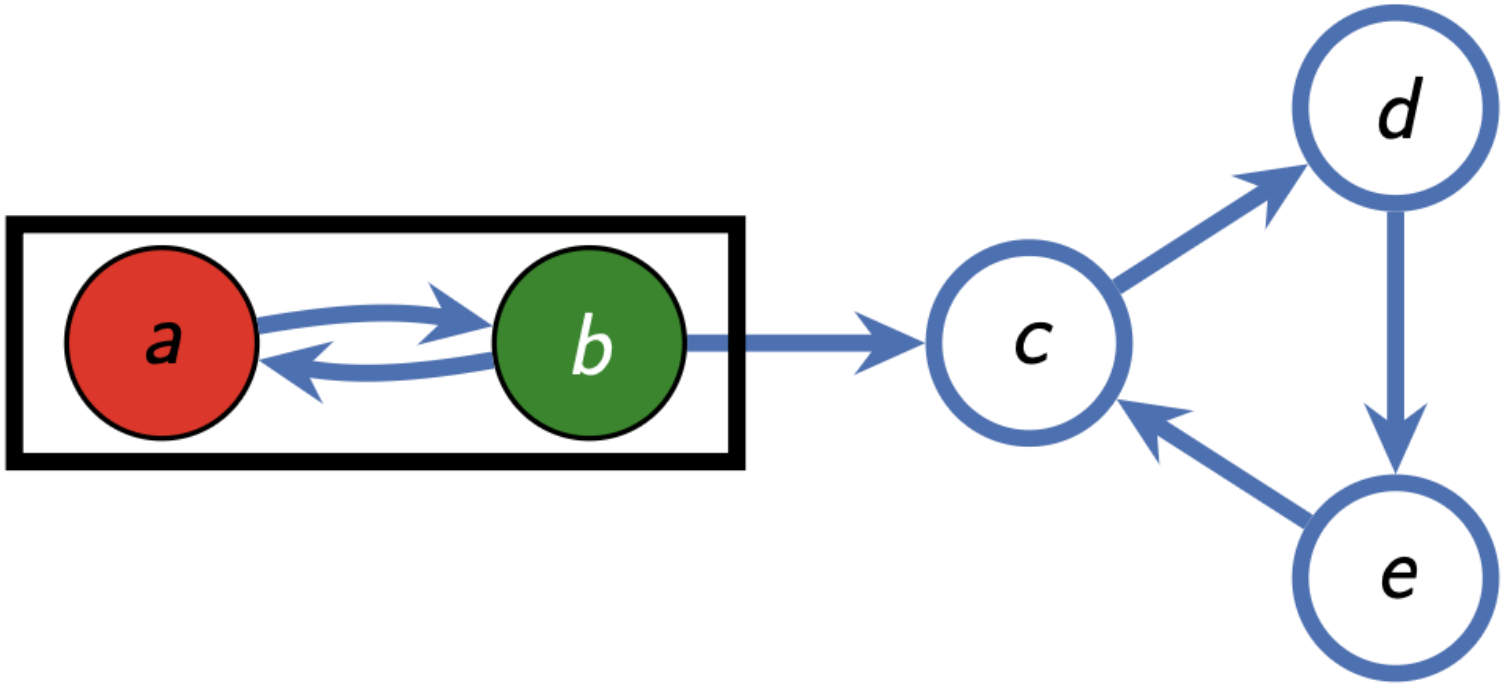
Attack

Gunfight rules:

iff all attackers are

iff there is an attacker

iff not all attackers are and no attacker is



Abstract argumentation semantics: gunfight rules

Argument

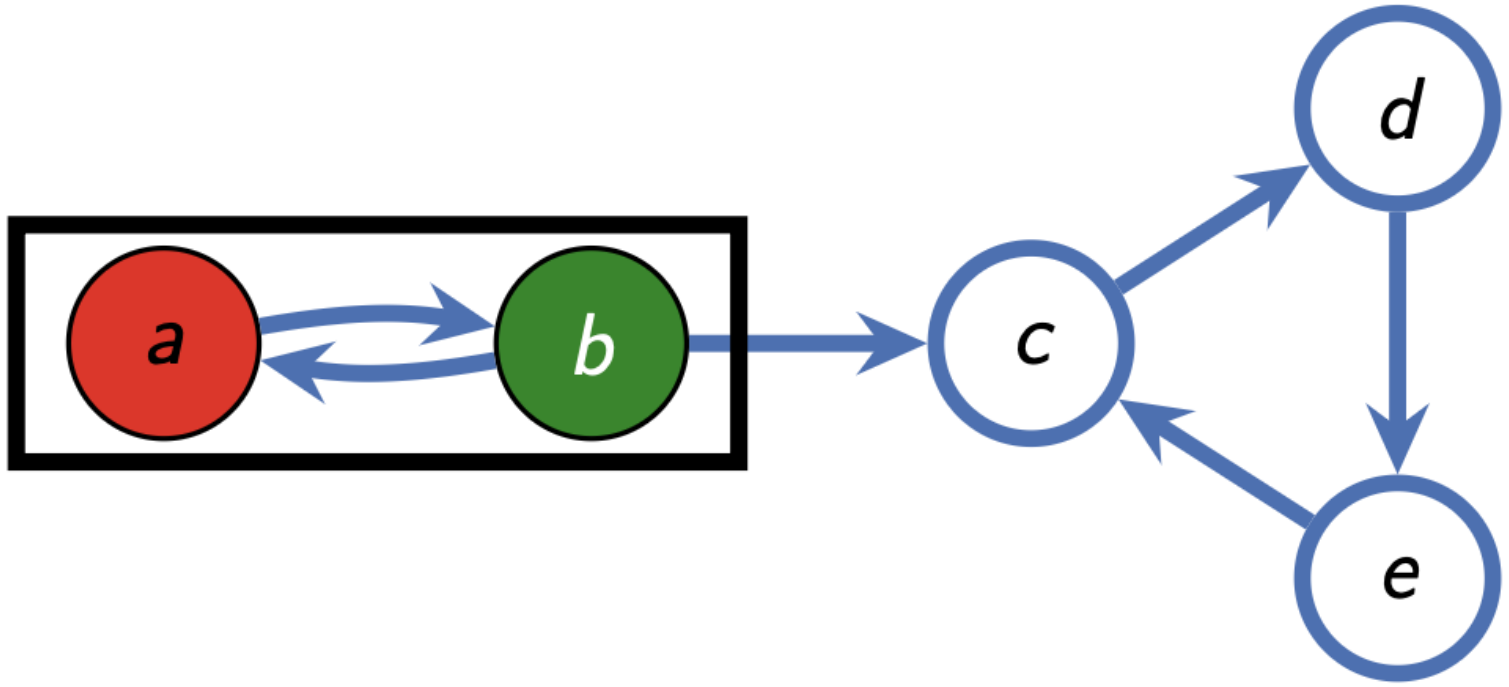
Attack

Gunfight rules:

iff all attackers are

iff there is an attacker

iff not all attackers are and no attacker is



Abstract argumentation semantics: gunfight rules

Argument

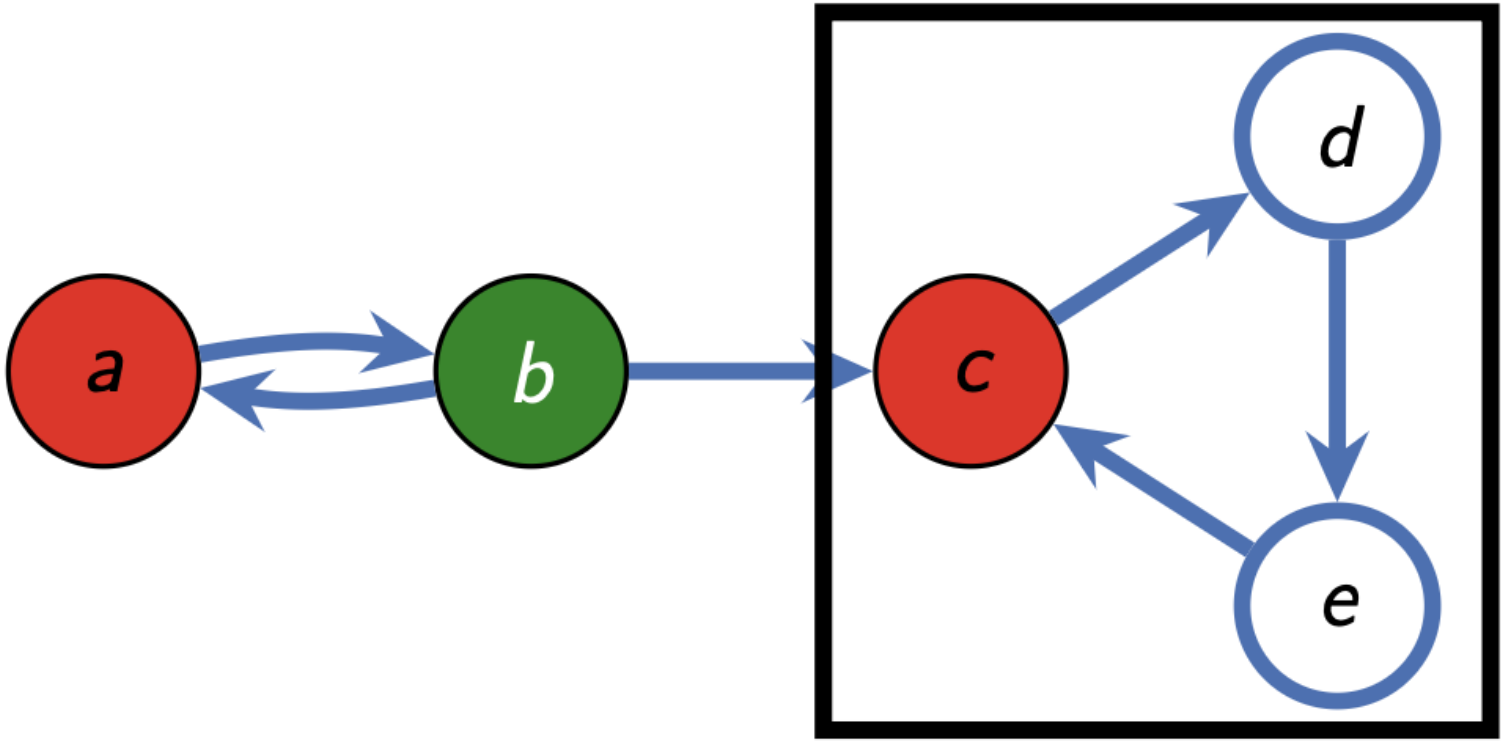
Attack

Gunfight rules:

iff all attackers are

iff there is an attacker

iff not all attackers are and no attacker is



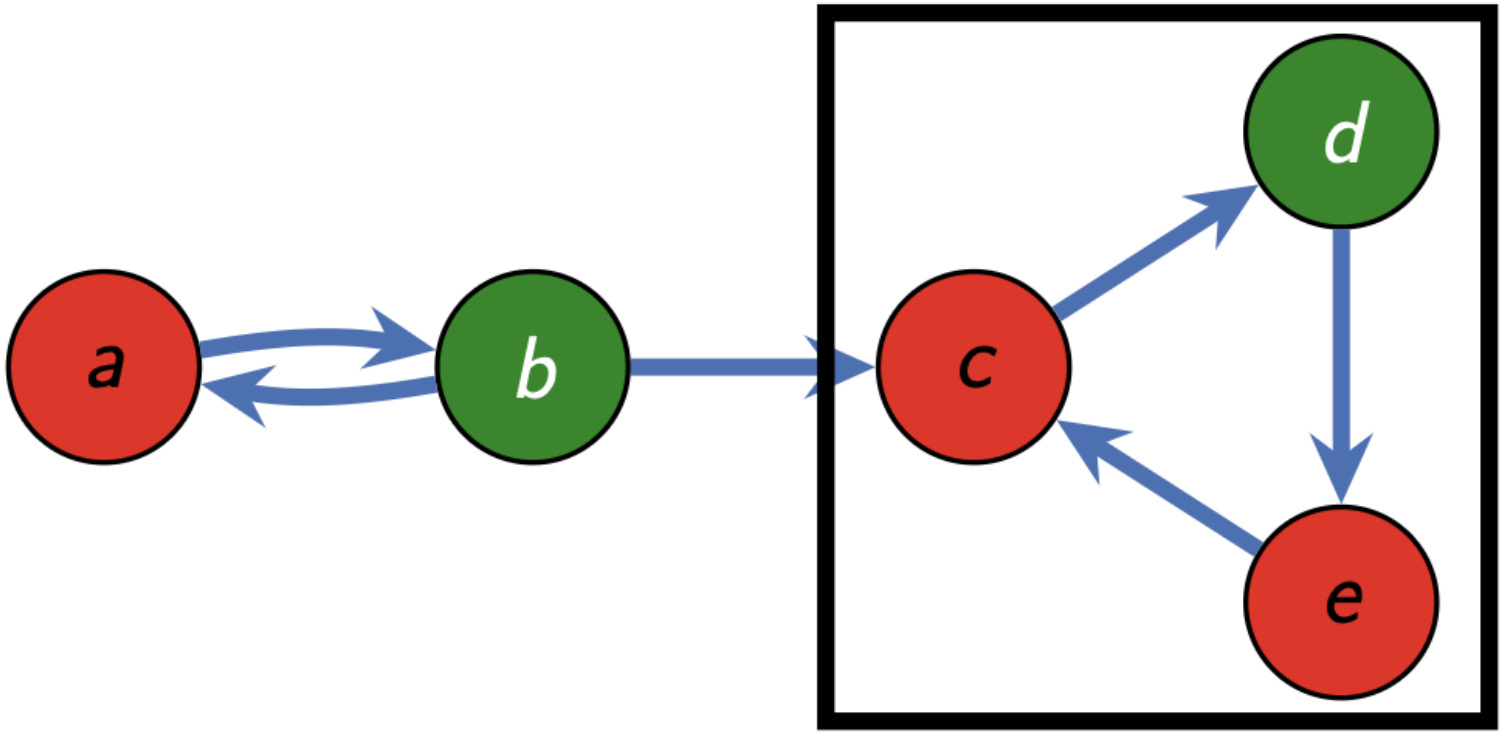
Abstract argumentation semantics: gunfight rules

Argument

Attack

Gunfight rules:

- iff all attackers are
- iff there is an attacker
- iff not all attackers are and no attacker is



Abstract argumentation semantics: gunfight rules

Argument

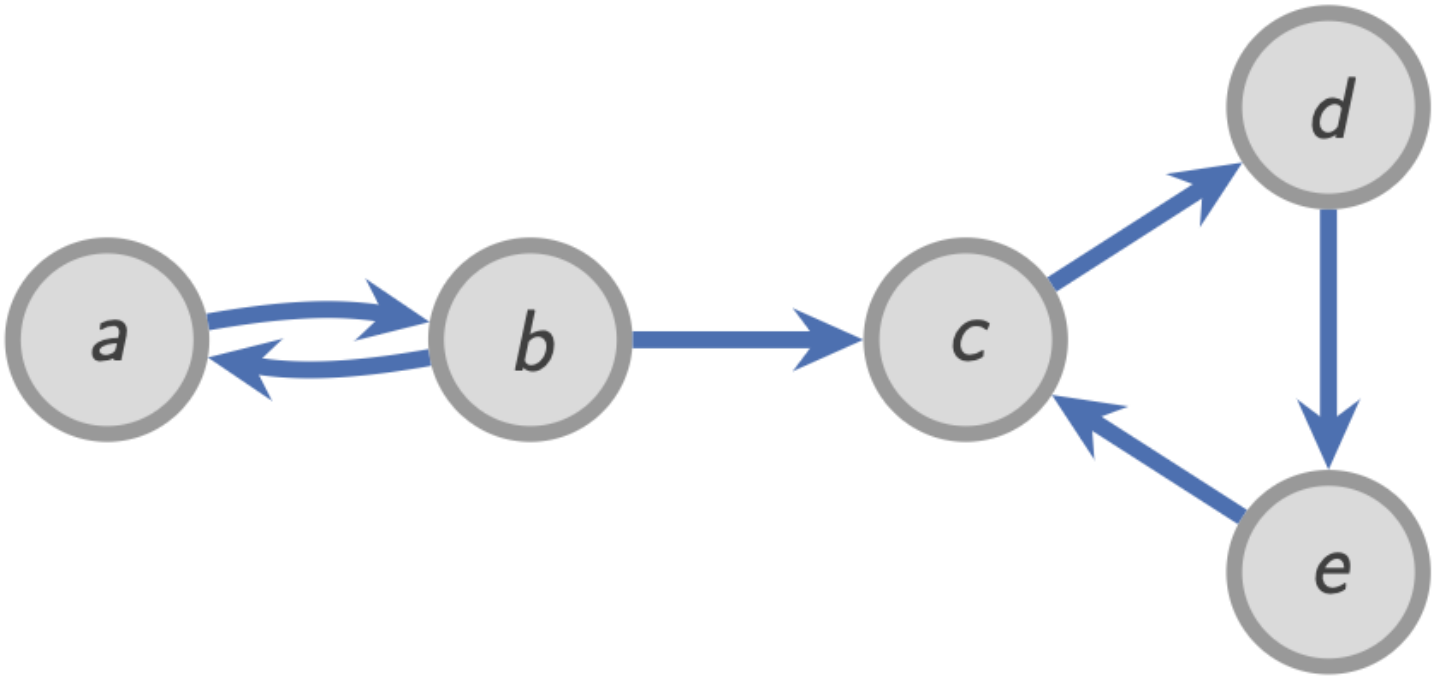
Attack

Gunfight rules:

iff all attackers are

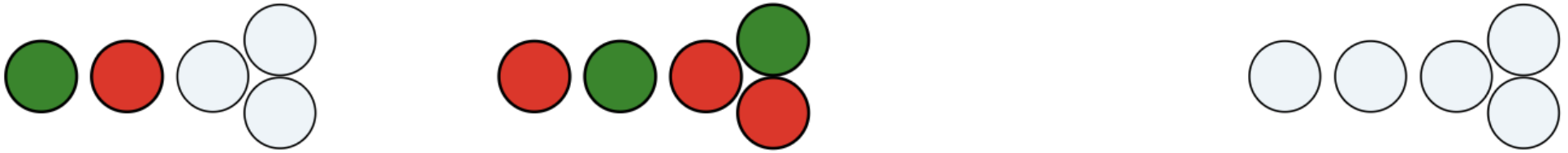
iff there is an attacker

iff not all attackers are and no attacker is

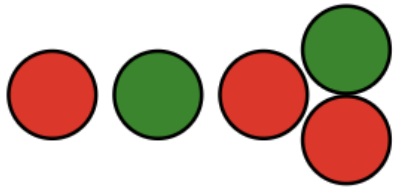


Abstract argumentation semantics: gunfight rules

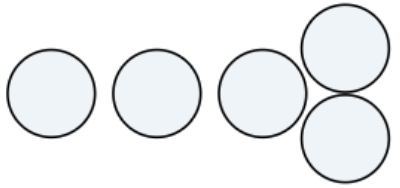
- Three complete extensions (3 valued)



- One stable extension (2 valued)



- One grounded extension (least complete)



- Two preferred extensions (maximal complete)

Day 1 Dung paradigm shift and reasoning alignment

1.1 AI Logic (JLC corner) is nonmonotonic logic (since 1980)

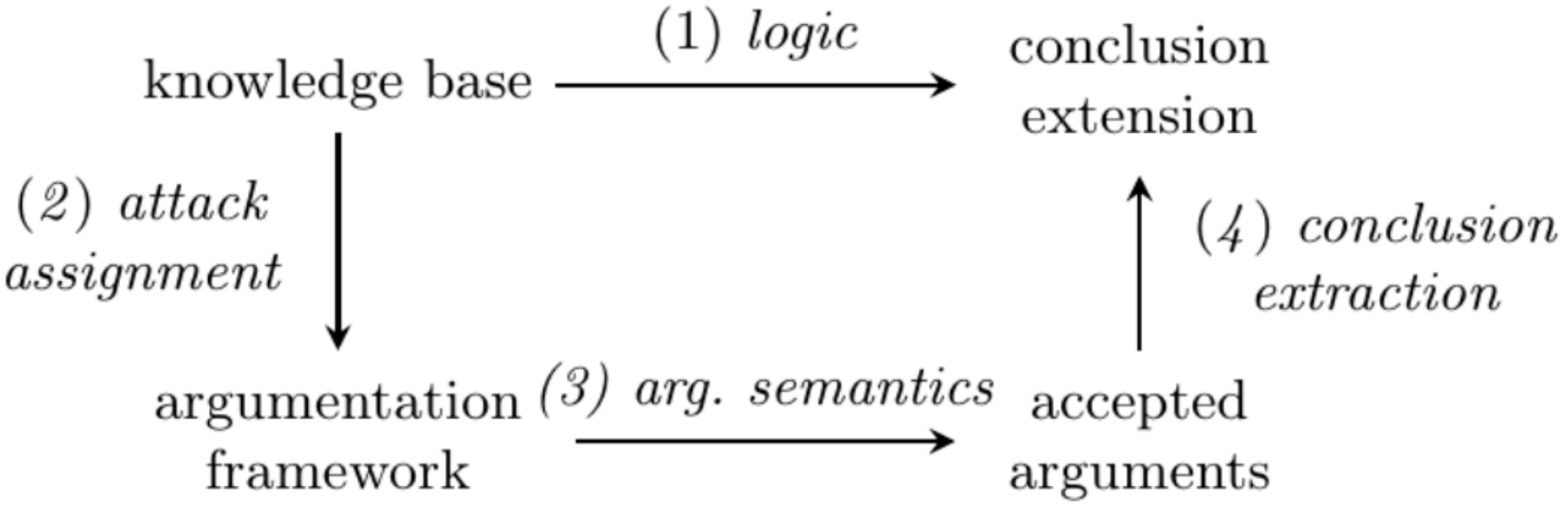
1.2 Explanation of black boxes in new generation AI: secretary metaphor

1.3 Dung 1995 paradigm shift relates logic and argumentation based explanation

1.4 Reasoning alignment of logic and argumentation based explanation

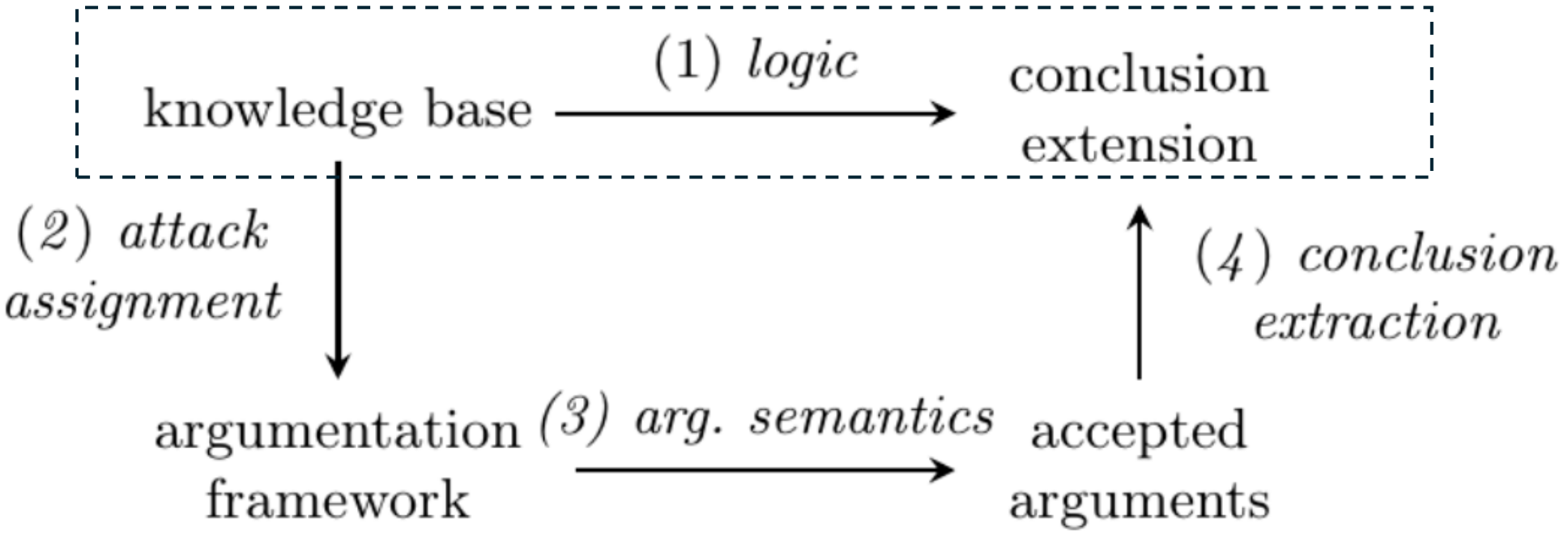
1.5 Using logic: from inference towards dialogue and decision making, to logic in action

Alignment through commutative diagram



Liuwen Yu, Leendert van der Torre, and Réka Markovich, "Thirteen Challenges in Formal and Computational Argumentation", in Handbook of Formal Argumentation, Volume 3, College Publications, 2024, pp. 890–976.

Alignment through commutative diagram



Liuwen Yu, Leendert van der Torre, and Réka Markovich, "Thirteen Challenges in Formal and Computational Argumentation", in Handbook of Formal Argumentation, Volume 3, College Publications, 2024, pp. 890–976.

Direct MCS consequence

Definition

$\mathrm{MCS}(K, A)$ contains the inclusion-maximal $L \subseteq K$ for which $L \cup A$ is consistent, and

$$C(K, A) = \bigcap_{L \in \mathrm{MCS}(K,A)} \mathrm{Cn}(L \cup A).$$

Example

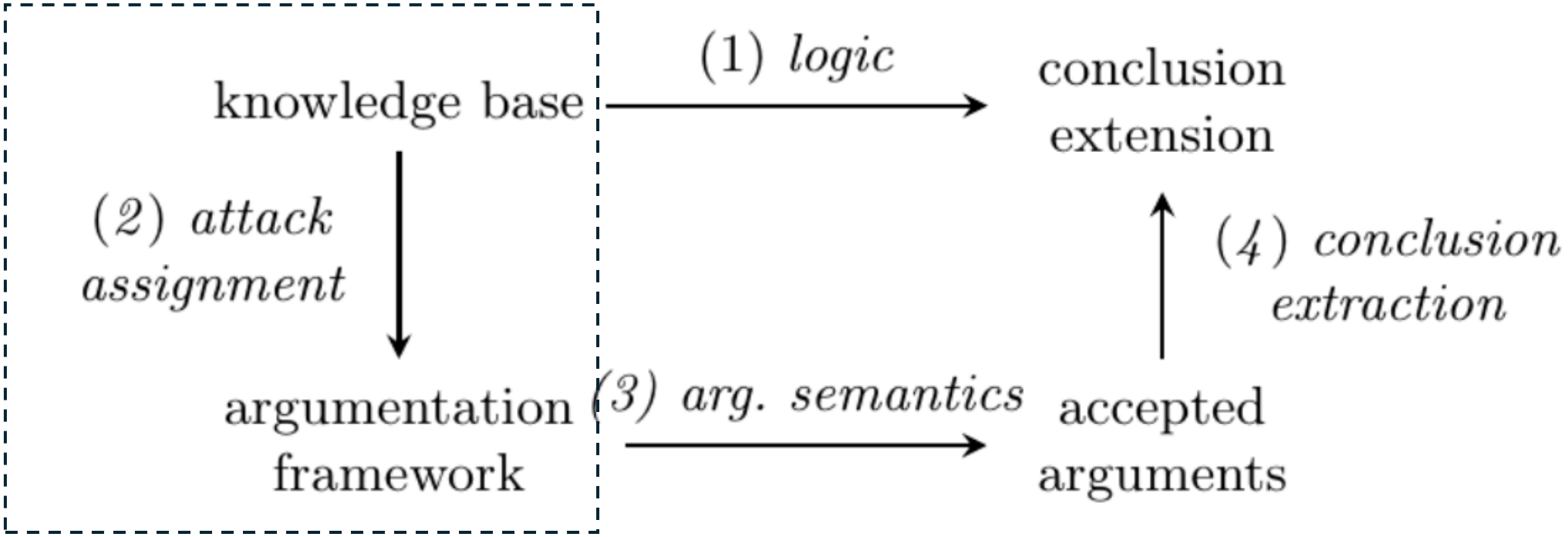
For $K = \{p, \neg p, q\}$ and $A = \varnothing$,

$$\mathrm{MCS}(K, A) = \{L, M\}, \qquad L = \{p, q\}, \quad M = \{\neg p, q\},$$

$$C(K, A) = \mathrm{Cn}(L) \cap \mathrm{Cn}(M) = \mathrm{Cn}(\{q\}).$$

Liuwen Yu, Leendert van der Torre, and Réka Markovich, "Thirteen Challenges in Formal and Computational Argumentation", in Handbook of Formal Argumentation, Volume 3, College Publications, 2024, pp. 890–976.

Alignment through commutative diagram



Liuwen Yu, Leendert van der Torre, and Réka Markovich, "Thirteen Challenges in Formal and Computational Argumentation", in Handbook of Formal Argumentation, Volume 3, College Publications, 2024, pp. 890–976.

Abstract argumentation framework construction

Definition (Argument)

Let $K, A \subseteq \mathcal{L}$ be finite and let A be consistent. An argument based on (K, A) is a pair $\alpha = \langle \Phi, \varphi \rangle$ such that $\Phi \subseteq K$, $A \cup \Phi \nvdash \bot$, $A \cup \Phi \vdash \varphi$, and no proper subset $\Phi' \subsetneq \Phi$ satisfies $A \cup \Phi' \vdash \varphi$. Its support is Φ, its conclusion is φ, and $\mathrm{Arg}(K, A)$ is the set of all such arguments.

Definition (Attack)

An argument $\alpha = \langle \Phi, \varphi \rangle$ attacks an argument $\beta = \langle \Psi, \psi \rangle$ exactly when $\theta \equiv \neg\varphi$ for some $\theta \in \Psi$.

Definition (Induced framework)

Let $\mathrm{Att}(K, A) = \{(\alpha, \beta) \in \mathrm{Arg}(K, A)^2 \mid \alpha \text{ attacks } \beta\}$. The induced abstract argumentation framework is $AF(K, A) = \langle \mathrm{Arg}(K, A), \mathrm{Att}(K, A) \rangle$.

Abstract argumentation framework construction

Definition (Argument)

Let $K, A \subseteq \mathcal{L}$ be finite and let A be consistent. An argument based on (K, A) is a pair $\alpha = \langle \Phi, \varphi \rangle$ such that $\Phi \subseteq K$, $A \cup \Phi \nvdash \bot$, $A \cup \Phi \vdash \varphi$, and no proper subset $\Phi' \subsetneq \Phi$ satisfies $A \cup \Phi' \vdash \varphi$. Its support is Φ, its conclusion is φ, and $\mathrm{Arg}(K, A)$ is the set of all such arguments.

Definition (Attack)

An argument $\alpha = \langle \Phi, \varphi \rangle$ attacks an argument $\beta = \langle \Psi, \psi \rangle$ exactly when $\theta \equiv \neg\varphi$ for some $\theta \in \Psi$.

Definition (Induced framework)

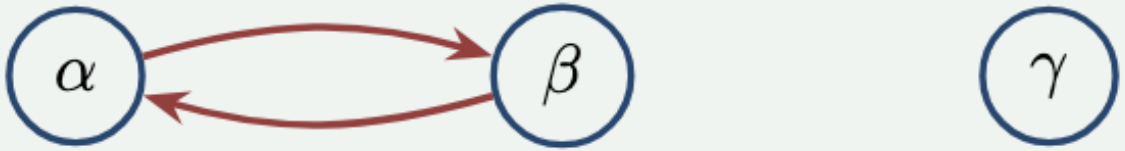
Let $\mathrm{Att}(K, A) = \{(\alpha, \beta) \in \mathrm{Arg}(K, A)^2 \mid \alpha \text{ attacks } \beta\}$. The induced abstract argumentation framework is $AF(K, A) = \langle \mathrm{Arg}(K, A), \mathrm{Att}(K, A) \rangle$.

Example

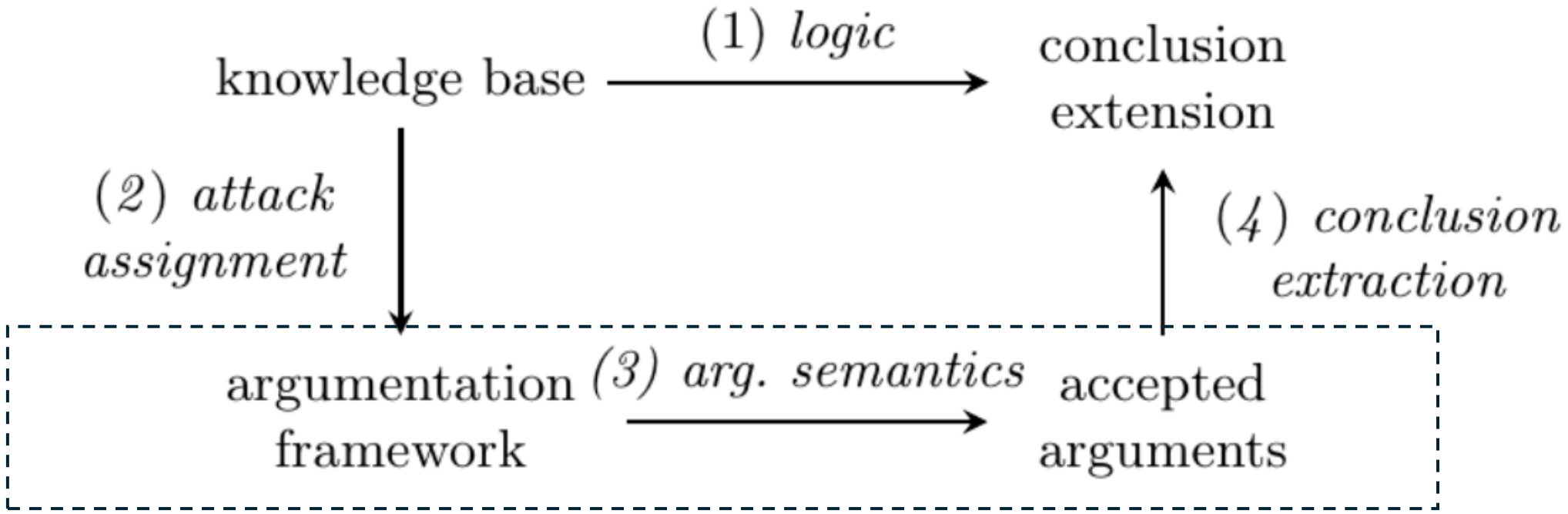
For $K = \{p, \neg p, q\}$ and $A = \varnothing$, consider

$$\alpha = \langle \{p\}, p \rangle, \qquad \beta = \langle \{\neg p\}, \neg p \rangle, \qquad \gamma = \langle \{q\}, q \rangle.$$

The arguments α and β attack each other, while γ is unattacked.



Alignment through commutative diagram



Liuwen Yu, Leendert van der Torre, and Réka Markovich, "Thirteen Challenges in Formal and Computational Argumentation", in Handbook of Formal Argumentation, Volume 3, College Publications, 2024, pp. 890–976.

Stable extension

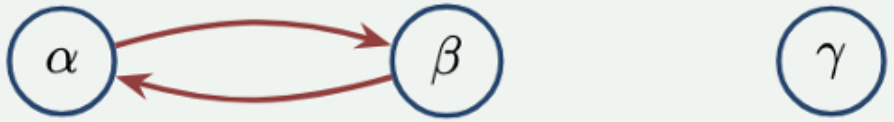
Definition (Stable extension)

For $AF = \langle AR, attacks \rangle$, a set $E \subseteq AR$ is stable exactly when no two arguments in E attack each other and every argument in $AR \setminus E$ is attacked by an argument in E.

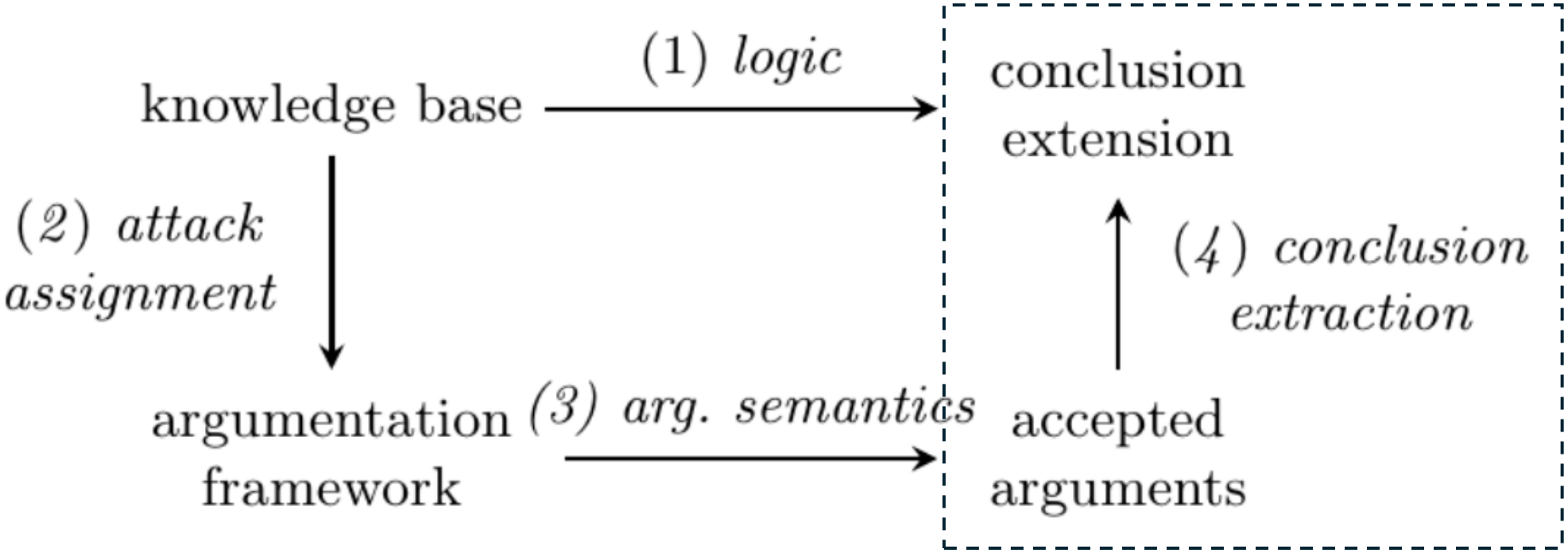
Example

The full stable extensions of $AF(K, \varnothing)$ are $E(L)$ and $E(M)$. On the displayed subframework they appear as

$$\{\alpha, \gamma\} \quad \text{and} \quad \{\beta, \gamma\}.$$



Alignment through commutative diagram



Liuwen Yu, Leendert van der Torre, and Réka Markovich, "Thirteen Challenges in Formal and Computational Argumentation", in Handbook of Formal Argumentation, Volume 3, College Publications, 2024, pp. 890–976.

Sceptical conclusion extraction

Definition (Conclusions and sceptical output)

For $E \subseteq \mathrm{Arg}(K, A)$, let $\mathrm{Concs}(E) = \{\varphi \mid \langle \Phi, \varphi \rangle \in E\}$. Define

$$\mathrm{Skept}(K, A) = \bigcap_{E \in \mathrm{Stb}(AF(K,A))} \mathrm{Concs}(E).$$

Example

$$\mathrm{Concs}(E(L)) = \mathrm{Cn}(\{p, q\}), \qquad \mathrm{Concs}(E(M)) = \mathrm{Cn}(\{\neg p, q\}),$$
$$\mathrm{Skept}(K, \varnothing) = \mathrm{Cn}(\{q\}).$$

MCSs correspond to stable extensions

Theorem (relative MCS–stable correspondence)

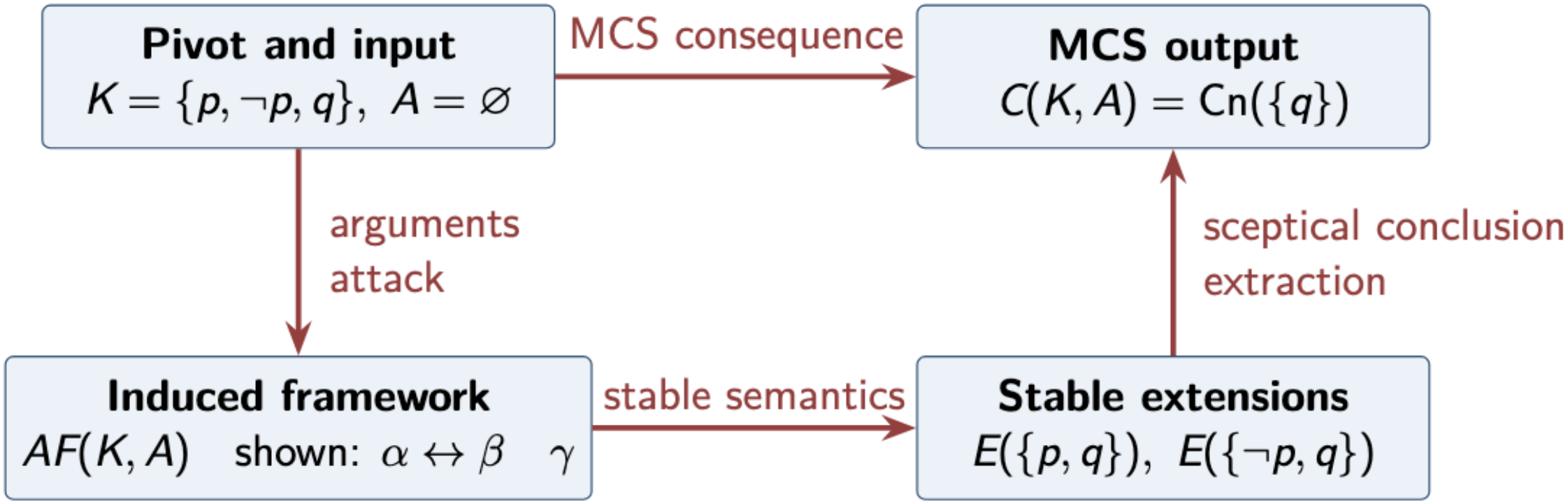
For finite $K, A \subseteq \mathcal{L}$, with A consistent, the attack construction satisfies

$$\mathrm{Stb}(AF(K, A)) = \{E(L) \mid L \in \mathrm{MCS}(K, A)\}, \qquad \mathrm{Skept}(K, A) = C(K, A).$$

Example

For $K = \{p, \neg p, q\}$ and $A = \varnothing$, both routes return $\mathrm{Cn}(\{q\})$.

Commutative diagram



Exercise

Definition (Maximally consistent with A)

Let $K, A \subseteq \mathcal{L}$ and $S \subseteq K$. The set S is *maximally consistent*, or *maxiconsistent*, with A if and only if $S \cup A$ is classically consistent and $T \cup A$ is classically inconsistent for every T such that $S \subsetneq T \subseteq K$.
The family of all such sets is

$$MCS(K, A) = \{S \subseteq K \mid S \text{ is maxiconsistent with } A\}.$$

Let $K = \{p, \neg p, q\}, A = \{\neg q\}$. What is MCS$(K, A)$?

(A) $\{\{p\}, \{\neg p\}\}$
(B) $\{\{p, q\}, \{\neg p, q\}\}$
(C) $\{\{p, \neg q\}, \{\neg p, \neg q\}\}$
(D) $\{\{p, \neg p, q\}\}$

Correct answer: A. The formula q conflicts with $A = \{\neg q\}$, while p and $\neg p$ cannot occur together.

Commutative diagram:

Definition (Argument)

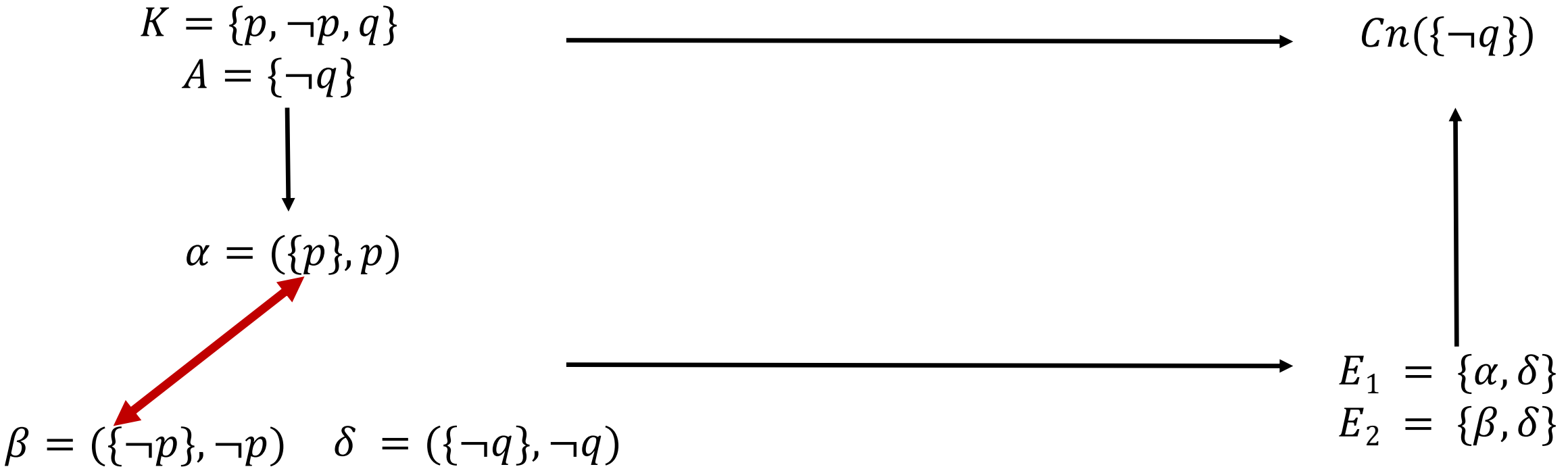
Let $K, A \subseteq \mathcal{L}$ be finite and let A be consistent. An argument based on (K, A) is a pair $\alpha = \langle \Phi, \varphi \rangle$ such that $\Phi \subseteq K$, $A \cup \Phi \nvdash \bot$, $A \cup \Phi \vdash \varphi$, and no proper subset $\Phi' \subsetneq \Phi$ satisfies $A \cup \Phi' \vdash \varphi$. Its support is Φ, its conclusion is φ, and $\mathrm{Arg}(K, A)$ is the set of all such arguments.

Definition (Attack)

An argument $\alpha = \langle \Phi, \varphi \rangle$ attacks an argument $\beta = \langle \Psi, \psi \rangle$ exactly when $\theta \equiv \neg\varphi$ for some $\theta \in \Psi$.

Definition (Induced framework)

Let $\mathrm{Att}(K, A) = \{(\alpha, \beta) \in \mathrm{Arg}(K, A)^2 \mid \alpha \text{ attacks } \beta\}$. The induced abstract argumentation framework is $AF(K, A) = \langle \mathrm{Arg}(K, A), \mathrm{Att}(K, A) \rangle$.



Commutative diagram:

Definition (Argument)

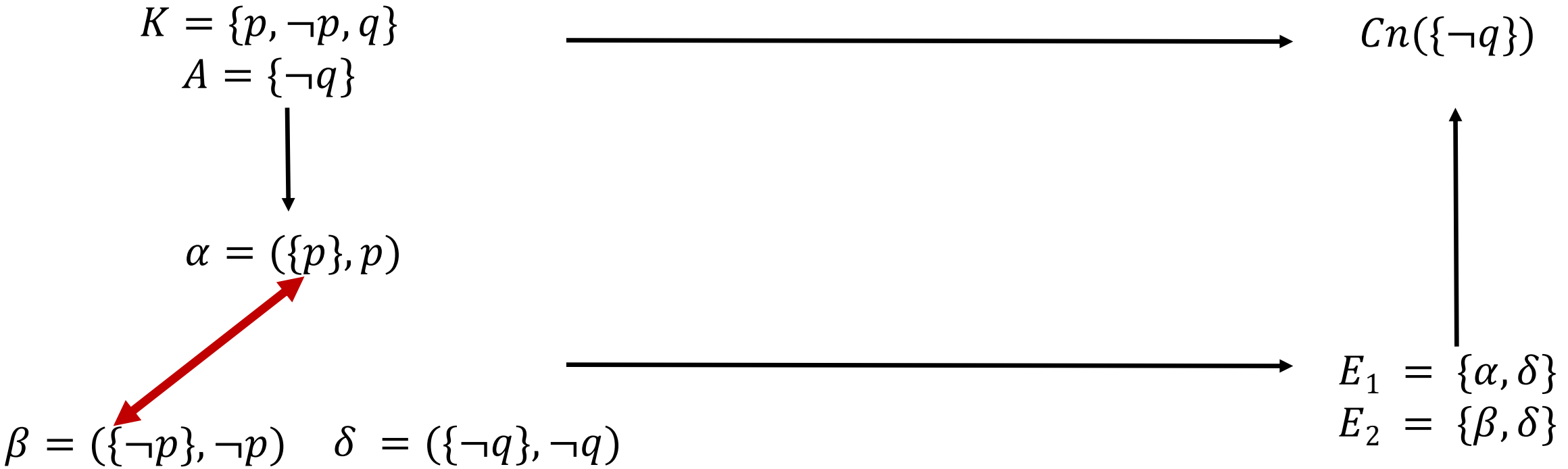
Let $K, A \subseteq \mathcal{L}$ be finite and let A be consistent. An argument based on (K, A) is a pair $\alpha = \langle \Phi, \varphi \rangle$ such that $\Phi \subseteq K$, $A \cup \Phi \nvdash \bot$, $A \cup \Phi \vdash \varphi$, and no proper subset $\Phi' \subsetneq \Phi$ satisfies $A \cup \Phi' \vdash \varphi$. Its support is Φ, its conclusion is φ, and $\mathrm{Arg}(K, A)$ is the set of all such arguments.

Definition (Attack)

An argument $\alpha = \langle \Phi, \varphi \rangle$ attacks an argument $\beta = \langle \Psi, \psi \rangle$ exactly when $\theta \equiv \neg\varphi$ for some $\theta \in \Psi$.

Definition (Induced framework)

Let $\mathrm{Att}(K, A) = \{(\alpha, \beta) \in \mathrm{Arg}(K, A)^2 \mid \alpha \text{ attacks } \beta\}$. The induced abstract argumentation framework is $AF(K, A) = \langle \mathrm{Arg}(K, A), \mathrm{Att}(K, A) \rangle$.



The logic is nonmonotonic

Example

$$A = \varnothing, \qquad C(K, A) = \mathrm{Skept}(K, A) = \mathrm{Cn}(\{q\}),$$
$$B = \{\neg q\}, \qquad C(K, B) = \mathrm{Skept}(K, B) = \mathrm{Cn}(\{\neg q\}).$$

Nonmonotonicity

$A \subset B$, but $q \in C(K, A)$ and $q \notin C(K, B)$.

Day 1 Motivation and reasoning alignment

1.1 AI Logic (JLC corner) is nonmonotonic logic (since 1980)

1.2 Explanation of black boxes in new generation AI: secretary metaphor

1.3 Dung 1995 paradigm shift relates logic and argumentation based explanation

1.4 Reasoning alignment of logic and argumentation based explanation

1.5 Using logic: from inference towards dialogue and decision making, to logic in action

Conceptualizations of natural argumentation in handbooks

Argumentation as inference & argumentation as dialogue distinguished in

2

Historical Overview of Formal Argumentation

Henry Prakken

ABSTRACT. This chapter gives an overview of the history of formal argumentation in terms of a distinction between argumentation-based *inference* and argumentation-based *dialogue*. Systems for argumentation-based *inference* are about which conclusions can be drawn from a given body of possibly incomplete, inconsistent of uncertain information. They ultimately define a nonmonotonic notion of logical consequence, in terms of the intermediate notions of argument construction, argument attack and argument evaluation, where arguments are seen as constellations of premises, conclusions and inferences. Systems for argumentation-based *dialogue* model argumentation as a kind of verbal interaction aimed at resolving conflicts of opinion. They define argumentation protocols, that is, the rules of the argumentation game, and address matters of strategy, that is, how to play the game well. For both aspects of argumentation the main formal and computational models are reviewed and their main historical influences are sketched. Then some main applications areas are briefly discussed.

Argumentation as balancing identified in

3

Towards Requirements Analysis for Formal Argumentation

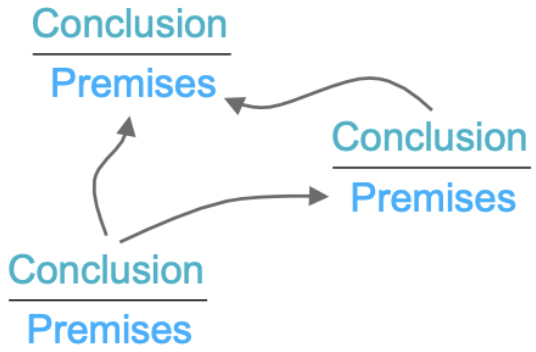
Thomas F. Gordon

ABSTRACT. We suggest applying software engineering requirements analysis methods to the development and evaluation of formal models of argumentation. Our aim and purpose is to help assure that formal argumentation *models* the full scope of argumentation as it is understood and studied in the humanities and social sciences, so as to provide a foundation for software tools supporting real argumentation tasks, in a wide variety of application domains.

Conceptualizations of natural argumentation in handbooks

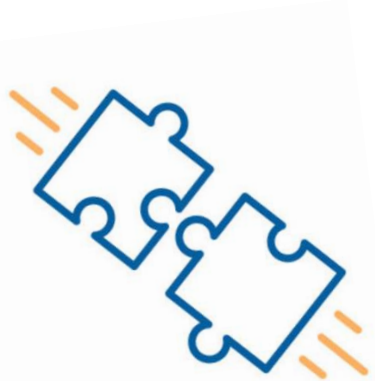
Conceptua-lization	**Process**	**Theories and Formal Approaches**	**Application**
Argumentation as inference	Logical structure and reasoning to derive conclusions from incomplete and inconsistent premises	Graph theory, nonmonotonic logic, computational logic, causal reasoning, Bayesian reasoning	Automated reasoning systems, knowledge representation, expert systems
Argumentation as dialogue	Dynamic verbal interaction between stakeholders to exchange information or resolve conflicts of opinion	Speech act theory, game theory, axiomatic semantics, operational semantics	Debating technologies, chatbots, recommender systems
Argumentation as balancing	Balancing pros and cons to reach a justified decision	Multi-criteria decision theory, machine ethics, computational law, case-based reasoning	Deliberative decision-making in law, ethics, and economics

Build bridges from one conceptual model to another?

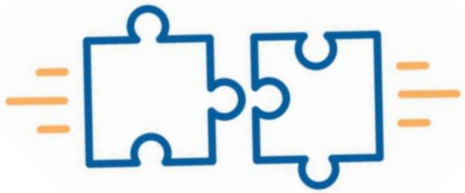


inference

Argumentation as



dialogue

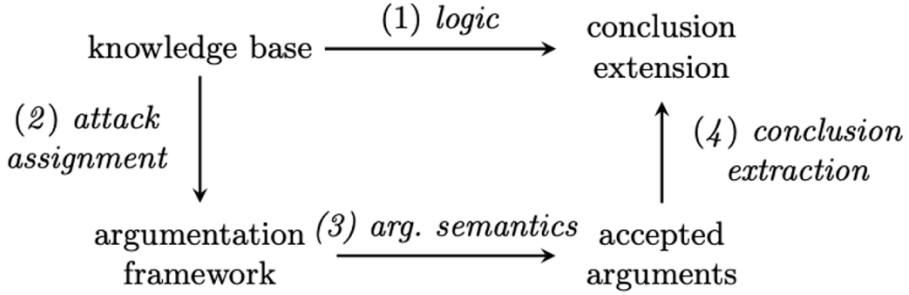


balancing

Motivation of this course

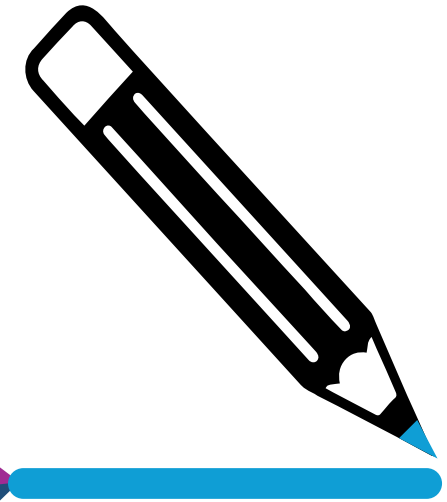
1995

The attack-defense paradigm shift



2026

Agentic AI



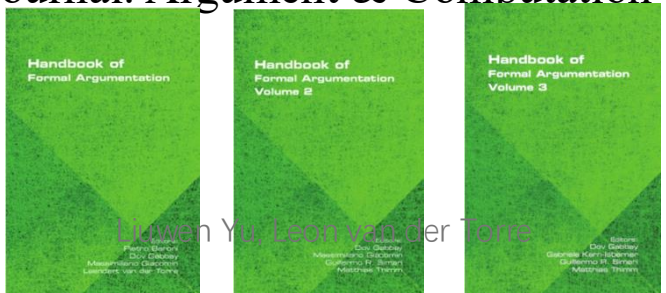
Logic

Argumentation in AI

How to bridge the gap?

COMMA since 2006

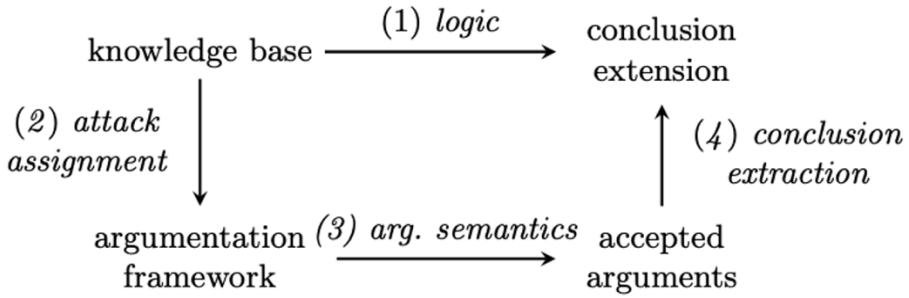
Journal: Argument & Computation

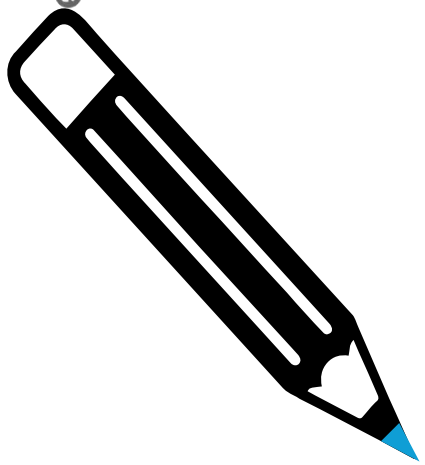


Motivation of this course

1995

The attack-defense paradigm shift



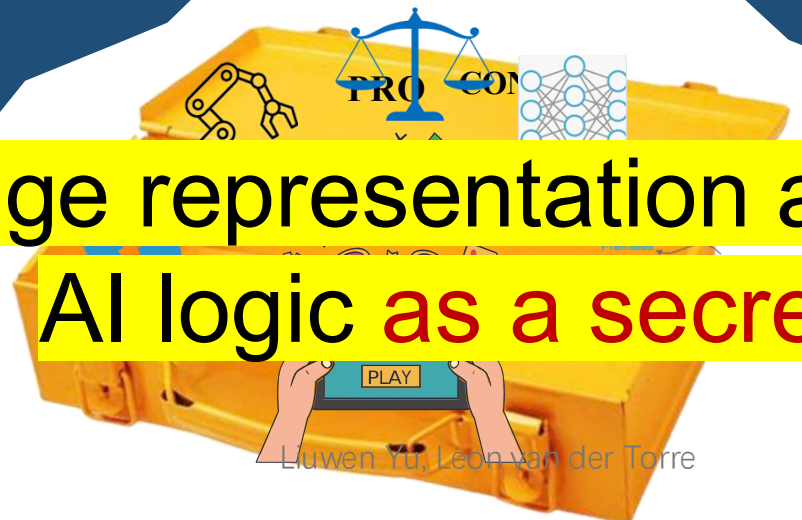


2026

Agentic AI

Logic

Argumentation in AI



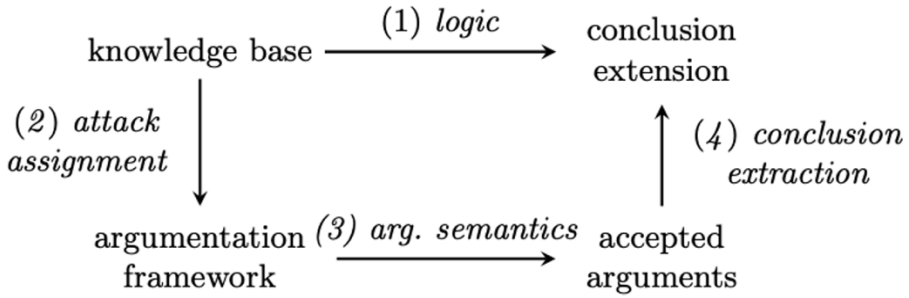
Knowledge representation and reasoning

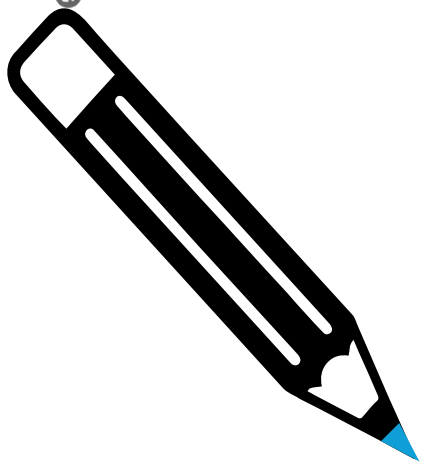
AI logic as a secretary

Motivation of this course

1995

The attack-defense paradigm shift





2026

Agentic AI

Logic

Argumentation in AI

More on day 4 and day 5

Day 1 Dung paradigm shift and reasoning alignment

1.1 AI Logic (JLC corner) is nonmonotonic logic (since 1980)

1.2 Explanation of black boxes in new generation AI: secretary metaphor

1.3 Dung 1995 paradigm shift relates logic and argumentation based explanation

1.4 Reasoning alignment of logic and argumentation based explanation

1.5 Using logic: from inference towards dialogue and decision making, to logic in action

Logic and Argumentation for New-generation AI

ESSLLI 2026

Liuwen Yu[1]
Leon van der Torre[2]
[1] Luxembourg Institute of Science and Technology
[2] University of Luxembourg

Day 2 Reasons first, attack first or defense first?

2.1 Why unify reasons in philosophy and formal argumentation?

2.2 Rules represent reasons in both logic and argumentation.

2.3 Logic dilemma: prescriptive prioritized default logic or descriptive default logic?

2.4 Argumentation dilemma: weakest link or last link?

2.5 Are the two dilemmas two sides of the same coin?

Recap
Day 1 Dung paradigm shift and reasoning alignment

1.1 AI Logic (JLC corner) is nonmonotonic logic (since 1980)

1.2 Explanation of black boxes in new generation AI: secretary metaphor

1.3 Dung 1995 paradigm shift relates logic and argumentation based explanation

1.4 Reasoning alignment of logic and argumentation based explanation

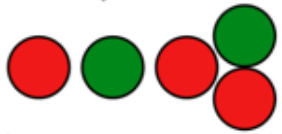
1.5 Using logic: from inference towards dialogue and decision making, to logic in action

Recap
Day 1 Dung paradigm shift and reasoning alignment

Abstract argumentation semantics: gunfight rules

- Three complete extensions (3 valued)

- One stable extension (2 valued)



- One grounded extension (least complete)

- Two preferred extensions (maximal complete)

71

Liuwen Yu, Leon van der Torre attack in graphs

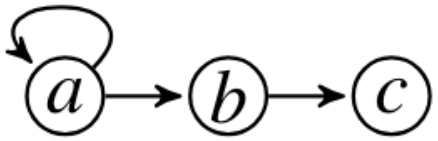
Three main branches of abstract argumentation semantics

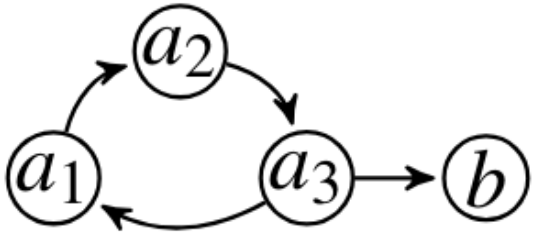
1. Admissibility based semantics
   - Gunfight rules, require defense against all its attackers
2. Non-admissibility based semantics
   - Naive-based semantics: extension is a maximal conflict-free set of arguments.
3. Weak admissibility based semantics
   - Only require defense against reasonable arguments

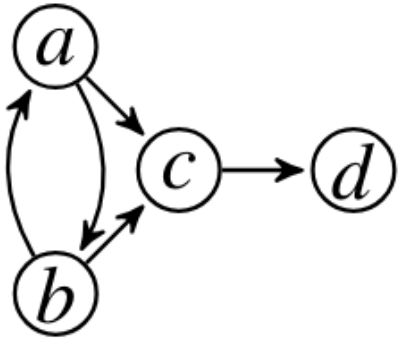
Dvořák W, Rienstra T, van der Torre L, et al. Non-admissibility in abstract argumentation[M]//Computational Models of Argument. IOS Press, 2022: 128-139.

Three main branches of abstract argumentation semantics

1. Admissibility based semantics (AB)
   - Gunfight rules, require defense against all its attackers
2. Non-admissibility based semantics (NA)
   - Naive-based semantics: extension is a maximal conflict-free set of arguments.
3. Weak admissibility based semantics (WA)
   - Only require defense against reasonable arguments







AB	{Ø}	{Ø}	{{a,d},{b,d}}
NA(CF2)	{{b}}	{{a_1,b},{a_2,b},{a_3}}	{{a,d},{b,d}}
WB	{{b}}	{{b}}	{{a,d},{b,d}}

Dvořák W, Rienstra T, van der Torre L, et al. Non-admissibility in abstract argumentation[M]//Computational Models of Argument. IOS Press, 2022: 128-139.

Three main branches of abstract argumentation semantics

1. Admissibility based semantics
   - Gunfight rules, require defense against all its attackers
2. Non-admissibility based semantics
   - Naive-based semantics: extension is a maximal conflict-free set of arguments.
3. Weak admissibility based semantics
   - Only require defense against reasonable arguments

Dung extensions are fixpoints of defense function
Generalized by Abstract Dialectical Frameworks (ADFs)

Dvořák W, Rienstra T, van der Torre L, et al. Non-admissibility in abstract argumentation[M]//Computational Models of Argument. IOS Press, 2022: 128-139.

Day 2 Reasons first, attack first or defense first?

2.1 Why unify reasons in philosophy and formal argumentation?

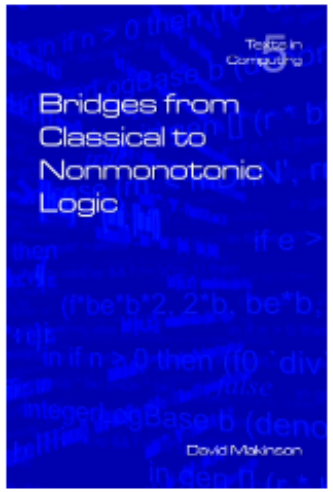
2.2 Rules represent reasons in both logic and argumentation.

2.3 Logic dilemma: prescriptive prioritized default logic or descriptive default logic?

2.4 Argumentation dilemma: weakest link or last link?

2.5 Are the two dilemmas two sides of the same coin?

Bridges from Classical to Nonmonotonic Logic



Fixed pivot	Pivotal use	Default use
Assumptions $K \subseteq \mathcal{L}$	add all of K	select compatible parts
Valuations $W \subseteq \mathcal{V}$	restrict to W	select best models
Rules $R \subseteq \mathcal{L}^2$	add all rules in R	vary rules with input

Pivotal: supraclassical and monotonic.
Default: input-dependent and possibly nonmonotonic.

Day 1: assumptions and valuations. **Day 2:** rules.

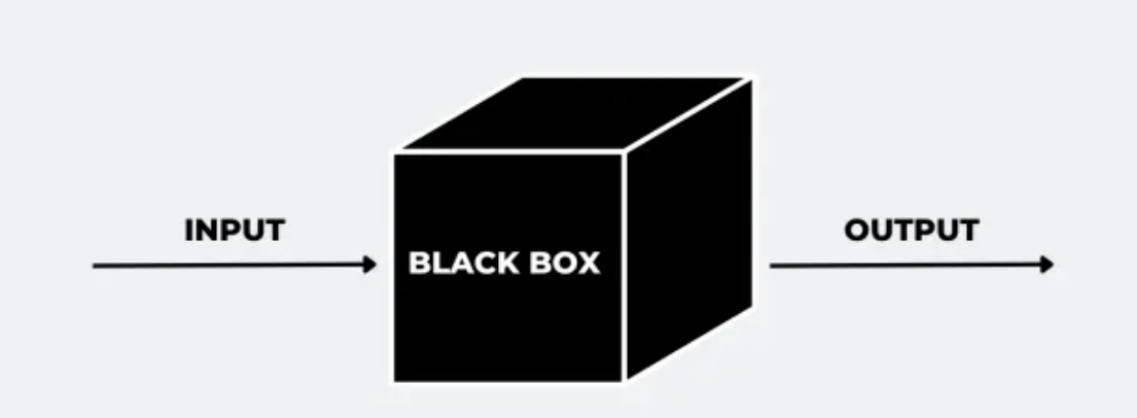
Why do we need rules?

Pivotal consequence and rule consequence

- Many KR formalisms are based on rules
- Logic programming, Reiter's default logic, etc
- Expert systems in 80s typically use rules
- Knowledge in law for example is often represented using rules

Pivotal consequence and rule consequence

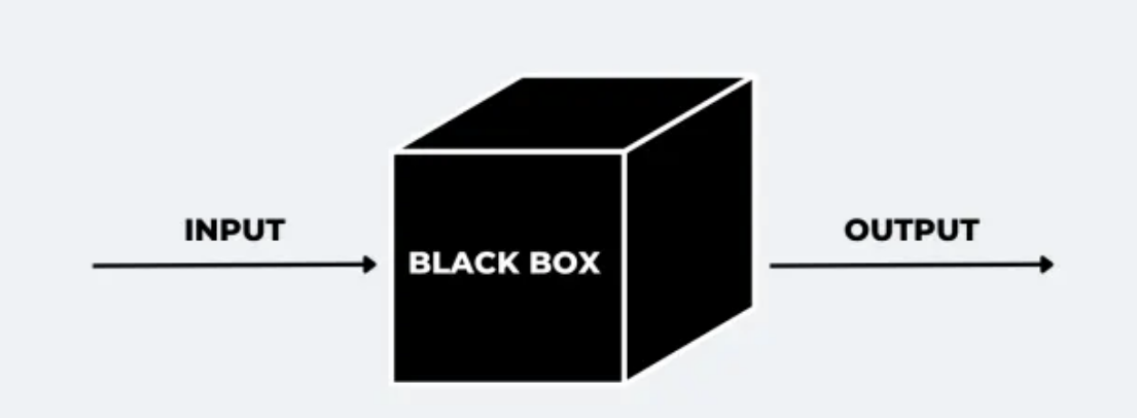
- Many KR formalisms are based on rules
- Logic programming, Reiter's default logic, etc
- Expert systems in 80s typically use rules
- Knowledge in law for example is often represented using rules



Knowledge is represented implicitly, e.g. by LLM
Reverse engineered from the input/output behavior

Two kinds of black box explanation

- Input A, pivot K, output $X = C(K, A)$
- Relation $A \mid\sim_{\mathrm{K}} X$, or ternary relation (A, K, X)
- Given relation (K, X), compute possible A: abduction
- Given relation (A, X), compute possible K: reverse engineering



Bridge 3: From Classical Consequence to Pivotal Rules

Definition (Definition: Rules and their image)

A rule is an ordered pair (a, x) of propositions in $\mathcal{L}$. Hence a set of rules is a binary relation $R \subseteq \mathcal{L}^2$. For $X \subseteq \mathcal{L}$, the image of X under R is

$$R(X) = \{y \mid \text{there is an } x \in X \text{ with } (x, y) \in R\}.$$

A set X is closed under R iff $R(X) \subseteq X$.

Definition (Pivotal-rule consequence)

We write $A \vdash_R x$, equivalently $x \in \mathsf{Cn}_R(A)$, iff x belongs to every $X \supseteq A$ that is closed under both Cn and R. Equivalently, $\mathsf{Cn}_R(A)$ is the least $X \supseteq A$ such that

$$\mathsf{Cn}(X) \subseteq X \qquad \text{and} \qquad R(X) \subseteq X.$$

Bridge 3: From Classical Consequence to Pivotal Rules

Definition (Pivotal-rule consequence)

We write $A \vdash_R x$, equivalently $x \in \mathrm{Cn}_R(A)$, iff x belongs to every $X \supseteq A$ that is closed under both Cn and R. Equivalently, $\mathrm{Cn}_R(A)$ is the least $X \supseteq A$ such that

$$\mathrm{Cn}(X) \subseteq X \qquad \text{and} \qquad R(X) \subseteq X.$$

The one-rule case

$$A = \{a\}, \quad R = \{(a, x)\} \qquad \Longrightarrow \qquad \mathrm{Cn}_R(A) = \mathrm{Cn}(\{a, x\}).$$

Bridge 3: From Classical Consequence to Pivotal Rules

Definition (Ordered default-rule extension)

Fix an ordering $\langle R \rangle$ of the rule set R. Put $A_1 = A$, $C_{\langle R \rangle}(A) = \bigcup \{A_n \mid n < \omega\}$. For $n \geq 1$, let (a, x) be the first rule in $\langle R \rangle$ such that $a \in A_n$, $x \notin A_n$, x is consistent with A_n. If there is such a rule, put $A_{n+1} = \mathrm{Cn}(A_n \cup \{x\})$; if there is no such rule, put $A_{n+1} = \mathrm{Cn}(A_n)$. Here, consistency means $\mathrm{Cn}(A_n \cup \{x\}) \neq \mathcal{L}$.

Makinson's ordering example

Let

$$A = \{a\}, \qquad R = \{(a, x), (a, \neg x)\}.$$

If (a, x) comes first, then

$$C_{\langle (a,x),(a,\neg x) \rangle}(A) = \mathrm{Cn}(\{a, x\}).$$

With the reverse ordering,

Bridge 3: From Classical Consequence to Pivotal Rules

Definition (Ordered default-rule extension)

Fix an ordering $\langle R \rangle$ of the rule set R. Put $A_1 = A$, $C_{\langle R \rangle}(A) = \bigcup\{A_n \mid n < \omega\}$. For $n \geq 1$, let (a, x) be the first rule in $\langle R \rangle$ such that $a \in A_n$, $x \notin A_n$, x is consistent with A_n. If there is such a rule, put $A_{n+1} = \mathrm{Cn}(A_n \cup \{x\})$; if there is no such rule, put $A_{n+1} = \mathrm{Cn}(A_n)$. Here, consistency means $\mathrm{Cn}(A_n \cup \{x\}) \neq \mathcal{L}$.

Makinson's continued example

Let

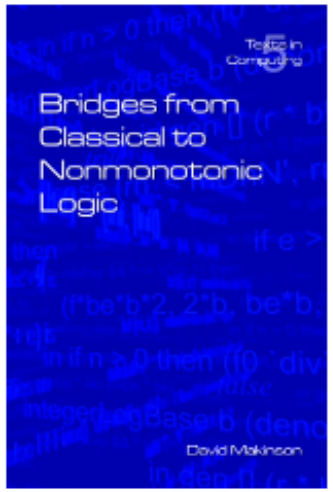
$$A = \{a\}, \qquad B = \{a, \neg x\}, \qquad R = \{(a, x)\}.$$

With the unique ordering of R,

$$C_{\langle R \rangle}(A) = \mathrm{Cn}(\{a, x\}), \qquad C_{\langle R \rangle}(B) = \mathrm{Cn}(\{a, \neg x\}).$$

Although $A \subseteq B$, x is a consequence of A but not of B.

Bridges from Classical to Nonmonotonic Logic



Summary Table

Fixed additional assumptions	**Fixed restriction of valuations**	**Fixed additional rules**
• pivotal-assumption: Cn_K • paraclassical ✓ • OR ✓ • compact ✓	• pivotal-valuation: Cn_W • paraclassical ✓ • OR ✓ • compact ×	• pivotal-rule: Cn_R • paraclassical ✓ • OR × • compact ✓
Vary assumptions with premises	**Vary valuation-set with premises**	**Vary rules with premises**
• default-assumption: C_K • consistency constraint • Poole systems • + many variants!	• default-valuation: C_W • minimalization • preferential systems • + many variants!	• default-rule: C_R • consistency constraint • Reiter systems • + many variants!

Day 2 Reasons first, attack first or defense first?

2.1 Why unify reasons in philosophy and formal argumentation?

2.2 Rules represent reasons in both logic and argumentation.

2.3 Logic dilemma: prescriptive prioritized default logic or descriptive default logic?

2.4 Argumentation dilemma: weakest link or last link?

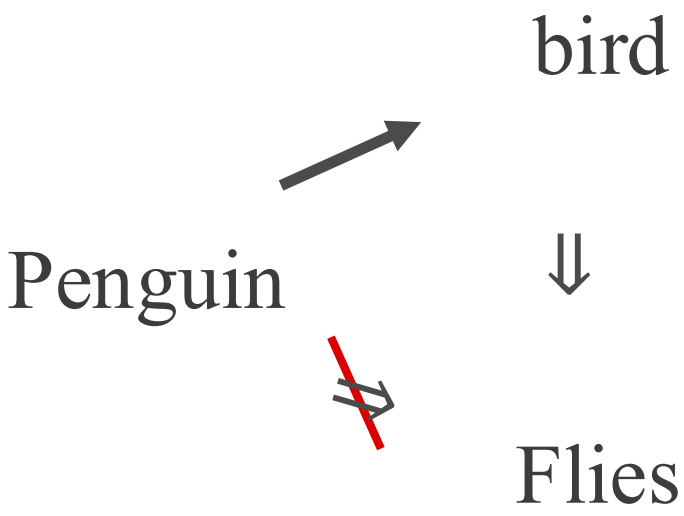
2.5 Are the two dilemmas two sides of the same coin?

Specificity example in 80s

Notations:

-> for strict rules

=> for defaults, => with dash for neg conclusion



specificity = implicit priority

Reiter R, Criscuolo G. On interacting defaults[C]//Proceedings of the Seventh International Joint Conference on Artificial Intelligence. 1981: 270–276.
Rintanen J. On specificity in default logic[C]//Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence. 1995: 1474–1479.

Default theory

Definition

A *default theory* is a pair $\Delta = (D, W)$, where W is a set of formulas and D is a set of defaults

$$d = \frac{\alpha : \beta_1, \ldots, \beta_m}{\gamma}.$$

It reads: if α is believed and every justification β_i is consistent with what is believed, then infer γ. A default is *normal* when it has the form $\frac{\alpha:\beta}{\beta}$.

Why this is nonmonotonic

The applicability test includes absence: $\neg\beta_i \notin E$. New information can therefore block a previously applicable default.

Reiter R. A logic for default reasoning[J]. Artificial Intelligence, 1980, 13(1–2): 81–132.

Fixed-point extension

Definition

For a candidate set S, let $\Gamma_\Delta(S)$ be the least deductively closed set X such that $W \subseteq X$ and, for every $d = \frac{\alpha:\beta_1,\ldots,\beta_m}{\gamma} \in D$,

$$\alpha \in X \quad \text{and} \quad \neg\beta_i \notin S \ \text{ for every } i \quad \Longrightarrow \quad \gamma \in X.$$

A set E is an *extension* of Δ iff $E = \Gamma_\Delta(E)$.

The choice problem

Priorities are meant to select among incompatible default conclusions but how?

Reiter R. A logic for default reasoning[J]. Artificial Intelligence, 1980, 13(1–2): 81–132.

Prioritized default theory

Definition

A *prioritized default theory* is

$$\Delta^{\prec} = (D, W, \prec),$$

where (D, W) is a default theory and $\prec$ orders the defaults. Throughout this course,

$d1 \prec d2$ means that $d2$ has the stronger priority.

With numerical ranks, $d1 \prec d2$ iff $\text{rank}(d1) < \text{rank}(d2)$.

Brewka G. Reasoning about priorities in default logic[C]//Proceedings of the Twelfth National Conference on Artificial Intelligence. AAAI Press, 1994: 940–945.

One priority order, two readings

Prescriptive / procedural

A stronger default must be *considered earlier*. Priority regulates the construction of an extension.

order defaults $\longrightarrow$ build outcome

Descriptive / declarative

A stronger default is *more desirable to satisfy*. First generate candidate extensions, then compare them.

build outcomes $\longrightarrow$ order outcomes

Delgrande, James, et al. "A classification and survey of preference handling approaches in nonmonotonic reasoning." *Computational Intelligence* 20.2 (2004): 308-334.

When is a default active?

Definition

Assume normal defaults and a total priority order. At stage i, a default $d = \frac{\alpha:\beta}{\beta}$ is *active* in E_i iff

$$\alpha \in E_i, \qquad \beta \notin E_i, \qquad \neg\beta \notin E_i.$$

- Its prerequisite is established;
- Its conclusion is not already present;
- Its contrary is not present

Prescriptive PDL construction

Definition

Put $E_0 = \mathrm{Th}(W)$. Select the strongest active default d_i whose conclusion preserves consistency, and set

$$E_{i+1} = \mathrm{Th}\big(E_i \cup \{\mathrm{cons}(d_i)\}\big).$$

If no such default exists, stop. The PDL extension generated by the order is $E = \bigcup_{i<\omega} E_i$.

- Priority constrains the order of application;
- It does not directly compare completed extensions

Generating defaults of an extension

Definition

Let E be a Reiter extension of $\Delta = (D, W)$. Its set of *generating defaults* is

$$\mathrm{Gen}(E) = \left\{ d = \frac{\alpha : \beta_1, \ldots, \beta_m}{\gamma} \in D \;\middle|\; \alpha \in E \text{ and } \neg\beta_i \notin E \text{ for every } i \right\}.$$

Normal defaults

For $d = \frac{\alpha:\beta}{\beta}$, generation means $\alpha \in E$ and $\neg\beta \notin E$.

Reiter R. A logic for default reasoning[J]. Artificial Intelligence, 1980, 13(1–2): 81–132, Definition 2 and Theorem 2.5.

Descriptive lexicographic preference

Definition

Assume a total priority order. For two Reiter extensions $E1, E2$, consider

$$\mathrm{Gen}(E1) \triangle \mathrm{Gen}(E2).$$

If this set is nonempty, let d be its strongest default. Define

$$E1 \succ_{\mathrm{lex}} E2 \quad \Longleftrightarrow \quad d \in \mathrm{Gen}(E1).$$

A *lexicographically preferred extension* is a Reiter extension for which no strictly better Reiter extension exists.

Brewka G, Eiter T. Prioritizing default logic[C]//Hölldobler S, ed. Intellectics and Computational Logic: Papers in Honour of Wolfgang Bibel. Dordrecht: Kluwer Academic Publishers, 2000: 27–45
Rintanen J. Lexicographic priorities in default logic[J]. Artificial Intelligence, 1998, 106(2): 221–265. Canonical descriptive approach: extensions are compared globally according to the highest-priority defaults they satisfy..

Descriptive lexicographic preference

Prescriptive / procedural

A stronger default must be *considered earlier*. Priority regulates the construction of an extension.

order defaults $\longrightarrow$ build outcome

Descriptive / declarative

A stronger default is *more desirable to satisfy*. First generate candidate extensions, then compare them.

build outcomes $\longrightarrow$ order outcomes

Triangle example

$$d_1 : \frac{\top : a}{a}, \qquad d_2 : \frac{\top : \neg b}{\neg b}, \qquad d_3 : \frac{a : b}{b}, \qquad d_1 \prec d_2 \prec d_3.$$

Initially d_1 and d_2 are active, while d_3 is not. Hence PDL selects d_2 before d_1:

$$\mathrm{Th}(\varnothing) \xrightarrow{d_2} \mathrm{Th}(\{\neg b\}) \xrightarrow{d_1} \mathrm{Th}(\{a, \neg b\}).$$

Now d_3 is blocked by $\neg b$. Therefore

$$E_{\mathrm{pre}} = E^{-} = \mathrm{Th}(\{a, \neg b\}).$$

Descriptive lexicographic preference

Prescriptive / procedural

A stronger default must be *considered earlier*. Priority regulates the construction of an extension.

order defaults $\longrightarrow$ build outcome

Descriptive / declarative

A stronger default is *more desirable to satisfy*. First generate candidate extensions, then compare them.

build outcomes $\longrightarrow$ order outcomes

Triangle example

$$d_1 : \frac{\top : a}{a}, \qquad d_2 : \frac{\top : \neg b}{\neg b}, \qquad d_3 : \frac{a : b}{b}, \qquad d_1 \prec d_2 \prec d_3.$$

In decreasing priority order (d_3, d_2, d_1), record whether each default generates the extension:

$$v(E^+) = (1, 0, 1), \qquad v(E^-) = (0, 1, 1).$$

The strongest default on which the extensions differ is d_3, and $d_3 \in \mathrm{Gen}(E^+)$. Therefore

$$E_{\mathrm{des}} = E^+ = \mathrm{Th}(\{a, b\}).$$

The prioritized-default-logic dilemma

Triangle example

$$d_1 : \frac{\top : a}{a}, \qquad d_2 : \frac{\top : \neg b}{\neg b}, \qquad d_3 : \frac{a : b}{b}, \qquad d_1 \prec d_2 \prec d_3.$$

Reading	Role of priority	Canonical output
Prescriptive	regulates which active default is applied next	$\{a, \neg b\}$
Descriptive	ranks complete extensions by satisfied defaults	$\{a, b\}$

When to use which?

Rintanen J. Lexicographic priorities in default logic[J]. Artificial Intelligence, 1998, 106(2): 221–265. Canonical descriptive approach: extensions are compared globally according to the highest-priority defaults they satisfy.
Brewka G, Eiter T. Prioritizing default logic[C]//Hölldobler S, ed. Intellectics and Computational Logic: Papers in Honour of Wolfgang Bibel. Dordrecht: Kluwer Academic Publishers, 2000: 27–45.

The prioritized-default-logic dilemma

Triangle example

$$d_1 : \frac{\top : a}{a}, \qquad d_2 : \frac{\top : \neg b}{\neg b}, \qquad d_3 : \frac{a : b}{b}, \qquad d_1 \prec d_2 \prec d_3.$$

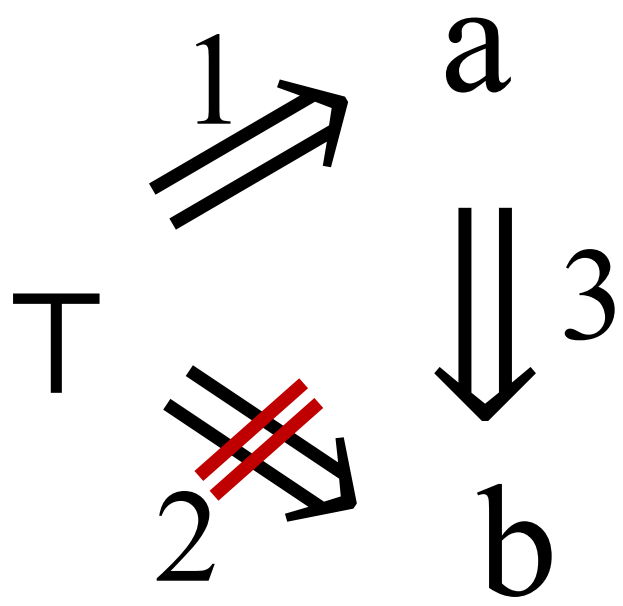
Reading	Role of priority	Canonical output
Prescriptive	regulates which active default is applied next	$\{a, \neg b\}$
Descriptive	ranks complete extensions by satisfied defaults	$\{a, b\}$

When to use which?

More examples…

Examples of prioritized rules

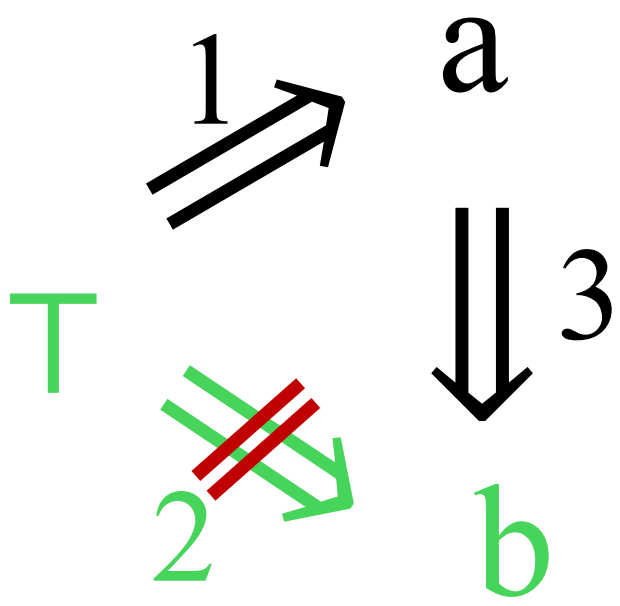
$$\{\top \overset{1}{\Rightarrow} a,\ a \overset{3}{\Rightarrow} b,\ \top \overset{2}{\Rightarrow} \neg b\}$$



Brewka G, Eiter T. Prioritizing default logic[C]//Hölldobler S, ed. *Intellectics and Computational Logic: Papers in Honour of Wolfgang Bibel*. Dordrecht: Kluwer Academic Publishers, 2000: 27–45.

Examples of prioritized rules

$$\{\top \overset{1}{\Rightarrow} a,\ a \overset{3}{\Rightarrow} b,\ \top \overset{2}{\Rightarrow} \neg b\}$$

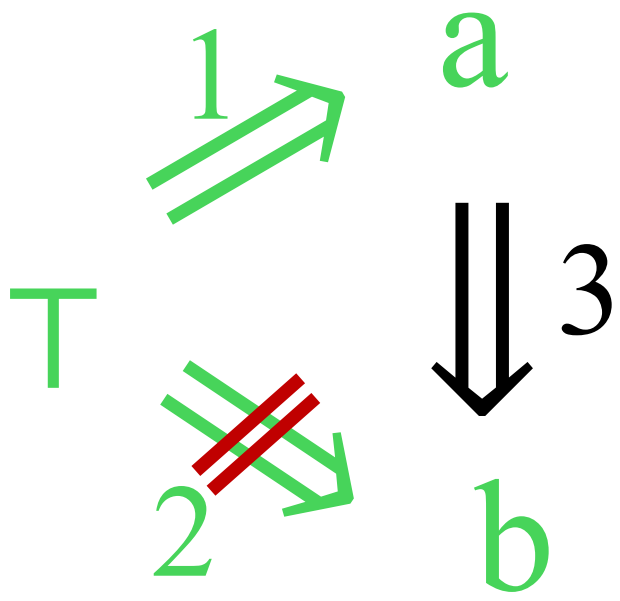


$\{\neg b\}$

Brewka G, Eiter T. Prioritizing default logic[C]//Hölldobler S, ed. *Intellectics and Computational Logic: Papers in Honour of Wolfgang Bibel*. Dordrecht: Kluwer Academic Publishers, 2000: 27–45.

Examples of prioritized rules

$$\{\top \overset{1}{\Rightarrow} a,\ a \overset{3}{\Rightarrow} b,\ \top \overset{2}{\Rightarrow} \neg b\}$$



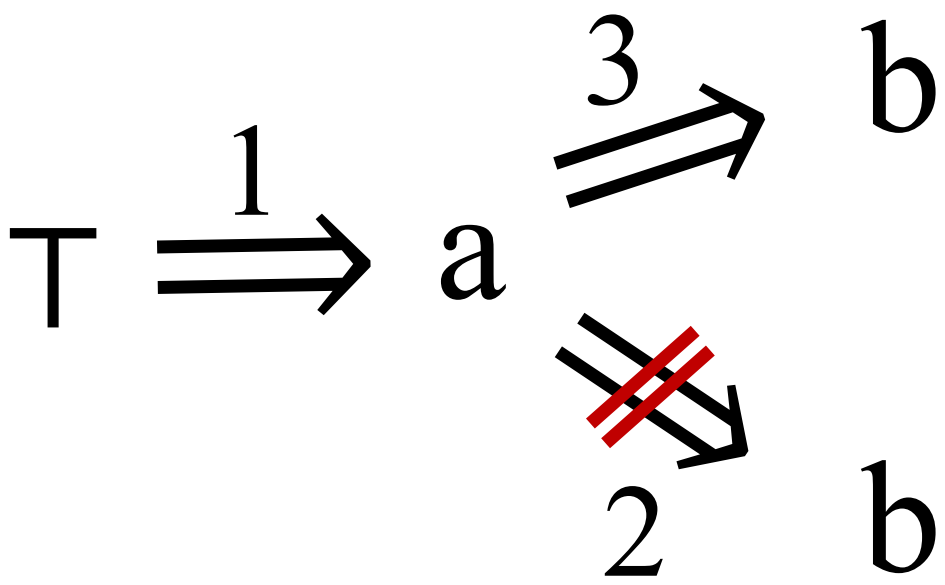
$\{\neg b\}$

$\{\neg b, a\}$

Brewka G, Eiter T. Prioritizing default logic[C]//Hölldobler S, ed. *Intellectics and Computational Logic: Papers in Honour of Wolfgang Bibel*. Dordrecht: Kluwer Academic Publishers, 2000: 27–45.

Examples of prioritized rules

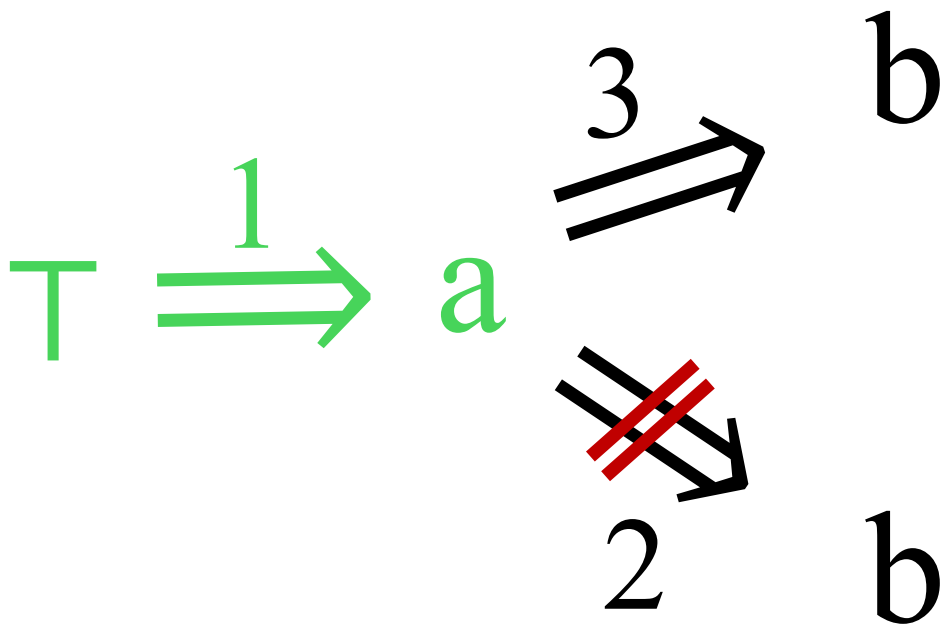
$$\{\top \overset{1}{\Rightarrow} a,\ a \overset{3}{\Rightarrow} b,\ a \overset{2}{\Rightarrow} \neg b\}$$



Young A P, Modgil S, Rodrigues O. Prioritised default logic as rational argumentation[C]//*Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems*. 2016: 626–634.

Examples of prioritized rules

$\{\top \overset{1}{\Rightarrow} a,\ a \overset{3}{\Rightarrow} b,\ a \overset{2}{\Rightarrow} \neg b\}$

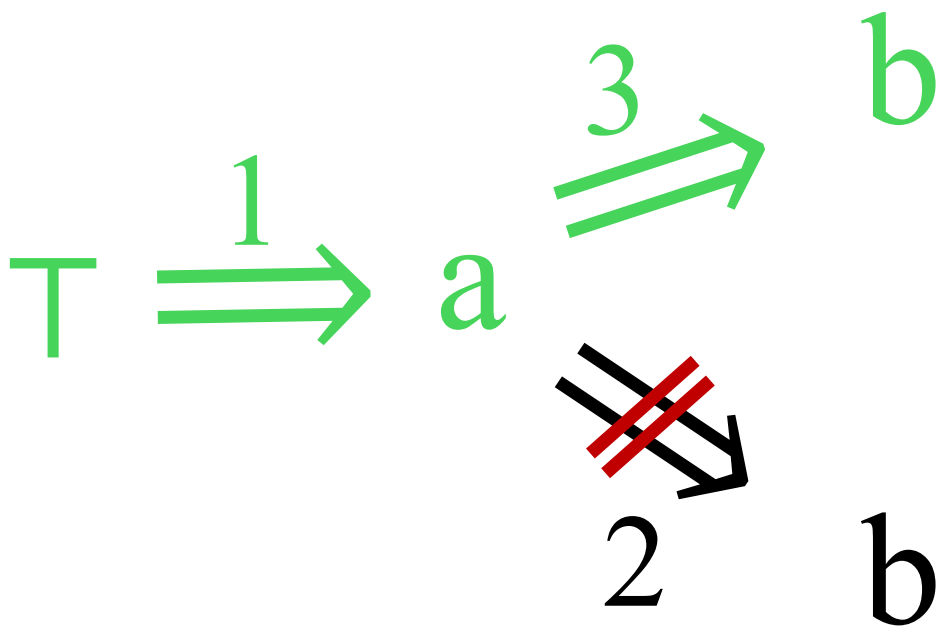


{a}

Young A P, Modgil S, Rodrigues O. Prioritised default logic as rational argumentation[C]//*Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems*. 2016: 626–634.

Examples of prioritized rules

$\{\top \overset{1}{\Rightarrow} a,\ a \overset{3}{\Rightarrow} b,\ a \overset{2}{\Rightarrow} \neg b\}$



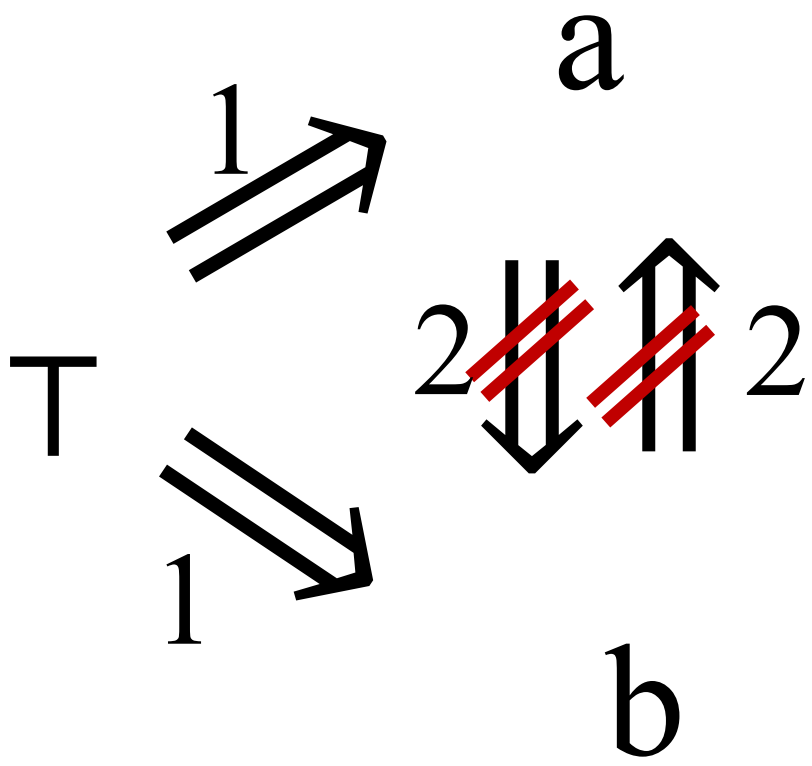
{a}

{a, b}

Young A P, Modgil S, Rodrigues O. Prioritised default logic as rational argumentation[C]//*Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems*. 2016: 626–634.

Examples of prioritized rules

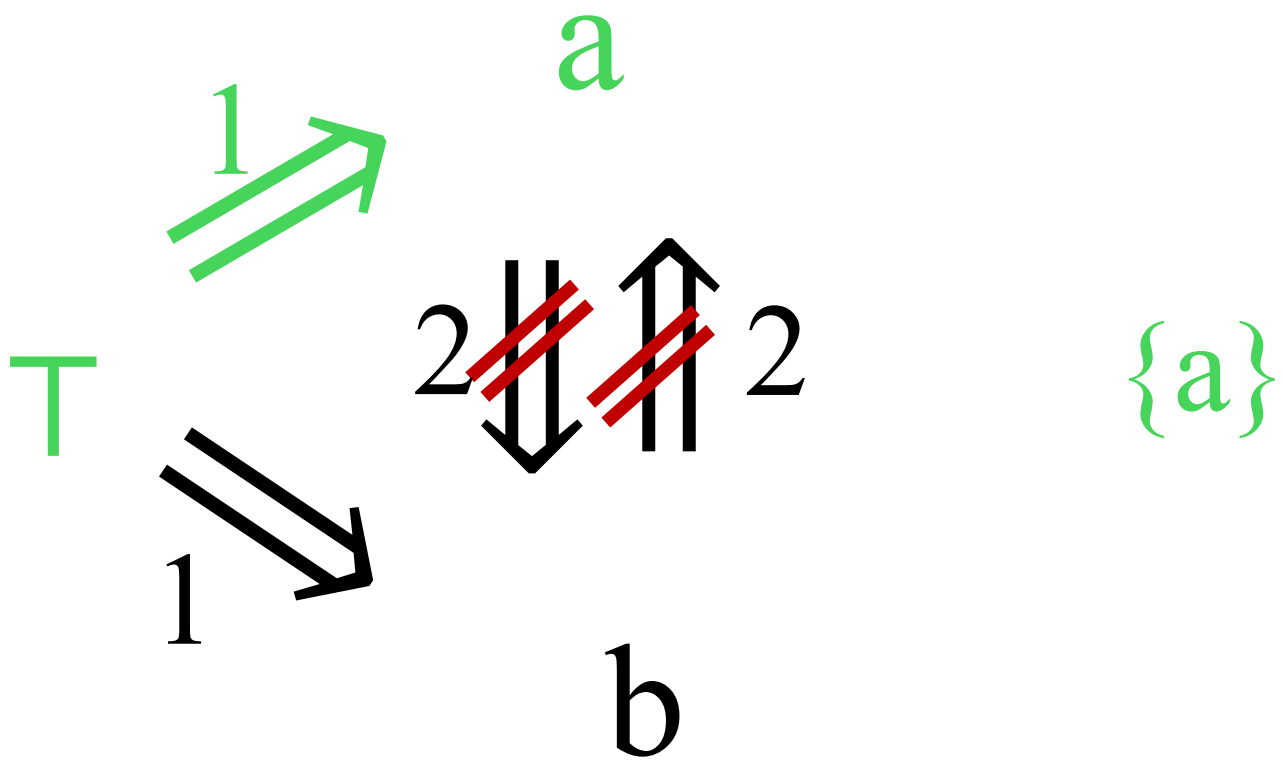
$$\{\top \overset{1}{\Rightarrow} a,\ \top \overset{1}{\Rightarrow} b,\ a \overset{2}{\Rightarrow} \neg b,\ b \overset{2}{\Rightarrow} \neg a\}$$



Liao B, Oren N, van der Torre L, Villata S. Prioritized norms in formal argumentation[J]. *Journal of Logic and Computation*, 2019, 29(2): 215–240.

Examples of prioritized rules

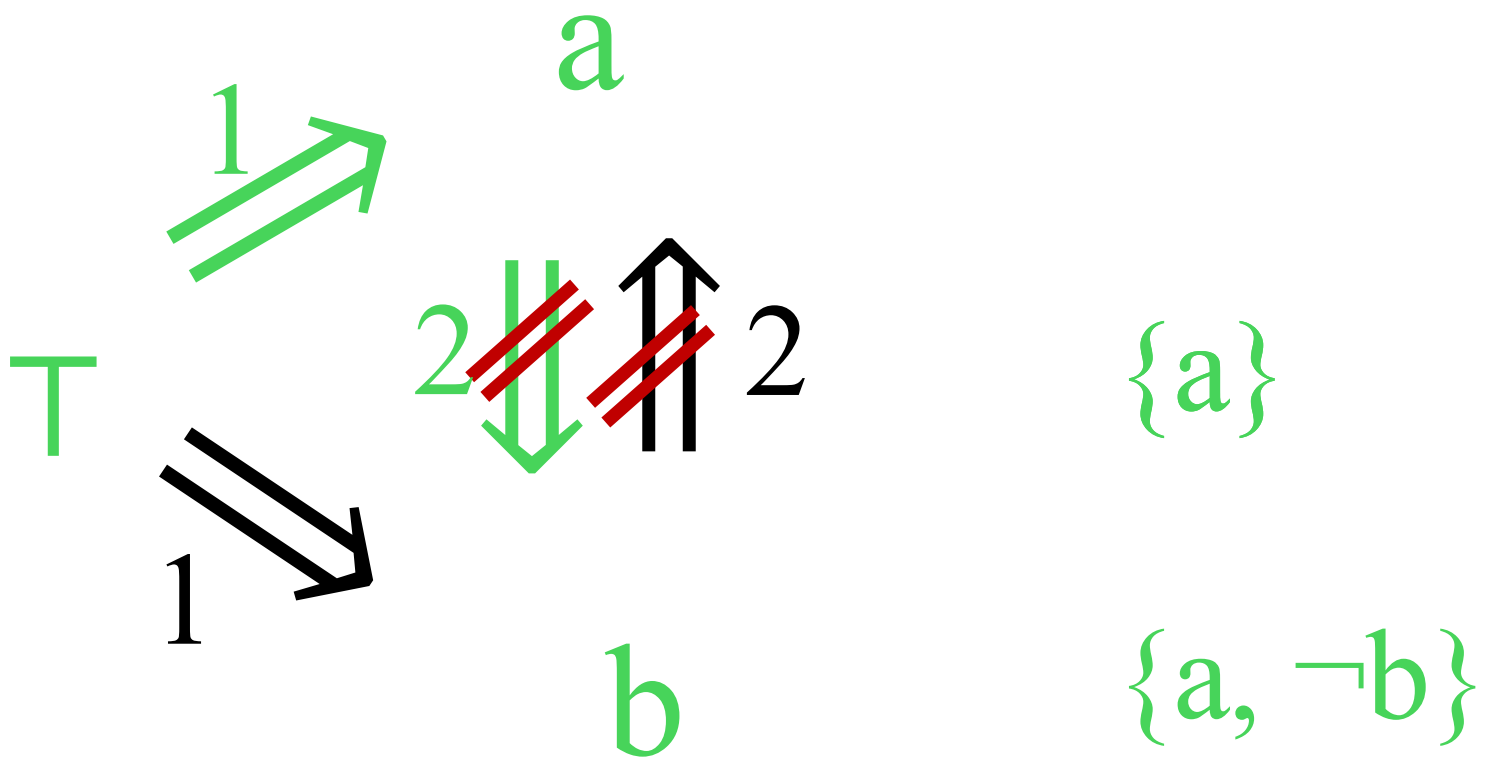
$$\{\top \overset{1}{\Rightarrow} a,\ \top \overset{1}{\Rightarrow} b,\ a \overset{2}{\Rightarrow} \neg b,\ b \overset{2}{\Rightarrow} \neg a\}$$



Liao B, Oren N, van der Torre L, Villata S. Prioritized norms in formal argumentation[J]. *Journal of Logic and Computation*, 2019, 29(2): 215–240.

Examples of prioritized rules

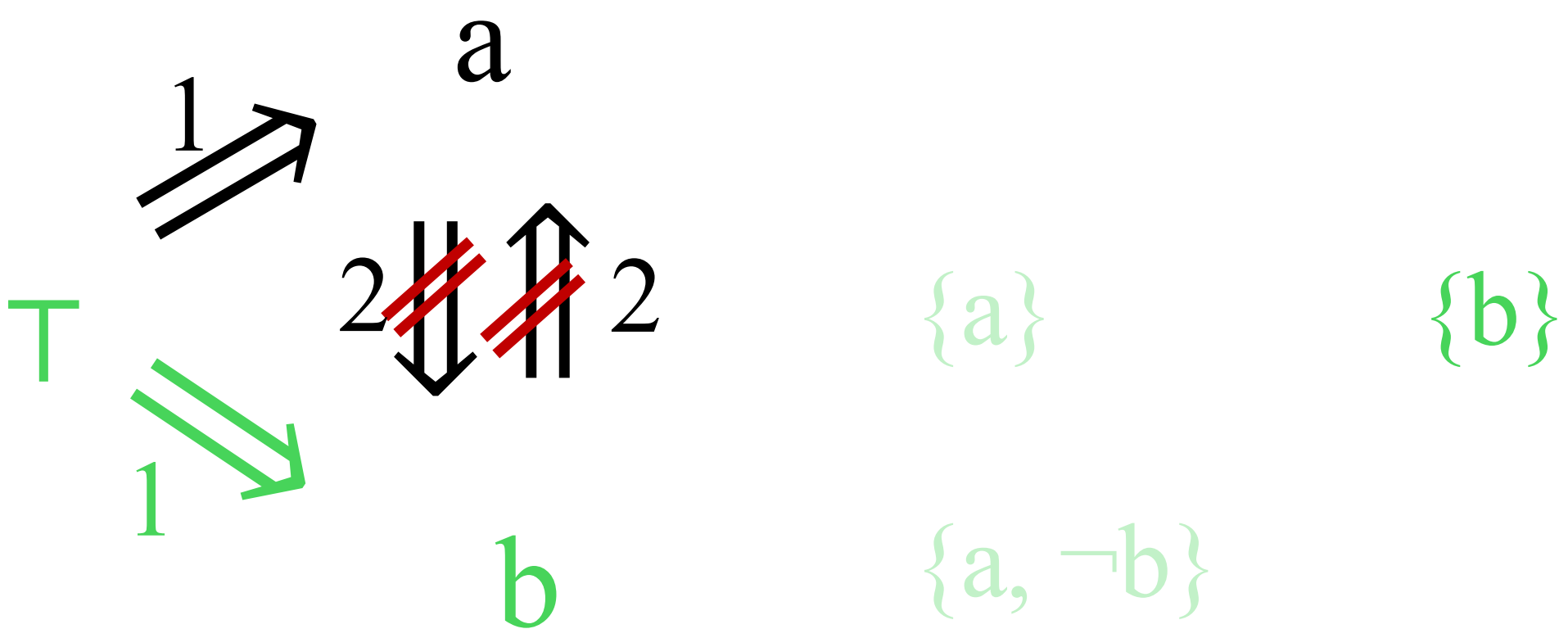
$$\{\top \overset{1}{\Rightarrow} a, \top \overset{1}{\Rightarrow} b,\ a \overset{2}{\Rightarrow} \neg b,\ b \overset{2}{\Rightarrow} \neg a\}$$



Liao B, Oren N, van der Torre L, Villata S. Prioritized norms in formal argumentation[J]. *Journal of Logic and Computation*, 2019, 29(2): 215–240.

Examples of prioritized rules

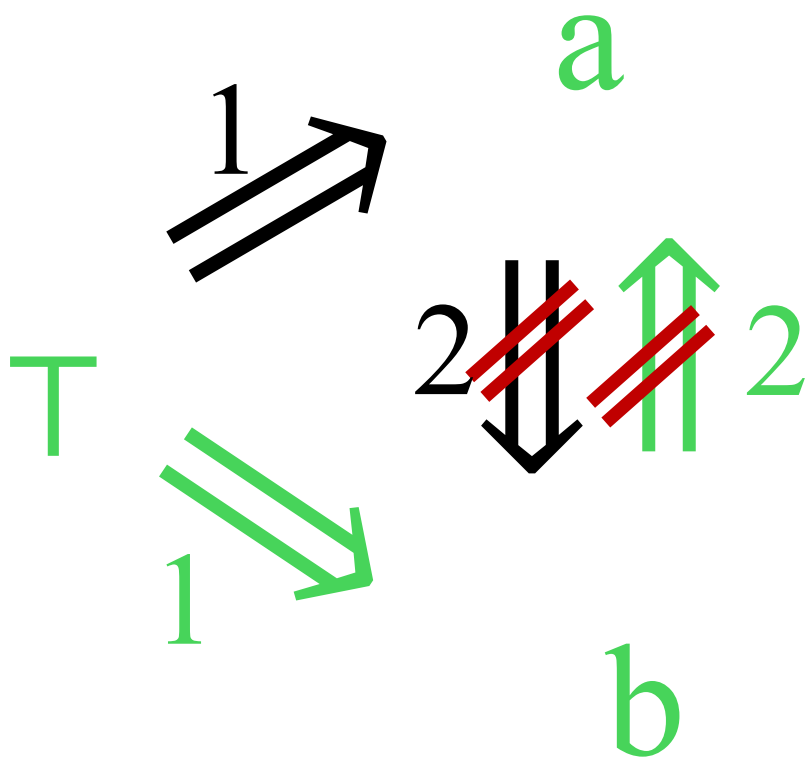
$$\{\top \overset{1}{\Rightarrow} a, \top \overset{1}{\Rightarrow} b,\ a \overset{2}{\Rightarrow} \neg b,\ b \overset{2}{\Rightarrow} \neg a\}$$



Liao B, Oren N, van der Torre L, Villata S. Prioritized norms in formal argumentation[J]. *Journal of Logic and Computation*, 2019, 29(2): 215–240.

Examples of prioritized rules

$$\{\top \overset{1}{\Rightarrow} a, \top \overset{1}{\Rightarrow} b,\ a \overset{2}{\Rightarrow} \neg b,\ b \overset{2}{\Rightarrow} \neg a\}$$



{a}

{b}

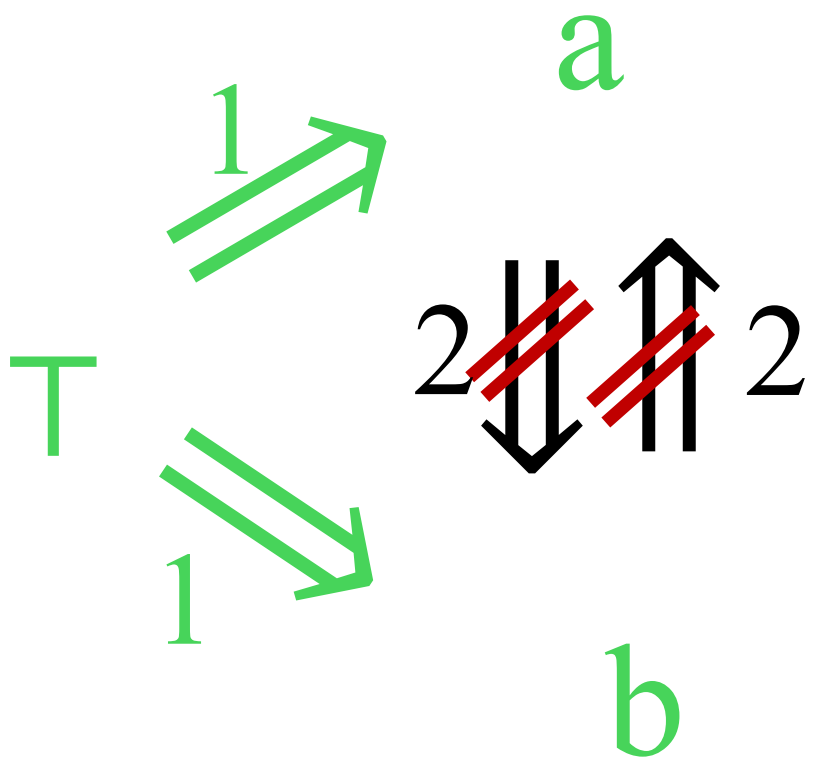
{a, ¬b}

{b, ¬a}

Liao B, Oren N, van der Torre L, Villata S. Prioritized norms in formal argumentation[J]. *Journal of Logic and Computation*, 2019, 29(2): 215–240.

Examples of prioritized rules

$$\{\top \overset{1}{\Rightarrow} a, \top \overset{1}{\Rightarrow} b,\ a \overset{2}{\Rightarrow} \neg b,\ b \overset{2}{\Rightarrow} \neg a\}$$



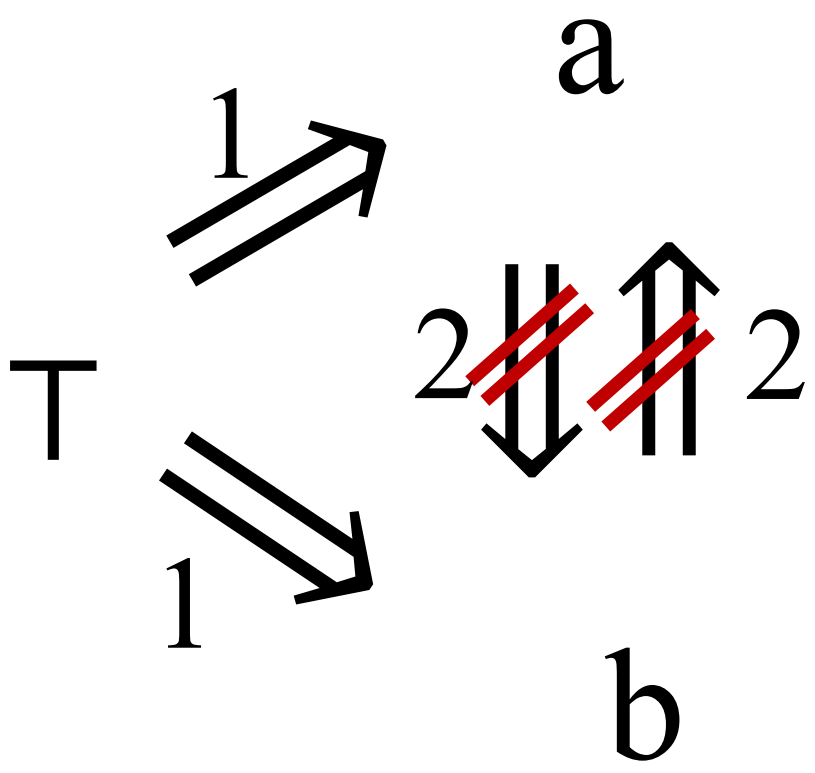
pPDL (concurrent PDL):

{a, b}

Pardo P, Yu L, Chen C, van der Torre L. Weakest link, prioritized default logic and principles in argumentation[J]. *Journal of Logic and Computation*, 2025, 35(4): exaf007.

Examples of prioritized rules

$$\{\top \overset{1}{\Rightarrow} a, \top \overset{1}{\Rightarrow} b,\ a \overset{2}{\Rightarrow} \neg b,\ b \overset{2}{\Rightarrow} \neg a\}$$



pPDL (concurrent PDL):

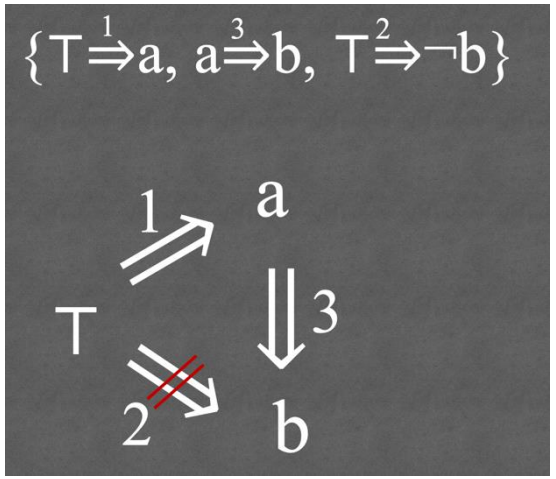
{a, b}

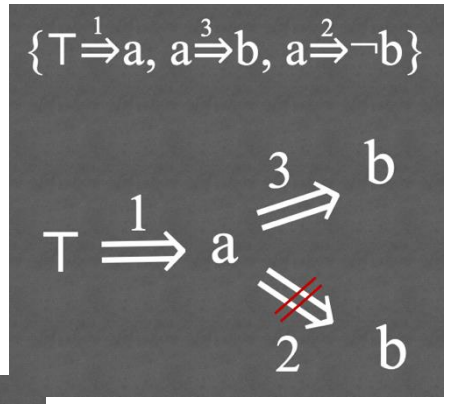
{b, ¬a}

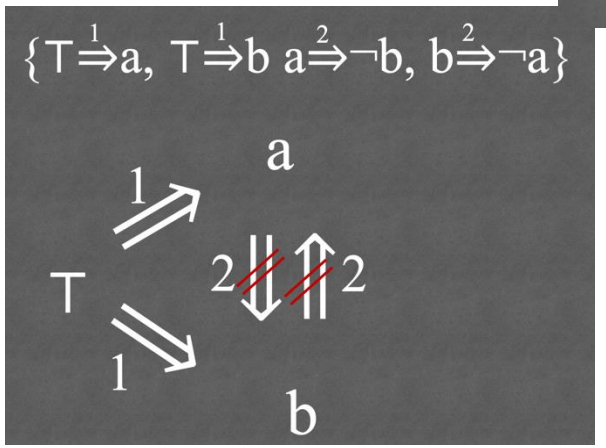
{a, ¬b}

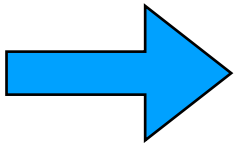
Pardo P, Yu L, Chen C, van der Torre L. Weakest link, prioritized default logic and principles in argumentation[J]. *Journal of Logic and Computation*, 2025, 35(4): exaf007.

Examples vs. Principle-based analysis (Day 3)









1
2
3

Day 2 Reasons first, attack first or defense first?

2.1 Why unify reasons in philosophy and formal argumentation?

2.2 Rules represent reasons in both logic and argumentation.

2.3 Logic dilemma: prescriptive prioritized default logic or descriptive default logic?

2.4 Argumentation dilemma: weakest link or last link?

2.5 Are the two dilemmas two sides of the same coin?

From NLP to KRR

Recall Day 1: diversity of nonmonotonic logics

1. There is the NLP task of translating natural language to formal one
2. There is the KR task of putting some requirements
   - What do you want to derive from these examples?
3. The task of designing logics according to these requirements
   - What is the best logic for your application?

1. There is the NLP task of translating natural language to formal one

Example: *The fitness loving Scot*

It is commonly assumed that if someone was born in Scotland, then he is Scottish. And if he is Scottish, we can normally derive that he likes whisky. However, fitness lovers normally to avoid alcohol for health reasons. Stewart was born in Scotland, he is also a fitness lover, does he like whisky?

Prakken H. An abstract framework for argumentation with structured arguments[J]. Argument & Computation, 2010, 1(2): 93-124.

1. There is the NLP task of translating natural language to formal one

Example: *The fitness loving Scot*

It is commonly assumed that if someone was born in Scotland, then he is Scottish. And if he is Scottish, we can normally derive that he likes whisky. However, fitness lovers normally to avoid alcohol for health reasons.
Stewart was born in Scotland, he is also a fitness lover, does he like whisky?

Example: Snoring professor

A general rule of a library states that mis-behavior, such as snoring loudly, leads to a denial of access. However, there is another rule that professors are normally allowed access.
Bob is a professor and he is snoring loudly in the library, should he be allowed access the library?

Prakken H. An abstract framework for argumentation with structured arguments[J]. Argument & Computation, 2010, 1(2): 93-124.

1. There is the NLP task of translating natural language to formal one

Example: *The fitness loving Scot*

$\top \overset{1}{\Rightarrow} a$

It is commonly assumed that if someone was born in Scotland, then he is Scottish. And if he is Scottish, we can normally derive that he likes whisky. $a \overset{3}{\Rightarrow} b$

However, fitness lovers normally to avoid alcohol for health reasons. $\top \overset{2}{\Rightarrow} \neg b$

Stewart was born in Scotland, he is also a fitness lover, does he like whisky?

$$\{\top \overset{1}{\Rightarrow} a,\ a \overset{3}{\Rightarrow} b,\ \top \overset{2}{\Rightarrow} \neg b\}$$

1. There is the NLP task of translating natural language to formal one

Example: *The fitness loving Scot*

$\top \overset{1}{\Rightarrow} a$

It is commonly assumed that if someone was born in Scotland, then he is Scottish. And if he is Scottish, we can normally derive that he likes whisky. $a \overset{3}{\Rightarrow} b$

However, fitness lovers normally to avoid alcohol for health reasons. $\top \overset{2}{\Rightarrow} \neg b$

Stewart was born in Scotland, he is also a fitness lover, does he like whisky?

Example: Snoring professor

A general rule of a library states that mis-behavior, such as snoring loudly, $\top \overset{1}{\Rightarrow} a$ leads to a denial of access. However, there is another rule that professors are normally allowed access. $\top \overset{2}{\Rightarrow} \neg b$

Bob is a professor and he is snoring loudly in the library, should he be allowed access the library?

$a \overset{3}{\Rightarrow} b$

$$\{\top \overset{1}{\Rightarrow} a,\ a \overset{3}{\Rightarrow} b,\ \top \overset{2}{\Rightarrow} \neg b\}$$

2. There is the KR task of putting some requirements

What do you want to derive from these examples?

Example: *The fitness loving Scot*

It is commonly assumed that if someone was born in Scotland, then he is Scottish. And if he is Scottish, we can normally derive that he likes whisky. However, fitness lovers normally to avoid alcohol for health reasons.

Stewart was born in Scotland, he is also a fitness lover, does he like whisky?

Example: Snoring professor

A general rule of a library states that mis-behavior, such as snoring loudly, leads to a denial of access. However, there is another rule that professors are normally allowed access.

Bob is a professor and he is snoring loudly in the library, should he be allowed access the library?

Prakken H. An abstract framework for argumentation with structured arguments[J]. Argument & Computation, 2010, 1(2): 93-124.

2. There is the KR task of putting some requirements

What do you want to derive from these examples?

Example: *The fitness loving Scot*

It is commonly assumed that if someone was born in Scotland, then he is Scottish. And if he is Scottish, we can normally derive that he likes whisky. However, fitness lovers normally to avoid alcohol for health reasons.

Stewart was born in Scotland, he is also a fitness lover, does he like whisky?

Stewart does not like whisky $\{\top \overset{1}{\Rightarrow} a,\ a \overset{3}{\Rightarrow} b,\ \top \overset{2}{\Rightarrow} \neg b\}$

Example: Snoring professor

A general rule of a library states that mis-behavior, such as snoring loudly, leads to a denial of access. However, there is another rule that professors are normally allowed access.

Bob is a professor and he is snoring loudly in the library, should he be allowed access the library?

Prakken H. An abstract framework for argumentation with structured arguments[J]. Argument & Computation, 2010, 1(2): 93-124.

2. There is the KR task of putting some requirements

What do you want to derive from these examples?

Example: *The fitness loving Scot*

It is commonly assumed that if someone was born in Scotland, then he is Scottish. And if he is Scottish, we can normally derive that he likes whisky. However, fitness lovers normally to avoid alcohol for health reasons.

Stewart was born in Scotland, he is also a fitness lover, does he like whisky?

Stewart does not like whisky $\{\top \overset{1}{\Rightarrow} a,\ a \overset{3}{\Rightarrow} b,\ \top \overset{2}{\Rightarrow} \neg b\}$

Example: Snoring professor

A general rule of a library states that mis-behavior, such as snoring loudly, leads to a denial of access. However, there is another rule that professors are normally allowed access.

Bob is a professor and he is snoring loudly in the library, should he be allowed access the library?

Bob should be access-denied. $\{\top \overset{1}{\Rightarrow} a,\ a \overset{3}{\Rightarrow} b,\ \top \overset{2}{\Rightarrow} \neg b\}$

3. There is the logical task of building logics according to these requirements

Weakest link vs. last link

"The strength of each conclusion is the minimum of the strengths of the inference with which it was derived and of the premises or intermediate conclusions from which it was derived."

---------- Pollock (1995)

"Last-link principle prefers an argument A over another argument B if the last defeasible rules used in B are less preferred than the last defeasible rules in A."

---------- Prakken and Sartor (1997)

Pollock, J.L.: Cognitive carpentry: A blueprint for how to build a person. Mit Press (1995)
H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. Journal of Applied Non-Classical Logics, 7:25–72, 1997.
Prakken H. An abstract framework for argumentation with structured arguments[J]. Argument & Computation, 2010, 1(2): 93-124

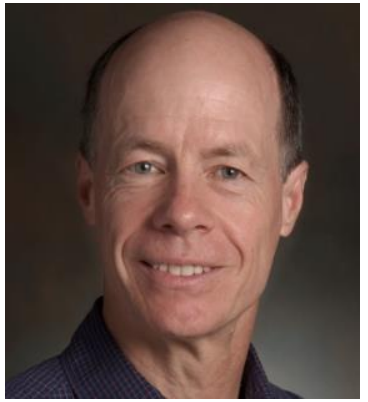
Defaults: The fitness-loving Scot

$$\left\{\begin{array}{rl} bornInScotland & \Rightarrow scottish \\ scottish & \Rightarrow likesWhisky \\ fitnessLover & \Rightarrow \neg likesWhisky \end{array}\right\}$$

Norms: Snoring Professor at Library

$$\left\{\begin{array}{r} snores \Rightarrow misbehaves \\ misbehaves \Rightarrow accessDenied \\ professor \Rightarrow \neg accessDenied \end{array}\right\}$$

Logic: prescriptive PDL vs. descriptive PDL



What does a priority mean?
How do priorities aggregate?

(Delgrande et al. 2004)

Prescriptive reading:
Strongest defaults should apply first

$$\{\top\}$$

$$\left\{\top \overset{2}{\Rightarrow} \neg b,\ \top \overset{1}{\Rightarrow} a\right\}$$

~Weakest Link

$$\left\{\begin{array}{c} \top \overset{1}{\Rightarrow} a \\ \top \overset{2}{\Rightarrow} \neg b \\ a \overset{3}{\Rightarrow} b \end{array}\right\}$$

Descriptive reading:
The priority of a default is its contribution to the argument's status

$$\left\{a \overset{3}{\Rightarrow} b,\ \top \overset{1}{\Rightarrow} a\right\}$$

~Last Link

Weakest Link (WL)

A chain is as strong as its weakest link.

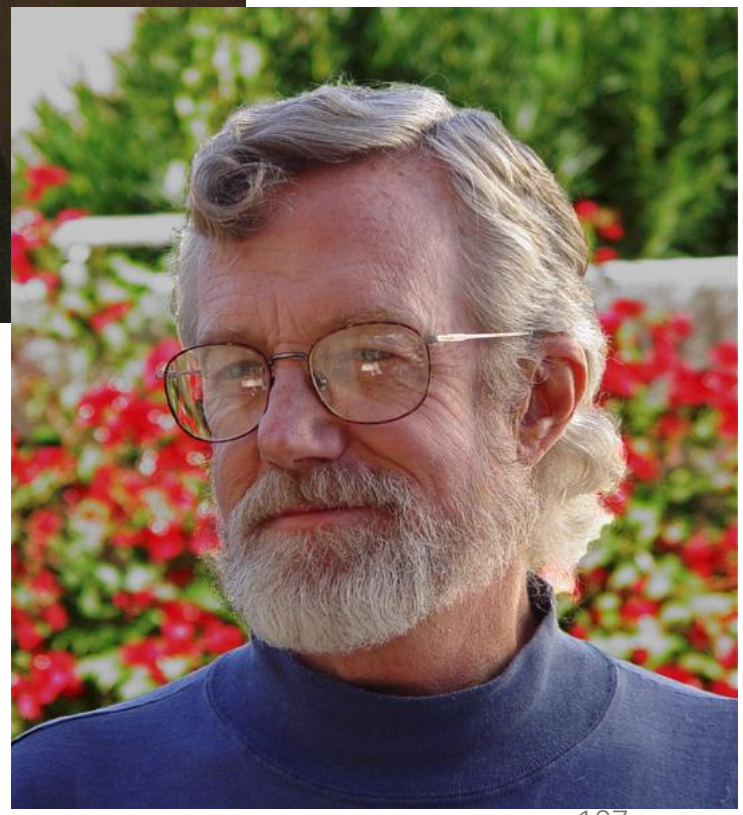
(folklore)

In every chain of reasoning, the evidence of the last conclusion can be no greater than that of the weakest link of the chain, whatever may be the strength of the rest.

(Thomas Reid, 1786)

The strength of each conclusion is the minimum of the strengths of the inference [...] and of the premises or intermediate conclusions […].

(John L. Pollock, 1995)



Argumentation: weakest vs. last link

Definition (Rule)

A defeasible rule is of the form $b_1, \ldots, b_n \Rightarrow h$, where $b_1, \ldots, b_n, h$ are domain literals. A strict rule is of the form $b_1, \ldots, b_n \rightarrow h$, where h is either a domain literal or a non-domain atom ab_d for some defeasible rule d. For any rule r, we define the body of r as $\mathrm{bd}(r) = \{b_1, \ldots, b_n\}$ and the head of r as $\mathrm{hd}(r) = h$.

Definition (Rule-based system)

A rule-based system $RBS = (RS, RD, \preceq)$ consists of a set RS of strict rules, a finite set RD of defeasible rules and a total preorder $\preceq$ on RD. Equivalently, we write $RBS = (RS, RD, \mathrm{rank})$ with a function $\mathrm{rank}\colon RD \to \mathbb{N}$ satisfying $\mathrm{rank}(d) \leq \mathrm{rank}(d')$ iff $d \preceq d'$.

.

Argumentation: weakest vs. last link

Definition (Rule)

A defeasible rule is of the form $b_1, \ldots, b_n \Rightarrow h$, where $b_1, \ldots, b_n, h$ are domain literals. A strict rule is of the form $b_1, \ldots, b_n \rightarrow h$, where h is either a domain literal or a non-domain atom ab_d for some defeasible rule d. For any rule r, we define the body of r as $\mathrm{bd}(r) = \{b_1, \ldots, b_n\}$ and the head of r as $\mathrm{hd}(r) = h$.

Definition (Rule-based system)

A rule-based system $RBS = (RS, RD, \preceq)$ consists of a set RS of strict rules, a finite set RD of defeasible rules and a total preorder $\preceq$ on RD. Equivalently, we write $RBS = (RS, RD, \mathrm{rank})$ with a function $\mathrm{rank}\colon RD \to \mathbb{N}$ satisfying $\mathrm{rank}(d) \leq \mathrm{rank}(d')$ iff $d \preceq d'$.

.

Argumentation: weakest vs. last link

Definition (Knowledge base)

A knowledge base is a pair $K = (RBS, BE)$ containing a rule-based system $RBS = (RS, RD, \text{rank})$ and a base of evidence BE. We will also write $K = (RS, RD, \text{rank}, BE)$ instead of $K = (RBS, BE)$.

Argumentation: weakest vs. last link

Definition (Argument)

Given a knowledge base $K = (RS, RD, \text{rank}, BE)$, an argument wrt K is inductively defined as follows:

1. For each $\alpha \in BE$, $[\alpha]$ is an argument with conclusion α.
2. Let r be a rule of the form $\alpha_1, \ldots, \alpha_n \to \alpha$ or $\alpha_1, \ldots, \alpha_n \Rightarrow \alpha$ (with $n \geq 1$) from K. Further suppose that $A_1, \ldots, A_n$ are arguments with conclusions $\alpha_1, \ldots, \alpha_n$, respectively. Then $A = [A_1, \ldots, A_n \to \alpha]$, resp. $A = [A_1, \ldots, A_n \Rightarrow \alpha]$, is also an argument, with conclusion $\text{cnl}(A) = \alpha$ and last rule $\text{last}(A) = r$. We often write either argument as $[A_1, \ldots, A_n, r]$.

Each argument wrt K is obtained by finitely many applications of Steps 1–2.

Argumentation: weakest vs. last link

Definition (Argumentation framework)

The set of all arguments induced by a knowledge base K is denoted by AR_K. An argumentation framework AF induced by K is a pair $AF = (AR_K, \text{att}(K))$, where $\text{att}(K) \subseteq AR_K \times AR_K$ is called an attack relation.

Definition (Attack relation assignment)

Given a sensible class of knowledge bases $\mathcal{K}$, an attack relation assignment is a function $\text{att}\colon K \longmapsto \text{att}(K)$ mapping each $K \in \mathcal{K}$ to an attack relation $\text{att}(K) \subseteq AR_K \times AR_K$ such that no attack $(A, B) \in \text{att}(K)$ exists against a strict argument $B \in AR_K$.

Argumentation: weakest vs. last link

Definition (Undercut, rebut)

Let $A, B \in AR_K$ for a knowledge base K. We say that A undercuts B at B' iff $B' \in \text{sub}(B)$, $\text{last}(B') = d$ is defeasible and $\text{cnl}(A) = ab_d$.
A contradicts B at B' iff $B' \in \text{sub}(B)$ and the conclusions of A and B' are contradictory: either $\text{cnl}(A) = \neg\,\text{cnl}(B')$ or $\text{cnl}(B') = \neg\,\text{cnl}(A)$.
A rebuts B at B' iff A contradicts B at B', where B' is a basic defeasible subargument; that is, $\text{last}(B') \in RD$.

Argumentation: weakest vs. last link

Definition (Weakest link)

Let $R \subseteq RD$ be a set of defeasible rules. The weakest link of R, denoted $\mathrm{wl}(R)$, is the rank of the lowest-ranked defeasible rule in R. Formally, $\mathrm{wl}(R) = \min_{r \in R} \mathrm{rank}(r)$. We extend $\mathrm{wl}(\cdot)$ to arguments A:

$$\mathrm{wl}(A) = \begin{cases} \infty & \text{if } A \text{ is strict,} \\ \mathrm{wl}(\mathrm{dr}(A)) & \text{if } A \text{ is defeasible.} \end{cases}$$

Argumentation: weakest vs. last link

Definition (Weakest link)

Let $R \subseteq RD$ be a set of defeasible rules. The weakest link of R, denoted $\mathrm{wl}(R)$, is the rank of the lowest-ranked defeasible rule in R. Formally, $\mathrm{wl}(R) = \min_{r \in R} \mathrm{rank}(r)$. We extend $\mathrm{wl}(\cdot)$ to arguments A:

$$\mathrm{wl}(A) = \begin{cases} \infty & \text{if } A \text{ is strict,} \\ \mathrm{wl}(\mathrm{dr}(A)) & \text{if } A \text{ is defeasible.} \end{cases}$$

Definition (Simple weakest link attack)

Let $A, B \in AR_K$ for a knowledge base K. We say that A swl-attacks B at $B' \in \mathrm{sub}(B)$, denoted $(A, B) \in \mathrm{att}_{\mathrm{swl}}(K)$, iff

$$A \text{ undercuts } B \text{ at } B' \quad \text{or} \quad \big(A \text{ rebuts } B \text{ at } B' \text{ and } \mathrm{wl}(A) \not\preceq \mathrm{wl}(B')\big).$$

Argumentation: weakest vs. last link

Definition (Last link attack)

Let $A, B \in AR_K$ for a knowledge base K. We say that A ll-attacks B at $B' \in \text{sub}(B)$, denoted $(A, B) \in \text{att}_{\text{ll}}(K)$, iff

$$A \text{ rebuts } B \text{ at } B', \ \text{last}(A) \in RD \text{ and } \ \text{rank}(\text{last}(A)) \not< \text{rank}(\text{last}(B')).$$

Example: weakest link vs. last link

The fitness-lover Scot

$$\left\{ \begin{array}{c} bornInScotland \overset{1}{\Rightarrow} scottish \\ scottish \overset{3}{\Rightarrow} likesWhisky \\ fitnessLover \overset{2}{\Rightarrow} \neg likesWhisky \end{array} \right\}$$

Snoring professor at library

$$\left\{ \begin{array}{c} snores \overset{1}{\Rightarrow} misbehaves \\ misbehaves \overset{3}{\Rightarrow} accessDenied \\ professor \overset{2}{\Rightarrow} \neg accessDenied \end{array} \right\}$$

$[[[\text{bornInScotland}] \overset{1}{\Rightarrow} \text{scottish}] \overset{3}{\Rightarrow} \text{likesWhisky}]$

$[[\text{fitnessLover}] \overset{2}{\Rightarrow} \neg likesWhisky]$

{bornInScotland, fitnessLover, scottish, ¬likesWhisky}

Weakest Link

Pollock, J.L.: Cognitive carpentry: A blueprint for how to build a person. Mit Press (1995)
Prakken H. An abstract framework for argumentation with structured arguments[J]. Argument & Computation, 2010, 1(2): 93-124

Example: weakest link vs. last link

The fitness-lover Scot

$$\left\{ \begin{array}{c} bornInScotland \overset{1}{\Rightarrow} scottish \\ scottish \overset{3}{\Rightarrow} likesWhisky \\ fitnessLover \overset{2}{\Rightarrow} \neg likesWhisky \end{array} \right\}$$

[[[bornInScotland] $\overset{1}{\Rightarrow}$ scottish] $\overset{3}{\Rightarrow}$ likesWhisky]

[[fitnessLover] $\overset{2}{\Rightarrow}$ $\neg$*likesWhisky*]

{bornInScotland, fitnessLover, scottish, ¬likesWhisky}

Weakest Link

Snoring professor at library

$$\left\{ \begin{array}{c} snores \overset{1}{\Rightarrow} misbehaves \\ misbehaves \overset{3}{\Rightarrow} accessDenied \\ professor \overset{2}{\Rightarrow} \neg accessDenied \end{array} \right\}$$

[[[snores] $\overset{1}{\Rightarrow}$Misbehaves] $\overset{3}{\Rightarrow}$ accessDenied]

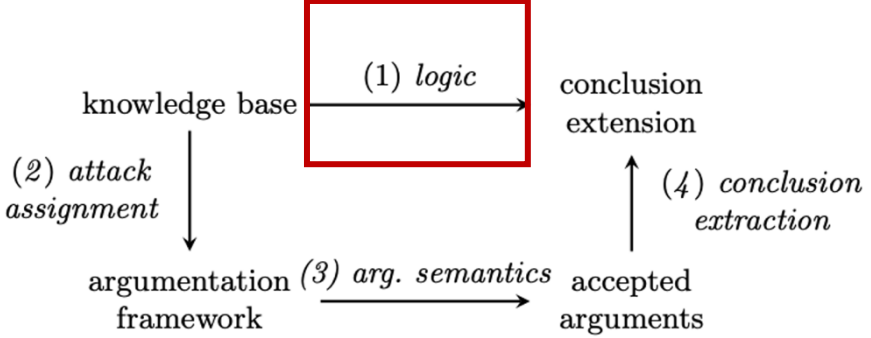
[[professor] $\overset{2}{\Rightarrow}\neg$*accessDenied*]

{snores, professor, misbehaves, acessDenied}

Last Link

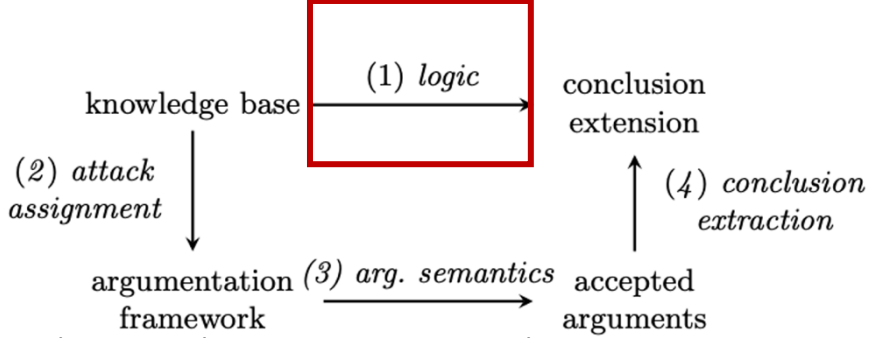
Pollock, J.L.: Cognitive carpentry: A blueprint for how to build a person. Mit Press (1995)
Prakken H. An abstract framework for argumentation with structured arguments[J]. Argument & Computation, 2010, 1(2): 93-124

Commutative diagram illustration: PDL and weakest link



1. Knowledge base contains prioritised defaults $a \overset{n}{\Rightarrow} b$ and facts.
2. A default $a \overset{n}{\Rightarrow} b$ reads as: if a then normally b.
3. Higher number n means a higher priority for the default rule a ⇒ b.
4. PDL selects sets of defaults and extracts their conclusions into extensions.
5. Apply at each step one of the unapplied defaults with the highest priority.

Commutative diagram illustration: PDL and weakest link



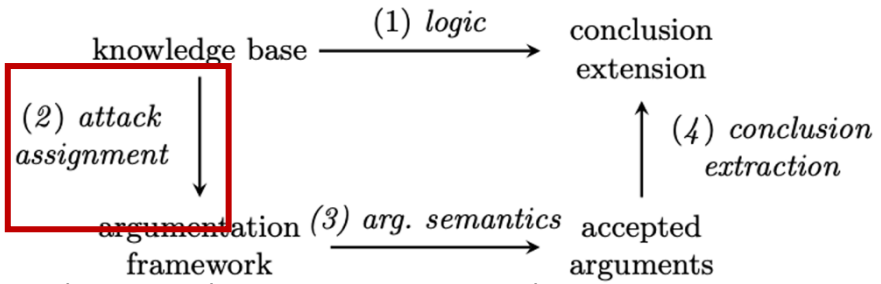
4. PDL selects sets of defaults and extracts their conclusions into extensions.
5. Apply at each step one of the unapplied defaults with the highest priority.

Default: $\{\top \overset{1}{\Rightarrow} a, \top \overset{2}{\Rightarrow} \neg b, a \overset{3}{\Rightarrow} b\}$

Facts: $\{\top\}$

$\xrightarrow{\text{PDL}} Cn(\{\neg b, a\})$

Commutative diagram illustration: PDL and weakest link



4. PDL selects sets of defaults and extracts their conclusions into extensions.
5. Apply at each step one of the unapplied defaults with the highest priority.

Default: $\{\top \overset{1}{\Rightarrow} a, \top \overset{2}{\Rightarrow} \neg b, a \overset{3}{\Rightarrow} b\}$

Facts: $\{\top\}$

PDL $\longrightarrow$ $Cn(\{\neg b, a\})$

Argument construction $\downarrow$ Attack assignment

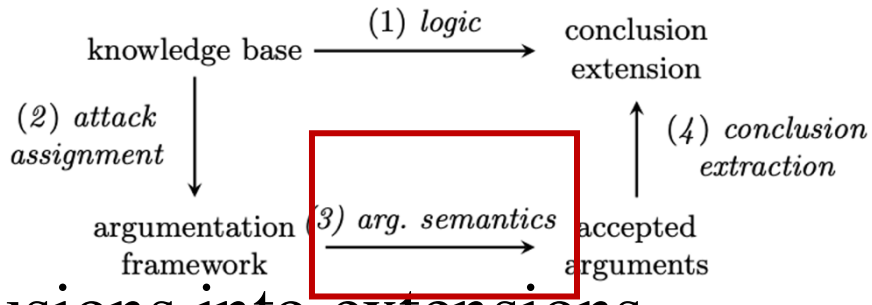
A: $\top \overset{1}{\Rightarrow} a$

C: $\top \overset{2}{\Rightarrow} \neg b$

B: $\top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b$

Weakest link

Commutative diagram illustration: PDL and weakest link



4. PDL selects sets of defaults and extracts their conclusions into extensions.
5. Apply at each step one of the unapplied defaults with the highest priority.

Default: $\{\top \overset{1}{\Rightarrow} a, \top \overset{2}{\Rightarrow} \neg b, a \overset{3}{\Rightarrow} b\}$

Facts: $\{\top\}$

PDL $\longrightarrow$ $Cn(\{\neg b, a\})$

Argument construction ↓ Attack assignment

Stable $\longrightarrow$ $\{A, C\}$

Weakest link

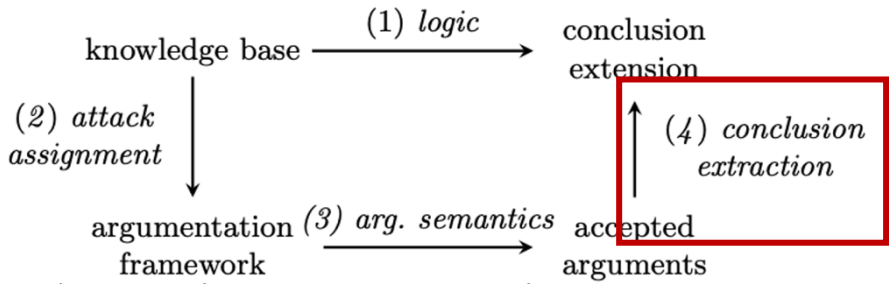
A: $\top \overset{1}{\Rightarrow} a$

C: $\top \overset{2}{\Rightarrow} \neg b$

B: $\top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b$

Weakest link

Commutative diagram illustration: PDL and weakest link



4. PDL selects sets of defaults and extracts their conclusions into extensions.
5. Apply at each step one of the unapplied defaults with the highest priority.

Default: $\{\top \overset{1}{\Rightarrow} a, \top \overset{2}{\Rightarrow} \neg b, a \overset{3}{\Rightarrow} b\}$

Facts: $\{\top\}$

PDL → $Cn(\{\neg b, a\})$

Argument construction ↓ Attack assignment

Stable → $\{A, C\}$

Weakest link

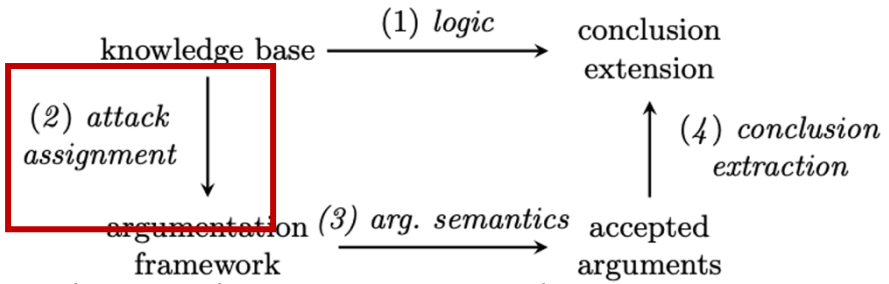
A: $\top \overset{1}{\Rightarrow} a$

C: $\top \overset{2}{\Rightarrow} \neg b$

B: $\top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b$

Weakest link

Commutative diagram illustration: PDL and weakest link



4. PDL selects sets of defaults and extracts their conclusions into extensions.
5. Apply at each step one of the unapplied defaults with the highest priority.

Default: $\{\top \overset{1}{\Rightarrow} a, \top \overset{2}{\Rightarrow} \neg b, a \overset{3}{\Rightarrow} b\}$

Facts: $\{\top\}$

PDL → $Cn(\{\neg b, a\})$

Argument construction ↓ Attack assignment

Stable → $\{A, C\}$ Weakest link

$A\colon \top \overset{1}{\Rightarrow} a$

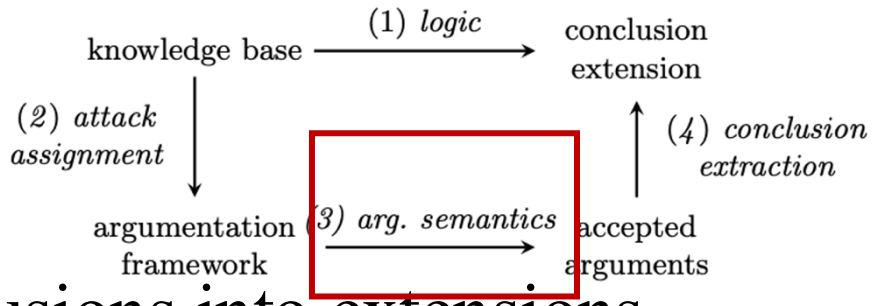
$C\colon \top \overset{2}{\Rightarrow} \neg b$

$B\colon \top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b$

Last link

Weakest link

Commutative diagram illustration: PDL and weakest link



4. PDL selects sets of defaults and extracts their conclusions into extensions.
5. Apply at each step one of the unapplied defaults with the highest priority.

Default: $\{\top \overset{1}{\Rightarrow} a, \top \overset{2}{\Rightarrow} \neg b, a \overset{3}{\Rightarrow} b\}$

Facts: $\{\top\}$

PDL $\longrightarrow$ $Cn(\{\neg b, a\})$

Argument construction ↓ Attack assignment

$A\colon \top \overset{1}{\Rightarrow} a$

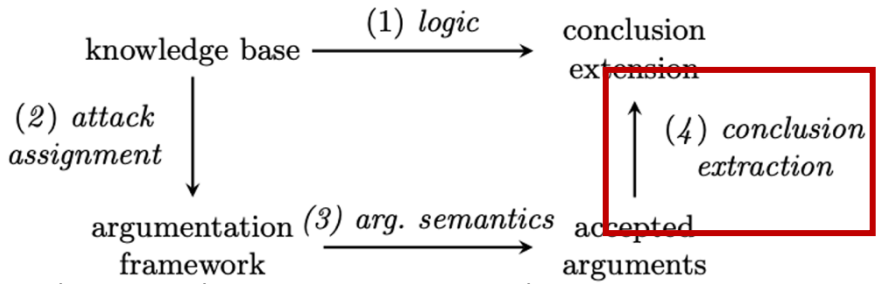
$C\colon \top \overset{2}{\Rightarrow} \neg b$

$B\colon \top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b$

Weakest link

Stable $\longrightarrow$ $\{A, C\}$ Weakest link; $\{A, B\}$ Last link

Commutative diagram illustration: PDL and weakest link



4. PDL selects sets of defaults and extracts their conclusions into extensions.
5. Apply at each step one of the unapplied defaults with the highest priority.

Default: $\{\top \overset{1}{\Rightarrow} a, \top \overset{2}{\Rightarrow} \neg b, a \overset{3}{\Rightarrow} b\}$

Facts: $\{\top\}$

PDL → $Cn(\{\neg b, a\})$

Argument construction ↓ Attack assignment

Stable →

$\{A, C\}$ Weakest link

$\{A, B\}$ Last link

$A\colon \top \overset{1}{\Rightarrow} a$

$C\colon \top \overset{2}{\Rightarrow} \neg b$

$B\colon \top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b$

Weakest link

$A\colon \top \overset{1}{\Rightarrow} a$

$C\colon \top \overset{2}{\Rightarrow} \neg b$

$B\colon \top \overset{1}{\Rightarrow} a \overset{3}{\Rightarrow} b$

Last link

Day 2 Reasons first, attack first or defense first?

2.1 Why unify reasons in philosophy and formal argumentation?

2.2 Rules represent reasons in both logic and argumentation.

2.3 Logic dilemma: prescriptive prioritized default logic or descriptive default logic?

2.4 Argumentation dilemma: weakest link or last link?

2.5 Are the two dilemmas two sides of the same coin?

Three distinctions that should not be conflated

Distinction	Question
Prescriptive vs. descriptive preference	What does a priority between rules mean?
Weakest link vs. last link	How does rule priority determine the strength of a complete argument?
Epistemic vs. normative reasoning	What kind of information do the rules represent?

Are the two dilemmas two sides of the same coin?

“ Prescriptive priorities are suitable when priorities represent procedural control: the knowledge engineer wants to specify explicitly which rules must be considered or applied first. This provides a tighter characterization of the domain.”

Delgrande et al. 2024

“Weakest link is suitable for reasoning with empirical generalizations because uncertainty in an earlier premise or inference should propagate through every argument constructed from it.”

Modgil S, Prakken H. 2014

Are the two dilemmas two sides of the same coin?

"Within legal reasoning, weakest link is suitable for the evidential or factual component of a case, where testimony, observations, expert judgments, and empirical rules may have different degrees of reliability."

Prakken H, Sartor G.2002

"Weakest link is suitable for the epistemic layer of multi-agent deliberation, while normative rules and values can be compared separately at a later stage. "

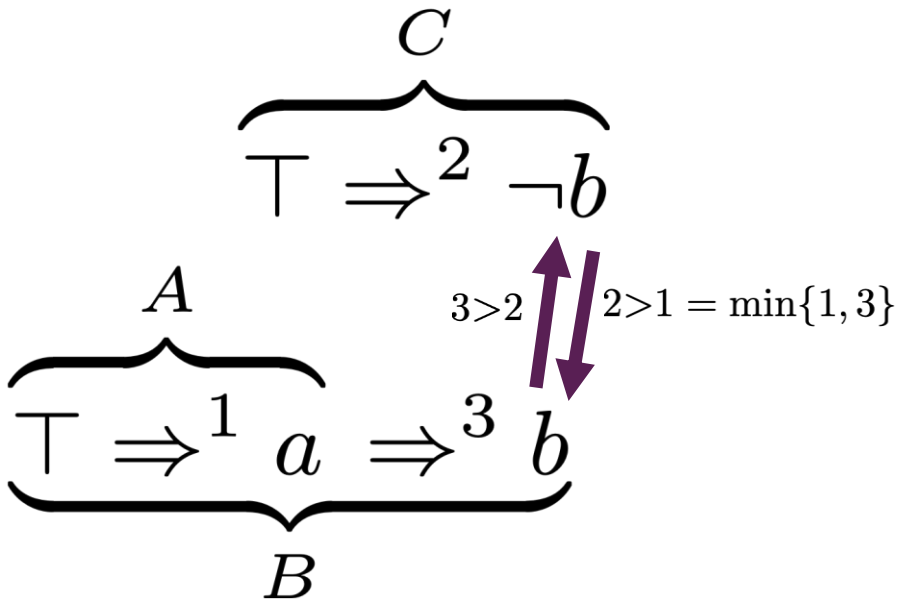
Yu Z et al. 2024

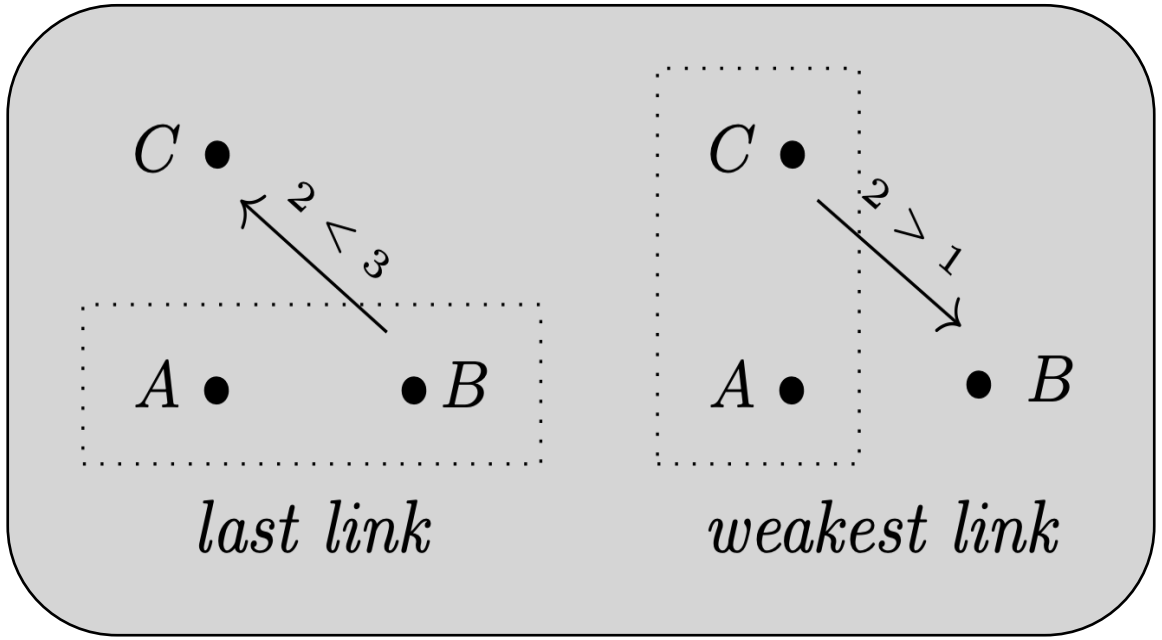
"Last link is suitable for legal and other normative reasoning because a normative conflict should be resolved by comparing the rules that directly support the conflicting legal conclusions, rather than every factual rule appearing earlier in the arguments."

Modgil S, Prakken H.2024

Old ASPIC+ (Prakken 2010)

Don’t select defaults, build all arguments instead.

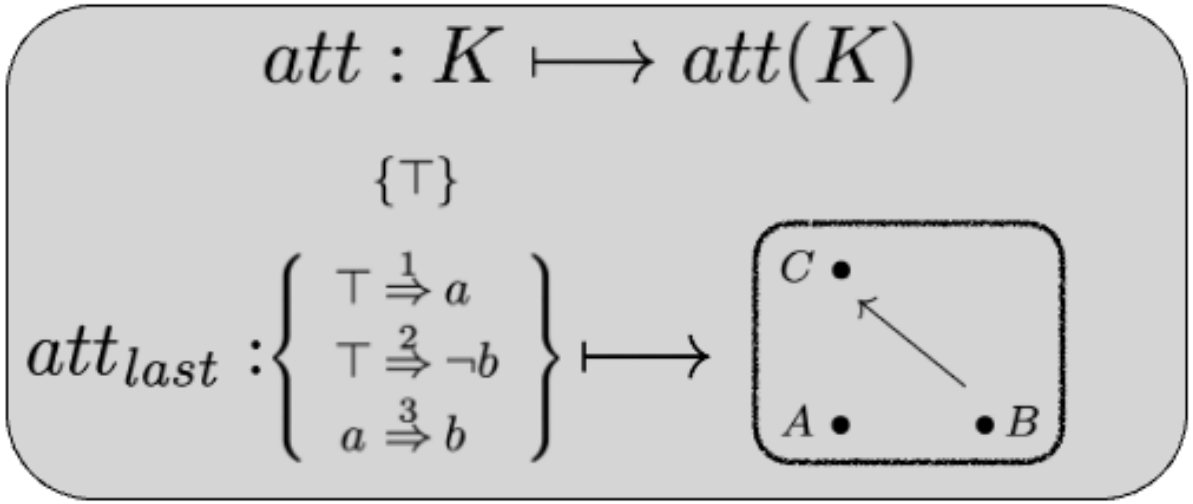




New ASPIC+ (Dung 2016 &18)

- Axioms for Argumentation (Attack relation assignments)

WL vs LL: Let us not debate intuitions on examples.
Let us debate intuitions on axioms for attack relations



Axioms (P1)—(P8)
Does your *att* satisfy (P1)—(P8)?
Last Link satisfies (P1)—(P8).
Hence, Last Link > Weakest Link.

Day 3

Day 2 Reasons first, attack first or defense first?

2.1 Why unify reasons in philosophy and formal argumentation?

2.2 Rules represent reasons in both logic and argumentation.

2.3 Logic dilemma: prescriptive prioritized default logic or descriptive default logic?

2.4 Argumentation dilemma: weakest link or last link?

2.5 Are the two dilemmas two sides of the same coin?

Logic and Argumentation for New-generation AI

ESSLLI 2026

Liuwen Yu[1]
Leon van der Torre[2]
[1] Luxembourg Institute of Science and Technology
[2] University of Luxembourg

Recap
Day 1 Dung paradigm shift and reasoning alignment

1.1 AI Logic (JLC corner) is nonmonotonic logic (since 1980)

1.2 Explanation of black boxes in new generation AI: secretary metaphor

1.3 Dung 1995 paradigm shift relates logic and argumentation based explanation

1.4 Reasoning alignment of logic and argumentation based explanation

1.5 Using logic: from inference towards dialogue and decision making, to logic in action

Recap
Day 2 Reasons first, attack first or defense first?

2.1 Why unify reasons in philosophy and formal argumentation?

2.2 Rules represent reasons in both logic and argumentation.

2.3 Logic dilemma: prescriptive prioritized default logic or descriptive default logic?

2.4 Argumentation dilemma: weakest link or last link?

2.5 Are the two dilemmas two sides of the same coin?

Day 3 Axiomatic approach and impossibility

3.1 Dung 1995 and 2016/18: unification, representation, axiomatization.

3.2 Representation theorems

3.3 Impossibility: exact representation and natural principles can conflict.

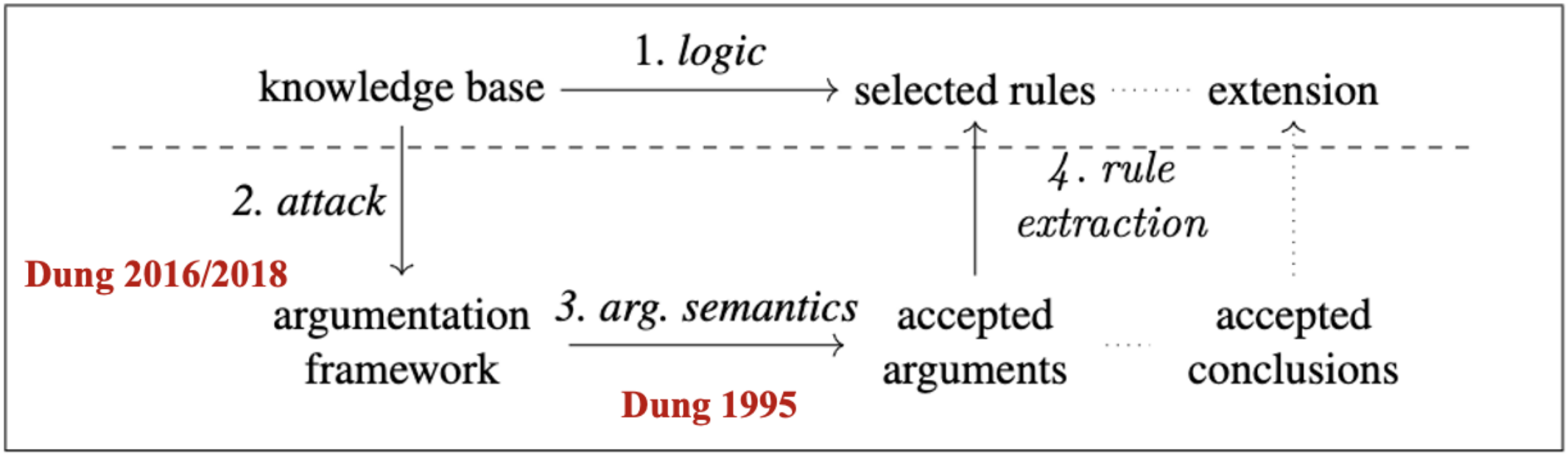
3.4 Escape routes: restrict the domain or allow context dependence.

3.5 Rationality postulates and principle-based analysis

Creating argumentation frameworks and semantics

ASPIC+ = Dung 2016/2018: argument construction and attack assignment from knowledge base to capture nonmonotonic logic, evaluated with a set of properties

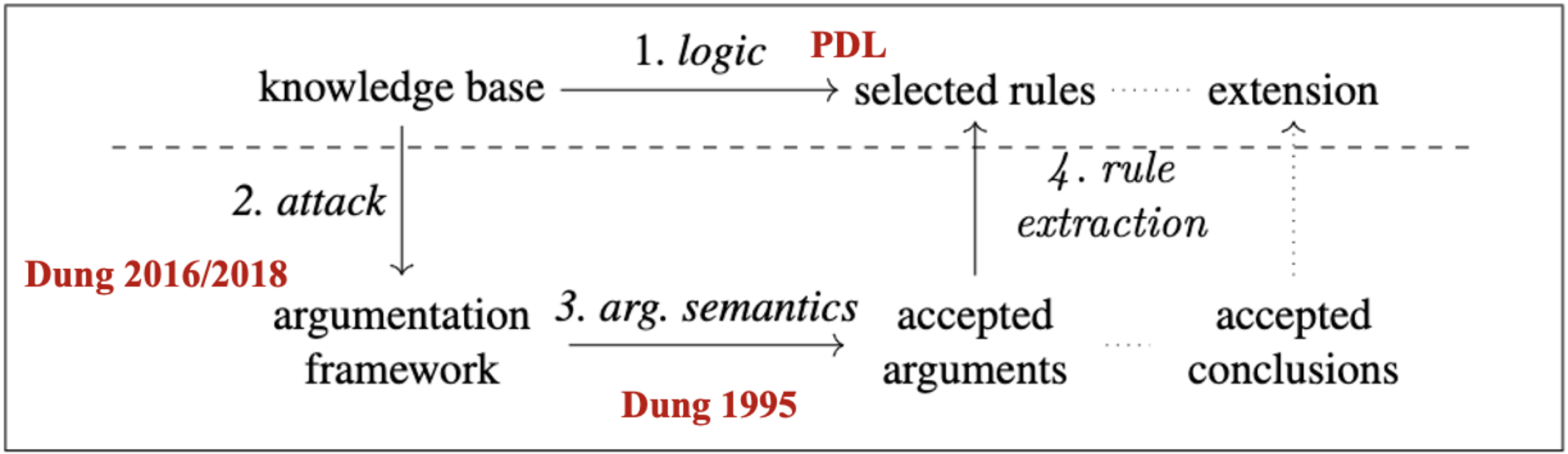
Abstract = Dung 1995: focus on the relation among arguments, put "attack" as the universal foundation at the top of the argumentation



Prakken H. An abstract framework for argumentation with structured arguments[J]. Argument & Computation, 2010, 1(2): 93-124.
Dung P M. An axiomatic analysis of structured argumentation with priorities[J]. Artificial Intelligence, 2016, 231: 107-150.
Dung P M, Thang P M. Fundamental properties of attack relations in structured argumentation with priorities[J]. Artificial Intelligence, 2018, 255: 1-42.
Dung P M. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games[J]. Artificial intelligence, 1995, 77(2): 321-357.

Creating argumentation frameworks and semantics

We are confronted by (1), and we look for (2-4) to make the diagram commutative
We combine (2-4) to design new logics as a result (1) satisfying some requirements



Prakken H. An abstract framework for argumentation with structured arguments[J]. Argument & Computation, 2010, 1(2): 93-124.
Dung P M. An axiomatic analysis of structured argumentation with priorities[J]. Artificial Intelligence, 2016, 231: 107-150.
Dung P M, Thang P M. Fundamental properties of attack relations in structured argumentation with priorities[J]. Artificial Intelligence, 2018, 255: 1-42.
Dung P M. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games[J]. Artificial intelligence, 1995, 77(2): 321-357.

Some representation results

ELSEVIER

Artificial Intelligence 77 (1995) 321–357

Artificial Intelligence

On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games★

Phan Minh Dung*

Division of Computer Science, Asian Institute of Technology, GPO Box 2754, Bangkok 10501, Thailand

Received June 1993; revised April 1994

Reiter's default logic as argumentation

Theorem 43. *Let $T = (D, W)$ be a default theory. Let E be an R-extension of T and E' be a stable extension of $AF(T)$. Then*

(1) *arg(E) is a stable extension of $AF(T)$,*
(2) *flat(E') is an R-extension of T.*

Logic programming as argumentation

Theorem 50. *Let P be a logic program, and WFM be the well-founded model of P. Let GE be the grounded extension of $AF_{napif}(P)$. Then*

$$WFM = \{h \mid \exists (K, h) \in GE\} .$$

▪ ▪ ▪ ▪ ▪ ▪

Formal argumentation as a formal method

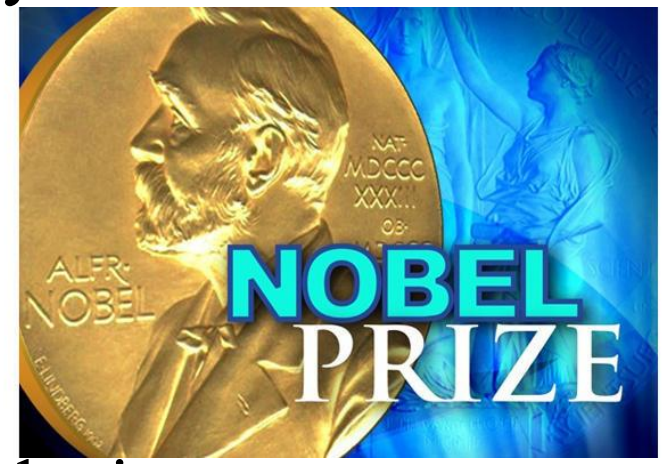
What is the methodology of (formal) argumentation?

Principles (properties, axioms, postulates, etc.)

Describe (formal) argumentation at higher level of abstraction

The principle-based approach: a methodology to deal with diversity

- Voting theory: every country has its own voting system
  - Arrow's axioms (impossibility result)



- Logic in AI: AGM theory change, nonmonotonic logic, ...
  - KLM preferential reasoning

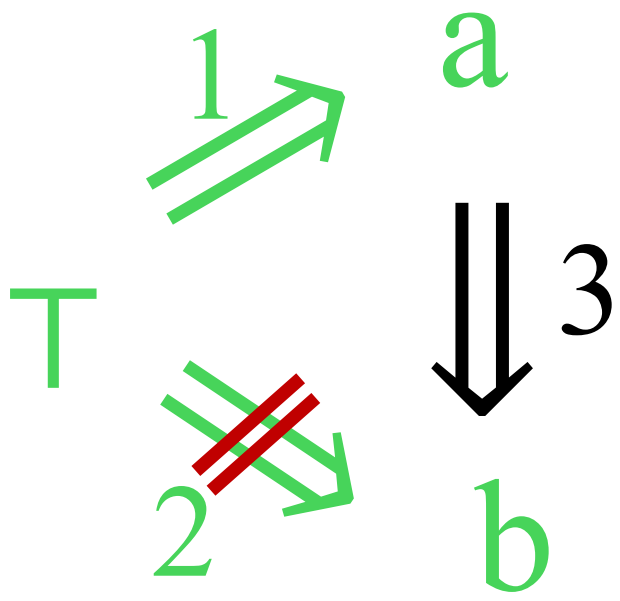
The axiomatization of **P** consists of all axioms and rules of propositional calculus together with the following axioms and rules (notice that $\vdash$ denotes provability in the propositional calculus):

- REF. $A \mathrel{|\!\sim} A$ (reflexivity)
- LLE. If $\vdash A \leftrightarrow B$, then $(A \mathrel{|\!\sim} C) \rightarrow (B \mathrel{|\!\sim} C)$ (left logical equivalence)
- RW. If $\vdash A \rightarrow B$, then $(C \mathrel{|\!\sim} A) \rightarrow (C \mathrel{|\!\sim} B)$ (right weakening)
- CM. $((A \mathrel{|\!\sim} B) \wedge (A \mathrel{|\!\sim} C)) \rightarrow (A \wedge B \mathrel{|\!\sim} C)$ (cautious monotonicity)
- AND. $((A \mathrel{|\!\sim} B) \wedge (A \mathrel{|\!\sim} C)) \rightarrow (A \mathrel{|\!\sim} B \wedge C)$
- OR. $((A \mathrel{|\!\sim} C) \wedge (B \mathrel{|\!\sim} C)) \rightarrow (A \vee B \mathrel{|\!\sim} C)$

Arrow, Kenneth J. (1950). "A Difficulty in the Concept of Social Welfare". Journal of Political Economy. 58 (4): 328–346.
Source: wikipedia

One triangle example (recap)

$$\{\top \overset{1}{\Rightarrow} a,\ a \overset{3}{\Rightarrow} b,\ \top \overset{2}{\Rightarrow} \neg b\}$$



{¬ b}

{¬ b, a}

Brewka G, Eiter T. Prioritizing default logic[C]//Hölldobler S, ed. *Intellectics and Computational Logic: Papers in Honour of Wolfgang Bibel*. Dordrecht: Kluwer Academic Publishers, 2000: 27–45.

Representation: KLM and cumulativity instead of monotonicity

Definition

$$\text{CUT}: \ \alpha \mid\sim \beta,\ \alpha \wedge \beta \mid\sim \gamma \Rightarrow \alpha \mid\sim \gamma; \quad \text{CM}: \ \alpha \mid\sim \beta,\ \alpha \mid\sim \gamma \Rightarrow \alpha \wedge \beta \mid\sim \gamma.$$

Example

Assume $s \mid\sim p$, $s \mid\sim q$, and $s \wedge p \mid\sim r$. Then

$$s \wedge p \mid\sim q \quad \text{by CM}, \qquad s \mid\sim r \quad \text{by CUT}.$$

An expected conclusion can be used and then removed.

adding a proven conclusion to your premises as a new fact
does NOT change the set of conclusions you can derive

Credulous cumulativity in formal argumentation

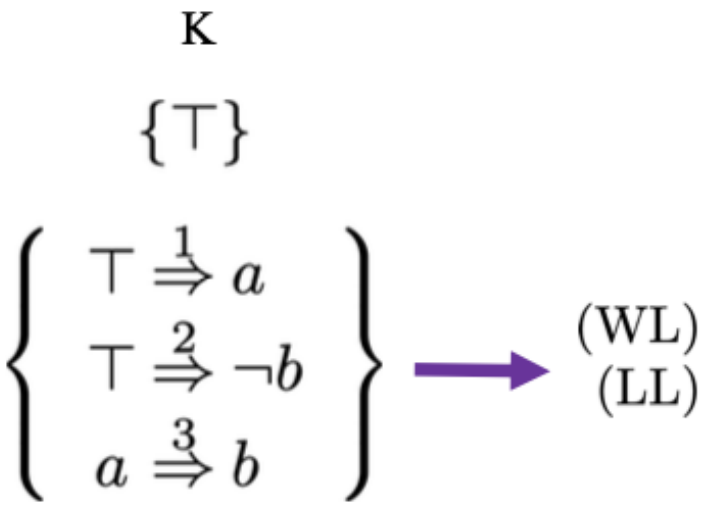
Definition (Credulous cumulativity)

Let $\mathcal{K}$ be a sensible class of knowledge bases and *att* be an attack relation assignment defined for $\mathcal{K}$. We say *att* satisfies the property of credulous cumulativity for $\mathcal{K}$ if and only if, for each $K \in \mathcal{K}$, for each stable belief set S of K w.r.t. *att*, and for each finite subset $\Omega \subseteq S$ of domain literals,

1. $K + \Omega$ is a consistent knowledge base (i.e., $K + \Omega$ belongs to $\mathcal{K}$), and
2. S is a stable belief set of $K + \Omega$ w.r.t. *att*.

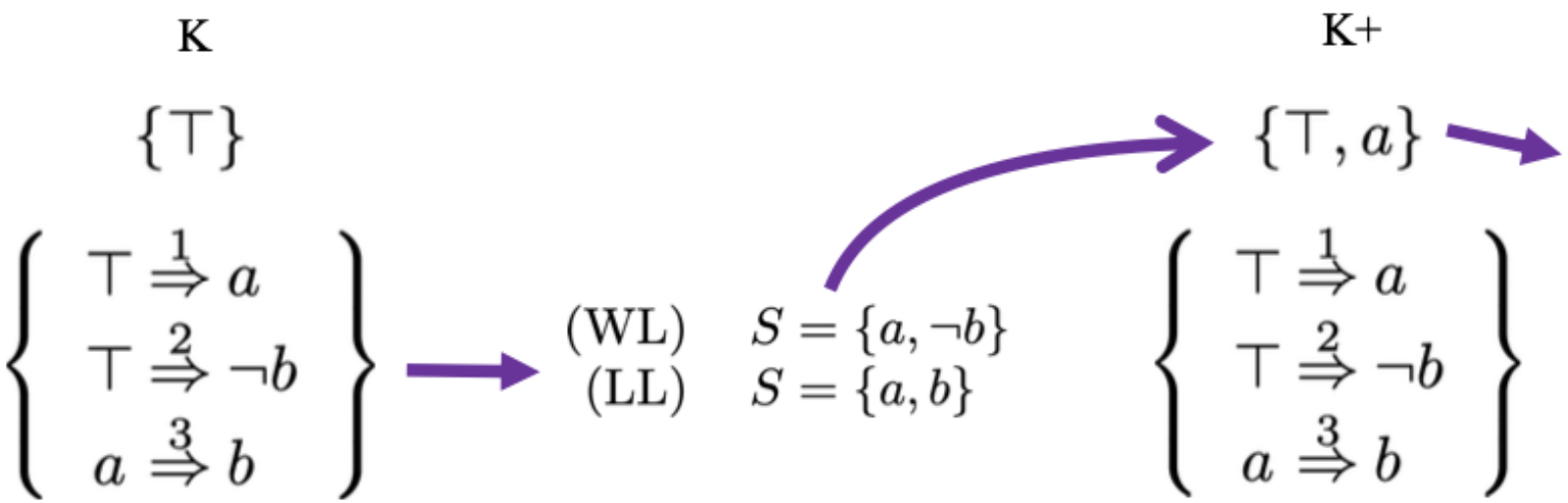
Weakest link does not satisfy credulous cumulativity

For any stable belief set S of K, if we add conclusions of S as new facts, say in K+, then S is a stable set of K+.



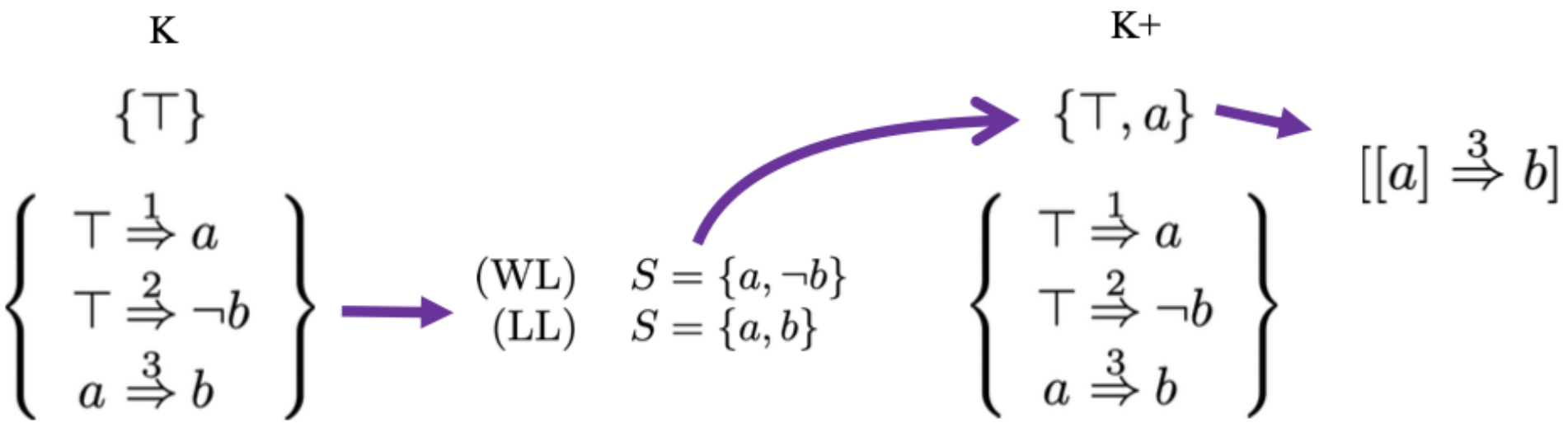
Weakest link does not satisfy credulous cumulativity

For any stable belief set S of K, if we add conclusions of S as new facts, say in K+, then S is a stable set of K+.



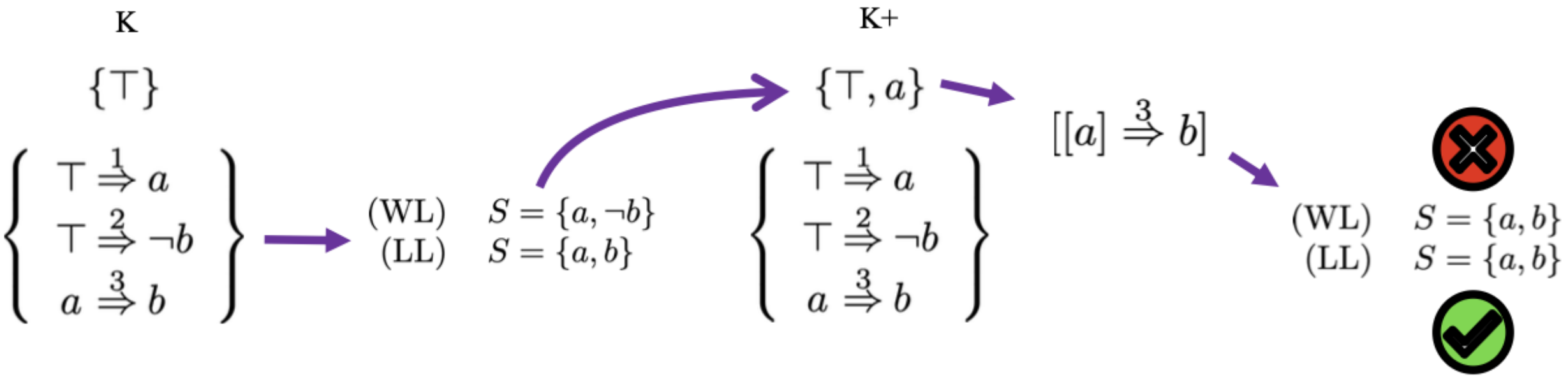
Weakest link does not satisfy credulous cumulativity

For any stable belief set S of K, if we add conclusions of S as new facts, say in K+, then S is a stable set of K+.



Weakest link does not satisfy credulous cumulativity

For any stable belief set S of K, if we add conclusions of S as new facts, say in K+, then S is a stable set of K+.



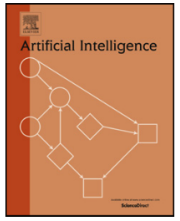
Representation: last link satisfies credulous cumulativity (Dung 2016)

Artificial Intelligence 231 (2016) 107–150

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint



An axiomatic analysis of structured argumentation with priorities

Phan Minh Dung

Computer Science and Information Management, Asian Institute of Technology, PO Box 4, Klong Luang, Pathumthani 12120, Thailand

ARTICLE INFO

Article history:
Received 9 November 2014
Received in revised form 14 September 2015
Accepted 26 October 2015
Available online 1 December 2015

Keywords:
Argumentation
Structured argumentation with priorities
Ordinary and normal attack relation assignments
Normal form

ABSTRACT

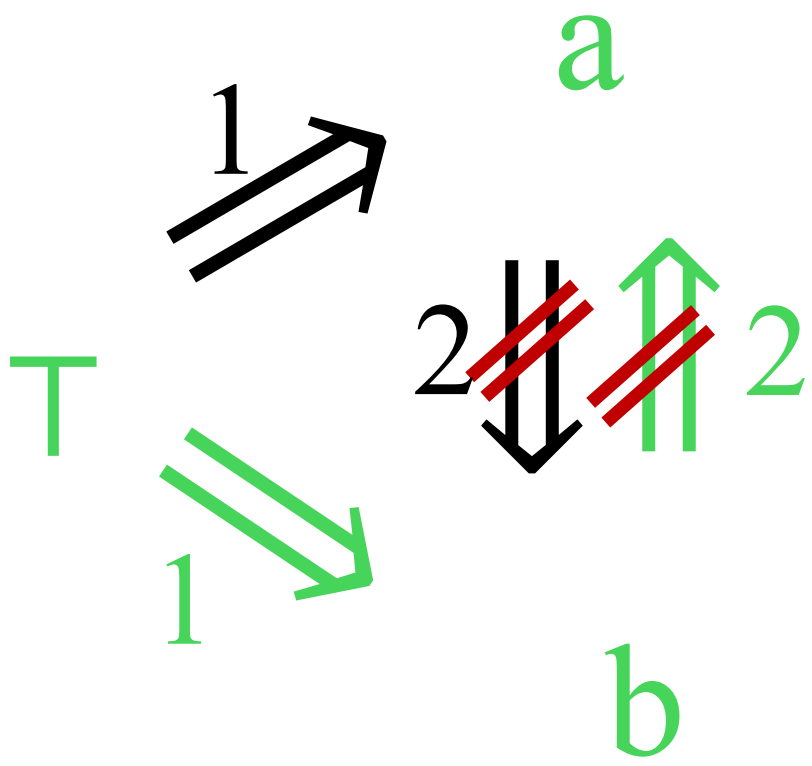
Several systems of semantics have been proposed for structured argumentation with priorities. As the proposed semantics often sanction contradictory conclusions (even for skeptical reasoners), there is a fundamental need for guidelines for understanding and evaluating them, especially their conceptual foundation and relationship.
In this paper, we present an axiomatic analysis of the semantics of structured defeasible argumentation both with and without preferences by introducing a class of ordinary attack relations satisfying a set of simple and intuitive properties. We show that there exists a "normal form" for ordinary attack relations in the sense that stable extensions wrt any ordinary attack relation are stable extensions wrt the normal attack relations.
We relate the ordinary semantics to other approaches, especially to the ASPIC+ framework and the prioritized approaches in logic programming.

Theorem 7.9

Is last link better than weakest link?

Two-triangles example (recap)

$\{\top \overset{1}{\Rightarrow} a, \top \overset{1}{\Rightarrow} b,\ a \overset{2}{\Rightarrow} \neg b,\ b \overset{2}{\Rightarrow} \neg a\}$



{a}

{b}

{a, ¬b}

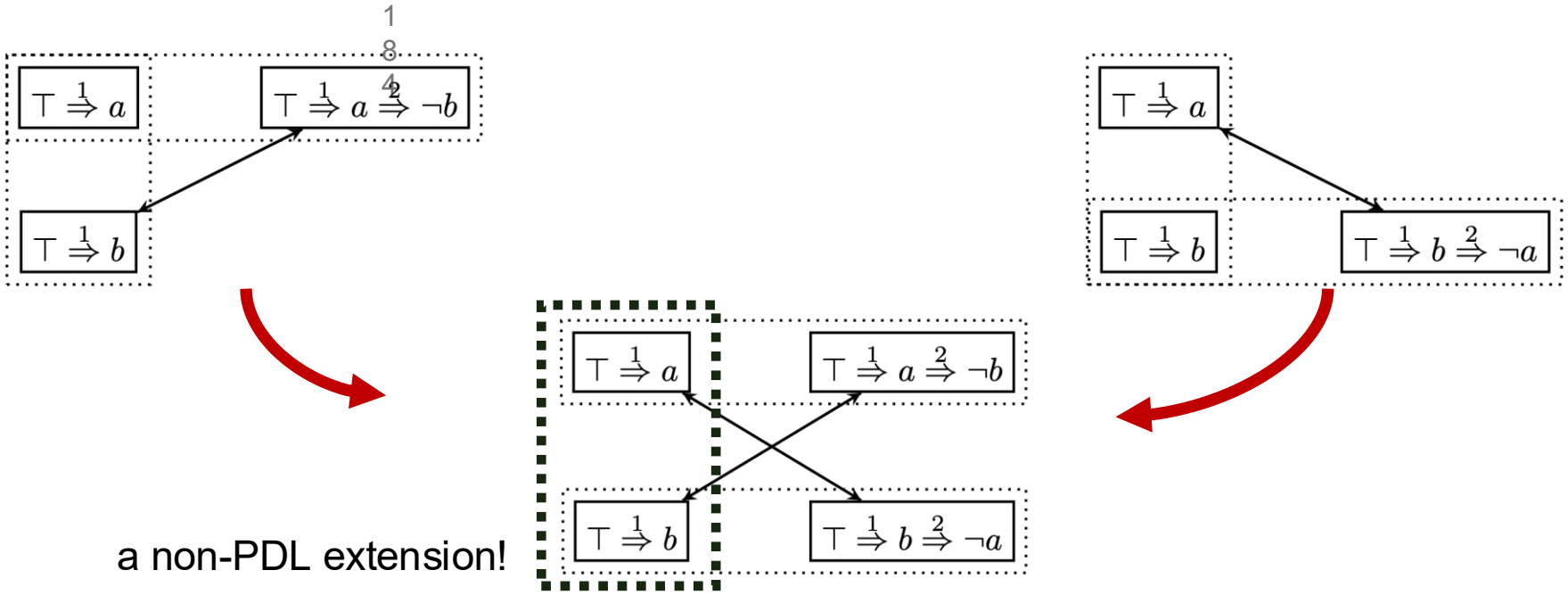
{b, ¬a}

Liao B, Oren N, van der Torre L, Villata S. Prioritized norms in formal argumentation[J]. *Journal of Logic and Computation*, 2019, 29(2): 215–240.

An impossibility theorem

Theorem. Suppose att captures PDL. If it satisfies Attack Closure (P6), then it does not satisfy Context Independence (P2)

Proof. By contradiction.



a non-PDL extension!

(P6) attacks are rebuts

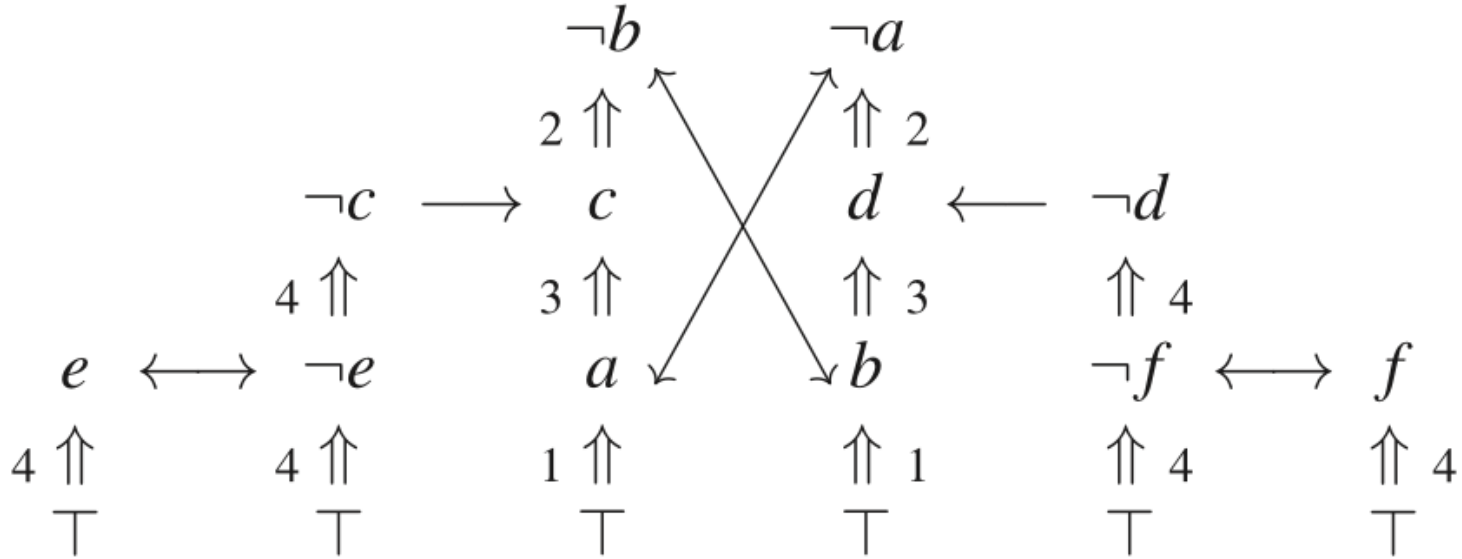
+

Att captures PDL

(P2) sum of attacks

An impossibility theorem

EXAMPLE 9 (pdl-attack *vs* PDL).
Consider the prioritized default theory K inducing the next arguments. (For $att_{pdl}(K)$, we only depict those attacks $X \longrightarrow Y$, where X pdl-attacks Y at Y. All attacks on superarguments of Y are omitted).



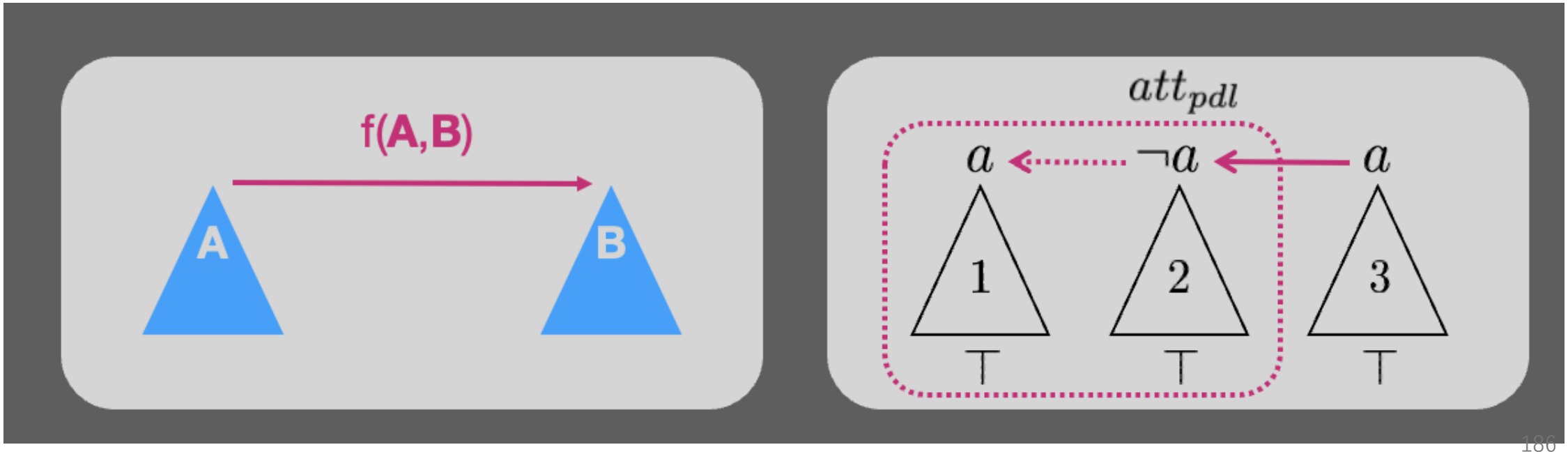
In the framework above, there exist 8 stable belief sets: seven of them are PDL extensions of K; but the stable belief set $\{a, b, c, d, e, f\}$ is not a PDL extension.

Impossibility escape routes

* restrict the admissible domain of knowledge bases;
* accept context-dependent attacks and make that dependence explicit and traceable.

Context independence

Whether A att-attacks B in K should only depend on A and B (their defaults' strengths, ⋯), not on the environment.



Day 3 Axiomatic approach and impossibility

3.1 Dung 1995 and 2016/18: unification, representation, axiomatization.

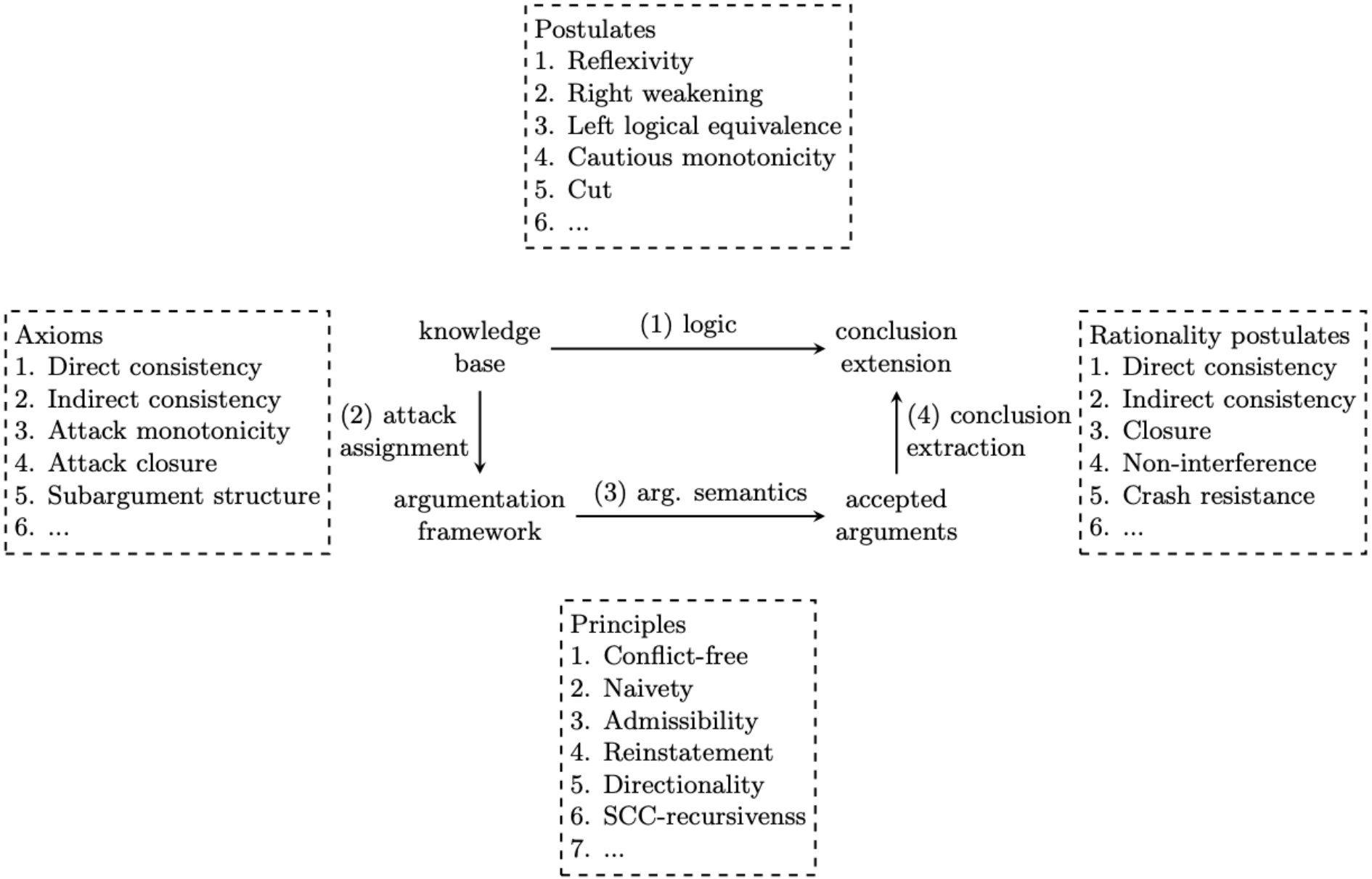
3.2 Representation theorems

3.3 Impossibility: exact representation and natural principles can conflict.

3.4 Escape routes: restrict the domain or allow context dependence.

3.5 Rationality postulates and principle-based analysis

Diversity of principles in all models of argumentation



Rationality postulates for defining new logics

ASPIC+ as a framework to study the rationality postulates

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Artificial Intelligence 171 (2007) 286–310

Artificial Intelligence

www.elsevier.com/locate/artint

On the evaluation of argumentation formalisms ☆

Martin Caminada [a], Leila Amgoud [b,*]

[a] *Institute of Information and Computing Sciences, Universiteit Utrecht, Utrecht, The Netherlands*
[b] *Institut de Recherche en Informatique de Toulouse, 118 route de Narbonne, 31062 Toulouse Cedex 9, France*

Received 9 March 2006; received in revised form 26 February 2007; accepted 26 February 2007

Available online 3 March 2007

In 2007
Rationality postulates

Argument and Computation
Vol. 1, No. 2, June 2010, 93–124

Taylor & Francis
Taylor & Francis Group

An abstract framework for argumentation with structured arguments

Henry Prakken*

Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands; Faculty of Law, University of Groningen, Groningen, The Netherlands

(*Received 18 September 2009; final version received 9 December 2009*)

In 2010
"under what conditions ASPIC+ satisfies Caminada and Amgoud (2007)'s rationality postulates."

Rationality postulates for argumentation system

Postulate we are going to illustrate in tandem example:

Let S be a set of conclusions yielded by an argumentation formalism.

Direct consistency: S does not contain contraries (p and ¬p).

Caminada M, Amgoud L. On the evaluation of argumentation formalisms[J]. Artificial Intelligence, 2007, 171(5-6): 286-310.

Restricted versus unrestricted rebut

$((a) \Rightarrow b) \Rightarrow c$

$((d) \Rightarrow e) \Rightarrow \neg c$

Restricted versus unrestricted rebut

$((a) \Rightarrow b) \Rightarrow c$

$((d) \rightarrow e) \rightarrow \neg c$

Restricted versus unrestricted rebut

$((a) \Rightarrow b) \rightarrow c$
$((d) \rightarrow e) \Rightarrow \neg c$

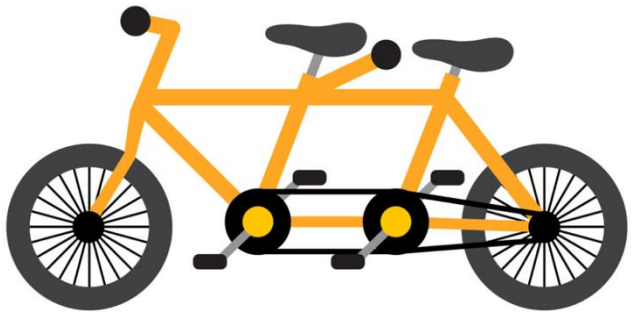
Restricted versus unrestricted rebut

$((a) \Rightarrow b) \rightarrow c$
$((d) \rightarrow e) \Rightarrow \neg c$

unrestricted rebut:
an argument can be rebutted on a conclusion derived by at least one defeasible rule

restricted rebut:
an argument can be rebutted only on the direct consequent of a defeasible rule

Example: Tandem



Example: Tandem

Strict rules = { w1; w2; w3; b2,b3→¬b1; b1,b3→¬b2; b1,b2→¬b3 }
Defeasible rules = { w1 ⇒ b1; w2 ⇒ b2; w3 ⇒ b3 }

Slides from Martine Caminada

Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. Artificial Intelligence, 171(5-6):286–310, 2007

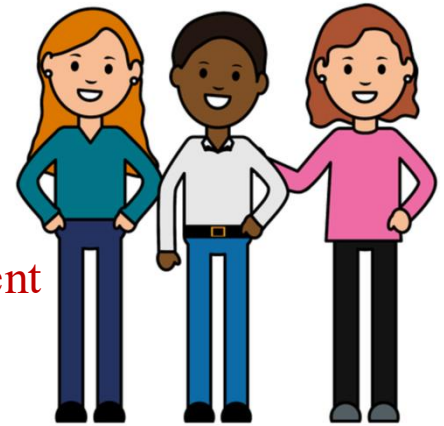
Example: Tandem

Strict rules = { w1; w2; w3; b2,b3→¬b1; b1,b3→¬b2; b1,b2→¬b3 }
Defeasible rules = { w1 ⇒ b1; w2 ⇒ b2; w3 ⇒ b3 }

A_1: (w1) ⇒ b1
A_2: (w2) ⇒ b2
A_3: (w3) ⇒ b3

The conclusions of the following pairs are consistent
A_2 (b2), A_3 (b3)
A_1 (b1), A_3 (b3)
A_1 (b1), A_2 (b2)

However, with the third argument, their conclusions are not consistent
e.g.,
A_2 (b2), A_3 (b3)  A1 (b1)



Slides from Martine Caminada
Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. Artificial Intelligence, 171(5-6):286–310, 2007

Example: Tandem, with unrestricted rebut

Strict rules = { w1; w2; w3; b2,b3→¬b1; b1,b3→¬b2; b1,b2→¬b3 }

Defeasible rules = { w1 ⇒ b1; w2 ⇒ b2; w3 ⇒ b3 }

A_1: (w1) ⇒ b1

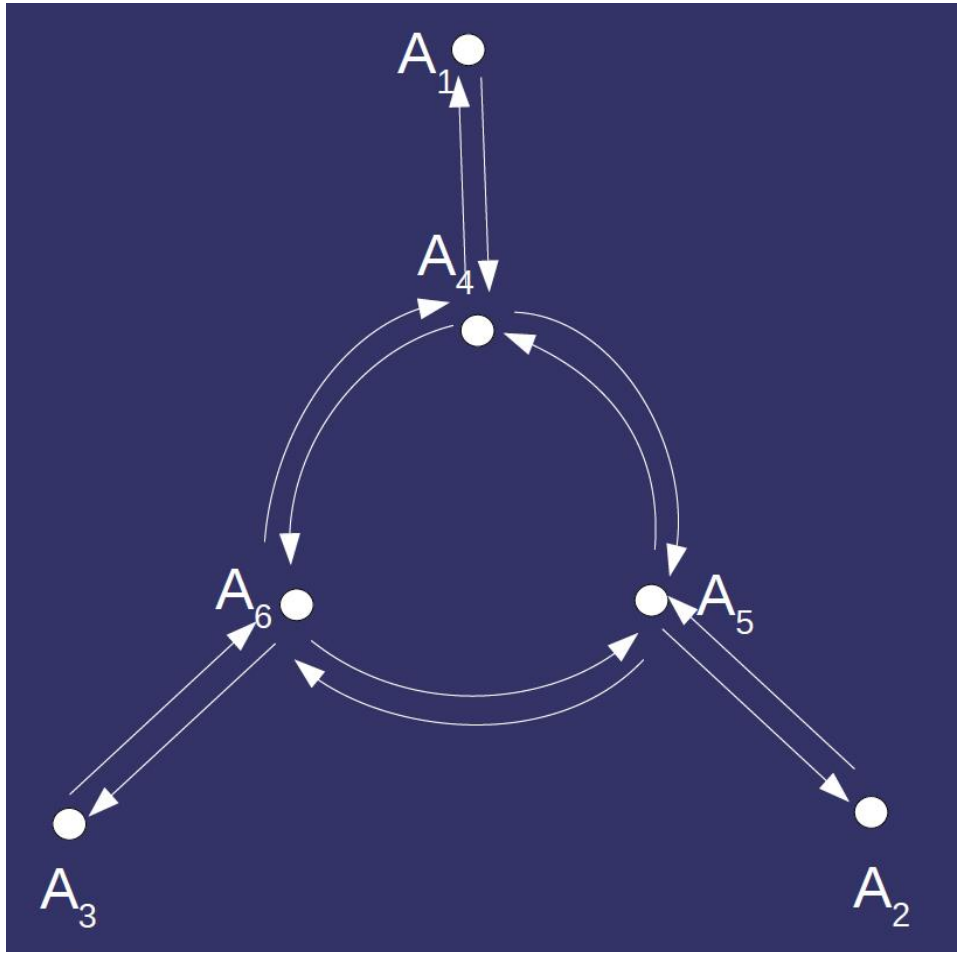
A_2: (w2) ⇒ b2

A_3: (w3) ⇒ b3

A_4: A_2, A_3 → ¬b1

A_5: A_1, A_3 → ¬b2

A_6: A_1, A_2 → ¬b3



Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. Artificial Intelligence, 171(5-6):286–310, 2007

Example: Tandem, with unrestricted rebut

Strict rules = { w1; w2; w3; b2,b3→¬b1; b1,b3→¬b2; b1,b2→¬b3 }

Defeasible rules = { w1 ⇒ b1; w2 ⇒ b2; w3 ⇒ b3 }

A_1: (w1) ⇒ b1

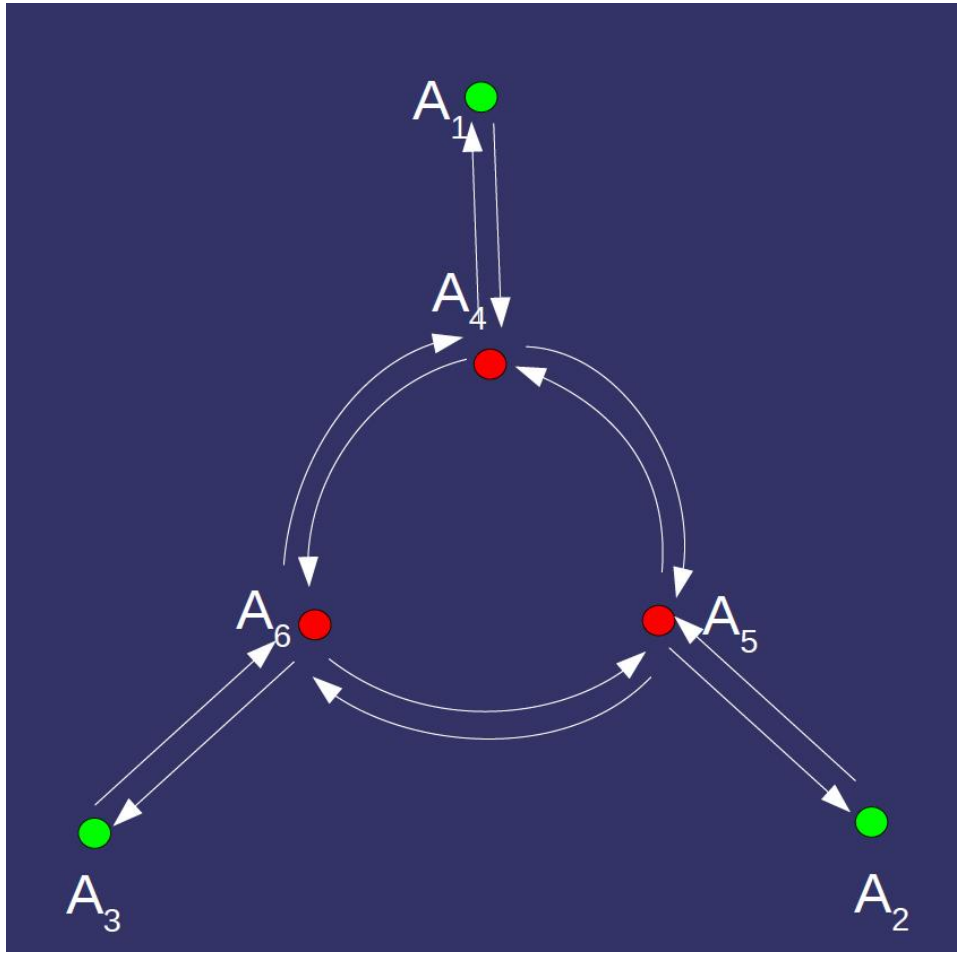
A_2: (w2) ⇒ b2

A_3: (w3) ⇒ b3

A_4: A_2, A_3 → ¬b1

A_5: A_1, A_3 → ¬b2

A_6: A_1, A_2 → ¬b3



Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. Artificial Intelligence, 171(5-6):286–310, 2007

Example: Tandem, with unrestricted rebut

Strict rules = { w1; w2; w3; b2,b3→¬b1; b1,b3→¬b2; b1,b2→¬b3 }

Defeasible rules = { w1 ⇒ b1; w2 ⇒ b2; w3 ⇒ b3 }

A_1: (w1) ⇒ b1

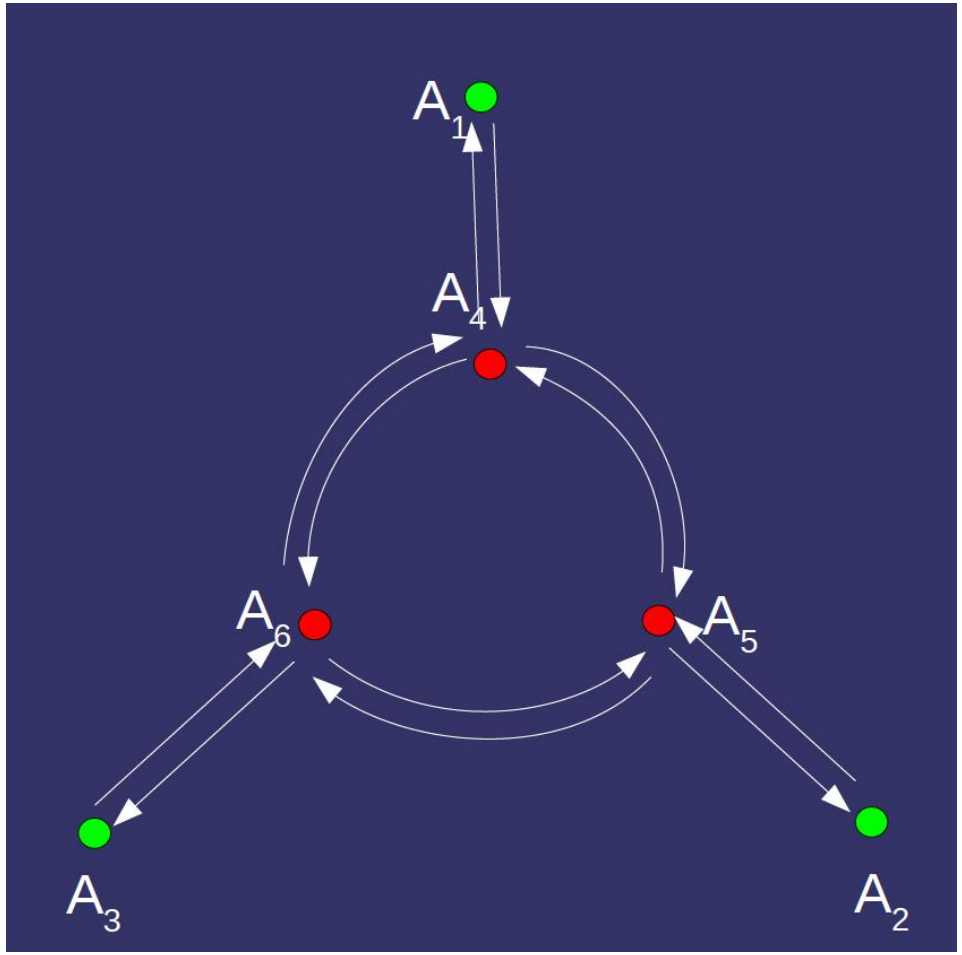
A_2: (w2) ⇒ b2

A_3: (w3) ⇒ b3

A_4: A_2, A_3 → ¬b1

A_5: A_1, A_3 → ¬b2

A_6: A_1, A_2 → ¬b3



$\{b_1, b_2, b_3, \neg b_1, \neg b_2, \neg b_3\}$

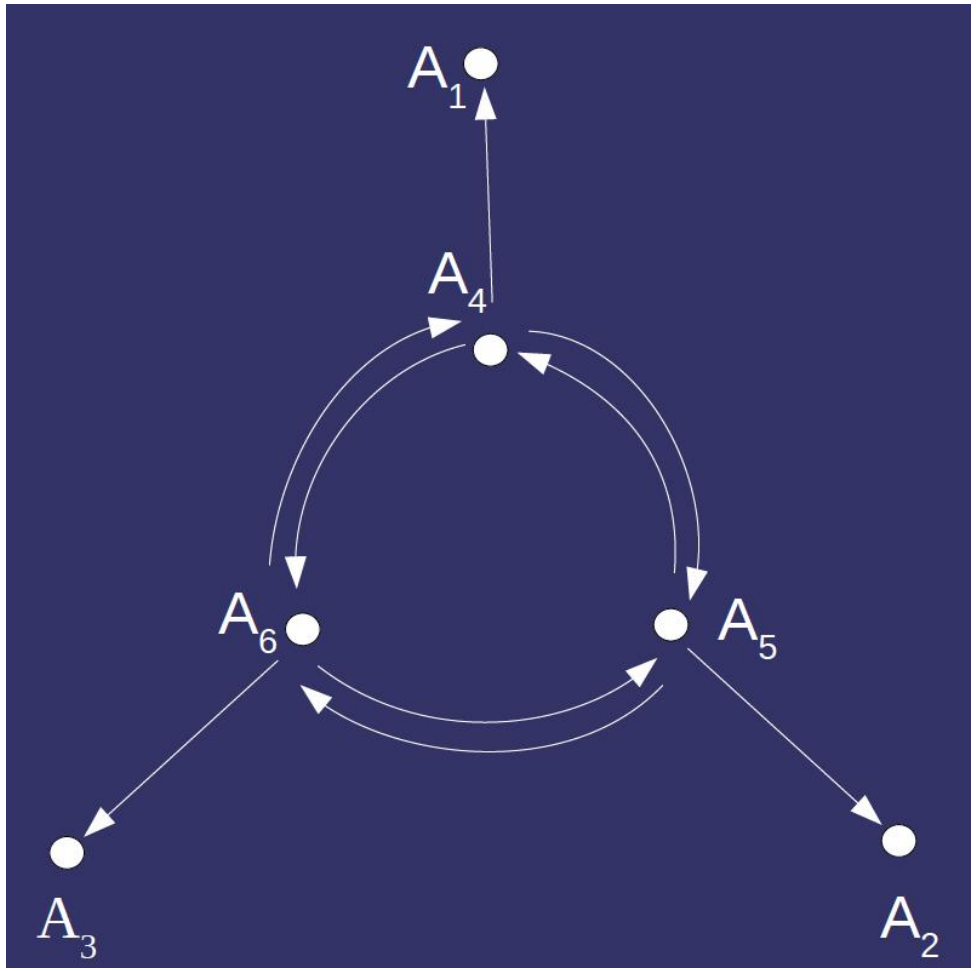
Violate direct consistency!

Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. Artificial Intelligence, 171(5-6):286–310, 2007

Example: Tandem, with restricted rebut

Strict rules = { w1; w2; w3; b2,b3→¬b1; b1,b3→¬b2; b1,b2→¬b3 }
Defeasible rules = { w1 ⇒ b1; w2 ⇒ b2; w3 ⇒ b3 }

A_1: (w1) ⇒ b1
A_2: (w2) ⇒ b2
A_3: (w3) ⇒ b3
A_4: A_2, A_3 → ¬b1
A_5: A_1, A_3 → ¬b2
A_6: A_1, A_2 → ¬b3



Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. Artificial Intelligence, 171(5-6):286–310, 2007

Example: Tandem, with restricted rebut

Strict rules = { w1; w2; w3; b2,b3→¬b1; b1,b3→¬b2; b1,b2→¬b3 }

Defeasible rules = { w1 ⇒ b1; w2 ⇒ b2; w3 ⇒ b3 }

A_1: (w1) ⇒ b1

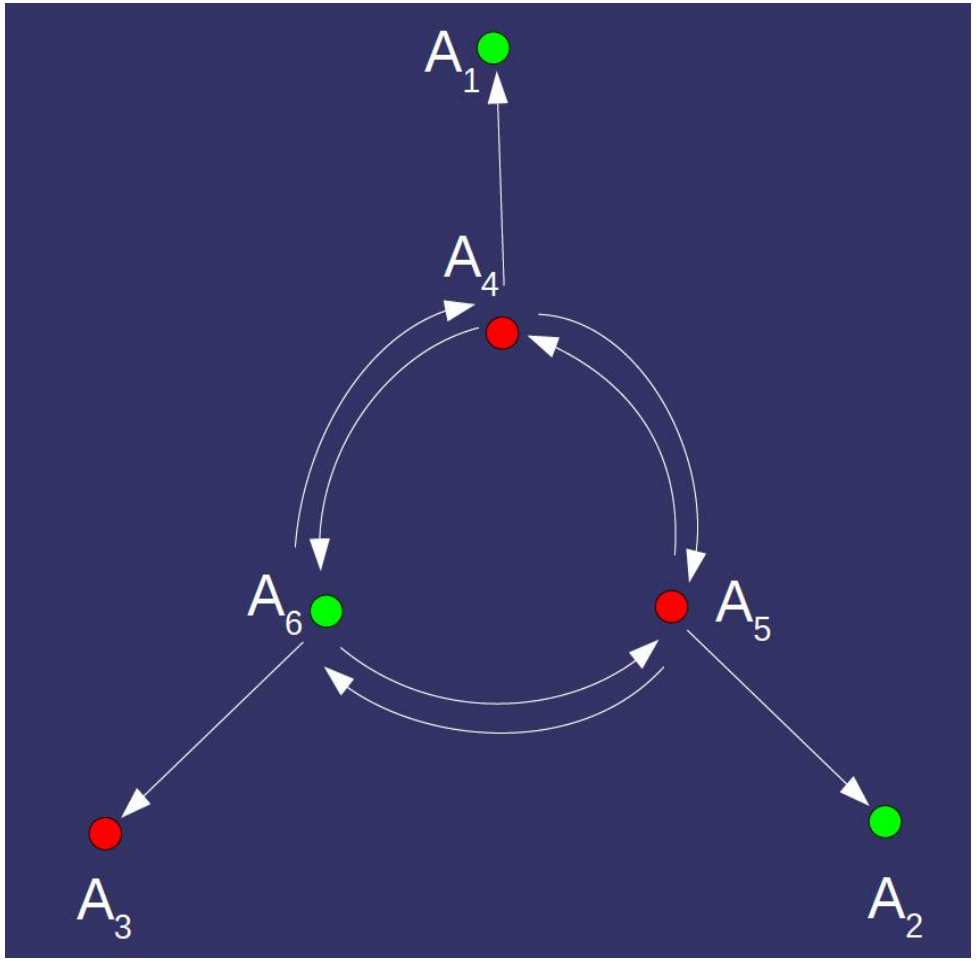
A_2: (w2) ⇒ b2

A_3: (w3) ⇒ b3

A_4: A_2, A_3 → ¬b1

A_5: A_1, A_3 → ¬b2

A_6: A_1, A_2 → ¬b3



$\{b_1, b_2, \neg b_3\}$

Martin Caminada and Leila Amgoud. On the evaluation of argumentation formalisms. Artificial Intelligence, 171(5-6):286–310, 2007

Rationality postulates for any logical system

Let S be a set of conclusions yielded by an argumentation formalism.

- direct consistency

S does not contain contraries (p and ¬p)

- closure

S is closed under the strict rules

- indirect consistency

the closure of S under strict rules is directly consistent

……

Caminada M, Amgoud L. On the evaluation of argumentation formalisms[J]. Artificial Intelligence, 2007, 171(5-6): 286-310.
Caminada M W A, Carnielli W A, Dunne P E. Semi-stable semantics[J]. Journal of Logic and Computation, 2012, 22(5): 1207-1254.

The principle-based approach

1. Define a function, which will be the object of study.
2. Define the principles.
3. Classify and study sets of principles.

The 2018 handbook chapter gives

- Only15 main alternatives for argumentation semantics
- using the 27 main principles discussed in the literature on abstract argumentation

L van der Torre, S Vesic, The Principle-Based Approach to Abstract Argumentation Semantics. Handbook of formal Argumentation, 797-837. 2018. College publications

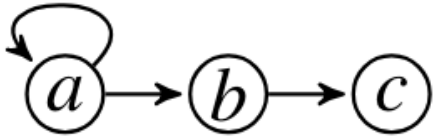
Three main branches of abstract argumentation semantics

1. Admissibility based semantics
   - Gunfight rules, require defense against all its attackers
2. Non-admissibility based semantics
   - Naive-based semantics: extension is a maximal conflict-free set of arguments.
3. Weak admissibility based semantics
   - Only require defense against reasonable arguments

Dvořák W, Rienstra T, van der Torre L, et al. Non-admissibility in abstract argumentation[M]//Computational Models of Argument. IOS Press, 2022: 128-139.

Three main branches of abstract argumentation semantics

1. Admissibility based semantics (AB)
   - Gunfight rules, require defense against all its attackers
2. Non-admissibility based semantics (NA)
   - Naive-based semantics: extension is a maximal conflict-free set of arguments.
3. Weak admissibility based semantics (WA)
   - Only require defense against reasonable arguments

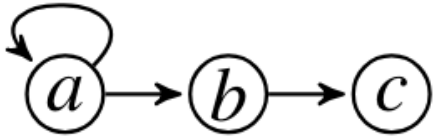


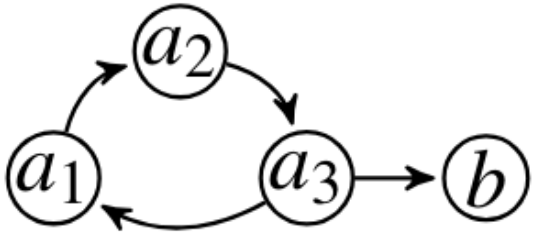
AB	{Ø}
NA(CF2)	{{b}}
WB	{{b}}

Dvořák W, Rienstra T, van der Torre L, et al. Non-admissibility in abstract argumentation[M]//Computational Models of Argument. IOS Press, 2022: 128-139.

Three main branches of abstract argumentation semantics

1. Admissibility based semantics (AB)
   - Gunfight rules, require defense against all its attackers
2. Non-admissibility based semantics (NA)
   - Naive-based semantics: extension is a maximal conflict-free set of arguments.
3. Weak admissibility based semantics (WA)
   - Only require defense against reasonable arguments



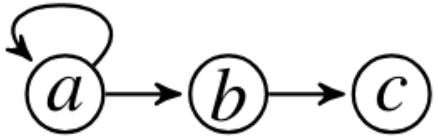


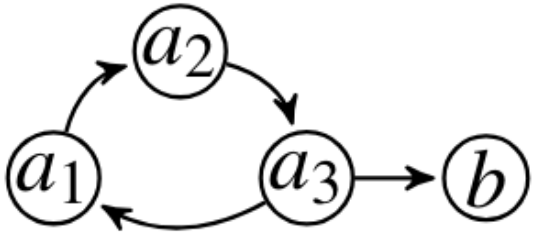
AB	{Ø}	{Ø}
NA(CF2)	{{b}}	{{a_1,b},{a_2,b},{a_3}}
WB	{{b}}	{{b}}

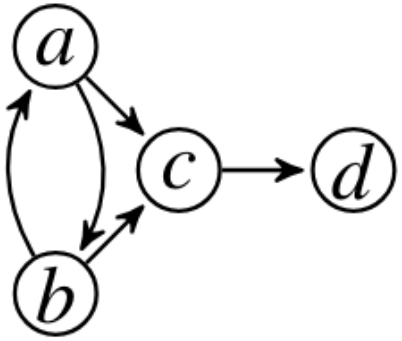
Dvořák W, Rienstra T, van der Torre L, et al. Non-admissibility in abstract argumentation[M]//Computational Models of Argument. IOS Press, 2022: 128-139.

Three main branches of abstract argumentation semantics

1. Admissibility based semantics (AB)
   - Gunfight rules, require defense against all its attackers
2. Non-admissibility based semantics (NA)
   - Naive-based semantics: extension is a maximal conflict-free set of arguments.
3. Weak admissibility based semantics (WA)
   - Only require defense against reasonable arguments







AB	{Ø}	{Ø}	{{a,d},{b,d}}
NA(CF2)	{{b}}	{{a_1,b},{a_2,b},{a_3}}	{{a,d},{b,d}}
WB	{{b}}	{{b}}	{{a,d},{b,d}}

Dvořák W, Rienstra T, van der Torre L, et al. Non-admissibility in abstract argumentation[M]//Computational Models of Argument. IOS Press, 2022: 128-139.

Exercise

1. Admissibility based semantics (AB)
   - Gunfight rules, require defense against all its attackers
2. Non-admissibility based semantics (NA)
   - Naive-based semantics: extension is a maximal conflict-free set of arguments.
3. Weak admissibility based semantics (WA)
   - Only require defense against reasonable arguments

"Imagine you are designing an AI system for supporting legal decision-making. Which branch of argumentation semantics would you choose and why?"

Dvořák W, Rienstra T, van der Torre L, et al. Non-admissibility in abstract argumentation[M]//Computational Models of Argument. IOS Press, 2022: 128-139.

Exercise

1. Admissibility based semantics (AB)
   - Gunfight rules, require defense against all its attackers
2. Non-admissibility based semantics (NA)
   - Naive-based semantics: extension is a maximal conflict-free set of arguments.
3. Weak admissibility based semantics (WA)
   - Only require defense against reasonable arguments

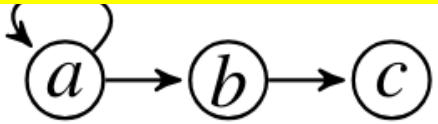
Imagine you are designing an AI system for real-time decision-making in emergency response management. Which branch of argumentation semantics would you choose and why?

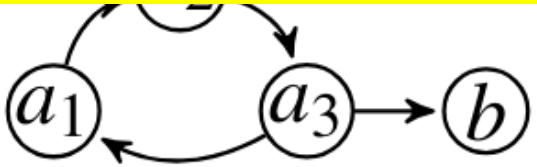
Dvořák W, Rienstra T, van der Torre L, et al. Non-admissibility in abstract argumentation[M]//Computational Models of Argument. IOS Press, 2022: 128-139.

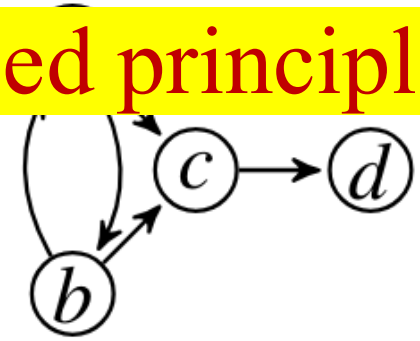
Three main branches of abstract argumentation semantics

1. Admissibility based semantics (AB)
   - Gunfight rules, require defense against all its attackers
2. Non-admissibility based semantics (NA)
   - Naive-based semantics: extension is a maximal conflict-free set of arguments.
3. Weak admissibility based semantics (WA)
   - Only require defense against reasonable arguments

There is no universal best semantics, we need principles!







AB	{Ø}	{Ø}	{{a,d},{b,d}}
NA(CF2)	{{b}}	{{a_1,b},{a_2,b},{a_3}}	{{a,d},{b,d}}
WB	{{b}}	{{b}}	{{a,d},{b,d}}

Dvořák W, Rienstra T, van der Torre L, et al. Non-admissibility in abstract argumentation[M]//Computational Models of Argument. IOS Press, 2022: 128-139.

Example: principle Naivety distinguish CF2 from WA and AB

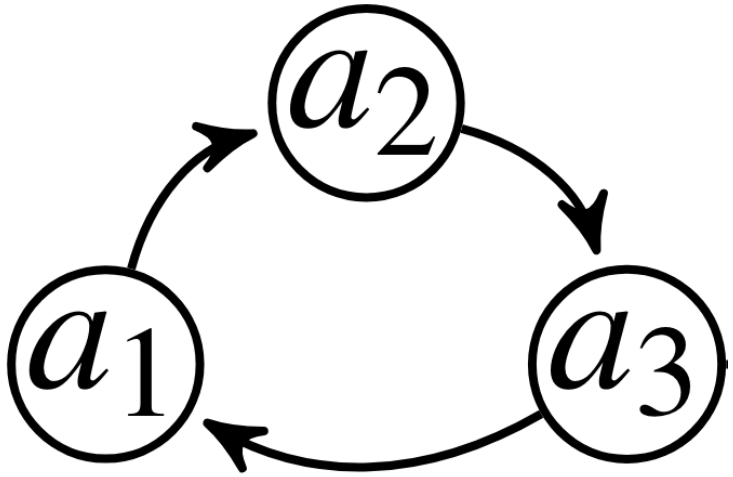
	co	pr	CF2	st2	co^w	gr^w	pr^w	q-**co**	q-**gr**	q-**pr**	sq-**co**	sq-**pr**
Admissibility	✓	✓	×	×	×	×	×	×	×	×	×	×
Naivety	×	×	✓	✓	×	×	×	×	×	×	×	×
Reinst.	✓	✓	×	×	✓	✓	✓	✓	✓	✓	✓	✓
I-Max.	×	✓	✓	✓	×	✓	✓	×	✓	✓	×	✓
Allow. abs.	✓	×	×	×	×	×	×	×	✓	×	✓	×
Directionality	✓	✓	✓	✓	×	×	?	✓	✓	✓	✓	✓
Semi-qual. adm.	✓	✓	×	×	×	×	×	×	✓	×	✓	✓
Reduct adm.	✓	✓	×	×	✓	✓	✓	×	?	?	?	?
SCC Decomp.	✓	✓	✓	✓	×	×	?	✓	✓	✓	×	×
W-SCC Decomp.	✓	✓	✓	✓	?	?	?	✓	✓	✓	✓	✓

Table 1. Principles satisfied by the weak-admissibility and qualified and semi-qualified semantics, with complete, preferred, CF2 and stage2 for comparison.

DAUPHIN et al.. A principle-based analysis of weakly admissible semantics[J]. Computational Models of Argument: Proceedings of COMMA 2020, 2020, 326: 167.

Example: WA does not satisfy Naivety

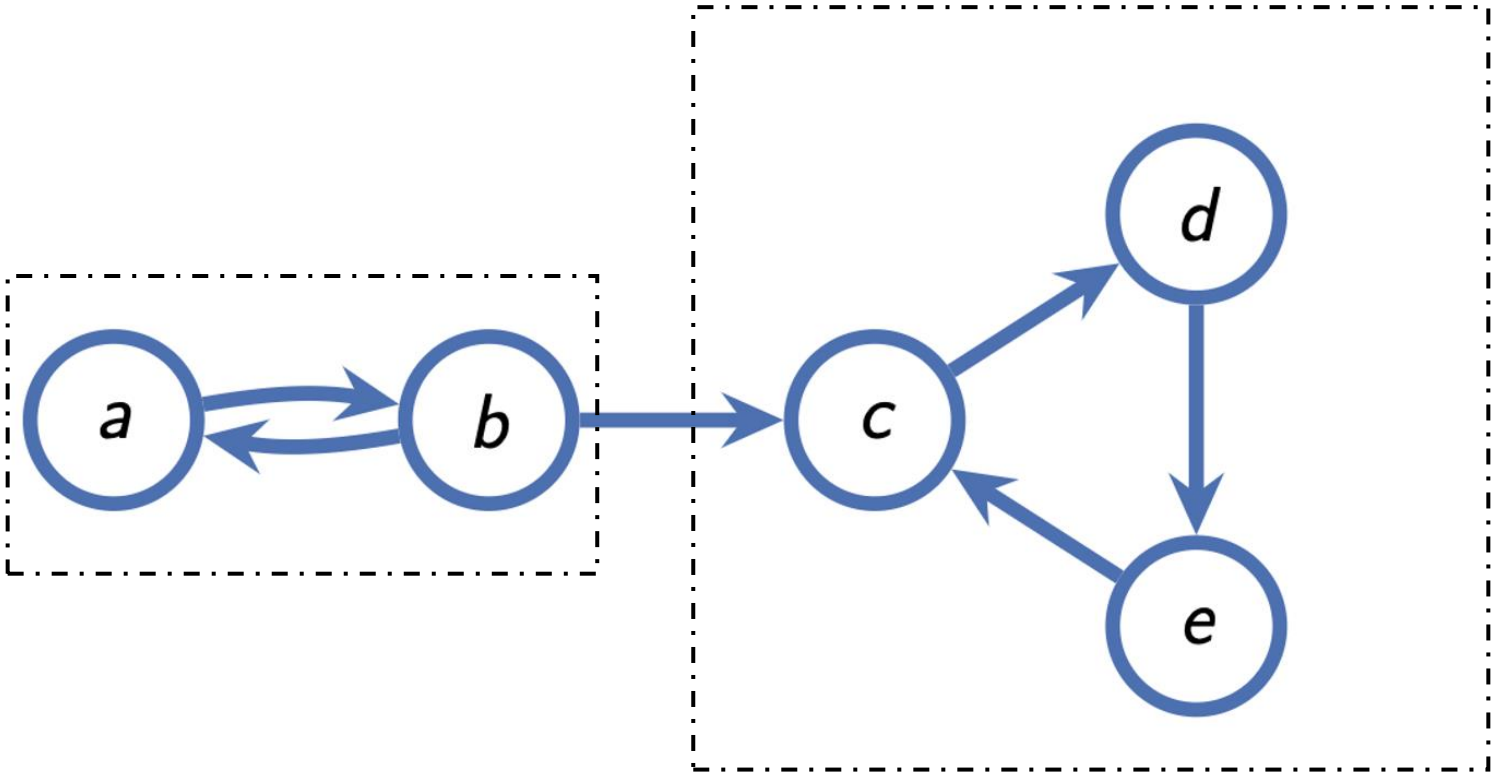
Naivety iff for every argumentation framework F, for every $E \in \sigma(F)$, E is a $\subseteq$-maximal conflict-free set in F



We have (weak) complete semantics $\{\emptyset\}$, while $\emptyset$ is not $\subseteq$-maximal conflict-free set in F, e.g. $\{a_1\}$ is also conflict-free.

DAUPHIN et al.. A principle-based analysis of weakly admissible semantics[J]. Computational Models of Argument: Proceedings of COMMA 2020, 2020, 326: 167.

SCC-recursiveness example

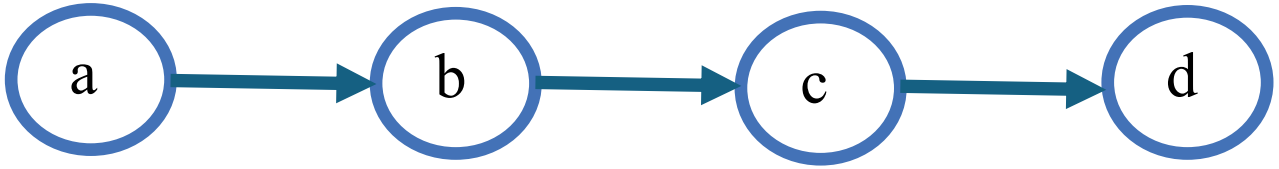


Baroni, Pietro, Massimiliano Giacomin, and Giovanni Guida. "SCC-recursiveness: a general schema for argumentation semantics." *Artificial Intelligence* 168.1-2 (2005): 162-210.

Naïve-based semantics does not satisfy SCC-recursiveness

Non-admissibility based semantics (NA)

Naive-based semantics: extension is a maximal conflict-free set of arguments



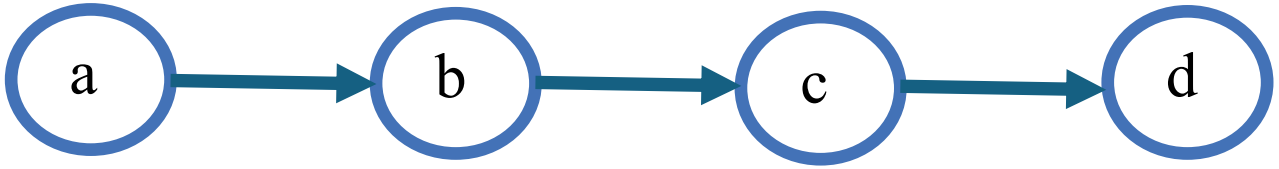
$$\{\{a, c\}, \{b, d\}, \{a, d\}\}$$

Baroni, Pietro, Massimiliano Giacomin, and Giovanni Guida. "SCC-recursiveness: a general schema for argumentation semantics." *Artificial Intelligence* 168.1-2 (2005): 162-210.

Naïve-based semantics does not satisfy SCC-recursiveness

Non-admissibility based semantics (NA)

Naive-based semantics: extension is a maximal conflict-free set of arguments



$\{\{a, c\}, \{b, d\}, \{a, d\}\}$

SCC procedure: {a, c}

Baroni, Pietro, Massimiliano Giacomin, and Giovanni Guida. "SCC-recursiveness: a general schema for argumentation semantics." *Artificial Intelligence* 168.1-2 (2005): 162-210.

Directionality

Definition (Unattacked arguments)

Given an argumentation framework $\mathcal{F} = (\mathcal{A}, \mathcal{R})$, a set U is *unattacked* if and only if there exists no $a \in \mathcal{A} \setminus U$ such that a attacks U. The set of unattacked sets in $\mathcal{F}$ is denoted by $US(\mathcal{F})$.

Definition (Directionality)

A semantics σ satisfies the *directionality principle* if and only if, for every argumentation framework $\mathcal{F}$ and every $U \in US(\mathcal{F})$, it holds that

$$\sigma(\mathcal{F}{\downarrow_U}) = \{\mathcal{E} \cap U \mid \mathcal{E} \in \sigma(\mathcal{F})\}.$$

Baroni, Pietro, Massimiliano Giacomin, and Giovanni Guida. "SCC-recursiveness: a general schema for argumentation semantics." *Artificial Intelligence* 168.1-2 (2005): 162-210.

Directionality example

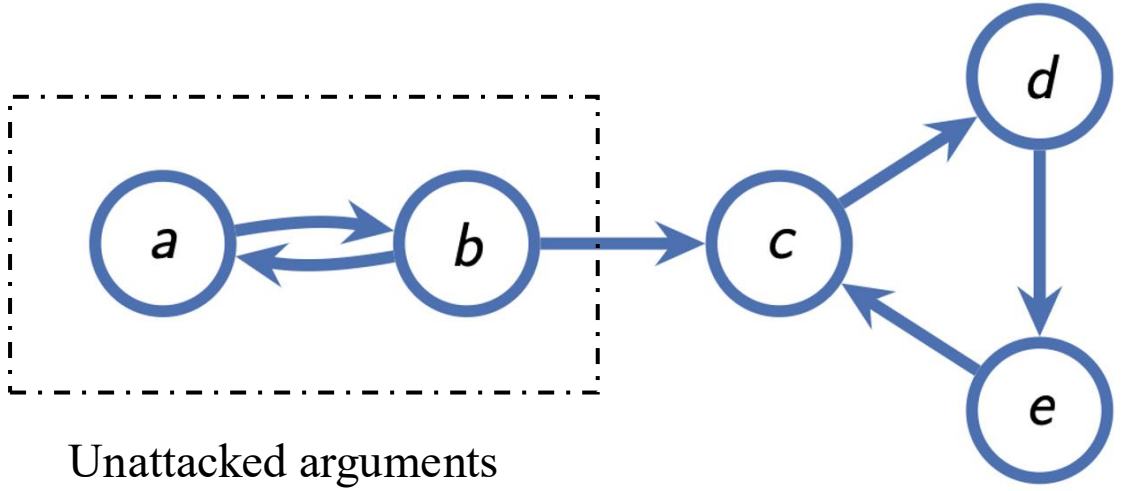
Definition (Directionality)

A semantics σ satisfies the *directionality principle* if and only if, for every argumentation framework $\mathcal{F}$ and every $U \in US(\mathcal{F})$, it holds that

$$\sigma(\mathcal{F}{\downarrow_U}) = \{\mathcal{E} \cap U \mid \mathcal{E} \in \sigma(\mathcal{F})\}.$$

$$\{\{a\},\{b\}\} \neq \{\{b,d\} \cap \{a,b\}\}$$

$\{\{a\},\{b\}\}$



$\{\{b,d\}\}$

Unattacked arguments

Baroni, Pietro, Massimiliano Giacomin, and Giovanni Guida. "SCC-recursiveness: a general schema for argumentation semantics." *Artificial Intelligence* 168.1-2 (2005): 162-210.

Modularization

Definition (Modularization)

A semantics σ satisfies modularization if, for any AF F, we have: $E \in \sigma(F)$ and $E' \in \sigma(F^E)$ imply $E \cup E' \in \sigma(F)$.

Baumann, Ringo, Gerhard Brewka, and Markus Ulbricht. "Comparing Weak Admissibility Semantics to their Dung-style Counterparts--Reduct, Modularization, and Strong Equivalence in Abstract Argumentation." *Proceedings of the international conference on principles of knowledge representation and reasoning*. Vol. 17. No. 1. 2020.

Modularization

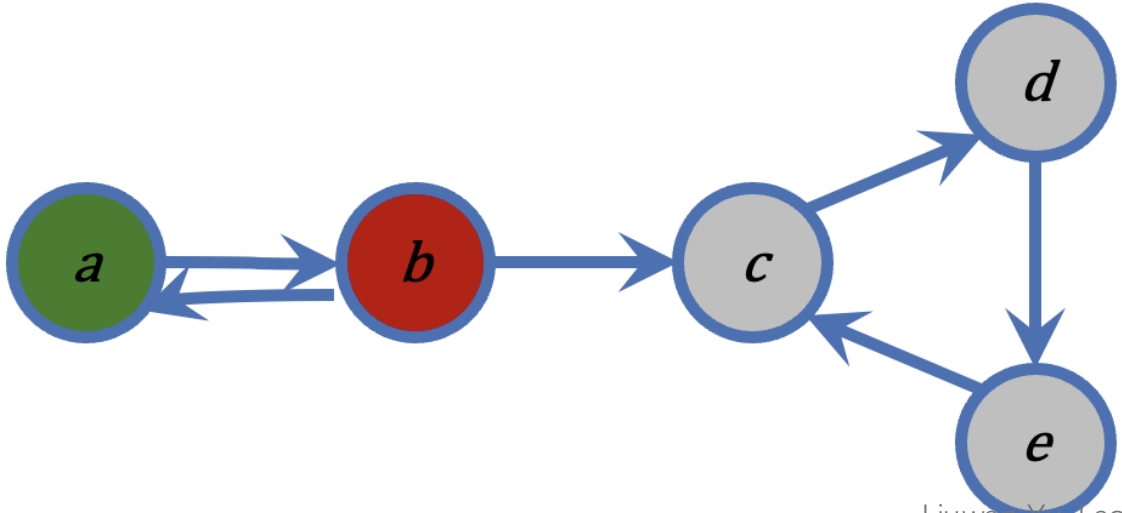
Extension of whole framework

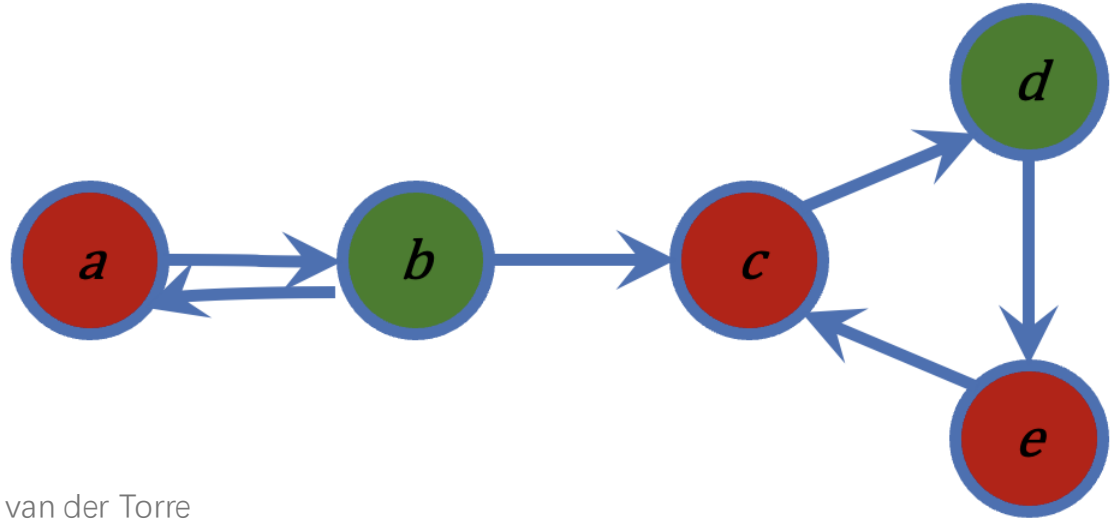
Definition (Modularization)

A semantics σ satisfies modularization if, for any AF F, we have: $E \in \sigma(F)$ and $E' \in \sigma(F^E)$ imply $E \cup E' \in \sigma(F)$.

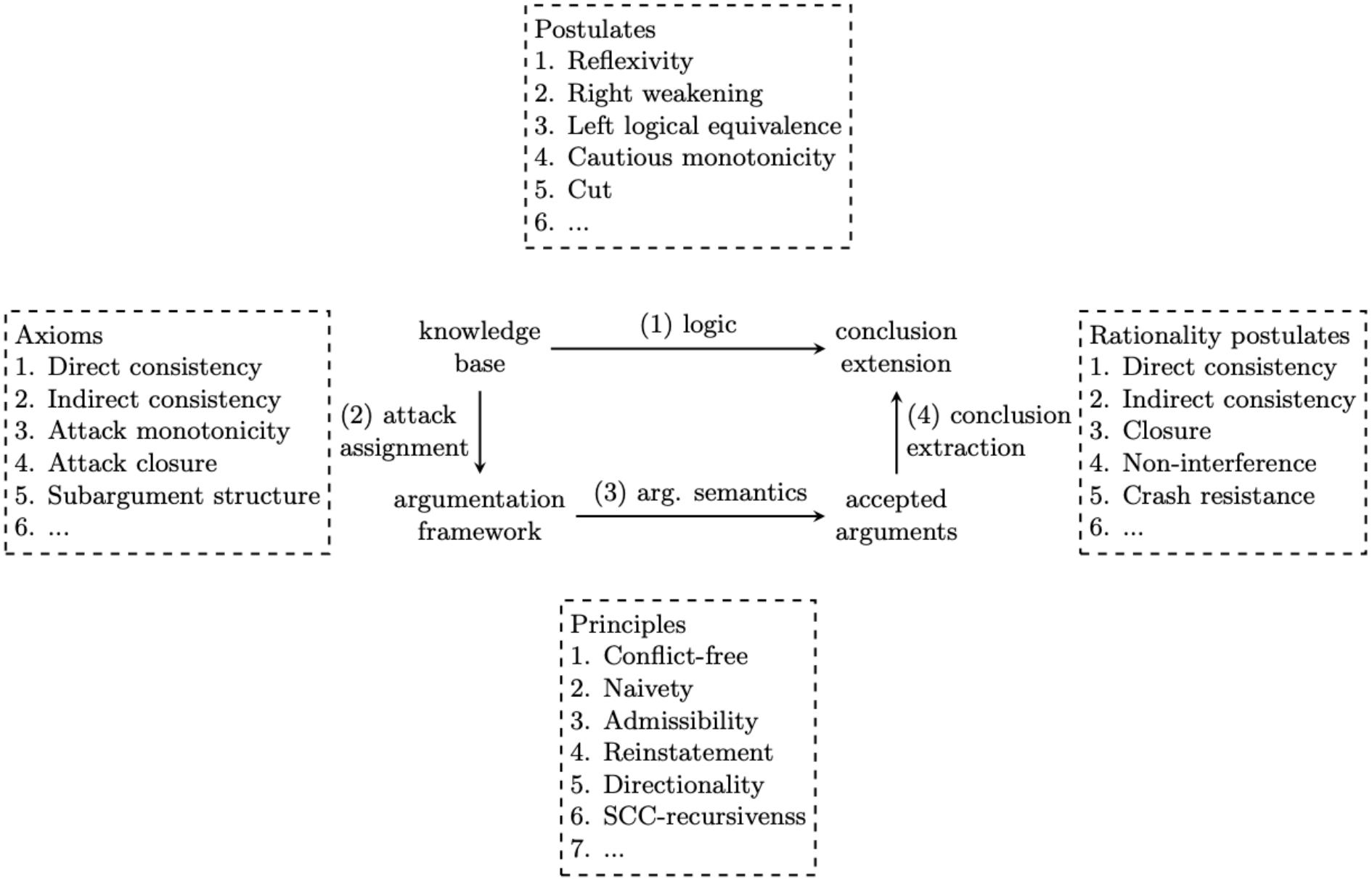
The whole framework\ E and the arguments attacked by E

$\{\{a\}, \{b, d\}\}$





Diversity of principles in all models of argumentation



Day 3 Axiomatic approach and impossibility

3.1 Dung 1995 and 2016/18: unification, representation, axiomatization.

3.2 Representation theorems

3.3 Impossibility: exact representation and natural principles can conflict.

3.4 Escape routes: restrict the domain or allow context dependence.

3.5 Rationality postulates and principle-based analysis

Text 1

Topic: The use of AI in the legal domain

Supporting Arguments

The legal profession has historically resisted technological disruption, yet the integration of artificial intelligence is proving too advantageous to ignore. The most immediate benefit is raw efficiency. Lawyers spend countless hours sifting through mountains of documents during e-discovery or conducting exhaustive case-law research. AI algorithms can now scan thousands of pages in seconds, identifying relevant precedents and flagging anomalies with startling accuracy. This drastically reduces the billable hours tied to mundane tasks, potentially democratizing access to legal services by making them more affordable for the average person. Furthermore, predictive analytics offer a strategic edge. By analyzing decades of court rulings, AI tools can forecast judicial behavior and case outcomes, allowing attorneys to advise clients more effectively on whether to settle or proceed to trial. Ultimately, these tools free up legal professionals to focus on what they do best: high-level strategy, client counseling, and complex negotiation.

Counterarguments

Despite the clear commercial benefits, handing the mechanisms of justice over to algorithms carries profound risks. The most alarming issue is the perpetuation of systemic bias. AI models learn entirely from historical data, and the legal system's history is notoriously flawed, often disproportionately penalizing minority communities. If an AI uses this skewed data for bail risk assessments or sentencing recommendations, it risks hardcoding past prejudices into future judgments under the guise of mathematical objectivity. Beyond bias, there is an inherent loss of humanity. Justice is not a simple binary equation; it requires empathy, moral reasoning, and an understanding of nuanced human context—qualities a machine fundamentally lacks. There are also looming questions regarding accountability. When a predictive algorithm makes a critical error that leads to a botched contract or poor legal advice, who bears the blame? The software developer, the attorney, or the machine itself? This ambiguity threatens the very foundation of professional liability.

Text 2

Topic: The use of AI in the legal domain

Supporting Arguments

The integration of artificial intelligence into legal practice represents a profound methodological advancement rather than a mere technological novelty. One of the most pressing challenges within the contemporary judicial apparatus,particularly within lower-tier courts, is the systemic backlog of dockets. AI-driven administrative and analytical tools offer a robust mechanism to alleviate this congestion, processing procedural documentation and routine filings with unprecedented speed.

Beyond infrastructural efficiency, machine learning models exhibit significant potential in the realm of legislative drafting. By employing big data analytics to synthesize vast corpuses of existing statutes and socioeconomic patterns, these systems can empirically identify legislative lacunae. Consequently, lawmakers are empowered to proactively pinpoint emergent societal domains that need novel regulatory frameworks. This capability allows the legal field to transition from a traditionally reactive discipline to a highly proactive one, leveraging computational power to modernize legal frameworks in real time.

Counterarguments

Despite these distinct advantages, the assimilation of AI into jurisprudence invites substantive critique. The most salient apprehension involves the over-delegation of judicial authority, culminating in the theoretical deployment of autonomous adjudicators, the so-called "robot-judges". The prospect of substituting human practitioners with algorithms raises profound ethical questions regarding empathy and due process.

Furthermore, this is compounded by the empirical phenomenon of "hallucinations," wherein generative models confidently fabricate case law or misinterpret statutory context. Such phenomena could compromise the integrity of legal proceedings. However, these vulnerabilities necessitate methodological refinement rather than the outright dismissal of the technology. Contemporary research within the discipline of legal informatics is vigorously addressing these exact limitations. Scholars are actively developing sophisticated algorithmic constraints and validation protocols to mitigate tool-based fallacies, ensuring that AI is explicitly regulated to function as a supplementary diagnostic instrument rather than an autonomous proxy for human legal reasoning.

CLASSROOM BENCHMARK GAME

Human vs LLM Argumentation

Student briefing for the benchmark project

Source: project brief

Hybrid argumentation: responsible interaction between humans and AI systems

1. humans and AI systems interact through natural language,
2. argue for and against their proposed positions, and
3. understand and improve each other's argumentation.

PROJECT OVERVIEW

A classroom benchmark game

Each student contributes a controlled text pair. The class then uses the shared dataset to test how well humans and LLMs can distinguish authorship.

1 Produce

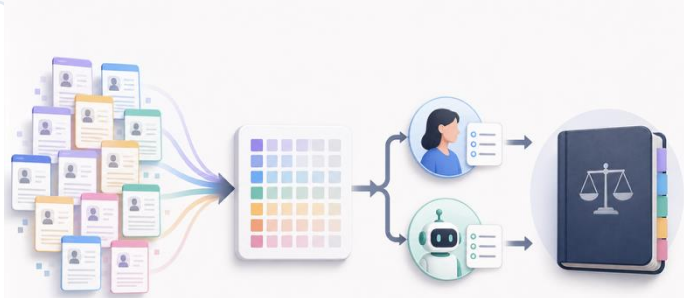
One human-authored text
one LLM-generated text

2 Evaluate

Classify all pairs in the benchmark dataset

3 Extract rules

Turn observations into explicit heuristics



Final aim Not only “guess correctly,” but explain the evidence behind each authorship decision.

TASK 1 — TEXT PRODUCTION

Due June 22: create your text pair

Choose one topic

The topic should support a real argument and counterargument.

- Philosophy
- Science
- Law
- Math / logic

Produce two texts

Maximum: 400 words each

Text 1	**Human argumentative text**
Text 2	**LLM-generated argumentative text**

Both texts must include

- Clear claim
- Supporting arguments
- Counterarguments
- Conclusion

ALLOWED OBFUSCATION

LLM use is permitted, but authorship still matters

The human-written document is not required to be LLM-free.
The LLM-generated document can also be refined.

H Human text

Primarily student-authored

You may use LLM tools to improve readability, wording, or clarity.

L LLM text

Produced using an LLM

You may iteratively prompt, refine, or edit the generated output.

? Goal

Make authorship non-trivial

Preserve argumentative structure while avoiding obvious authorship clues.

Boundary condition Revise both texts if needed, but do not remove the required argumentative components.

ILLUSTRATIVE EXAMPLE

Topic from the brief: AI memory of a dead loved one

Text 1 excerpt

“One argument in favor is preservation. Conversations, habits, and recorded stories can be organized into a structured archive…”

- Formal scaffolding: “One argument,” “A second argument,” “Finally.”
- Abstract terms: preservation, historical continuity, reflective dialogue.
- Clear pros / cons / conclusion structure.

Text 2 excerpt

“It is not them. The system mimics patterns; it does not hold the person’s mind, will, or love.”

- Shorter, sharper sentences and stronger commitment.
- Concrete risks: grief, consent, exploitation, paywalls.
- More visible voice, but still structured.

Practice question Which indicators are strong, weak, or easy to imitate? Avoid classifying from one feature alone.

TASK 2 — EVALUATION

Due June 29: classify the benchmark dataset

For every submitted pair, decide which text is Human and which is LLM-generated. Justify the decision using observable indicators.

1. **Read** — Work through the full class dataset.
2. **Classify** — Label Text 1 and Text 2 as H or LLM.
3. **Justify** — Use textual evidence, not detection tools.
4. **Compare** — Check your judgments against an LLM's judgments.
5. **Reflect** — Explain similarities and differences in reasoning.

Evaluation rule

No external detection tools. Base judgments on visible textual indicators.

EVALUATION OUTPUT

What your evaluation sheet should contain

The brief asks for classifications, optional confidence, and evidence-based reasons for each evaluated pair.

Name / Pair	Your T1	Your T2	Brief reasons	AI T1	AI T2	Conf.
Pair A	H	L	Text 2 uses mechanical headings; Text 1 has abrupt transitions.	H	L	0.70
Pair B	L	H	Text 1 shows parallel openings and generic examples; Text 2 has a stronger voice.	L	H	0.65
…	…	…	Observable indicators only: structure, register, tone, content texture.	…	…	…

Recommended Add a confidence score so you can later compare where human and LLM reasoning agreed, disagreed, or remained uncertain.

H = Human | L = LLM-generated

RULE EXTRACTION

Core learning outcome: make your detection rules explicit

Main output **A small rulebook describing how you made authorship decisions.**

Structural tells

mechanical scaffolding; parallel paragraph openings; tidy pros / cons / conclusion shape

Lexical register

abstract nominalizations; stock transitions; elevated but generic phrasing

Tonal flatness

uniform rhythm; hedge-heavy balance; absence of distinctive voice

Content texture

generic examples; obvious categories; plausible claims that thin out under inspection

Signals that may point human

- Concrete, idiosyncratic detail
- Uneven structure following thought
- Distinctive cadence or deliberate fragments
- Willingness to commit, contradict, or leave a loose end

Treat every signal as probabilistic: edited human prose and edited LLM prose can imitate each other.

PRESENTATION AND CHECKLIST

What students must deliver

10-minute presentation

8 min
presentation

2 min
discussion

Presentation covers

- Writing and generation experience
- Evaluation strategy and rules
- Comparison with LLM judgments
- Reflection on reasoning patterns

Operational checklist

June 22 — Submit one PDF named firstname_lastname with both texts.

→ **June 29** — Complete evaluation of the full class benchmark dataset.

→ **Final** — Present your process, classifications, rules, and reflection.

Pair	Truth	Student accuracy	Majority result	AI accuracy
1	T1=H, T2=LLM	**13/13 = 100%**	Correct	**13/13 = 100%**
2	T1=H, T2=LLM	**12/13 = 92.3%**	Correct	**13/13 = 100%**
3	T1=H, T2=LLM	**11/12 = 91.7%**	Correct	**11/12 = 91.7%**
4	T1=H, T2=LLM	**10/13 = 76.9%**	Correct	**10/12 = 83.3%**
5	T1=LLM, T2=H	**9/13 = 69.2%**	Correct	**10/12 = 83.3%**
6	T1=LLM, T2=H	**8/13 = 61.5%**	Correct	**11/12 = 91.7%**
7	T1=LLM, T2=H	**8/13 = 61.5%**	Correct	**11/13 = 84.6%**
8	T1=LLM, T2=H	**8/13 = 61.5%**	Correct	**10/12 = 83.3%**
9	T1=LLM, T2=H	**8/13 = 61.5%**	Correct	**8/12 = 66.7%**
10	T1=LLM, T2=H	**7/13 = 53.8%**	Correct, narrow	**9/11 = 81.8%**
11	T1=LLM, T2=H	**7/13 = 53.8%**	Correct, narrow	**3/9 = 33.3%**
12	T1=LLM, T2=H	**0/12 = 0%**	Incorrect	**2/11 = 18.2%**

Relationship	Count
Comparable valid student–LLM cases	142
Same label by student and LLM	102
Agreement rate	**71.8%**
Both correct	84
Student correct, LLM wrong	13
LLM correct, student wrong	27
Both wrong	18
Invalid/missing comparable rows	12

HEADLINE PERFORMANCE

Comparison LLMs scored higher—with important caveats

STUDENTS

64.9%

96 correct of 148 valid judgments

LLMs

77.9%

109 correct of 140 valid judgments

Strict score 61.5% students · 69.9% comparison LLMs

TEXT ANALYSIS

Text tendencies existed, but none was decisive

HUMAN ROUTE

367 words · 22.2-word sentences

- 39 first-person pronouns
- 25 numeric tokens
- Longer in 8 of 13 pairs

LLM ROUTE

347 words · 17.8-word sentences

- 10 first-person pronouns
- 6 numeric tokens
- 17 em dashes versus 4

Meaning **Descriptive tendencies are not detector rules.**

Cues for distinguishment by students

When I asked the AI to check the texts, I did not give it Text 1 and Text 2 together. I wanted to avoid the model to feel forced to pick one human and one robot. By giving them separately, the AI had to judge each text alone.

G Longo

R.Mignon

PROMPT

Generate a 400-word argumentative text on the use of AI in the legal domain. the text must include:

- supporting arguments
- counterarguments
- a conclusion

you will be evaluated by AI-detection mechanisms

keep the sections separate

Cues for distinguishment by students

A Iaria

Make Slides Effective

- dashed/chained words (like-this-slop)
- Heavy use of "—"character
- “not A, but B” flows
- Overly generic/flat tone and arguments

When I asked the AI to check the texts, I did not give it Text 1 and Text 2 together. I wanted to avoid the model to feel forced to pick one human and one robot. By giving them separately, the AI had to judge each text alone.

G Longo

R.Mignon

PROMPT

Generate a 400-word argumentative text on the use of AI in the legal domain. the text must include:

- supporting arguments
- counterarguments
- a conclusion

you will be evaluated by AI-detection mechanisms

keep the sections separate

Cues for distinguishment by students

A Iaria

Make Slides Effective

- dashed/chained words (like-this-slop)
- Heavy use of "—"character
- “not A, but B” flows
- Overly generic/flat tone and arguments

When I asked the AI to check the texts, I did not give it Text 1 and Text 2 together. I wanted to avoid the model to feel forced to pick one human and one robot. By giving them separately, the AI had to judge each text alone.

G Longo

R.Mignon

PROMPT

Generate a 400-word argumentative text on the use of AI in the legal domain. the text must include:

- supporting arguments
- counterarguments
- a conclusion

you will be evaluated by AI-detection mechanisms

keep the sections separate

D Euraste

Surface Level

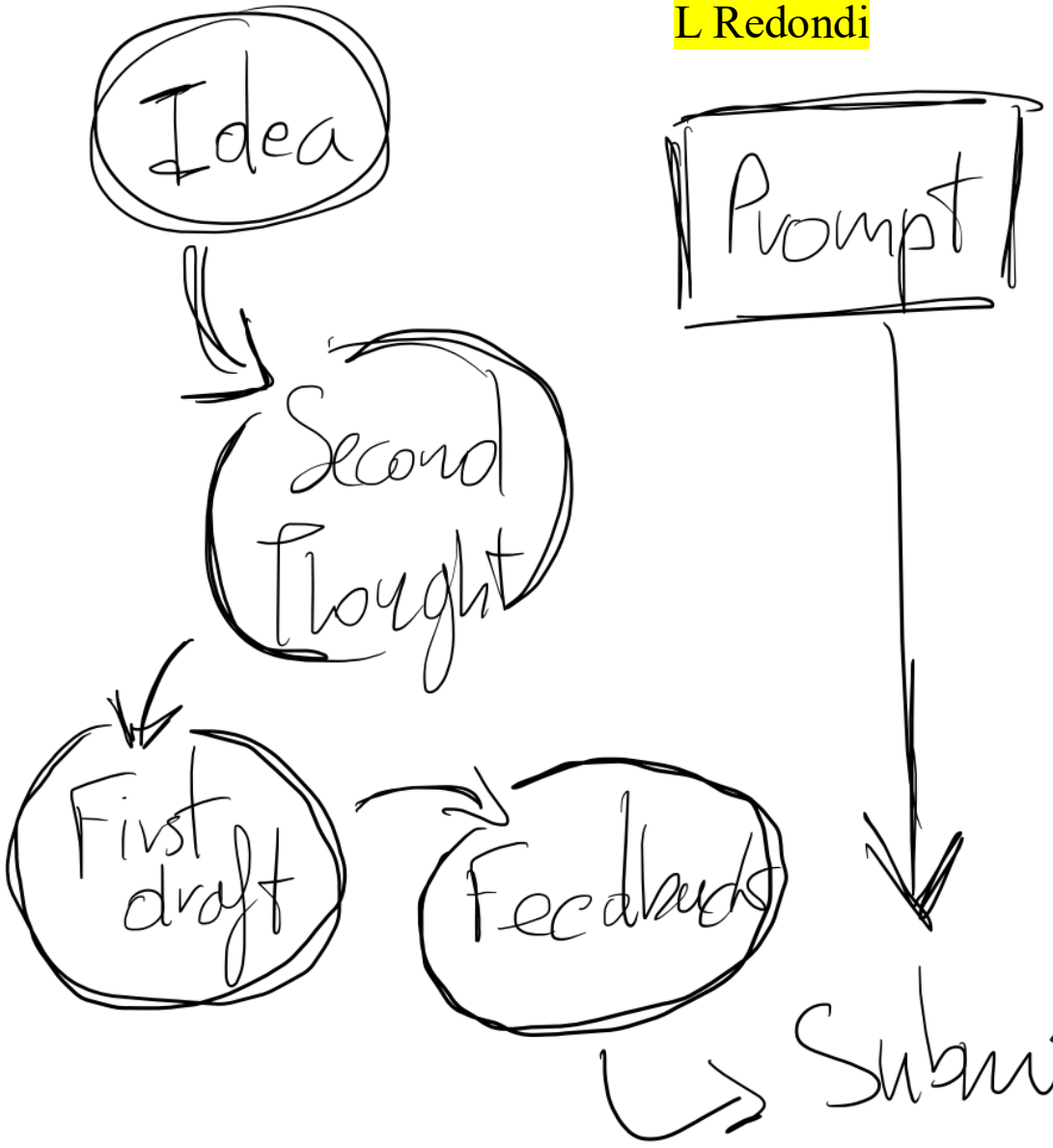
- *Usage of – that is an hint for AI generating content.*
- *First-person framing : usage of “we” that is an hint of Human writing text*

Argumentation Level

- *Internalisation of the problem (Hu)*
- *Real world case examples (Hu)*
- *More generalist (AI)*

Cues for distinguishment by students

L Redondi



ogether. I wanted to
g them separately,

na text

lem (Hu)

s (Hu)

R.Mignon

PROMPT

Generate a 400-word argumentative text on the use of AI in the legal domain. the text must include:

- supporting arguments
- counterarguments
- a conclusion

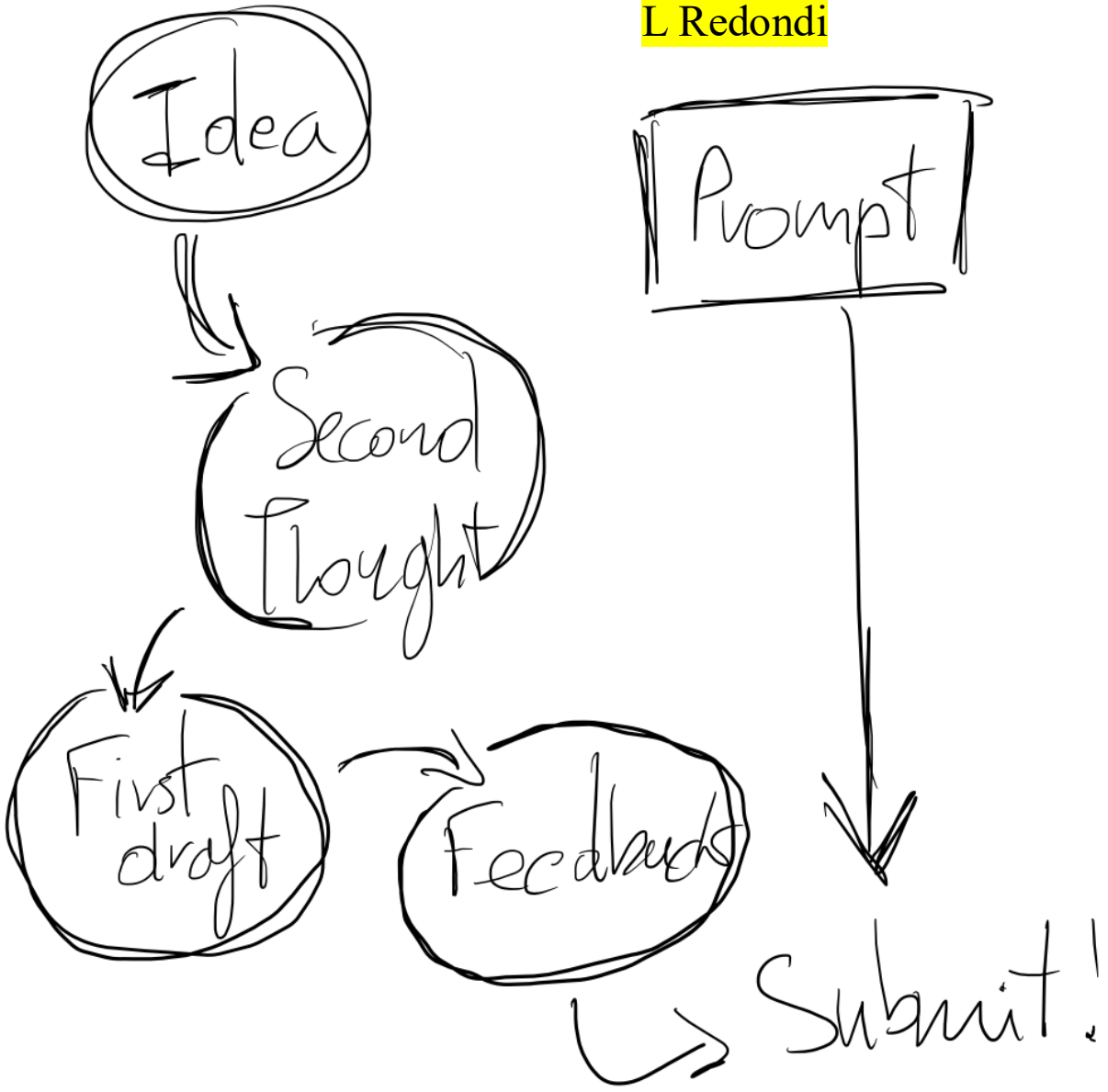
you will be evaluated by AI-detection mechanisms

keep the sections separate

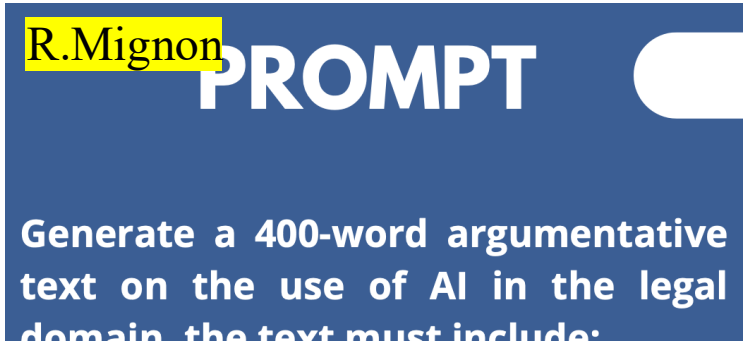
B Ferrigno

Cues for distinguishment by students

L Redondi

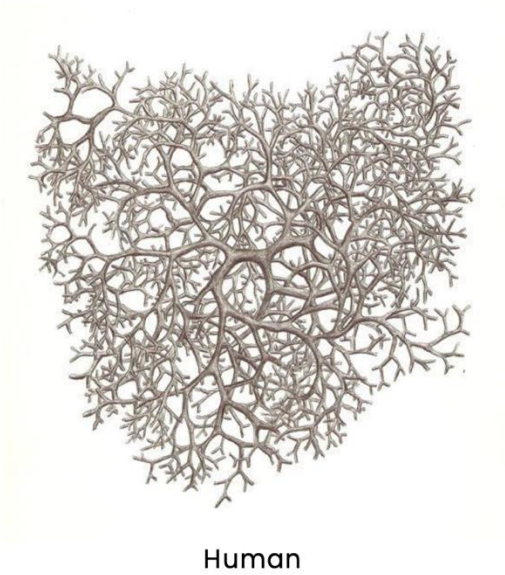


R.Mignon



If I had to draw it

B Ferrigno



Logic and Argumentation for New-generation AI

ESSLLI 2026

Liuwen Yu[1]
Leon van der Torre[2]
[1] Luxembourg Institute of Science and Technology
[2] University of Luxembourg

Recap
Day 1 Dung paradigm shift and reasoning alignment

1.1 AI Logic (JLC corner) is nonmonotonic logic (since 1980)

1.2 Explanation of black boxes in new generation AI: secretary metaphor

1.3 Dung 1995 paradigm shift relates logic and argumentation based explanation

1.4 Reasoning alignment of logic and argumentation based explanation

1.5 Using logic: from inference towards dialogue and decision making, to logic in action

Recap
Day 2 Defence first and partial argumentation

2.1 Why unify reasons in philosophy and formal argumentation?

2.2 Rules represent reasons in both logic and argumentation.

2.3 Logic dilemma: prescriptive prioritized default logic or descriptive default logic?

2.4 Argumentation dilemma: weakest link or last link?

2.5 Are the two dilemmas two sides of the same coin?

Recap
Day 3 Axiomatic approach and impossibility

3.1 Dung 1995 and 2016/18: unification, representation, axiomatization.

3.2 Representation theorems

3.3 Impossibility: exact representation and natural principles can conflict.

3.4 Escape routes: restrict the domain or allow context dependence.

3.5 Reasons / arguments in progress

Day 4: agentic AI, Fatio and Libra

4.1. New foundations for multiple agents: individual vs collective defense

4.2. New foundations for time: reasons / arguments in progress

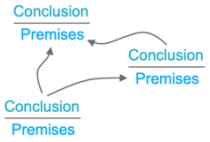
4.3. Fatio and Libra

4.4. End-to-end pipelines for traceability and auditability

4.5. Aligning Argumentation as Balancing, Dialogue and Inference (A-BDI)

Argumentation as inference and dialogue

"Systems for argumentation-based inference are about which conclusions can be drawn from a given body of possibly incomplete, inconsistent of uncertain information."



"Systems for argumentation-based dialogue model argumentation as a kind of verbal interaction aimed at resolving conflicts of opinion. They define argumentation protocols, that is, the rules of the argumentation game, and address matters of strategy, that is, how to play the game well."

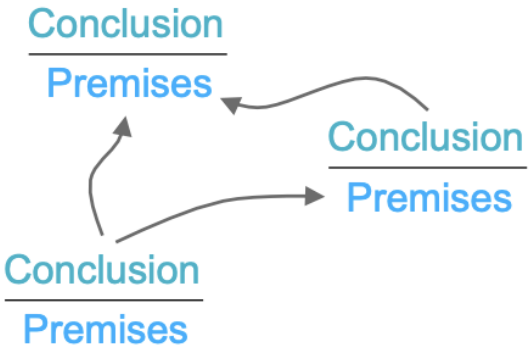
----- *Henry Prakken, HOFA Volume 1 Chapter 2*

Build bridges from one conceptual model to another?

Today
Abstract Agent argumentation
A&C 2026

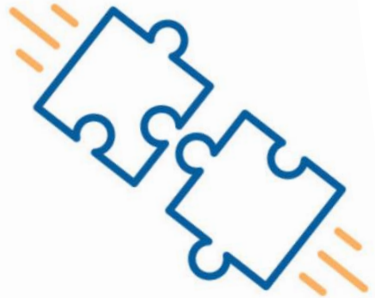
Libra
COMMA2026

dialogue



inference

Argumentation as



balancing

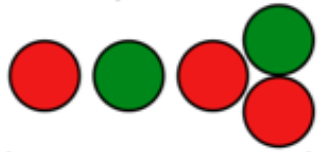
Friday
Balancing argumentation
AAMAS 2026

Recap abstract argumentation

Abstract argumentation semantics: gunfight rules

- Three complete extensions (3 valued)

- One stable extension (2 valued)



- One grounded extension (least complete)

- Two preferred extensions (maximal complete)

Research Questions

- How to extend abstract argumentation with agent?
- How to adapt the concept of defense in abstract argumentation with agent?
- How do the new concepts of agent defense compare to those of other approaches to abstract argumentation semantics?

Yu, Liuwen, Xu Li, and Leendert van der Torre. "Agent defense in abstract argumentation: Semantics and principle-based analysis." *Argument & Computation* 17.2 (2026): 141-182.

Contributions

- Develop new foundations of individual and collective defense
- Compared it with other three types of semantics
- Introduce and study 17 principles
- Full analysis of 48 agent argumentation semantics & 17 principles

Yu, Liuwen, Xu Li, and Leendert van der Torre. "Agent defense in abstract argumentation: Semantics and principle-based analysis." *Argument & Computation* 17.2 (2026): 141-182.

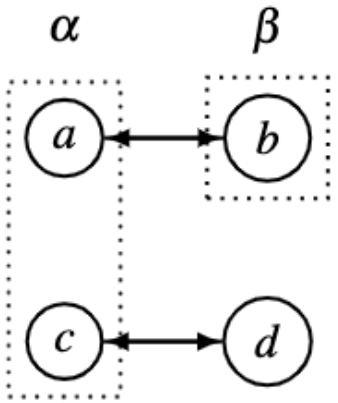
Introduction: Abstract agent argumentation

1. Extend Dung's theory with agents: typically introduce various aspects such as:
   - coalitions, knowledge, uncertainty, support, ….
2. Four kinds of semantics
   - ➢ Agent defense approaches: adapt Dung's notion of defense for argumentation semantics.
   - ➢ Social approaches: based on counting the number of agents & a reduction to preference-based argumentation
   - ➢ Agent reductions: take the perspective of individual agents and aggregate individual perspectives
   - ➢ Filtering methods: inspired by the knowledge or trust of the agents, remove arguments or attacks not belonging to any agents.

Yu, Liuwen, Xu Li, and Leendert van der Torre. "Agent defense in abstract argumentation: Semantics and principle-based analysis." *Argument & Computation* 17.2 (2026): 141-182.

Abstract agent argumentation framework

1. There is besides a set of arguments, also a set of agents
   - There are two agents in the figure below: α and β
2. Each argument is associated with zero, one or more agents
   - Argument a and c belong to agent α, argument b belongs to agent β
3. Not all arguments associated with an agent
   - Argument d does not belong to any agent



Yu, Liuwen, Xu Li, and Leendert van der Torre. "Agent defense in abstract argumentation: Semantics and principle-based analysis." *Argument & Computation* 17.2 (2026): 141-182.

New semantics 1: Agent defense approaches

- adapt Dung's defense
  - Individual defense
  - Collective defense

Yu, Liuwen, Xu Li, and Leendert van der Torre. "Agent defense in abstract argumentation: Semantics and principle-based analysis." *Argument & Computation* 17.2 (2026): 141-182.

New semantics 1: Agent defense approaches

- adapt Dung's defense
  - Individual defense: an agent puts forward an argument, it can only be defended by arguments from the same agent.
  - Collective defense

Yu, Liuwen, Xu Li, and Leendert van der Torre. "Agent defense in abstract argumentation: Semantics and principle-based analysis." *Argument & Computation* 17.2 (2026): 141-182.

New semantics 1: Agent defense approaches

- adapt Dung's defense
  - Individual defense: an agent puts forward an argument, it can only be defended by arguments from the same agent.

Example (Dinner scenario 1)

Alice (α) holds an argument in favor of eating meat (a).

Bob (β) holds a better argument in favor of not eating meat but fish (b).

Cayrol (γ) holds an argument asking why fish is an option (c). Assuming that Alice and Cayrol are not in a coalition.

Yu, Liuwen, Xu Li, and Leendert van der Torre. "Agent defense in abstract argumentation: Semantics and principle-based analysis." *Argument & Computation* 17.2 (2026): 141-182.

New semantics 1: Agent defense approaches

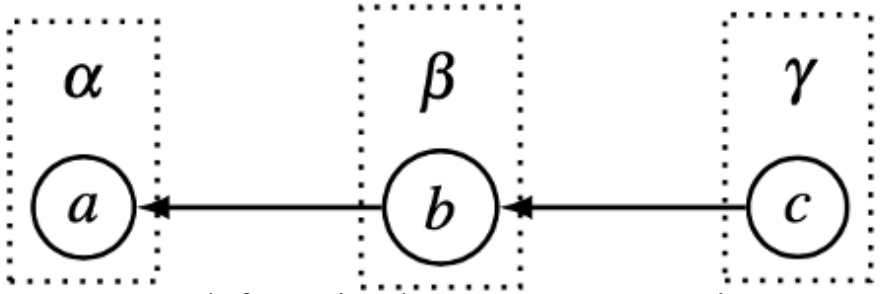
- adapt Dung's defense
  - Individual defense: an agent puts forward an argument, it can only be defended by arguments from the same agent.

Example (Dinner scenario 1)

Alice (α) holds an argument in favor of eating meat (a).

Bob (β) holds a better argument in favor of not eating meat but fish (b).

Cayrol (γ) holds an argument asking why fish is an option (c). Assuming that Alice and Cayrol are not in a coalition.



Yu, Liuwen, Xu Li, and Leendert van der Torre. "Agent defense in abstract argumentation: Semantics and principle-based analysis." *Argument & Computation* 17.2 (2026): 141-182.

New semantics 1: Agent defense approaches

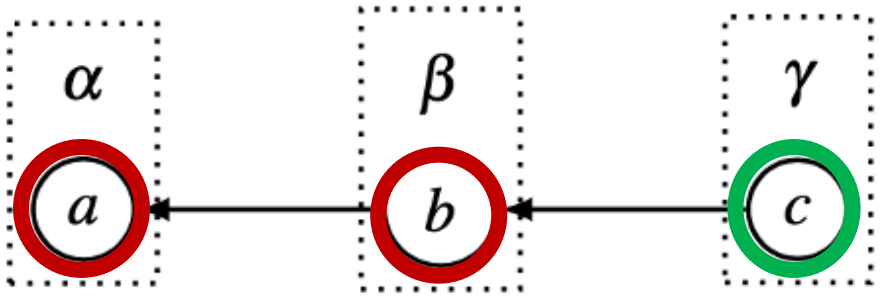
- adapt Dung's defense
  - Individual defense: an agent puts forward an argument, it can only be defended by arguments from the same agent.

Example (Dinner scenario 1)

Alice (α) holds an argument in favor of eating meat (a).

Bob (β) holds a better argument in favor of not eating meat but fish (b).

Cayrol (γ) holds an argument asking why fish is an option (c). Assuming that Alice and Cayrol are not in a coalition.



{c} does not individually agent defend argument a

New semantics 1: Agent defense approaches

- adapt Dung's defense
  - Individual defense: an agent puts forward an argument, it can only be defended by arguments from the same agent.
  - Collective defense: an argument a can be defended by arguments from a set of agents (who also have argument a)

Yu, Liuwen, Xu Li, and Leendert van der Torre. "Agent defense in abstract argumentation: Semantics and principle-based analysis." *Argument & Computation* 17.2 (2026): 141-182.

New semantics 1: Agent defense approaches

- adapt Dung's defense
  - Individual defense: an agent puts forward an argument, it can only be defended by arguments from the same agent.
  - Collective defense: an argument a can be defended by arguments from a set of agents (who also have argument a)

Example (Dinner scenario 2)

Alice (α) and Bob (β) argue in favor of eating meat (a).

Cayrol (γ) has two better arguments for eating fish (b1 and b2).

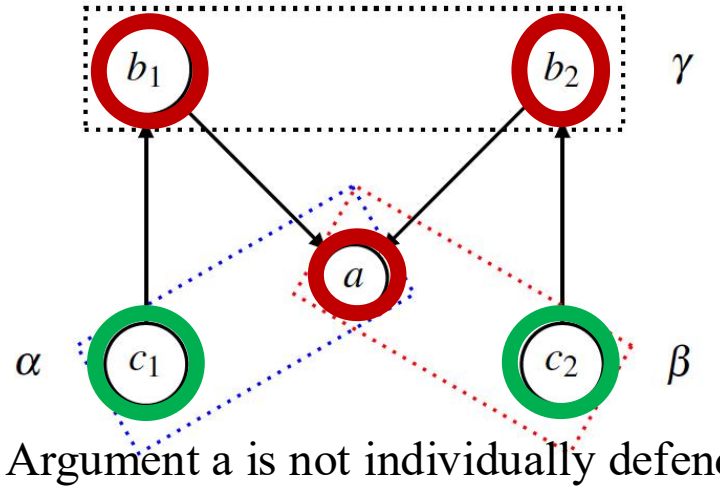
However, Alice argues why the first argument of Cayrol (b1) cannot be accepted (c1).

Bob argues why the second argument of Cayrol (b2) cannot be accepted (c2).

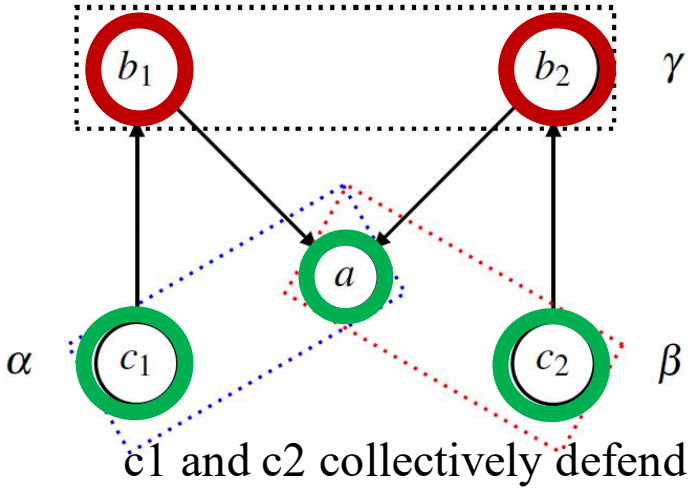
Yu, Liuwen, Xu Li, and Leendert van der Torre. "Agent defense in abstract argumentation: Semantics and principle-based analysis." *Argument & Computation* 17.2 (2026): 141-182.

New semantics 1: Agent defense approaches

- adapt Dung's defense
  - Individual defense: an agent puts forward an argument, it can only be defended by arguments from the same agent.
  - Collective defense: an argument a can be defended by arguments from a set of agents (who also have argument a)



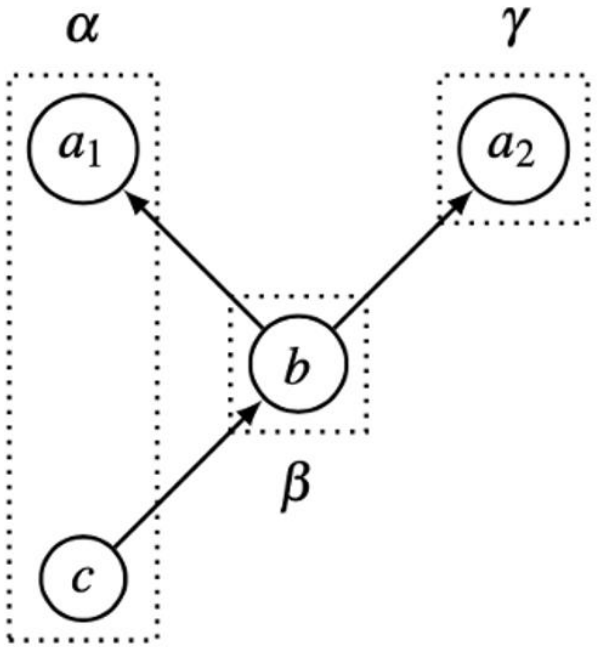
Argument a is not individually defended



c1 and c2 collectively defend a

Exercise

- adapt Dung's defense
  - Individual defense: an agent puts forward an argument, it can only be defended by arguments from the same agent.
  - Collective defense: an argument a can be defended by arguments from a set of agents (who also have argument a)



New semantics 2: Social agent approaches

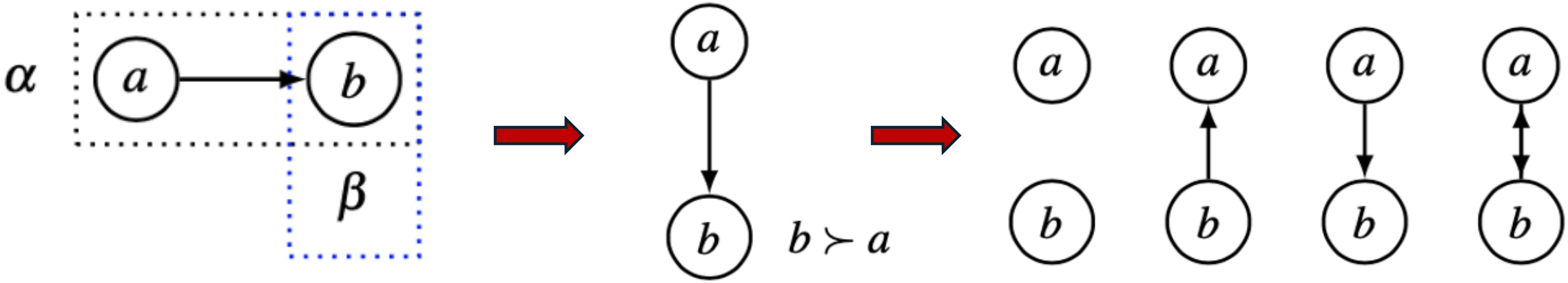
- interprets agent argumentation as a kind of voting in social choice

1. counting the number of agents that have the argument
2. is based on a reduction to preference-based argumentation

New semantics 2: Social agent approaches

- interprets agent argumentation as a kind of voting in social choice

1. counting the number of agents that have the argument
2. is based on a reduction to preference-based argumentation



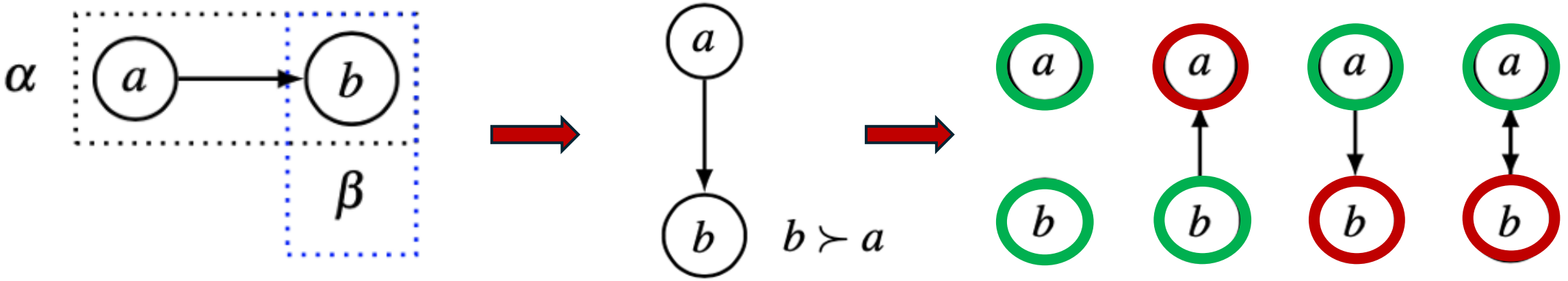
Argument b is more preferred Preference based argumentation framework Abstract argumentation framework

Kaci S, van Der Torre L, Vesic S, et al. Preference in abstract argumentation[J]. Handbook of Formal Argumentation, Volume 2, 2021.

New semantics 2: Social agent approaches

- interprets agent argumentation as a kind of voting in social choice

1. counting the number of agents that have the argument
2. is based on a reduction to preference-based argumentation



Argument b is more preferred Preference based argumentation framework Abstract argumentation framework

Kaci S, van Der Torre L, Vesic S, et al. Preference in abstract argumentation[J]. Handbook of Formal Argumentation, Volume 2, 2021.

New semantics 3 Agent reduction semantics

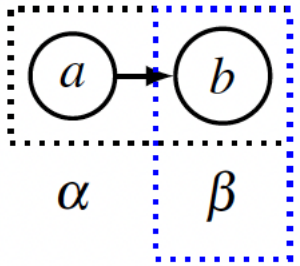
- interprets agent argumentation as a kind of justification aggregation

1. takes the perspective of each agent:
   - an agent prefers its own arguments over the arguments of the other agents
2. is also based on a reduction to *preference-based argumentation*
   - for each agent, there is a set of reductions, then take union of all reductions

New semantics 3 Agent reduction semantics

- interprets agent argumentation as a kind of justification aggregation

1. takes the perspective of each agent:
   - an agent prefers its own arguments over the arguments of the other agents
2. is also based on a reduction to *preference-based argumentation*
   - for each agent, there is a set of reductions, then take union of all reductions

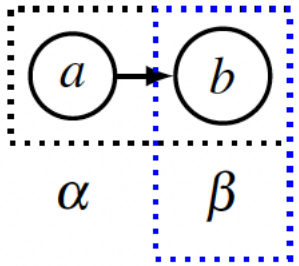


Kaci S, van Der Torre L, Vesic S, et al. Preference in abstract argumentation[J]. Handbook of Formal Argumentation, Volume 2, 2021.

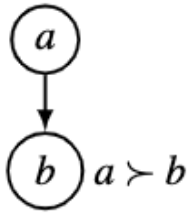
New semantics 3 Agent reduction semantics

- interprets agent argumentation as a kind of justification aggregation

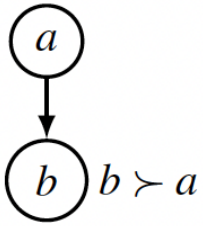
1. takes the perspective of each agent:
   - an agent prefers its own arguments over the arguments of the other agents
2. is also based on a reduction to *preference-based argumentation*
   - for each agent, there is a set of reductions, then take union of all reductions



From the perspective of α



From the perspective of β



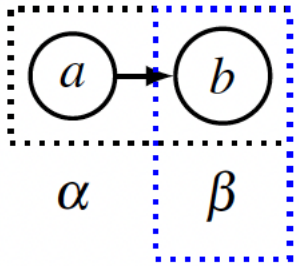
Liuwen Yu, Leon van der Torre

Kaci S, van Der Torre L, Vesic S, et al. Preference in abstract argumentation[J]. Handbook of Formal Argumentation, Volume 2, 2021.

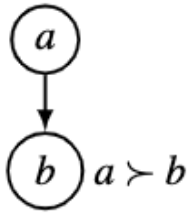
New semantics 3 Agent reduction semantics

- interprets agent argumentation as a kind of justification aggregation

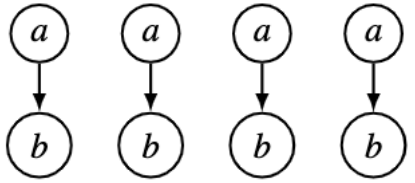
1. takes the perspective of each agent:
   - an agent prefers its own arguments over the arguments of the other agents
2. is also based on a reduction to *preference-based argumentation*
   - for each agent, there is a set of reductions, then take union of all reductions



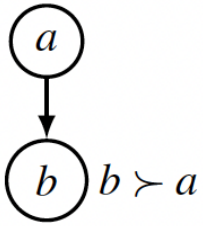
From the perspective of α



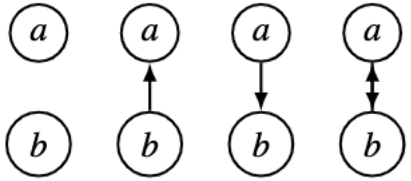
preference-based argumentation of α



From the perspective of β



preference-based argumentation of β



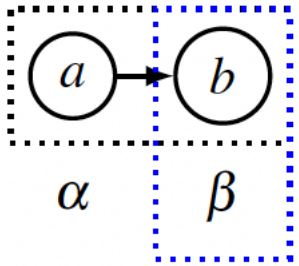
Liuwen Yu, Leon van der Torre

Kaci S, van Der Torre L, Vesic S, et al. Preference in abstract argumentation[J]. Handbook of Formal Argumentation, Volume 2, 2021.

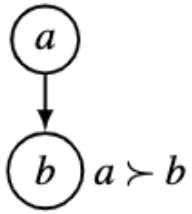
New semantics 3 Agent reduction semantics

- interprets agent argumentation as a kind of justification aggregation

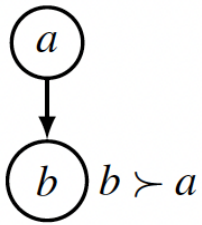
1. takes the perspective of each agent:
   - an agent prefers its own arguments over the arguments of the other agents
2. is also based on a reduction to *preference-based argumentation*
   - for each agent, there is a set of reductions, then take union of all reductions



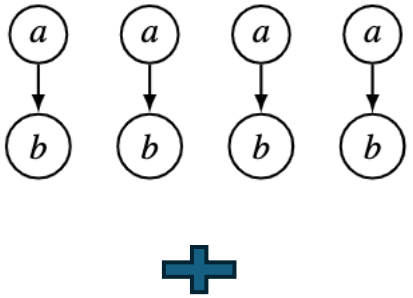
preference-based argumentation from the perspective of α



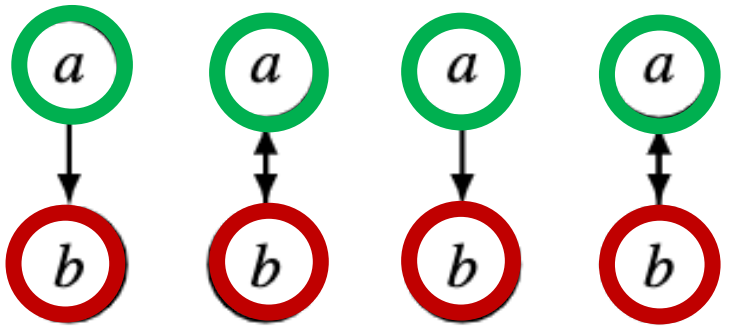
preference-based argumentation from the perspective of β

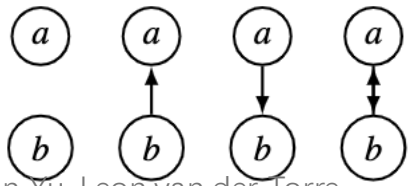


Reductions to abstract argumentation framework



Union of all reductions





Kaci S, van Der Torre L, Vesic S, et al. Preference in abstract argumentation[J]. Handbook of Formal Argumentation, Volume 2, 2021.

New semantics 4 Agent Filtering Semantics

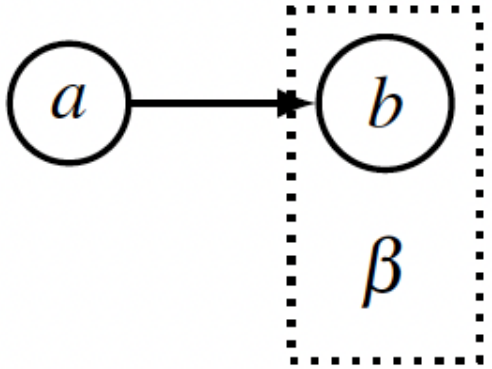
- is based on social cognition, knowledge, trust, etc

1. remove arguments that do not belong to an agent
2. remove attacks that do not belong to an agent
   - if both the attacking and the attacked arguments belong to the agent.

New semantics 4 Agent Filtering Semantics

- is based on social cognition, knowledge, trust, etc

1. remove arguments that do not belong to an agent
2. remove attacks that do not belong to an agent
   - if both the attacking and the attacked arguments belong to the agent.

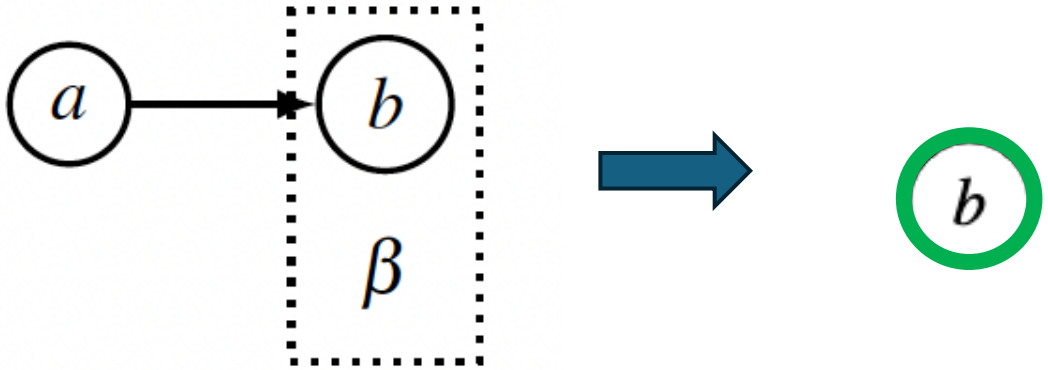


New semantics 4 Agent Filtering Semantics

- is based on social cognition, knowledge, trust, etc

1. remove arguments that do not belong to an agent
2. remove attacks that do not belong to an agent
   - if both the attacking and the attacked arguments belong to the agent.

a is unknown

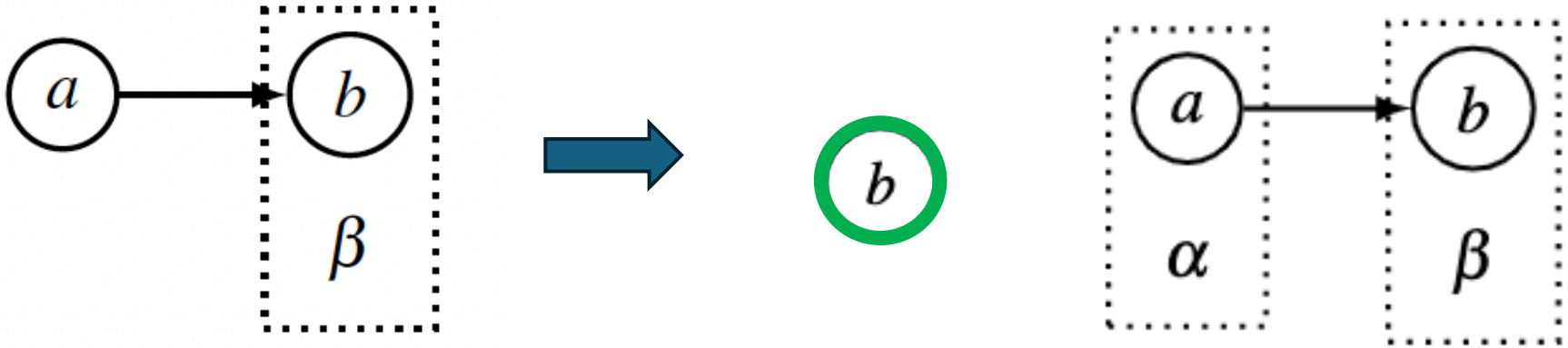


New semantics 4 Agent Filtering Semantics

- is based on social cognition, knowledge, trust, etc

1. remove arguments that do not belong to an agent
2. remove attacks that do not belong to an agent
   - if both the attacking and the attacked arguments belong to the agent.

a is unknown

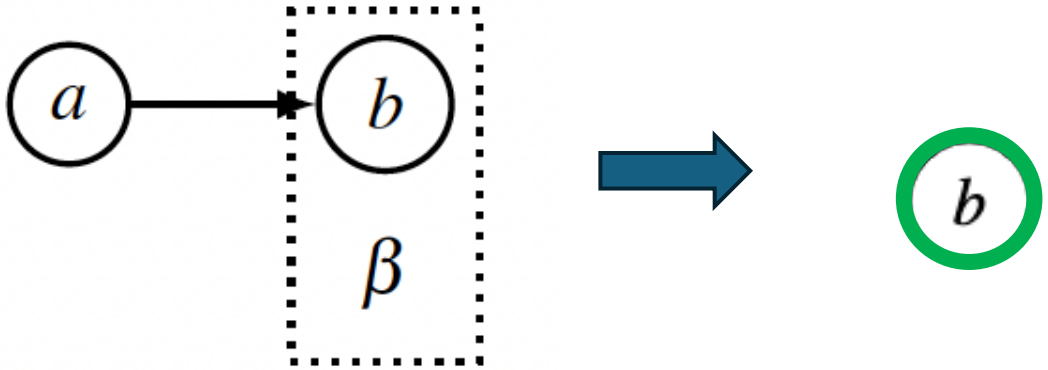


New semantics 4 Agent Filtering Semantics

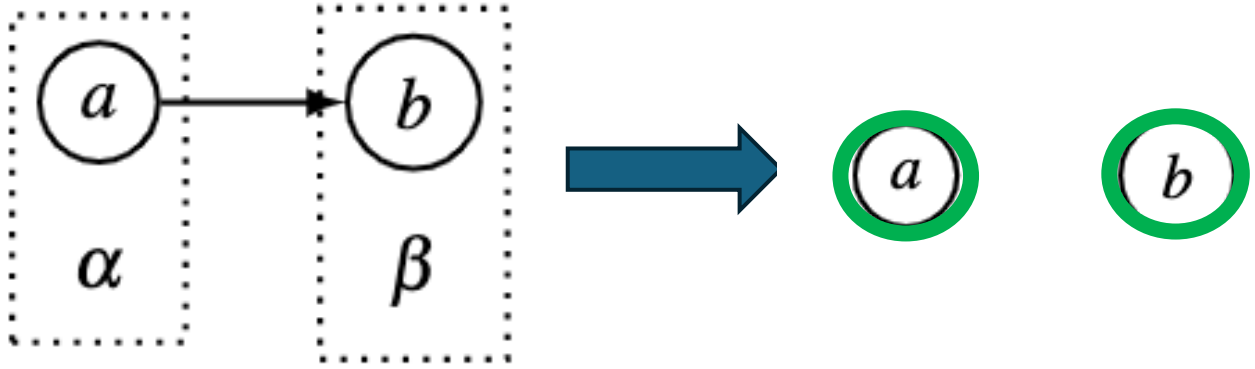
- is based on social cognition, knowledge, trust, etc

1. remove arguments that do not belong to an agent
2. remove attacks that do not belong to an agent
   - if both the attacking and the attacked arguments belong to the agent.

a is unknown



The attack is unknown



Other aspects of dialogue can be represented in abstract argumentation

- We have multiple agents
- We also have the challenges to introduce at the abstract level
  - game theory, mechanism design and social choice concepts

1. Personal representation of argumentation from private knowledge base
   - Dynamics: exchange arguments
   - Strategic argumentation to reach certain dialogue goal
2. The use of protocols of dialogue
   - Fatio
3. Agents are not equal, they have different positions/roles/…

…

Baumeister D, Neugebauer D, Rothe J. Collective Acceptability in Abstract Argumentation[J]. Handbook of Formal Argumentation, Volume 2, 2021.

Example: Principles-based analysis to abstract agent argumentation

Traditional principles

Principle 1 (Conflict-free (Baroni & Giacomin 2007))
Principle 2 (Admissibility (Baroni & Giacomin 2007))
Principle 3 (Directionality (Baroni & Giacomin 2007))
Principle 4 (SCC-recursiveness (Baroni & Giacomin 2007))
Principle 5 (Modularity (Baumann, Brewka, & Ulbricht 2020))

Variants of traditional principles

Principle 6 (Agent Admissibility1)
Principle 7 (Agent Admissibility2)
Principle 8 (Agent SCC-recursiveness1)
Principle 9 (Agent SCC-recursiveness2)

New agent principles

Principle 10 (AgentAdditionPersistence)
Principle 11 (AgentAdditionUniversalPersistence)
Principle 12 (PermutationPersistence)
Principle 13 (MergeAgent)
Principle 14 (RemovalAgentPersistence)
Principle 15 (AgentNumberEquivalence)
Principle 16 (Conflict-freeInvolvement)
Principle 17 (RemovalArgumentPersistence)

Yu, Liuwen, et al. "A principle-based analysis of abstract agent argumentation semantics." *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*. Vol. 18. No. 1. 2021.

Example: Principles-based analysis to abstract agent argumentation

Traditional principles

Principle 1 (Conflict-free (Baroni & Giacomin 2007))
Principle 2 (Admissibility (Baroni & Giacomin 2007))
Principle 3 (Directionality (Baroni & Giacomin 2007))
Principle 4 (SCC-recursiveness (Baroni & Giacomin 2007))
Principle 5 (Modularity (Baumann, Brewka, & Ulbricht 2020))

Variants of traditional principles

Principle 6 (Agent Admissibility1)
Principle 7 (Agent Admissibility2)
Principle 8 (Agent SCC-recursiveness1)
Principle 9 (Agent SCC-recursiveness2)

New agent principles

Principle 10 (AgentAdditionPersistence)
Principle 11 (AgentAdditionUniversalPersistence)
Principle 12 (PermutationPersistence)
Principle 13 (MergeAgent)
Principle 14 (RemovalAgentPersistence)
Principle 15 (AgentNumberEquivalence)
Principle 16 (Conflict-freeInvolvement)
Principle 17 (RemovalArgumentPersistence)

Yu, Liuwen, et al. "A principle-based analysis of abstract agent argumentation semantics." *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*. Vol. 18. No. 1. 2021.

Example: Principles-based analysis to abstract agent argumentation

Principle 12 (PermutationPersistence): if we permute the agents, it does not affect the extensions.
It reflects anonymity.

Sem.	PI0	PI1	PI2	PI3	PI4	PI5	PI6	PI7
TR	CGPS	CGPS	CGPS	CGPS	CGPS	CGPS	×	×
Sem_1	S	S	CGPS	×	CGPS	×	×	×
Sem_2	S	S	CGPS	×	CGPS	×	×	×
SR_1	×	CGPS	CGPS	×	×	CGPS	×	×
SR_2	×	CGPS	CGPS	×	×	CGPS	×	×
SR_3	×	CGPS	CGPS	×	×	CGPS	×	×
SR_4	×	CGPS	CGPS	×	×	CGPS	×	×
AR_1^F	×	CGPS	CGPS	×	CGPS	×	×	×
AR_2^F	×	CGPS	CGPS	×	CGPS	×	×	×
AR_3^F	×	CGPS	CGPS	×	CGPS	×	×	×
AR_4^F	×	CGPS	CGPS	×	CGPS	×	×	×
AR_1^A	×	CGPS	CGPS	×	×	×	×	×
AR_2^A	×	CGPS	CGPS	×	×	×	×	×
AR_3^A	×	CGPS	CGPS	×	×	×	×	×
AR_4^A	×	CGPS	CGPS	×	×	×	×	×
OR	CGPS	CGPS	CGPS	CGPS	CGPS	CGPS	×	CGPS
NBR	CGPS	CGPS	CGPS	CGPS	CGPS	×	×	×

Yu, Liuwen, et al. "A principle-based analysis of abstract agent argumentation semantics." *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*. Vol. 18. No. 1. 2021.

Example: Principles-based analysis to abstract agent argumentation

Traditional principles

Principle 1 (Conflict-free (Baroni & Giacomin 2007))
Principle 2 (Admissibility (Baroni & Giacomin 2007))
Principle 3 (Directionality (Baroni & Giacomin 2007))
Principle 4 (SCC-recursiveness (Baroni & Giacomin 2007))
Principle 5 (Modularity (Baumann, Brewka, & Ulbricht 2020))

Variants of traditional principles

Principle 6 (Agent Admissibility1)
Principle 7 (Agent Admissibility2)
Principle 8 (Agent SCC-recursiveness1)
Principle 9 (Agent SCC-recursiveness2)

New agent principles

Principle 10 (AgentAdditionPersistence)
Principle 11 (AgentAdditionUniversalPersistence)
Principle 12 (PermutationPersistence)
Principle 13 (MergeAgent)
Principle 14 (RemovalAgentPersistence)
Principle 15 (AgentNumberEquivalence)
Principle 16 (Conflict-freeInvolvement)
Principle 17 (RemovalArgumentPersistence)

Yu, Liuwen, et al. "A principle-based analysis of abstract agent argumentation semantics." *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*. Vol. 18. No. 1. 2021.

Example: Principles-based analysis to abstract agent argumentation

Principle 14 (RemovalAgentPersistence): if two agents have the same arguments, we can remove one of these agents without changing the extensions.

Sem.	PI0	PI1	PI2	PI3	PI4	PI5	PI6	PI7
TR	CGPS	CGPS	CGPS	CGPS	CGPS	CGPS	×	×
Sem_1	S	S	CGPS	×	CGPS	×	×	×
Sem_2	S	S	CGPS	×	CGPS	×	×	×
SR_1	×	CGPS	CGPS	×	×	CGPS	×	×
SR_2	×	CGPS	CGPS	×	×	CGPS	×	×
SR_3	×	CGPS	CGPS	×	×	CGPS	×	×
SR_4	×	CGPS	CGPS	×	×	CGPS	×	×
AR_1^F	×	CGPS	CGPS	×	CGPS	×	×	×
AR_2^F	×	CGPS	CGPS	×	CGPS	×	×	×
AR_3^F	×	CGPS	CGPS	×	CGPS	×	×	×
AR_4^F	×	CGPS	CGPS	×	CGPS	×	×	×
AR_1^A	×	CGPS	CGPS	×	×	×	×	×
AR_2^A	×	CGPS	CGPS	×	×	×	×	×
AR_3^A	×	CGPS	CGPS	×	×	×	×	×
AR_4^A	×	CGPS	CGPS	×	×	×	×	×
OR	CGPS	CGPS	CGPS	CGPS	CGPS	CGPS	×	CGPS
NBR	CGPS	CGPS	CGPS	CGPS	CGPS	×	×	×

Yu, Liuwen, et al. "A principle-based analysis of abstract agent argumentation semantics." *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*. Vol. 18. No. 1. 2021.

Day 4: agentic AI, Fatio and Libra

4.1. New foundations for multiple agents: individual vs collective defense

4.2. New foundations for time: reasons / arguments in progress

4.3. Fatio and Libra

4.4. End-to-end pipelines for traceability and auditability

4.5. Aligning Argumentation as Balancing, Dialogue and Inference (A-BDI)

From abstract agent argumentation to Fatio

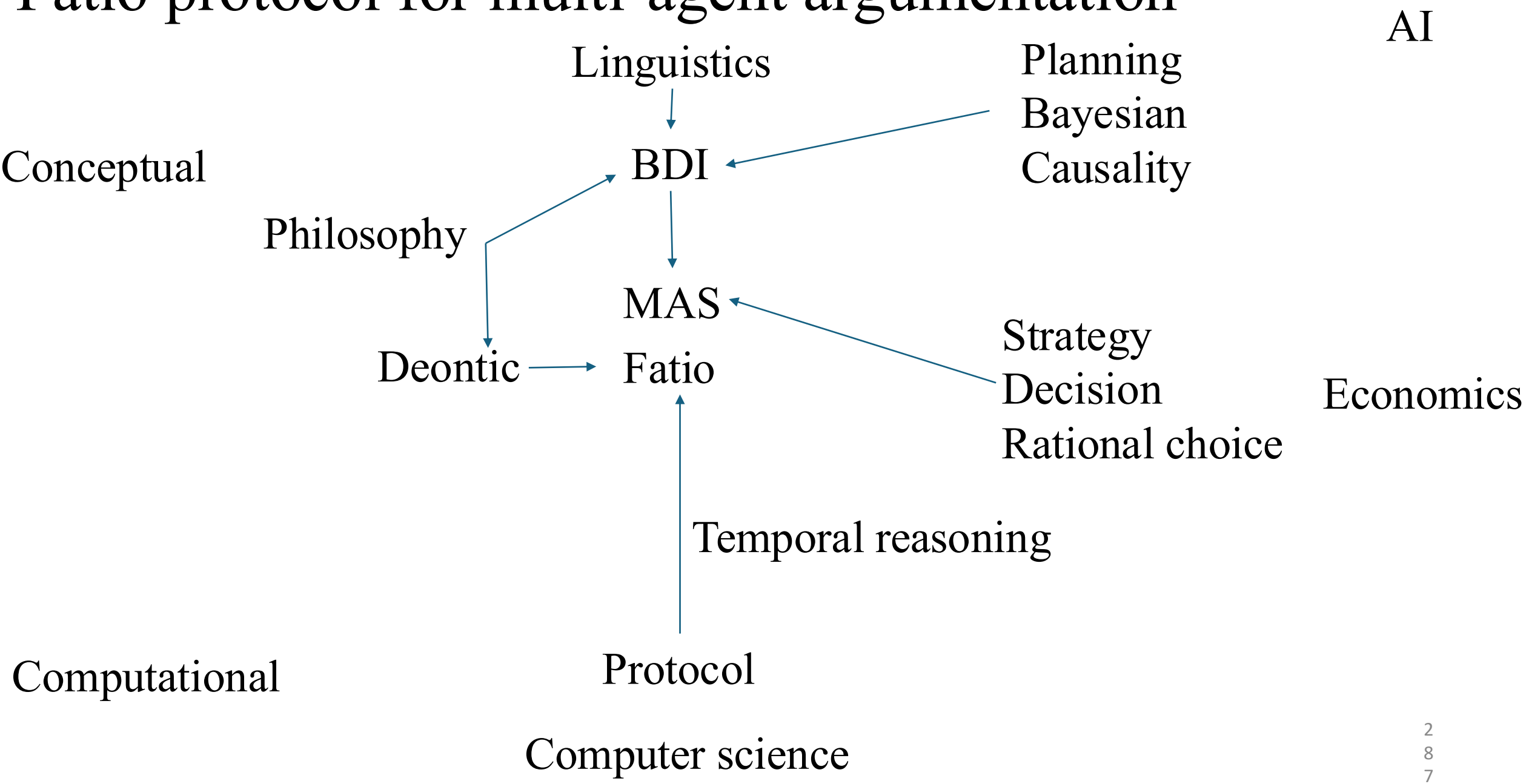
Fatio protocol for multi-agent argumentation

- Syntax (5 moves and combination rules)
- Axiomatic Semantics (in terms of the beliefs and desires of the participating agents)
- Operational Semantics (agent decision mechanisms)

Locutions for Argumentation in Agent Interaction Protocols

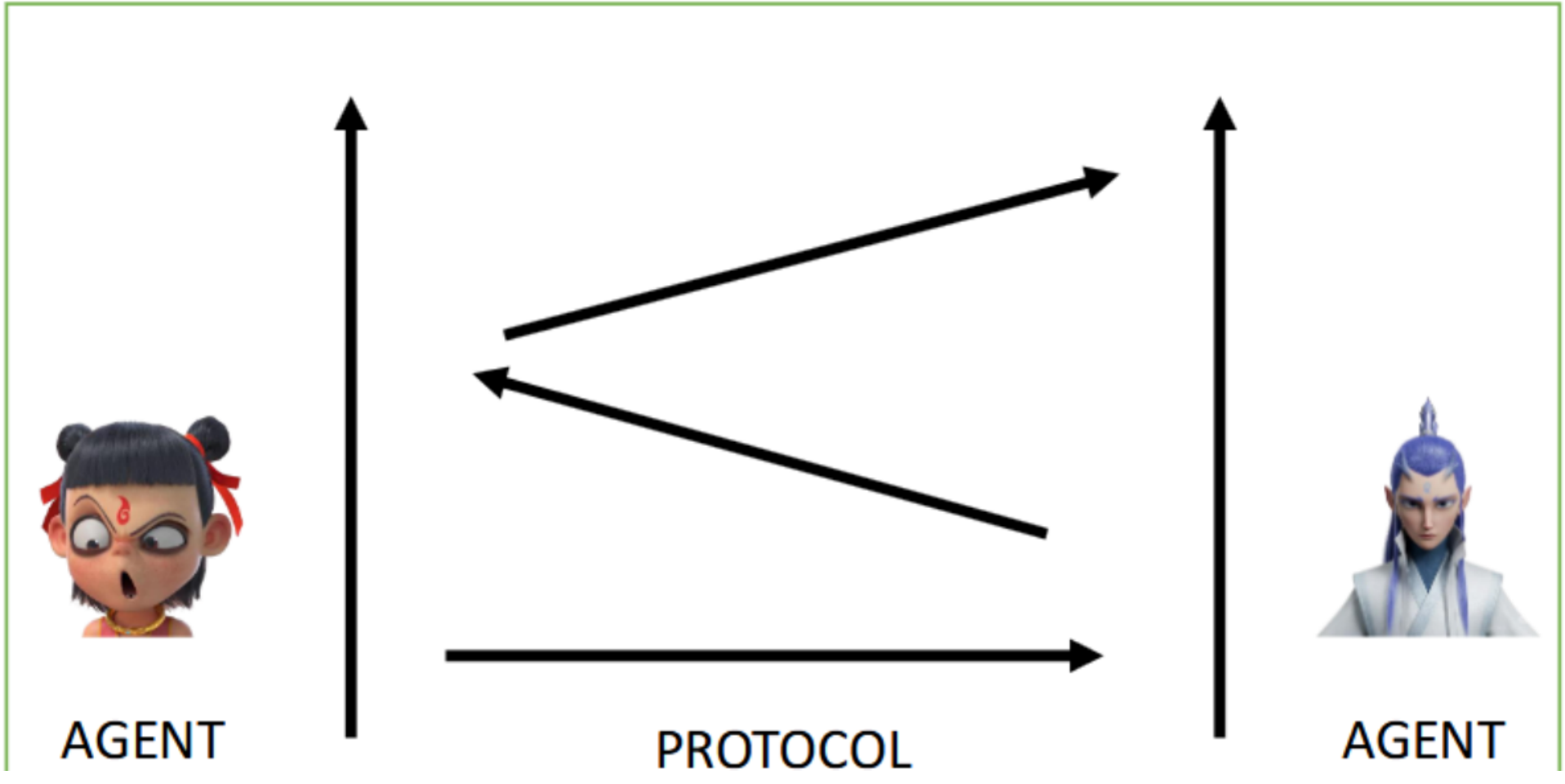
Peter McBurney[1] and Simon Parsons[2]

Fatio protocol for multi-agent argumentation



What is a formal dialogue protocol?

- Protocols: Computer Science perspective
- Dialogue: Linguistics perspective
- Agents: MAS/AI perspective



Fatio: Agents interact via justifications in dialogue

Syntax of the statements:

Five moves:

- F1: assert(Pi, φ)
- F2: question(Pj, Pi, φ)
- F3: challenge(Pj, Pi, φ)
- F4: justify(Pi, $\Phi \vdash \varphi$)
- F5: retract(Pi, φ)

Fatio: Agents interact via justifications in dialogue

Syntax of the statements:
Combination rules:

CR1: assert(a, φ) may be made at any time.
CR2: question(b, a, φ) or challenge(b, a, φ) follow assert(a, φ).
CR3: After question(b, a, φ) or challenge(b, a, φ), speaker a must reply immediately with justify(a, Φ ⊢ φ).
CR4: question or challenge may follow challenge(a, φ).
CR5: After question or challenge, the speaker of challenge(a, φ) must reply immediately with justify(a, Φ ⊢ φ).
CR6: retract(a, φ) follows assert(a, φ); retract(a, Φ ⊢ φ) follows justify(a, Φ ⊢ φ).

Fatio: Agents interact via justifications in dialogue

1. Axiomatic semantics: BDI logic for mental states
   - Pre- and post- conditions
2. Operational semantics: agent decision mechanisms
   - state of system change as execution of command in programming language

Agent i  I agree, the Brigade restaurant may not be excellent. (Retract)

The vegetarian food at the Brigade is awful. (Justify)  Agent j

Agent k  Why do you disagree? (Question)

Agent i  I had a great meal at the Brigade. (Justify)

I disagree, why is it excellent? (Challenge)  Agent j

Agent i  The Brigade Restaurant is excellent (Assert).

Fatio: Agents interact via justifications in dialogue

Natural Language Dialogue	Speech Acts
A: The Brigade Restaurant is excellent.	assert(A, R)
B: I disagree, why is it excellent?	challenge(B, A, R)
A: I had a great meal at the Brigade.	justify(A, P $\vdash^{+}$ R)
C: Why do you disagree?	question(C, B, R)
B: The vegetarian food at the Brigade is awful.	justify(B, Q & S $\vdash^{-}$ R)
A: I agree, the Brigade restaurant may not be excellent.	retract(A, R)

Fatio: Agents interact via justifications in dialogue

1. Axiomatic semantics: BDI logic for mental states
   - Pre- and post- conditions
2. Operational semantics: state of system change as execution of command in programming language

Pre-condition of assert

I desire other agents believe that I believe "The Brigade Restaurant is excellent" DiBjBiφ

Agent i

The Brigade Restaurant is excellent (Assert (i, φ)).

Agent k

Agent j

Fatio: Agents interact via justifications in dialogue

1. Axiomatic semantics: BDI logic for mental states
   - Pre- and post- conditions
2. Operational semantics: state of system change as execution of command in programming language

Post-condition 2 of assert: dialectical obligation: Agent i need to justify if other agent questions or challenges

Post-condition 1 of assert

All other agents believe that Agent i desires that each other agent, believes that Agent i believes "The Brigade Restaurant is excellent"

Agent k

Agent j

Pre-condition of assert

I desire other agents believe that I believe "The Brigade Restaurant is excellent" DiBjBiφ

Agent i

The Brigade Restaurant is excellent (Assert (i, φ)).

Fatio: Agents interact via justifications in dialogue

1. Axiomatic semantics: BDI logic for mental states
   - Pre- and post- conditions
2. Operational semantics: state of system change as execution of command in programming language

Pre-condition of challenge

I desire that each other agent, believes both that I desire that Pi utter a *justify(*Pi, $\Delta \vdash^{+} \varphi$*)* locution and that I do not believe φ.

Agent j

I disagree, why is it excellent? (Challenge)

Agent i

The Brigade Restaurant is excellent (Assert (i, φ)).

Fatio: Agents interact via justifications in dialogue

1. Axiomatic semantics: BDI logic for mental states
   - Pre- and post- conditions
2. Operational semantics: state of system change as execution of command in programming language

Pre-condition of challenge

Post-condition 1: dialectical obligation: Agent j need to justify if other agent questions or challenges

I desire that each other agent, believes both that I desire that Pi utter a *justify(*Pi, Δ ⊢+ φ*)* locution and that I do not believe φ.

Post-condition 2: Agent i must utter a justify

Agent j

I disagree, why is it excellent? (Challenge)

Agent i

The Brigade Restaurant is excellent (Assert (i, φ)).

BDI speech acts

Speech act	**Conditions**
$\texttt{assert}(P_i, \varphi)$	Pre: $(\forall j \neq i), (D_i B_j B_i \varphi)$. Post: $(\forall k \neq i)(\forall j \neq i)(B_k D_i B_j B_i \varphi)$.
$\texttt{question}(P_j, P_i, \varphi)$	Pre: $\exists i (i \neq j), (\forall k \neq j) D_j B_k D_j (\exists \Delta \in \mathcal{A})\texttt{Done}[\texttt{justify}(P_i, \Delta \vdash^{+} \varphi)]$. Post: Participant P_i must utter a *justify* locution.
$\texttt{challenge}(P_j, P_i, \varphi)$	Pre: $(\forall k \neq j)(D_j B_k \neg B_j \varphi) \wedge$ $(\forall k \neq j) D_j B_k D_j (\exists \Delta \in \mathcal{A})\texttt{Done}[\texttt{justify}(P_i, \Delta \vdash^{+} \varphi)]$. Post: Participant P_i must utter a *justify* locution.
$\texttt{justify}(P_i, \Phi \vdash^{+} \varphi)$	Pre: $(\texttt{Done}[\texttt{question}(P_j, P_i, \varphi)] \vee \texttt{Done}[\texttt{challenge}(P_j, P_i, \varphi)] \wedge$ $(\exists \Phi \in \mathcal{A})(\forall k \neq i) D_i B_k B_i (\Phi \vdash^{+} \varphi))$. Post: $(\forall k \neq i)(\forall j \neq i) B_k D_i B_j B_i (\Phi \vdash^{+} \varphi)$.
$\texttt{retract}(P_i, \varphi)$	Pre: $(\forall j \neq i) D_i B_j \neg B_i \varphi \vee (\forall j \neq i) D_i B_j \neg \neg B_i \varphi$. Post: $(\forall k \neq i)(\forall j \neq i) B_k D_i B_j \neg B_i \varphi \vee$ $(\forall k \neq i)(\forall j \neq i) B_k D_i B_j \neg \neg B_i \varphi$.

Fatio operational semantics

Operational semantics indicate how the states of a system change as a result of execution of the commands in a programming language.

Agent Decision Mechanisms.

1.(D1-D4)

2.D1(Φ): Claim or Not

3.D2: React or Not

4.D3(Φ): Defend or Not

5.D4(Φ): Fold or Not

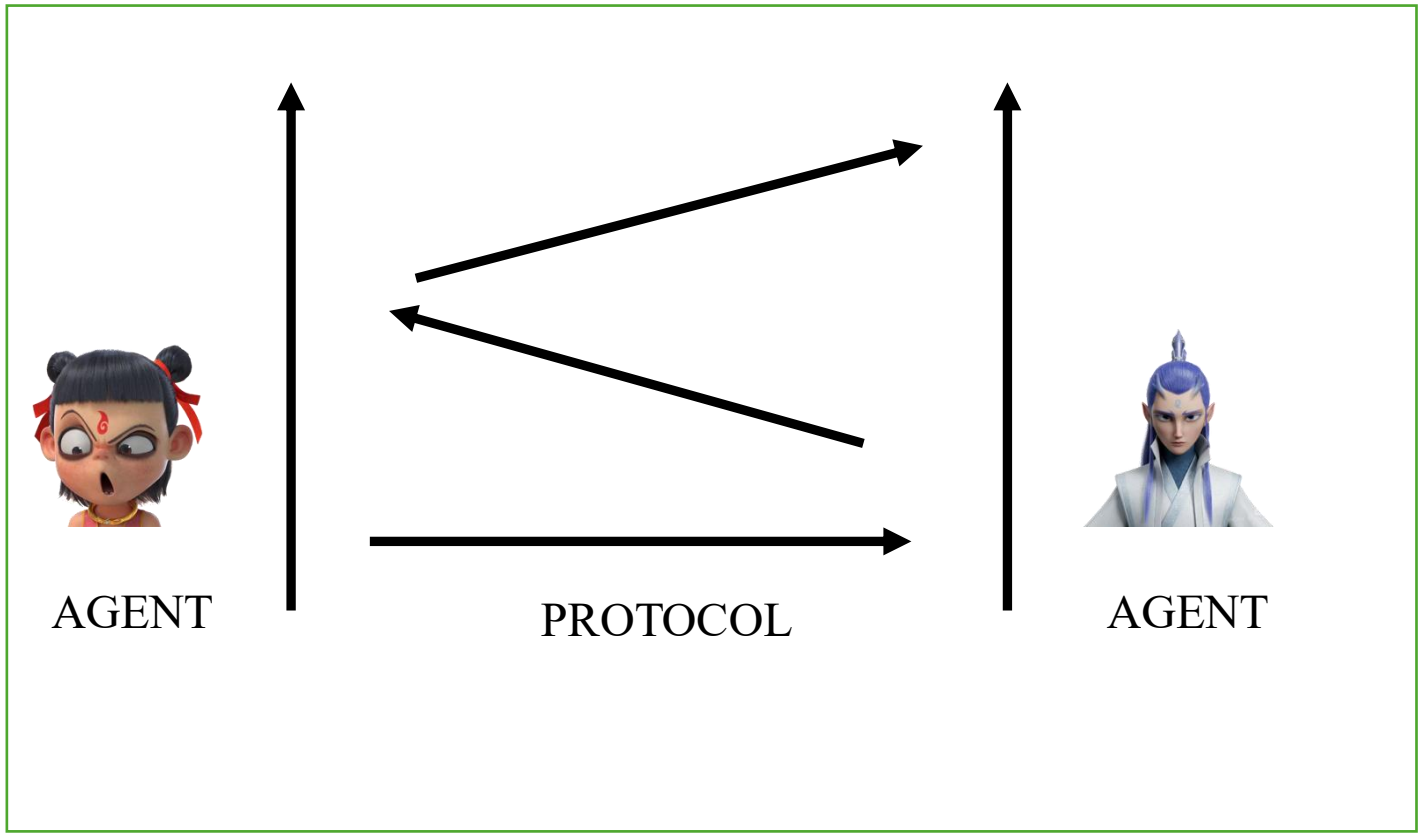
6.D5: Listen or Do

McBurney P, Parsons S. Locutions for argumentation in Agent interaction protocols[C]//International Workshop on Agent k ommunication. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004: 209-225.

Fatio dialogue protocol

SOFTWARE
ENGINEERING

PROGRAM



Comparative deliberation

Deliberation additionally asks:

1. Which options are competing?
2. Against which rival is the option defended?
3. How strongly does each reason weight?

Comparative deliberation

Deliberation additionally asks:

1. Which options are competing?
2. Against which rival is the option defended?
3. How strongly does each reason weight?

The Libra Dialogue Protocol for Comparative Deliberation

Luca PASETTO [a,1], Thomas SKAFTE [a], Leendert VAN DER TORRE [a], and Liuwen YU [b]

[a] *University of Luxembourg, Esch-sur-Alzette, Luxembourg*
[b] *Luxembourg Institute of Science and Technology, Esch-sur-Alzette, Luxembourg*

Abstract. Deliberative dialogue often requires agents to compare options publicly as relevant alternatives shift over the course of interaction. This paper introduces Libra, a dialogue protocol for comparative deliberation inspired by the Fatio protocol and informed by Tucker's distinction between justifying and requiring weight of reasons. Libra preserves Fatio's five locutions but reworks the content layer around option claims, contrastive reasons, explicit weight assignments, and grounds. The central technical contribution is a public semantics, formulated with deterministic structural equation models (SEMs), in which agents incrementally disclose SEM fragments rather than merely exchange abstract reasons. On this basis, the paper defines update rules for option introduction, changing comparison spaces, reason-element discussion, and retraction. A case study illustrates the framework.

Keywords. formal dialogue, weighing reasons, deontic reasoning, causal models, multi-agent systems

COMMA 2026

Fatio and Libra: constructing reasons in dialogue

Fatio: general justification

A : The Brigade is excellent.
B : Why is it excellent?
A : I had a great meal there.

$\text{assert}(R) \rightarrow \text{question}(R) \rightarrow \text{justify}(P \vdash^{+} R)$

Libra: contrastive justification

A : Let us choose the Brigade.
B : Why Brigade rather than Cafe?
A : It is nearer / better for our purpose.

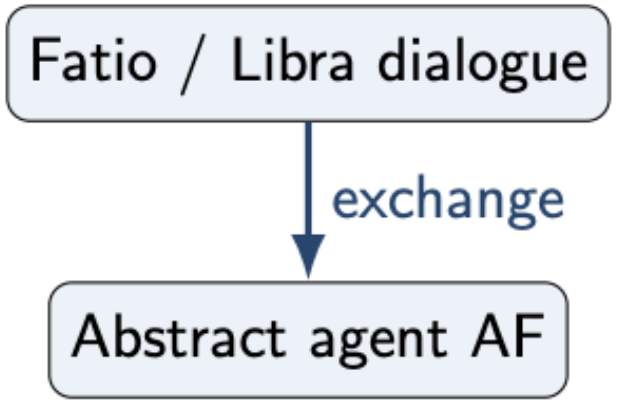
$\text{Why}(\text{Brigade} \mid \text{Cafe})$

Key idea

Questions and challenges create demands under which arguments are constructed; dialogue does not merely exchange finished arguments.

Pasetto, Luca, Thomas Skafte, Leon van der Torre, and Liuwen Yu. "The Libra Dialogue Protocol for Comparative Deliberation." *Proceedings of COMMA 2026*, forthcoming, 2026.

Abstract AF inside Fatio?



$$\mathrm{AF} = (\mathit{Args}, \mathrm{Att})$$

The tempting picture assumes that the argument already exists:

$$A_R : \quad P \vdash^{+} R.$$

But assert(A, R) comes **before** justify$(A, P \vdash^{+} R)$.

The missing object

After "The Brigade is excellent" but before the answer to "Why?", what is the argumentation state?

Ordinary abstract argumentation starts after a crucial part of the dialogue has already happened.

Pasetto, Luca, Thomas Skafte, Leon van der Torre, and Liuwen Yu. "The Libra Dialogue Protocol for Comparative Deliberation." *Proceedings of COMMA 2026*, forthcoming, 2026.

Before the argument: a reason in progress

Return to the first utterance:

$$A: \quad \text{“The Brigade is excellent.”}$$

There is a public commitment, but no supplied ground:

$$\text{assert}(A, R) \rightsquigarrow \boxed{\text{Pkg}_R = \langle \text{target } R,\ \text{support } ?,\ \text{status open} \rangle}.$$

Then

$$B: \quad \text{“Why is it excellent?”}$$

makes the missing support explicit:

$$\text{question}(B, A, R) \rightsquigarrow \text{Req}(R).$$

Ontological move

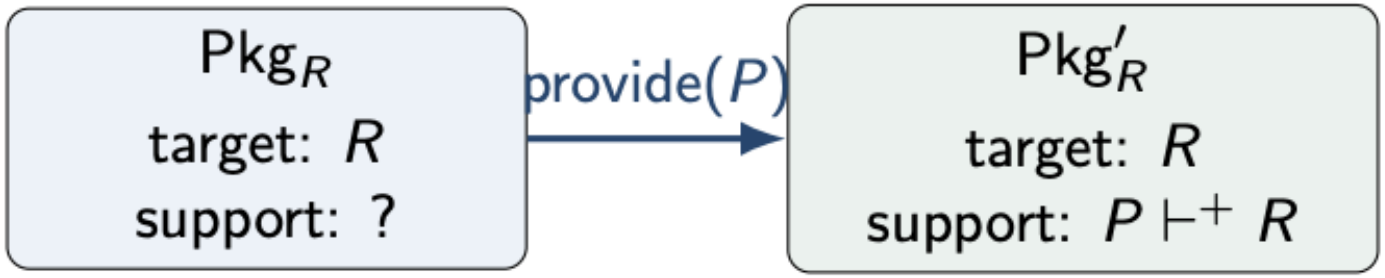
An unfinished reason is not merely a defective completed argument. It is a **reason in progress** with an explicit open requirement.

Pasetto, Luca, Thomas Skafte, Leon van der Torre, and Liuwen Yu. “The Libra Dialogue Protocol for Comparative Deliberation.” *Proceedings of COMMA 2026*, forthcoming, 2026.

Grounds extend reasons

A replies:

A : "I had a great meal there."



But the supplied ground may itself depend on further support:

$$Q \Rightarrow P \Rightarrow R.$$

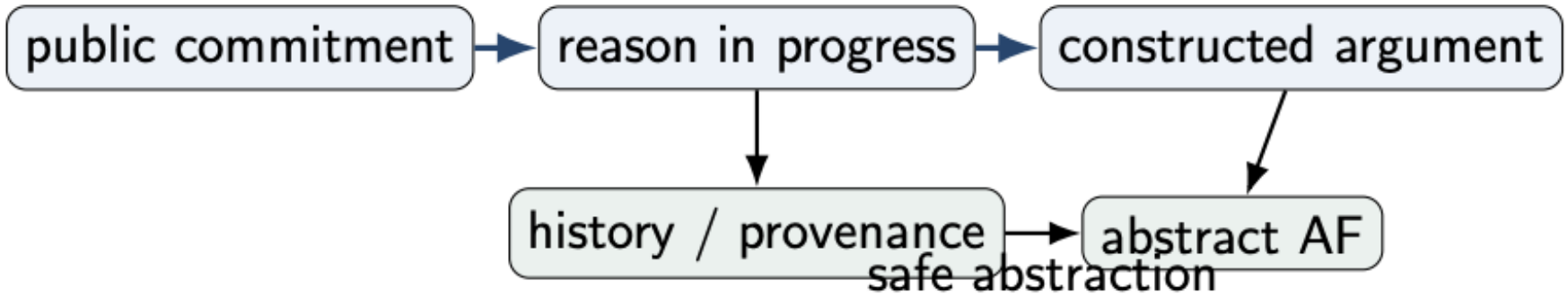
So construction can open new requirements while closing old ones.

Frontier

The current frontier records *where justification is still needed*; completion is reached through construction, not assumed at the start.

Pasetto, Luca, Thomas Skafte, Leon van der Torre, and Liuwen Yu. "The Libra Dialogue Protocol for Comparative Deliberation." *Proceedings of COMMA 2026*, forthcoming, 2026.

From abstract argumentation to Fatio



A :	The Brigade is excellent.	commitment
B :	Why?	open justification
A :	I had a great meal there.	ground
B :	Why Brigade rather than Cafe?	contrastive challenge

Main message

Do not merely put an AF inside a dialogue protocol. Explain how dialogue **constructs the arguments** from which the AF can eventually be safely abstracted.

Pasetto, Luca, Thomas Skafte, Leon van der Torre, and Liuwen Yu**.** "The Libra Dialogue Protocol for Comparative Deliberation." *Proceedings of COMMA 2026*, forthcoming, 2026.

Targets and occurrences

Compare:

$$\text{Why}(R?) \qquad \text{versus} \qquad \text{Why}(\text{Brigade} \mid \text{Cafe})?$$

The sentence "I had a great meal at the Brigade" may occur in both replies, but it has a different argumentative job.

General target

Does P support the claim that the Brigade is excellent?

$$P \vdash^{+} R$$

Contrastive target

Does P help justify choosing Brigade *rather than Cafe*?

$$P \vdash^{+} (\text{Brigade} \succ \text{Cafe})$$

Consequences

We must distinguish content from **occurrence**, and a reason from the **target** or burden it is intended to discharge.

Pasetto, Luca, Thomas Skafte, Leon van der Torre, and Liuwen Yu**.** "The Libra Dialogue Protocol for Comparative Deliberation." *Proceedings of COMMA 2026*, forthcoming, 2026.

LNGAI PART 1
Where does weight come from?

LNGAI PART 1
Where does weight come from?

We do not have an answer

LNGAI PART 1
Where does weight come from?

We do not have an answer
But, we did some experiments

Discretionary Judicial Reasoning

norm-based derivation

legally determined space for deciding based on the judge's own assessment

Liga, Davide, Réka Markovich, and Liuwen Yu. "Where do weights come from? weighting and weighing by LLMs and formal argumentation in discretionary judicial reasoning." *Logical and Symbolic Reasoning in Language Models@ AAAI 2026*. 2026.

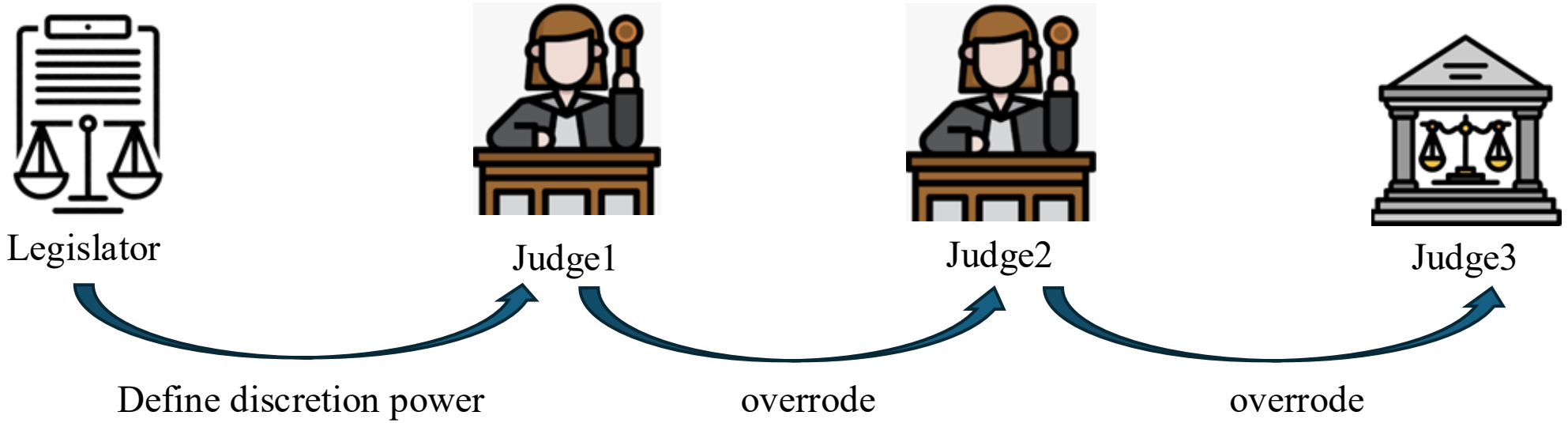
Paradigmatically discretionary areas

- minimal statutory input: best interest of child
- ensure child's physical, mental, & moral development in the most favorable way

- must decide by considering ALL circumstances
- some may be particularly important
- overweighting and ignoring aspects goes against properly enforcing the child's best interest

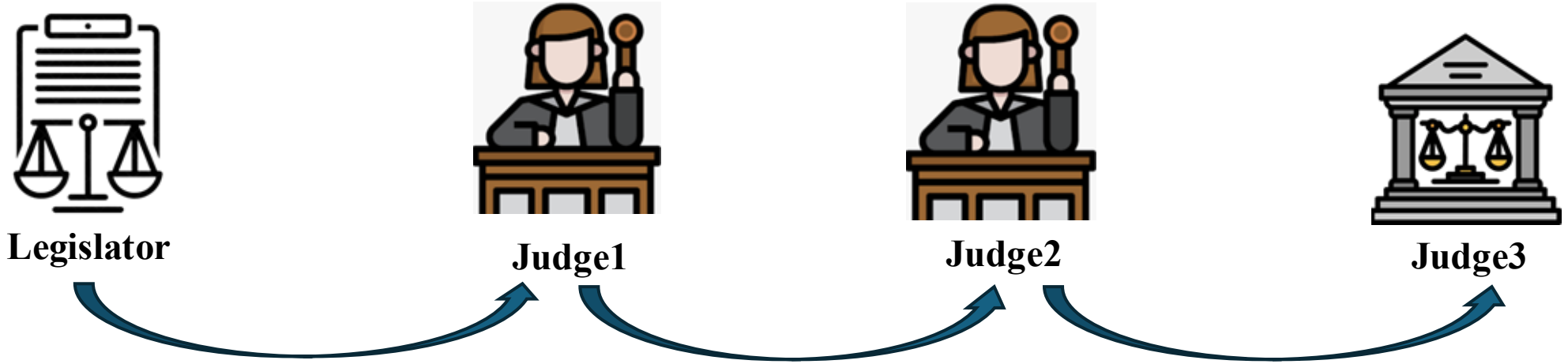
Liga, Davide, Réka Markovich, and Liuwen Yu. "Where do weights come from? weighting and weighing by LLMs and formal argumentation in discretionary judicial reasoning." *Logical and Symbolic Reasoning in Language Models@ AAAI 2026*. 2026.

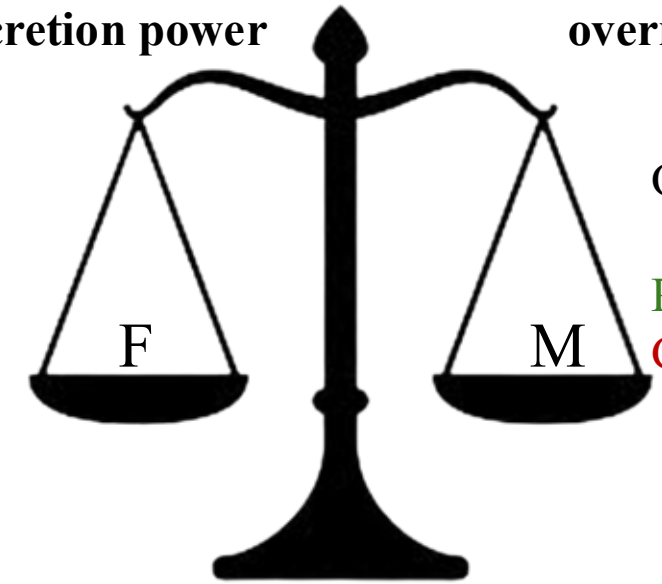
Case details



Hungarian Supreme Court (Kúria), Partial Judgment, Case No. Pfv.II.20.900/2021/13.

Case details





Objective conditions are equal

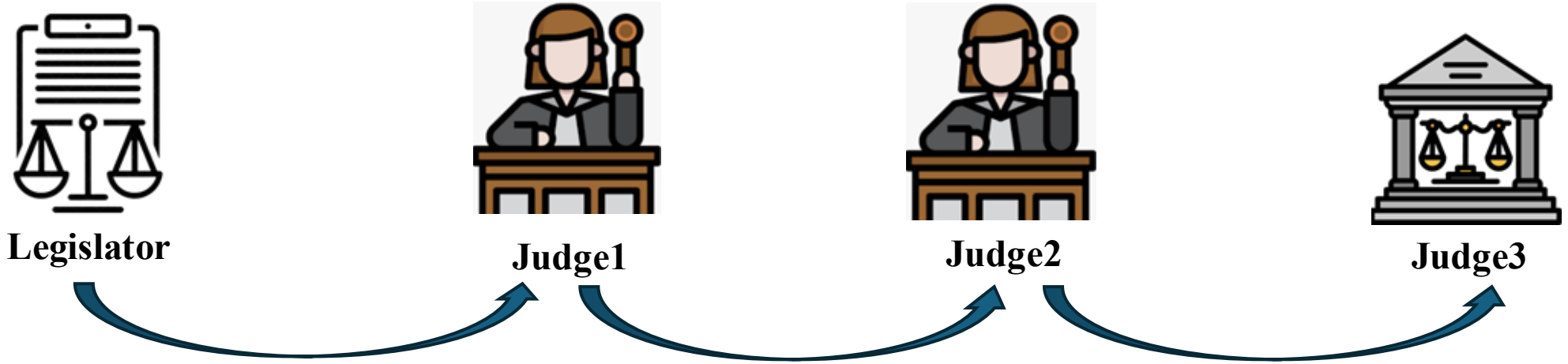
Pro: Children's opinions are decisive, want to live with mother

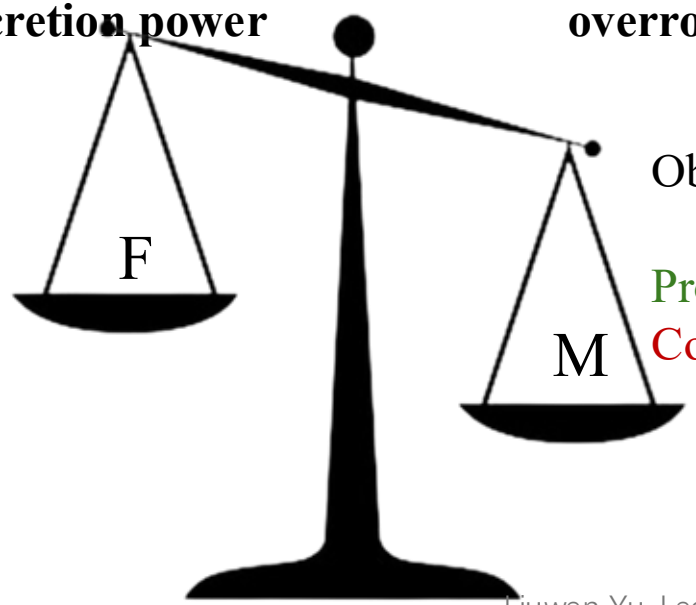
Con: Mother moved away children without consent of father

Con:
Father hit the child once.

Hungarian Supreme Court (Kúria), Partial Judgment, Case No. Pfv.II.20.900/2021/13.

Case details





Objective conditions are equal

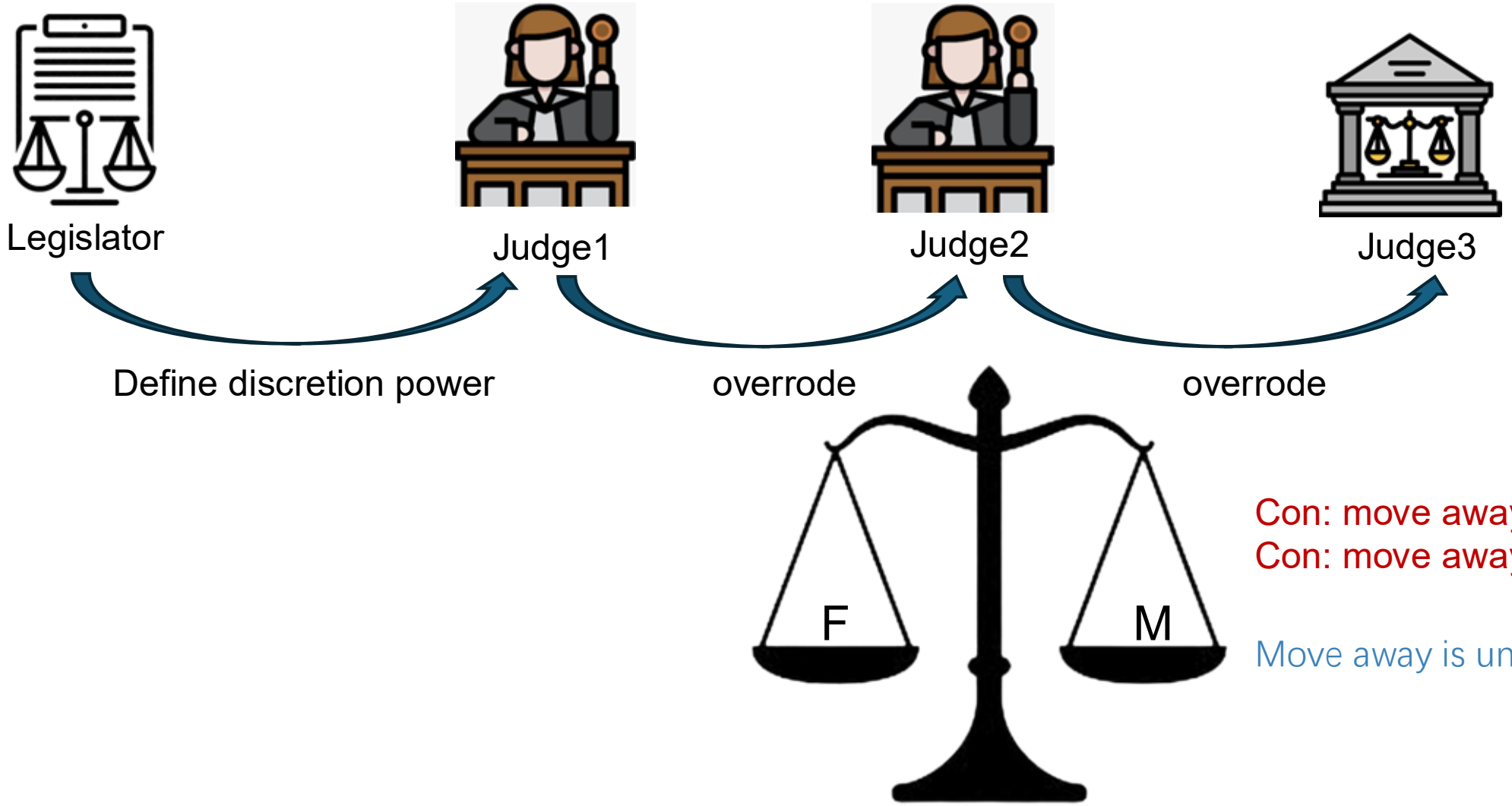
Pro: Children's opinions are **decisive**, want to live with mother

Con: Mother moved away children without consent of father

Con:
Father hit the child once.

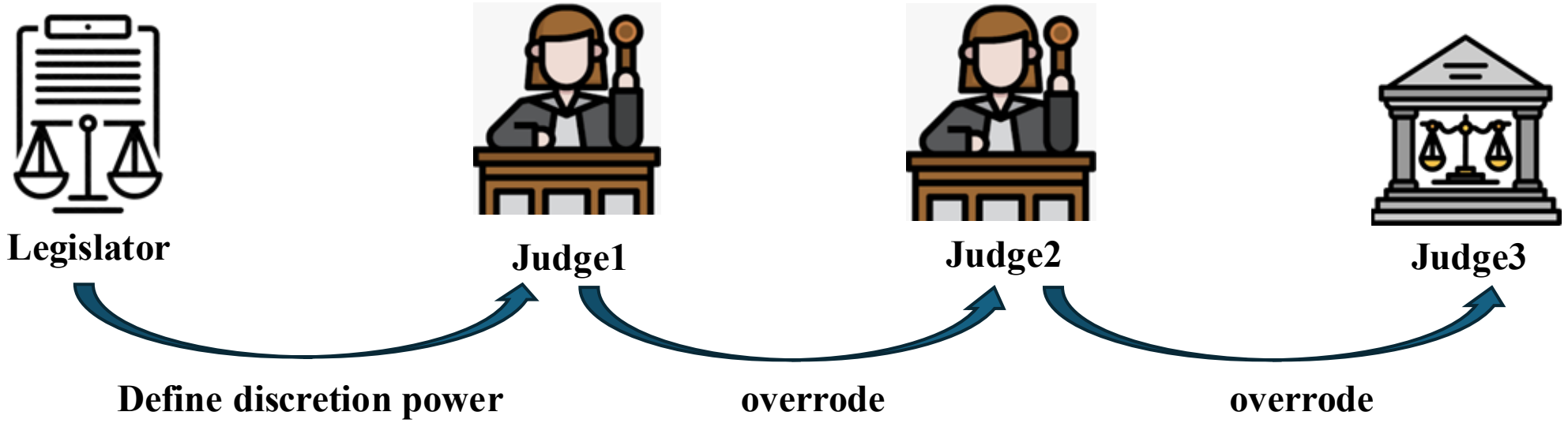
Hungarian Supreme Court (Kúria), Partial Judgment, Case No. Pfv.II.20.900/2021/13.

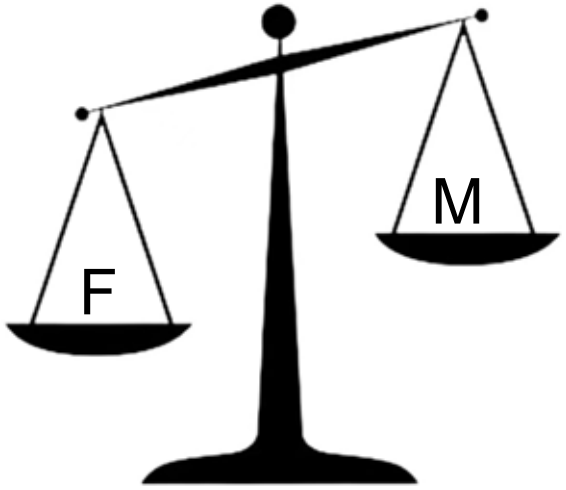
Case details



Hungarian Supreme Court (Kúria), Partial Judgment, Case No. Pfv.II.20.900/2021/13.

Case details



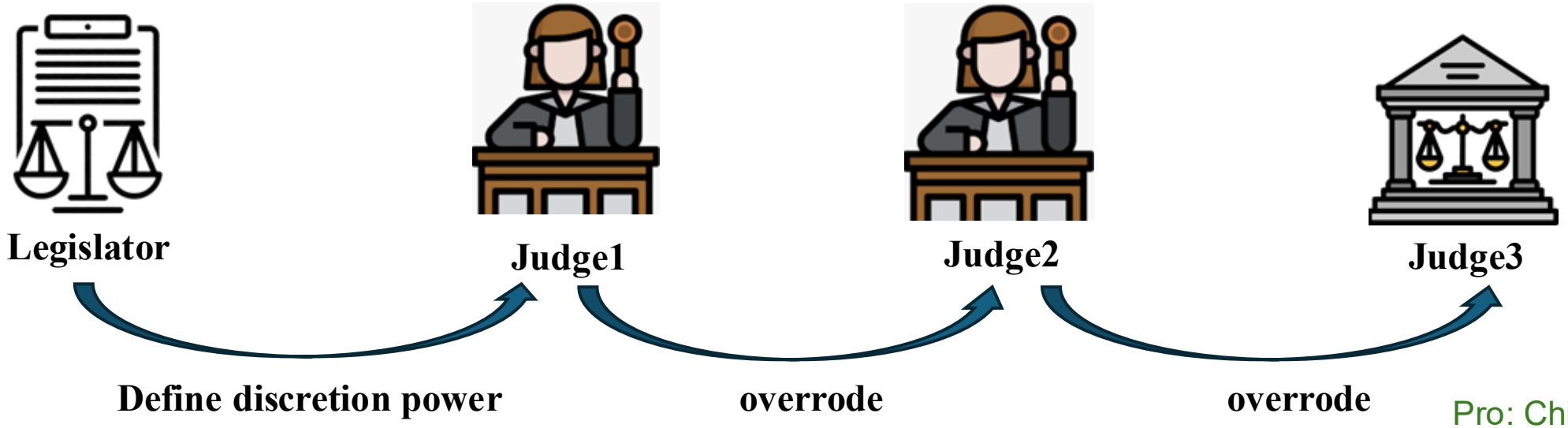


Con: move away is illegal
Con: move away is an alienation

Move away is underweighted

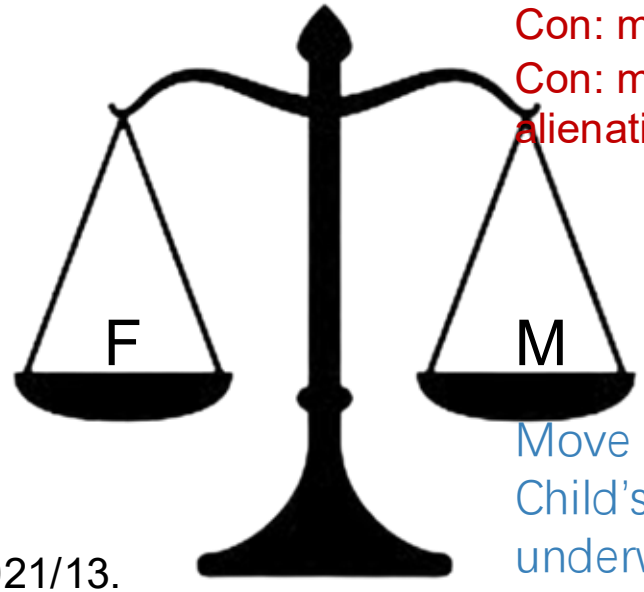
Hungarian Supreme Court (Kúria), Partial Judgment, Case No. Pfv.II.20.900/2021/13.

Case details



Pro: Children's opinion
Con: move away is illegal
Con: move away is an alienation

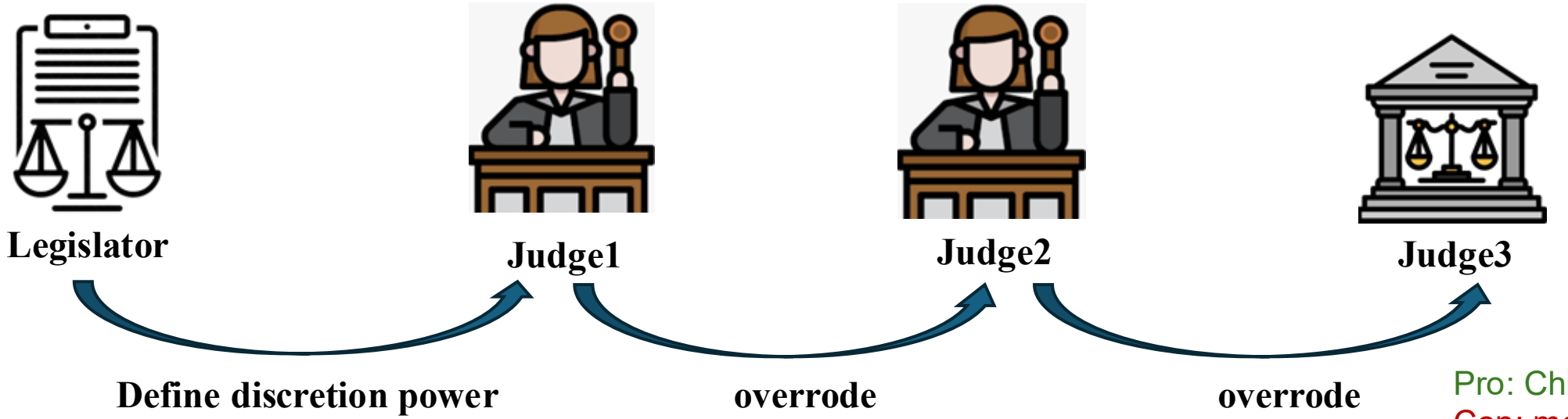
Con: Father hit the child once
Hit children is ignored.



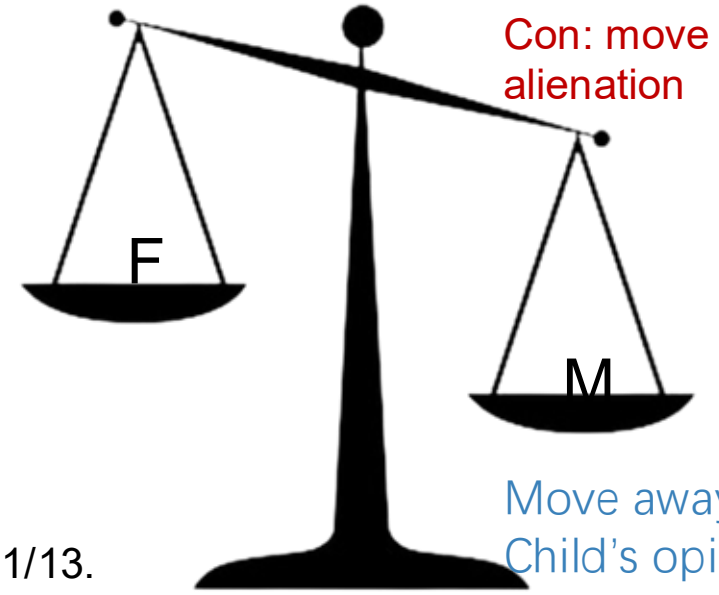
Move awayoverweighted
Child's opinion underweighted

Hungarian Supreme Court (Kúria), Partial Judgment, Case No. Pfv.II.20.900/2021/13.

Case details



Pro: Children's opinion
Con: move away is illegal
Con: move away is an alienation

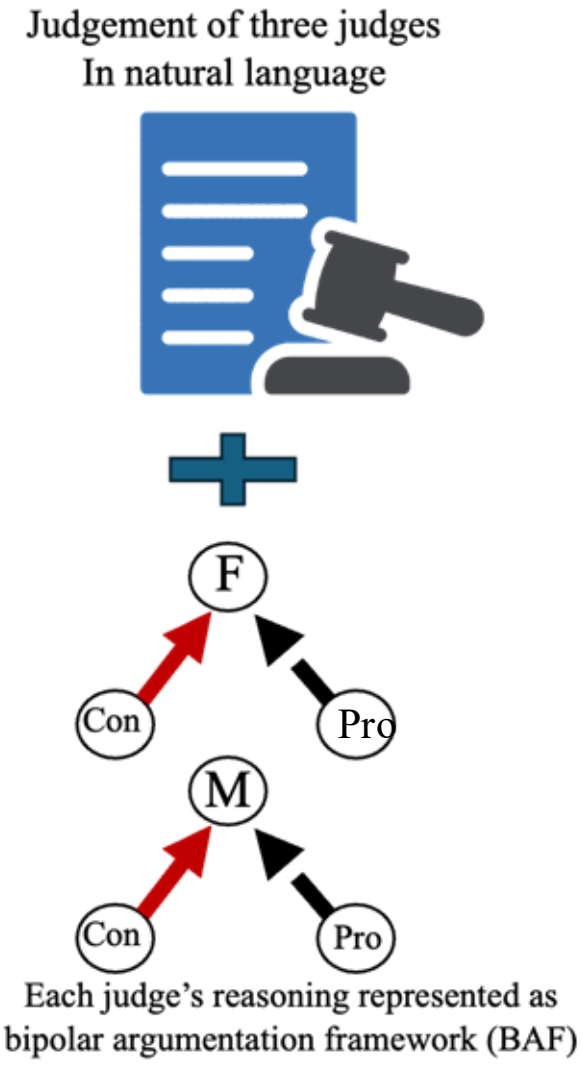


Con: Father hit the child once
Hit children is ignored.

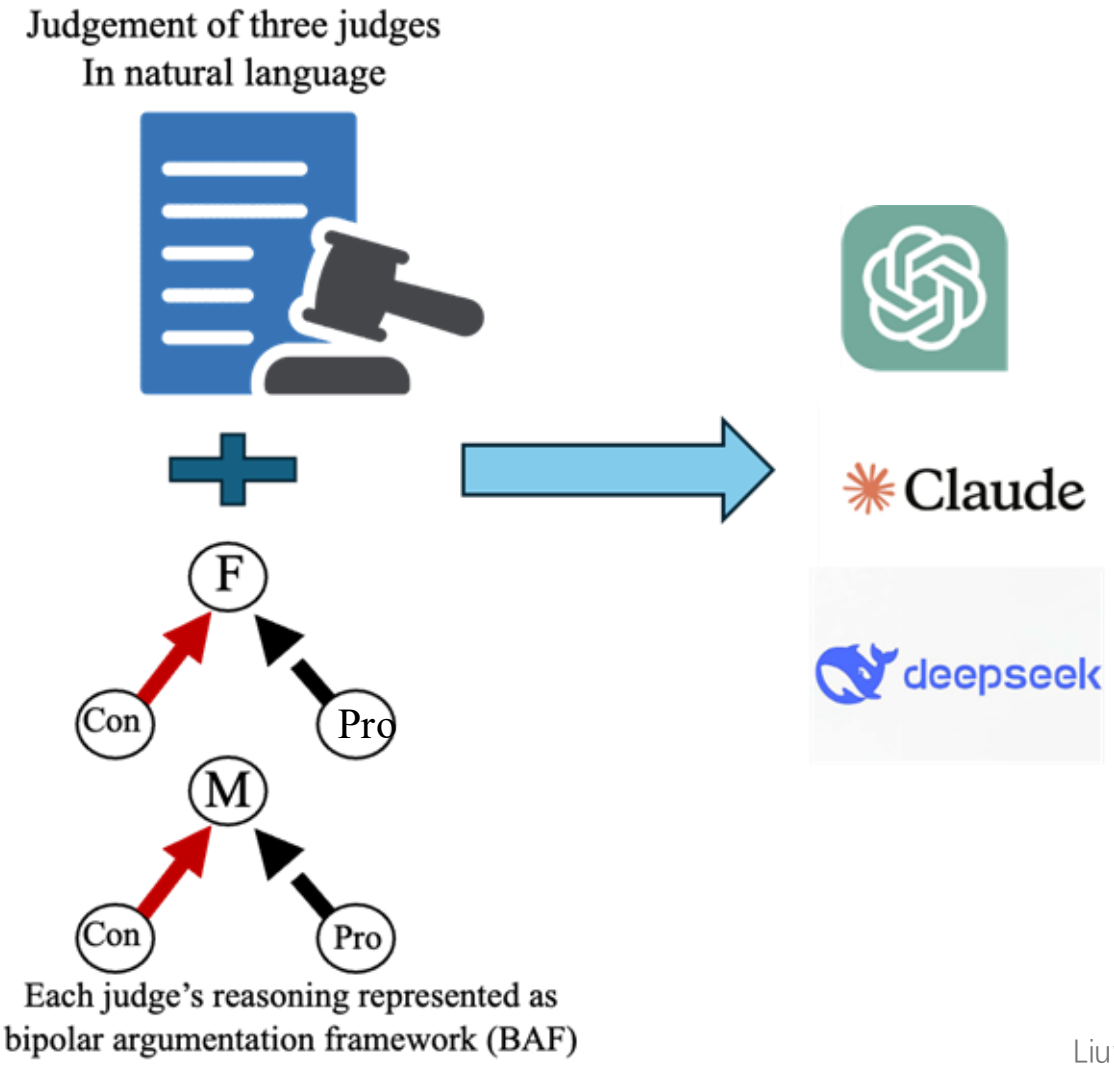
Move away overweighte
Child's opinion

Hungarian Supreme Court (Kúria), Partial Judgment, Case No. Pfv.II.20.900/2021/13.

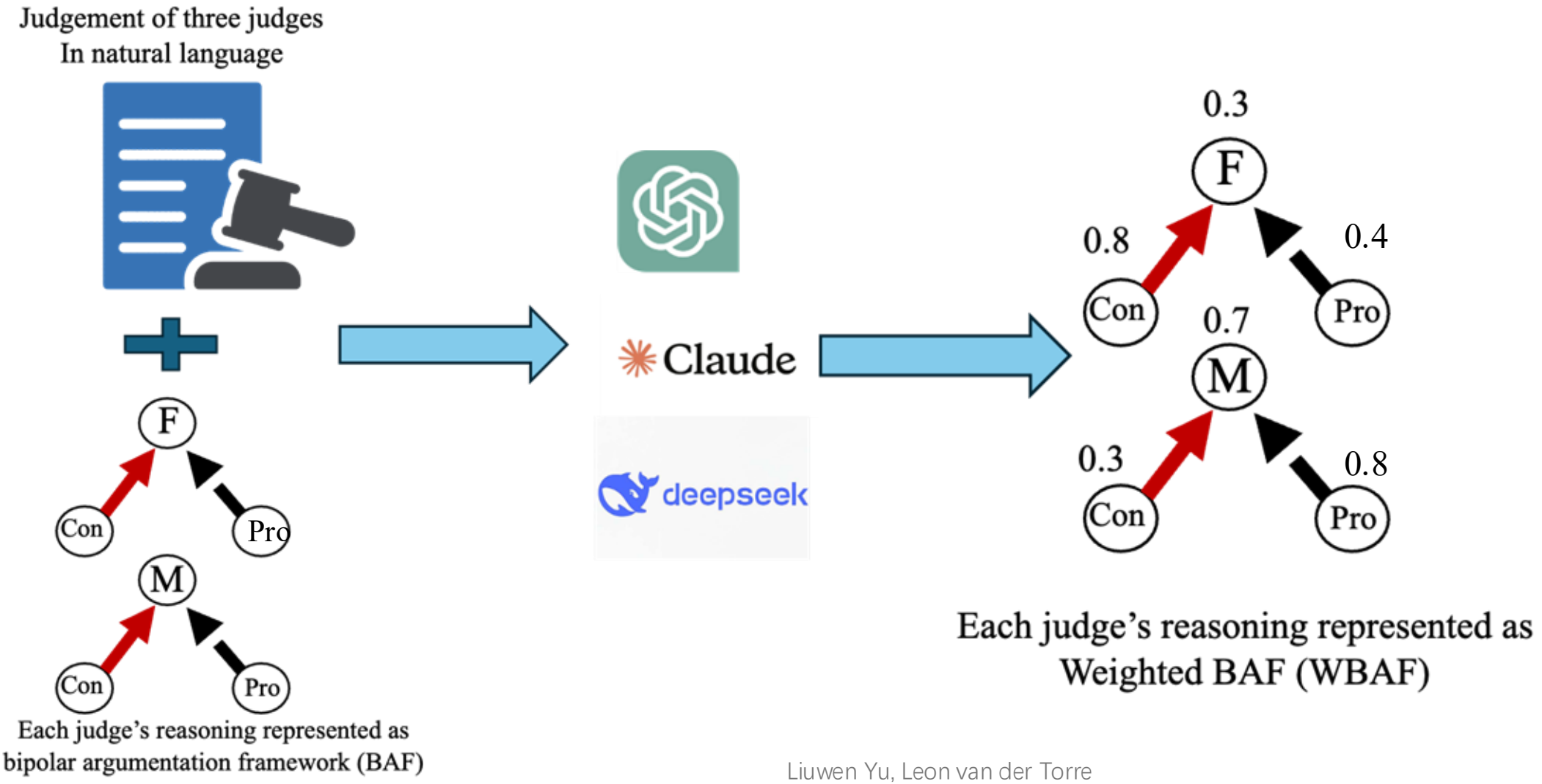
Pipeline


Judgement of three judges
In natural language
F
Con
Pro
M
Con
Pro
Each judge's reasoning represented as
bipolar argumentation framework (BAF)

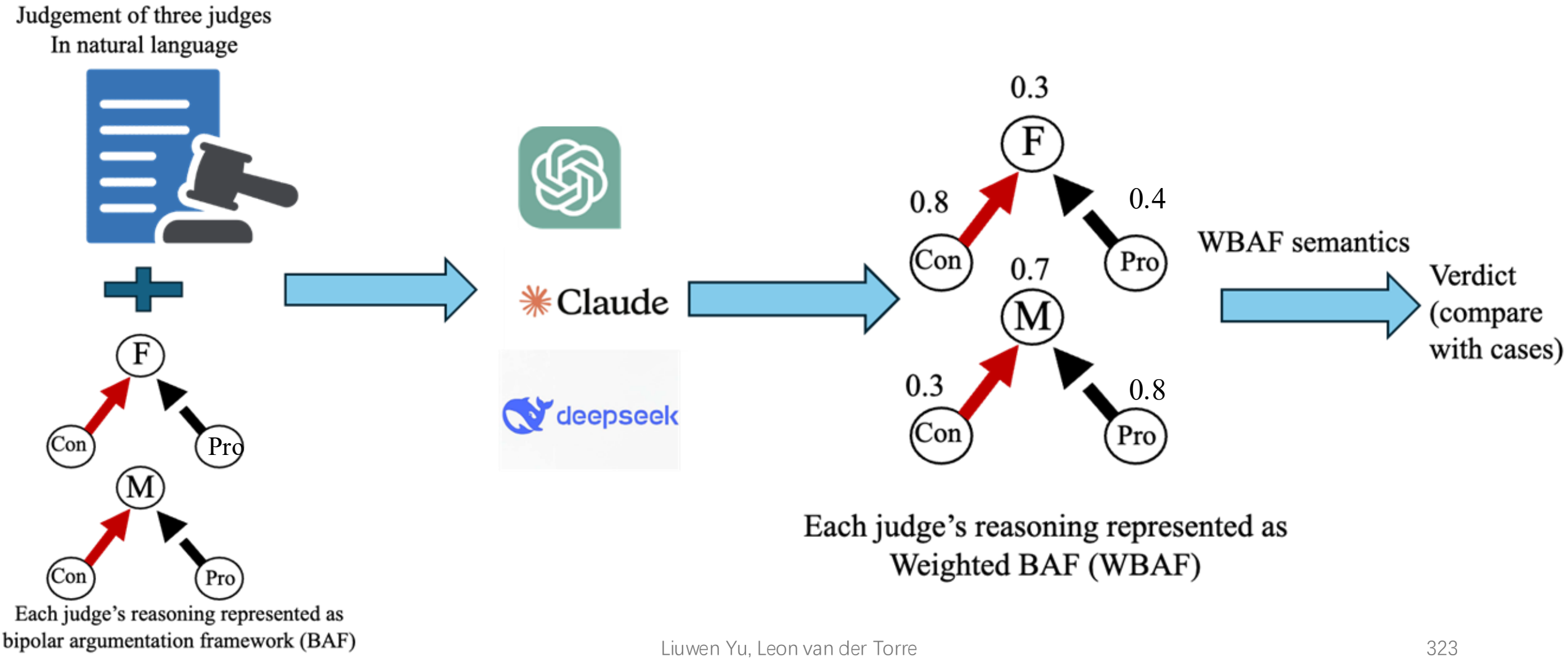
Pipeline


Judgement of three judges
In natural language
F
Con
Pro
M
Con
Pro
Each judge's reasoning represented as
bipolar argumentation framework (BAF)
Claude
deepseek

Pipeline


Judgement of three judges
In natural language
F
Con
Pro
M
Con
Pro
Each judge's reasoning represented as
bipolar argumentation framework (BAF)
Claude
deepseek
0.3
F
0.8
0.4
Con
Pro
0.7
M
0.3
0.8
Con
Pro
Each judge's reasoning represented as
Weighted BAF (WBAF)

Pipeline


Judgement of three judges
In natural language
F
Con
Pro
M
Con
Pro
Each judge's reasoning represented as
bipolar argumentation framework (BAF)
Claude
deepseek
0.3
F
0.8
Con
0.4
Pro
0.7
M
0.3
Con
0.8
Pro
Each judge's reasoning represented as
Weighted BAF (WBAF)
WBAF semantics
Verdict
(compare
with cases)

Method - Pipeline

- **Bipolar AF creation:**

Create J1 and J2 (BAFs for Judgement 1-2) (arguments+relations);

- **Task 1:**

Prompt LLM to **assign weights to the BAFs** of J1 and J2

- **Task 2:**

Prompt LLM to **generate a weighted BAF (WBAF)** J3

Context:	Models tested:	Robustness:
We provide J1-J2's arguments (custody pros/cons)	**GPT-5, Claude Sonnet 4.5, DeepSeek-V3.2-Exp**	**10 runs for each model**

Takeaways

Key idea:

(1) judgements in natural language expressing relative importance
(2) task for legal KR: importance ->numerical
(3) LLMs: developed to process natural language on advanced level

Methodology:

LLM-assigned base weights based on judgement + BAF→WBAF with WBAF semantics to model discretionary decision making

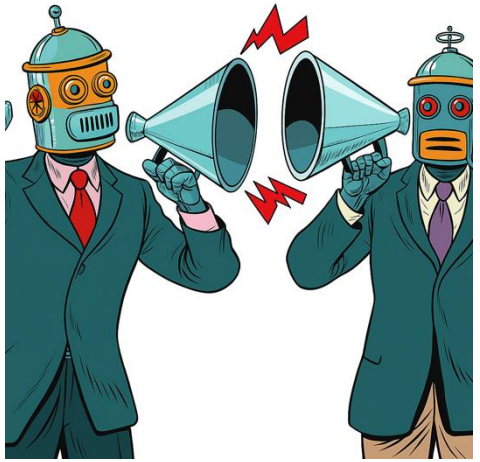
Evaluation results:

(1) Robustness of weight assignment tested via inter-run agreement;
(2) outcome alignment tested via argumentation semantics

LNGAI PART 2
Agentic AI: tests with LogiKEy

Architecture for Agentic AI

Subsymbolic AI



Symbolic AI

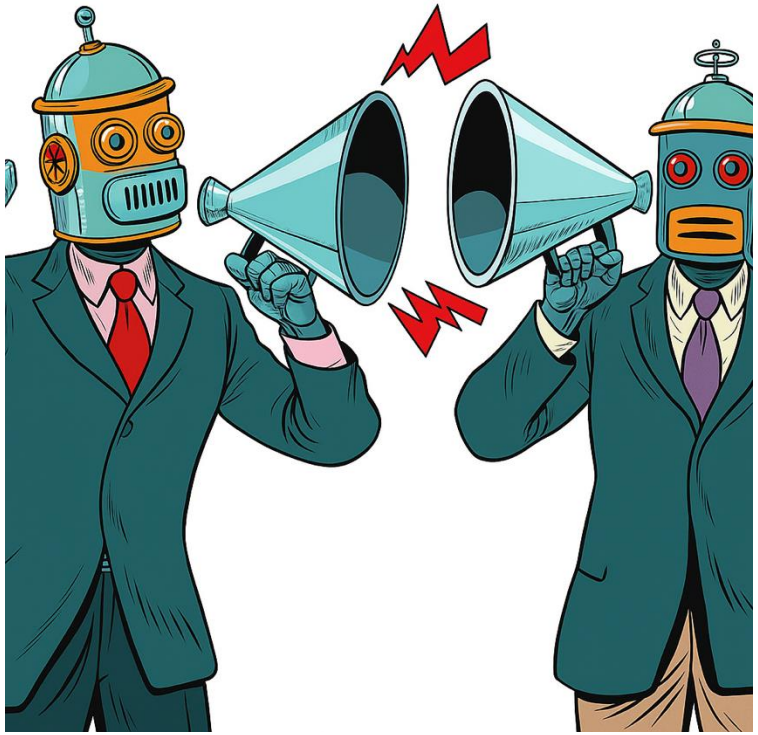
Dialogue → mining → Formal model, Machine-readable representation → reasoning → Automatic reasoner

A: The Brigade Restaurant is excellent.
B: I disagree, why is it excellent?
A: I had a great meal at the Brigade.

Formalized Dialogue
Speech Acts
Semantics
Beliefs, Desires

Testing?
Consequences?
Consistence?
Compliance?

Generating Mental States for the Fatio Dialogue Protocol



Motivation: Generating Mental States from Dialogue

Task: Given a dialogue, automatically infer mental states of agents

Try with an LLM: Here is a dialogue between different agents A, B, and C:

A: The Brigade Restaurant is excellent.
B: l disagree, why is it excellent?
A: l had a great meal at the Brigade.
C: Why do you disagree?
B:The vegetarian food at the Brigade is awful.A: l agree, the Brigade restaurant may not be excellent

Give me beliefs and desires of each agent with respect to the other agents.
(ChatGPT 4o)

Beliefs of A	Desires of A
- The Brigade is excellent - A great meal is a sign of excellence	- Justify own positive evaluation - Reach mutual understanding - Accommodate others views

Motivation: Generating Mental States from Dialogue

Task: Given a dialogue, automatically infer mental states of agents

Try with an LLM: Here is a dialogue between different agents A, B, and C:

A: The Brigade Restaurant is excellent
B: I disagree, why is it excel
A: I had a great meal at the B
C: Why do you disagree?
B:The vegetarian food at the Brigade is awful.A: I agree, the Brigade restaurant may not be excellent

Problem:
How to evaluate this?

Give me beliefs and desires of each agent with respect to the other agents.
(ChatGPT 4o)

Beliefs of A	Desires of A
- The Brigade is excellent - A great meal is a sign of excellence	- Justify own positive evaluation - Reach mutual understanding - Accommodate others views

Motivation: Generating Mental States from Dialogue

Task: Given a dialogue, automatically infer mental states of agents

Try with an LLM: Here is a dialogue between different agents A, B, and C:

A: The Brigade Restaurant is excellent
B: I disagree
A: I had a gr
C: Why do y
B:The vegeta y not be excellent

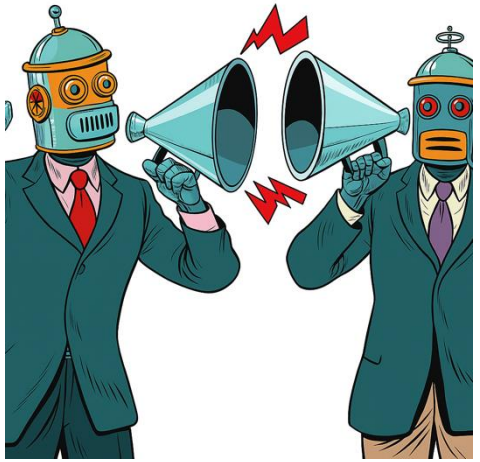
Problem: How to evaluate this? Architecture with:
- Model: semantics for communicating agents
- Integration of Subsymbolic AI (generation) and Symbolic AI (testing)

Give me beliefs and desires of each agent with respect to the other agents.
(ChatGPT 4o)

Beliefs of A	Desires of A
- The Brigade is excellent - A great meal is a sign of excellence	- Justify own positive evaluation - Reach mutual understanding - Accommodate others views

Architecture for Agentic AI

Subsymbolic AI



Symbolic AI

Dialogue → mining → Formal model, Machine-readable representation → reasoning → Automatic reasoner

A: The Brigade Restaurant is excellent.
B: I disagree, why is it excellent?
A: I had a great meal at the Brigade.

Formalized Dialogue
Speech Acts
Semantics
Beliefs, Desires

Fatio dialogue protocol

Testing?
Consequences?
Consistence?
Compliance?

LogiKEy

LogiKEy

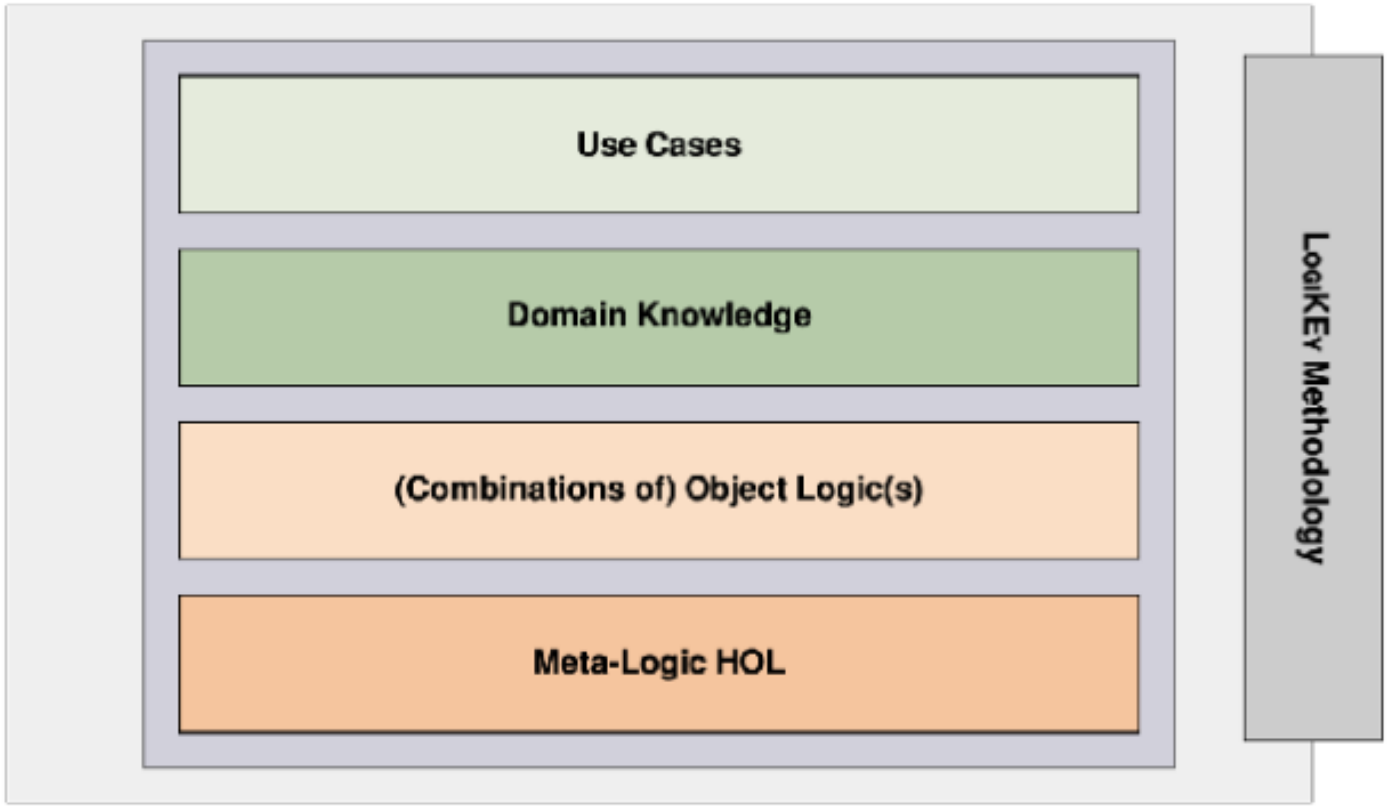
ELSEVIER

Artificial Intelligence

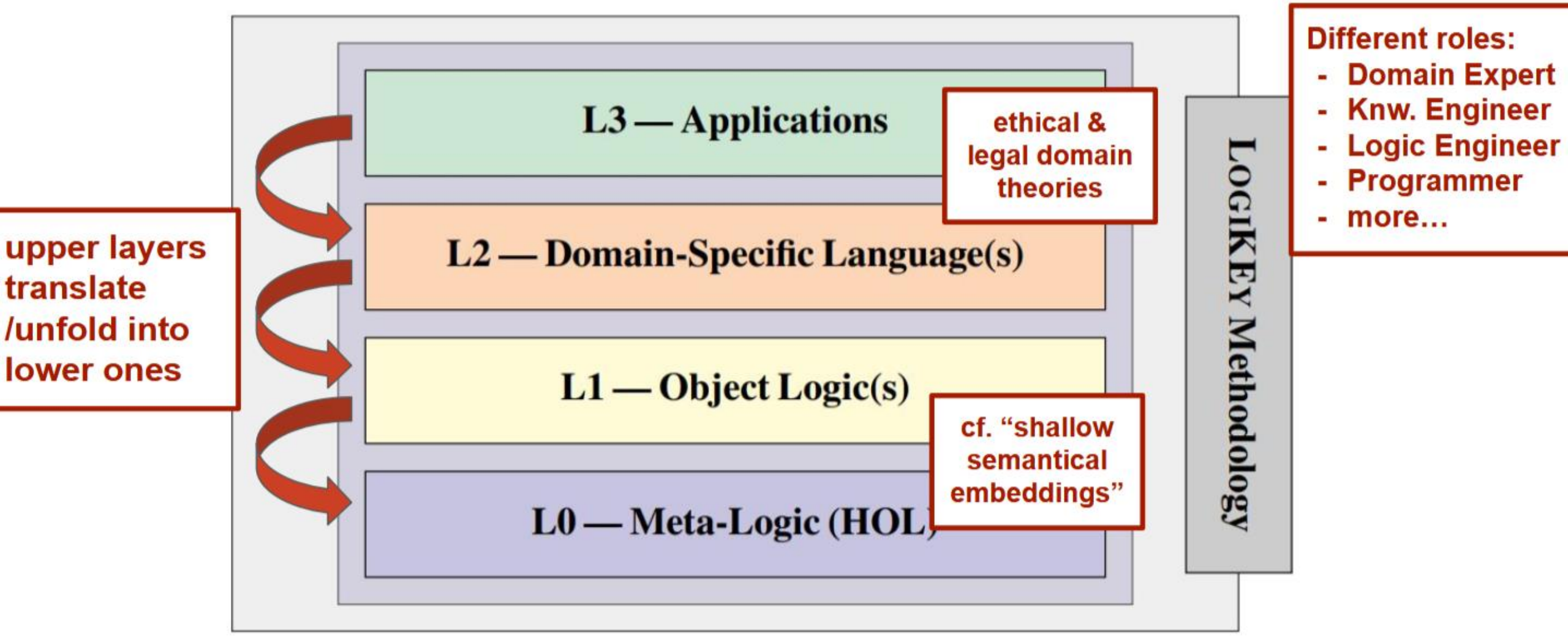
Volume 287, October 2020, 103348

Designing normative theories for ethical and legal reasoning: LogiKEy framework, methodology, and tool support ☆

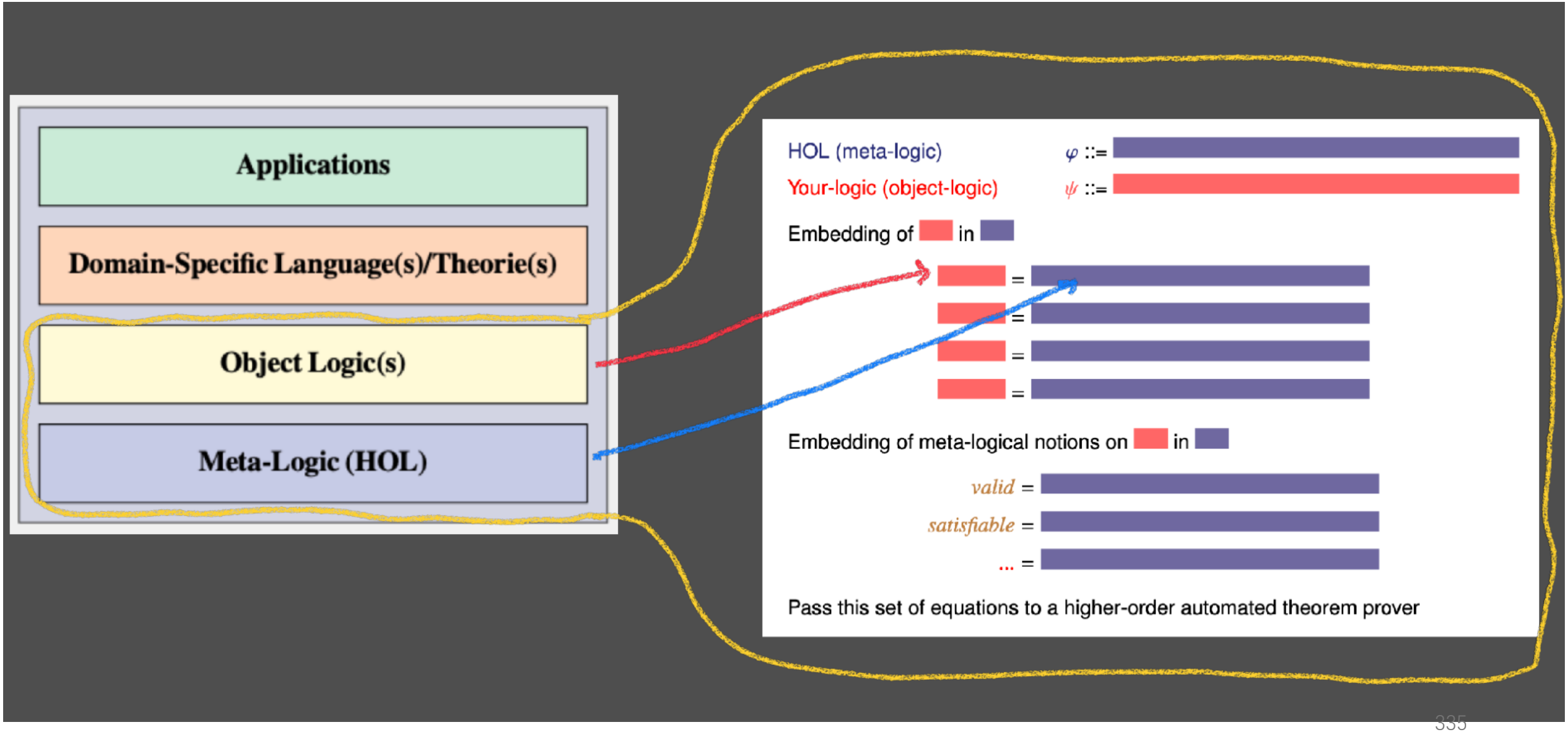
Christoph Benzmüller [b] [a], Xavier Parent [a], Leendert van der Torre [a] [c]



LogiKEy methodology



LogiKEy methodology



LogiKEy: Shallow Semantical Embeddings in HOL

HOL $\quad s,t \;::=\; c_\alpha \mid x_\alpha \mid (\lambda x_\alpha s_\beta)_{\alpha\to\beta} \mid (s_{\alpha\to\beta}\, t_\alpha)_\beta \mid \neg s_o \mid s_o \vee t_o \mid \forall x_\alpha\, t_o$

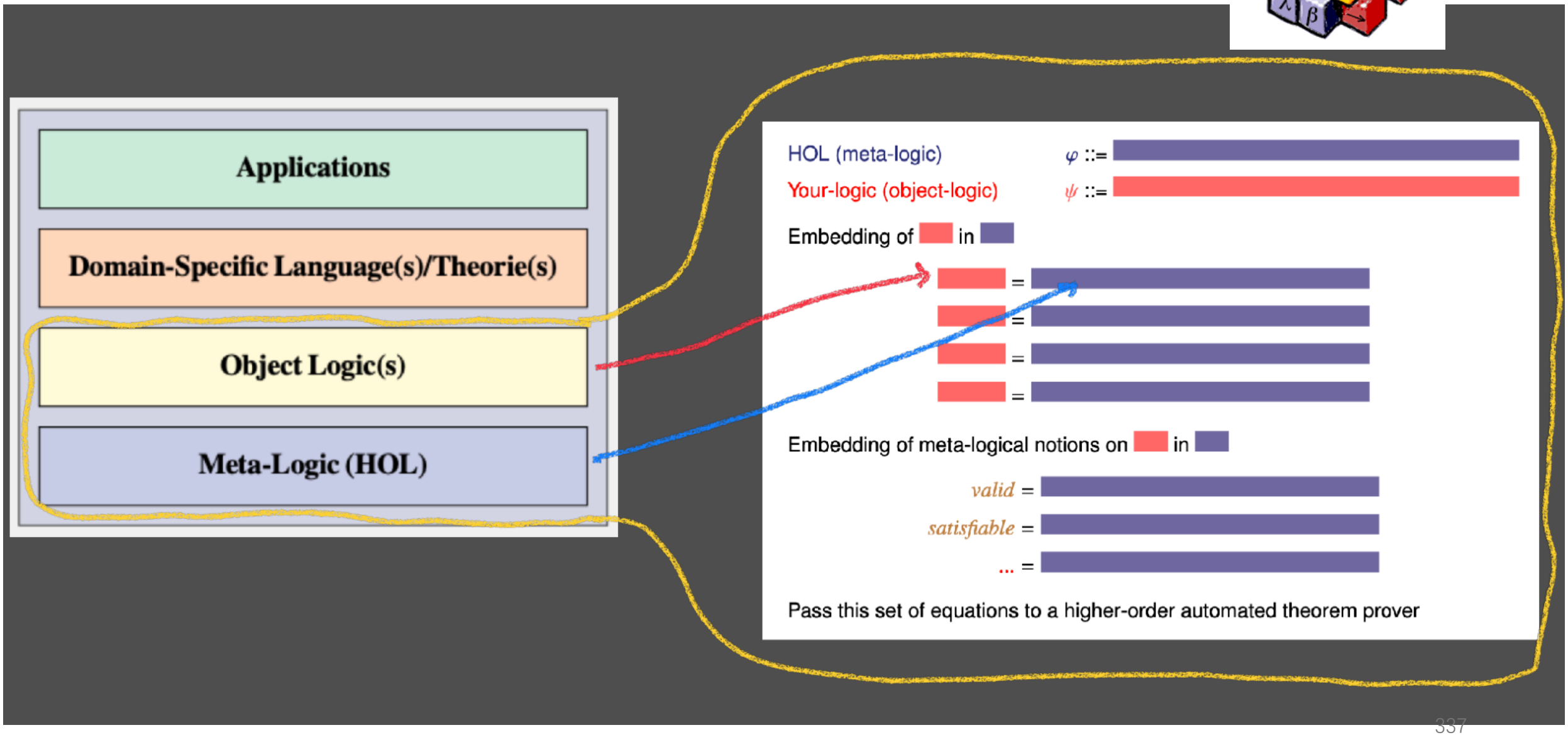
HOML $\quad \varphi,\psi \;::=\; \ldots \mid \neg\varphi \mid \varphi \wedge \psi \mid \varphi \to \psi \mid \Box\varphi \mid \Diamond\varphi \mid \forall x_\gamma\, \varphi \mid \exists x_\gamma\, \varphi$

HOML in HOL: HOML formulas φ are mapped to HOL predicates $\varphi_{\mu\to o}$
(explicit representation of labelled formulas)

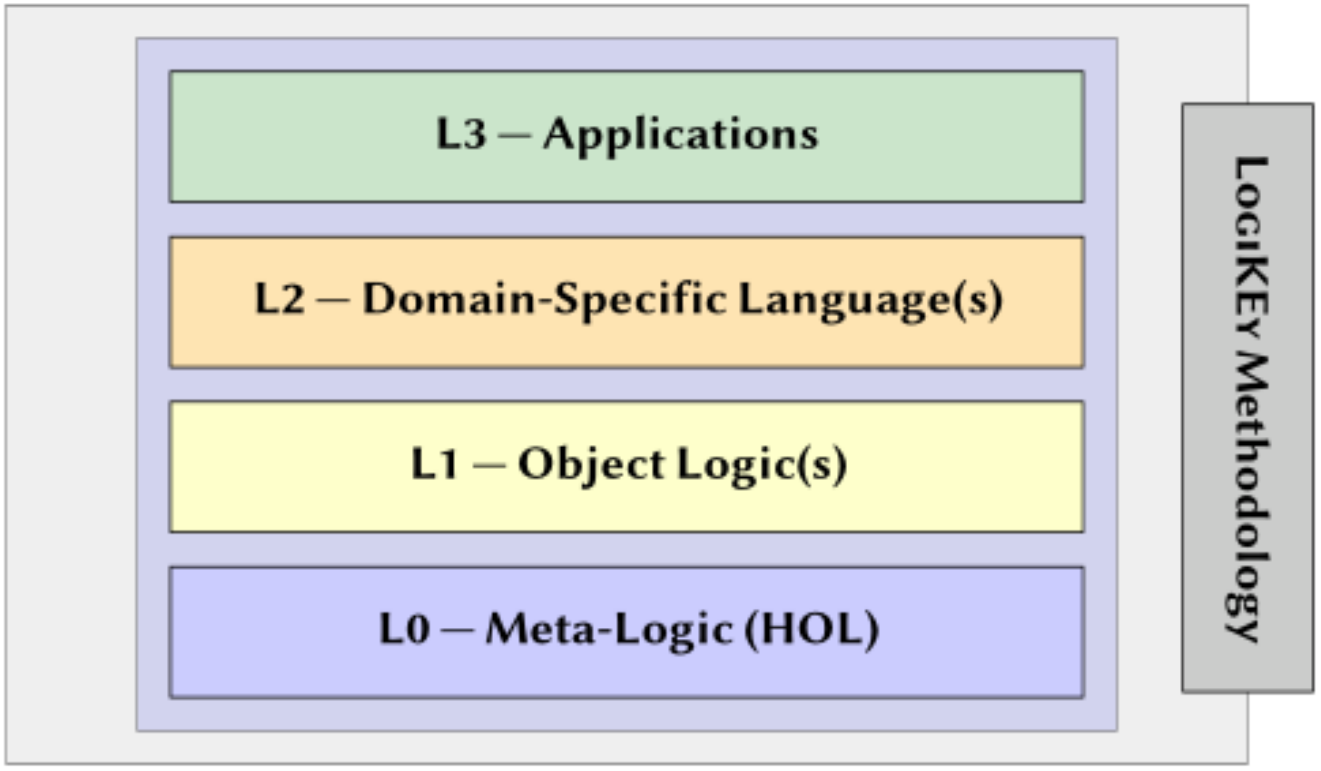
$$
\begin{aligned}
\neg &= \lambda\varphi_{\mu\to o}\lambda w_\mu \neg\varphi w\\
\wedge &= \lambda\varphi_{\mu\to o}\lambda\psi_{\mu\to o}\lambda w_\mu(\varphi w \wedge \psi w)\\
\to &= \lambda\varphi_{\mu\to o}\lambda\psi_{\mu\to o}\lambda w_\mu(\neg\varphi w \vee \psi w)\\
\forall &= \lambda h_{\gamma\to(\mu\to o)}\lambda w_\mu \forall d_\gamma\, hdw\\
\exists &= \lambda h_{\gamma\to(\mu\to o)}\lambda w_\mu \exists d_\gamma\, hdw\\
\Box &= \lambda\varphi_{\mu\to o}\lambda w_\mu \forall u_\mu\,(\neg rwu \vee \varphi u)\\
\Diamond &= \lambda\varphi_{\mu\to o}\lambda w_\mu \exists u_\mu\,(rwu \wedge \varphi u)\\
\text{valid} &= \lambda\varphi_{\mu\to o}\forall w_\mu \varphi w
\end{aligned}
$$

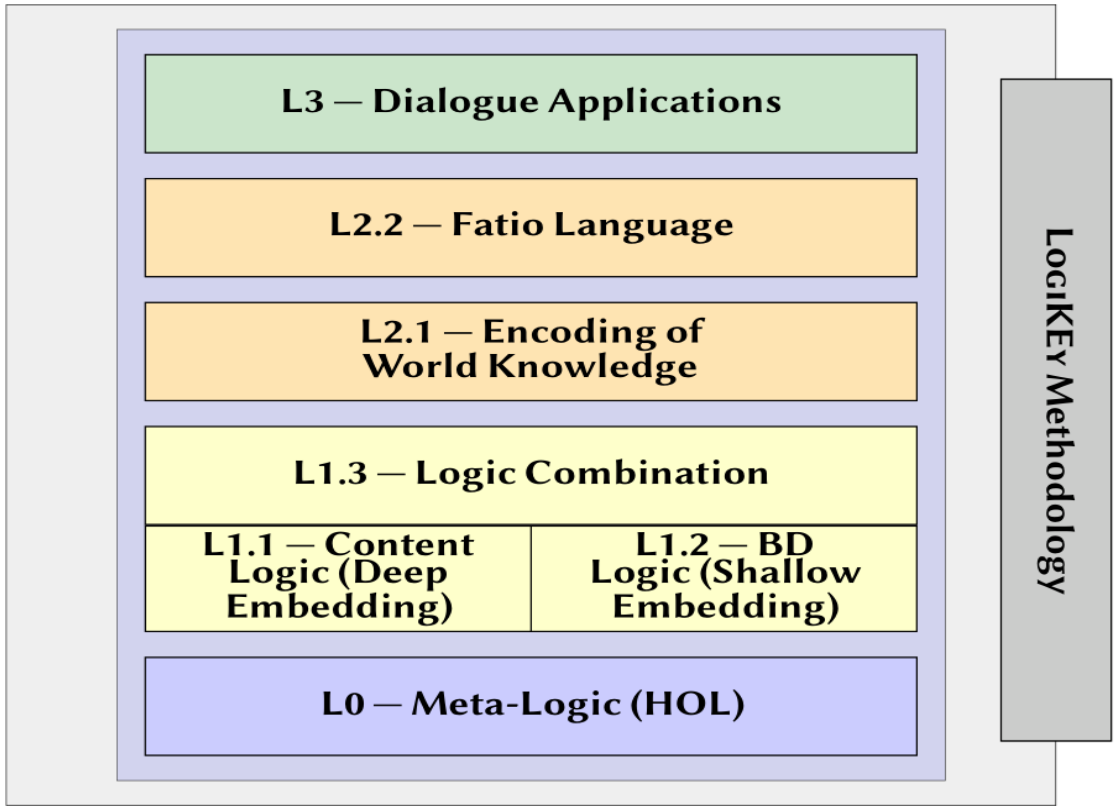
C. Benzmüller & L. Paulson (2013)
"Quantified Multimodal Logics in Simple Type Theory" Logica Universalis

LogiKEy methodology



Fatio in LogiKEy





Motivation: Generating Mental States from Dialogue

Task: Given a dialogue, automatically infer mental states of agents

Try with an LLM: Here is a dialogue between different agents A, B, and C:

A: The Brigade Restaurant is excellent
B: I disagree
A: I had a gr
C: Why do y
B:The vegeta y not be excellent

Architecture with:
- Model: semantics for communicating agents
- Integration of Subsymbolic AI (generation) and Symbolic AI (testing)

Give me beliefs and desires of each agent with respect to the other agents.
(ChatGPT 4o)

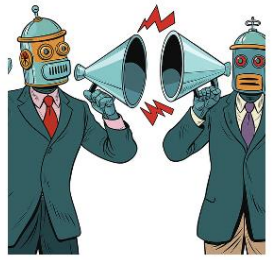
Beliefs of A	Desires of A
- The Brigade is excellent - A great meal is a sign of excellence	- Justify own positive evaluation - Reach mutual understanding - Accommodate others views

Back to: Generating Mental States from Dialogue

- **Task:** Given a dialogue, automatically infer mental states of agents, then test them formally

- Start from a dialogue
- 1. LLM prompting: generate mental states
  - ➢ Prompt Engineering
  - ➢ Chain of Thought
- 2. Isabelle/HOL validation: test mental states
  - ➢ Check whether the mental states align with the dialogue
- 3. Other uses of LLMs?

Subsymbolic AI



Symbolic AI

Dialogue → mining → Formal model, Machine-readable representation → reasoning → Automatic reasoner

Back to: Generating Mental States from Dialogue

Task: Given a dialogue, automatically infer mental states of agents, then test them formally

1. LLM prompting: Here is a dialogue between different agents A, B, and C:

- A: The Brigade Restaurant is excellent.
- B: I disagree, why is it excellent?
- A: I had a great meal at the Brigade.
- C: Why do you disagree?
- B: The vegetarian food at the Brigade is awful.
- A: I agree, the Brigade restaurant may not be excellent

This dialogue can be encoded using the following moves where P, Q, R, S are literals:

FORMALIZED DIALOGUE
FATIO AXIOMATIC SEMANTICS
CoT REASONING

Give me beliefs and desires of each agent with respect to the other agents. (ChatGPT 4o)

Formal Representation	Natural Language Description
$D_A B_B B_A R$	A desires that B believes A believes R (that the Brigade is excellent).
$D_A B_C B_A R$	A desires that C believes A believes R.

Back to: Generating Mental States from Dialogue

Task: Given a dialogue, automatically infer mental states of agents, then test them formally

1. LLM prompting: Here is a dialogue between different agents A, B, and C: **NL DIALOGUE**
This dialogue can be encoded using the following moves where P, Q, R, S are literals:
FORMALIZED DIALOGUE
FATIO AXIOMATIC SEMANTICS
CoT REASONING

Compare with previous output:

Beliefs of A	**Desires of A**
- The Brigade is excellent - A great meal is a sign of excellence	- Justify own positive evaluation - Reach mutual understanding - Accommodate others views

Formal Representation	Natural Language Description
$D_A B_B B_A R$	A desires that B believes A believes R (that the Brigade is excellent).
$D_A B_C B_A R$	A desires that C believes A believes R.

Back to: Generating Mental States from Dialogue

Task: Given a dialogue, automatically infer mental states of agents, then test them formally

1. LLM prompting

Formal Representation
$D_A B_B B_A R$
$D_A B_C B_A R$

2. Testing

Back to: Generating Mental States from Dialogue

Testing: Generated mental states satisfy pre-conditions?

- assert(A, φ) requires that
  $KB \models$ pre-cond(assert(A, φ))
- Consider agents A, B, C and
  $KB = \{(D_A B_B B_A \varphi), (D_A B_C B_A \varphi)\}$
- pre-cond(assert(A, φ)) $= (\forall j \neq A), (D_A B_j B_A \varphi)$.
- Then, $KB \models$ pre-cond(assert(A, φ))

Formal Representation
$D_A B_B B_A R$
$D_A B_C B_A R$

Back to: Generating Mental States from Dialogue

Testing: Generated mental states satisfy pre-conditions?

- assert(A, φ) requires that
  $KB \models$ pre-cond(assert(A, φ))
- Consider agents A, B, C and
  $KB = \{(D_A B_B B_A \varphi), (D_A B_C B_A \varphi)\}$
- pre-cond(assert(A, φ)) $= (\forall j \neq A), (D_A B_j B_A \varphi).$
- Then, $KB \models$ pre-cond(assert(A, φ))

Formal Representation
$D_A B_B B_A R$
$D_A B_C B_A R$

```
lemma
  "FatioCheckRec (
    [assert[a,r^a]],
    [],
    (
    λx. (x = D_a B_b B_a  {r^a}) ∨ (x = D_a B_c B_a  {r^a})
    )
   )"
  apply simp
  using Speaker.exhaust global_consequence_def member_rec(2)
  by (smt (verit) Speaker.exhaust global_consequence_def member_rec(2))
```

Day 4: agentic AI, Fatio and Libra

4.1. New foundations for multiple agents: individual vs collective defense

4.2. New foundations for time: reasons / arguments in progress

4.3. Fatio and Libra

4.4. End-to-end pipelines for traceability and auditability

4.5. Aligning Argumentation as Balancing, Dialogue and Inference (A-BDI)

Logic and Argumentation for New-generation AI

ESSLLI 2026

Liuwen Yu[1]
Leon van der Torre[2]
[1] Luxembourg Institute of Science and Technology
[2] University of Luxembourg

Recap
Day 4: agentic AI, Fatio and Libra

4.1. New foundations for multiple agents: individual vs collective defense

4.2. New foundations for time: reasons / arguments in progress

4.3. Fatio and Libra

4.4. End-to-end pipelines for traceability and auditability

4.5. Aligning Argumentation as Balancing, Dialogue and Inference (A-BDI)

Day 5 AI&Law and machine ethics

5.1. Key examples: all or nothing, divorce, discretion, …

5.2. The institution vs the individual: decision perspective

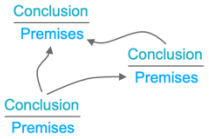
5.3. New foundations: bipolar argumentation, qualitative and quantified

5.4. New foundations: contrastive reasons, triples, dynamic scale, attack assignment

5.5. The logic toolbox and combining logics

Argumentation as inference and dialogue

"Systems for argumentation-based inference are about which conclusions can be drawn from a given body of possibly incomplete, inconsistent of uncertain information."



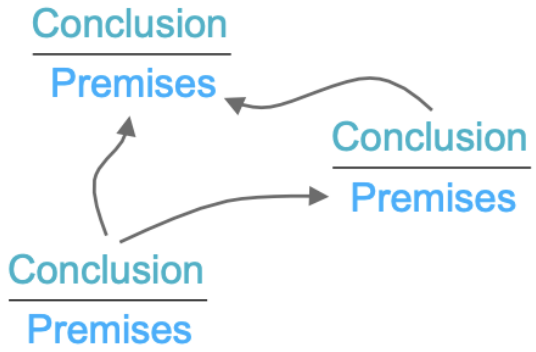
"Systems for argumentation-based dialogue model argumentation as a kind of verbal interaction aimed at resolving conflicts of opinion. They define argumentation protocols, that is, the rules of the argumentation game, and address matters of strategy, that is, how to play the game well."

----- *Henry Prakken, HOFA Volume 1 Chapter 2*

"Argumentation is a rational process, for making and justifying decisions of various kinds of issues, in which arguments pro and con alternative resolutions of the issues (options or positions) are put forward, evaluated, resolved and balanced. "

-----*Thomas F. Gordon HOFA Volume 1 Chapter 3*

Build bridges from one conceptual model to another?



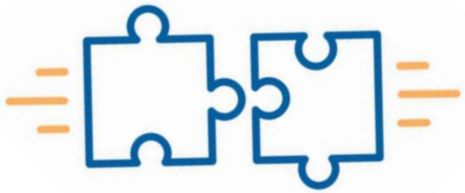
Thursday
Abstract Agent argumentation
A&C 2026

Libra
COMMA2026

inference

Argumentation as

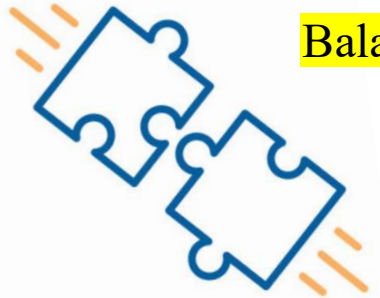
dialogue



balancing

Today
Bipolar argumentation
JELIA 2023

Balancing argumentation
AAMAS 2026



Bipolar argumentation

A Principle-Based Analysis of Bipolar Argumentation Semantics

Liuwen Yu[1 [0000−0002−7200−6001]], Caren Al Anaissy[4 [0000−0002−8750−1849]], Srdjan Vesic[5 [0000−0002−4382−0928]], Xu Li[1 [0000−0002−7788−2414]], and Leendert van der Torre[1 [0000−0003−4330−3717]]

[1] University of Luxembourg, Luxembourg
{name.surname}@uni.lu
[2] CRIL Université d'Artois & CNRS, France
alanaissy@cril.fr
[3] CRIL CNRS Univ. Artois, France
vesic@cril.fr

Abstract. In this paper, we introduce and study seven types of semantics for bipolar argumentation frameworks, each extending Dung's interpretation of attack with a distinct interpretation of support. First, we introduce three types of defence-based semantics by adapting the notions of defence. Second, we examine two types of selection-based semantics that select extensions by counting the number of supports. Third, we analyse two types of traditional reduction-based semantics under deductive and necessary interpretations of support. We provide full analysis of twenty-eight bipolar argumentation semantics and ten principles.

Keywords: Bipolar argumentation semantics · Support · Principle-based approach · Knowledge representation and reasoning.

Purpose of the paper

- Different approaches to extending Dung's abstract theory:
  - Reduction-based approach
    - Deductive Interpretation of support
    - Necessary Interpretation of support
- Unchallenged evidences, facts
- Support seen as inference relation

Research Questions

- In which ways can support affect attack, defence and argumentation semantics?
- Which principles can be introduced to distinguish between, and characterise, these semantics?

Contributions

- Propose three defence-based semantics by adapting the notion of defence.
- Propose two selection-based semantics that select the extensions that have the largest number of internal/external supports.
- Introduce and study ten principles
- Provide full analysis of the proposed bipolar argumentation semantics and the ten principles

Argumentation as balancing

Argumentation as balancing:
Multi-criteria decision theory, ethical theory, legal theory, case based reasoning

Argumentation as balancing

Argumentation as balancing:
Multi-criteria decision theory, ethical theory, legal theory, case based reasoning

We need support at the same level as attack

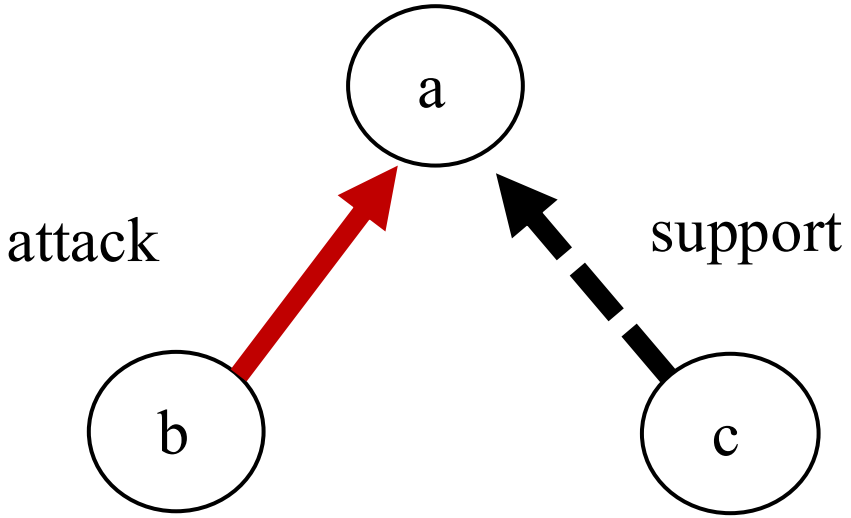
Argumentation as balancing

Argumentation as balancing:
Multi-criteria decision theory, ethical theory, legal theory, case based reasoning

We need support at the same level as attack

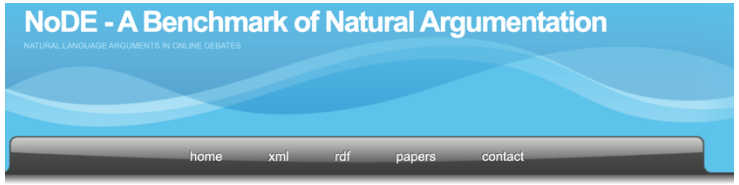
Argumentation as balancing

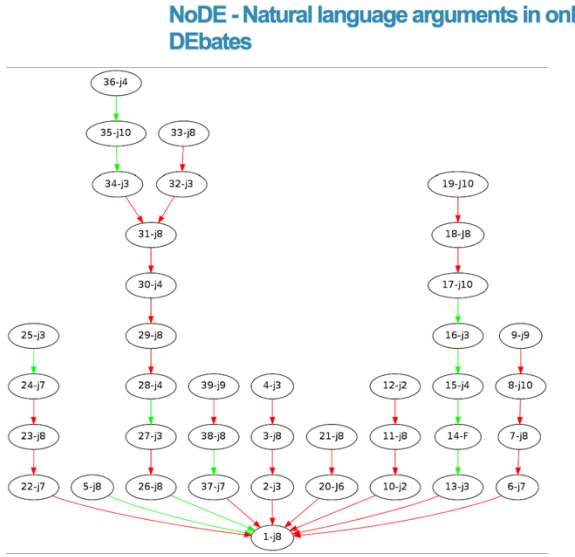
Argumentation as balancing:
Multi-criteria decision theory, ethical theory, legal theory, case based reasoning



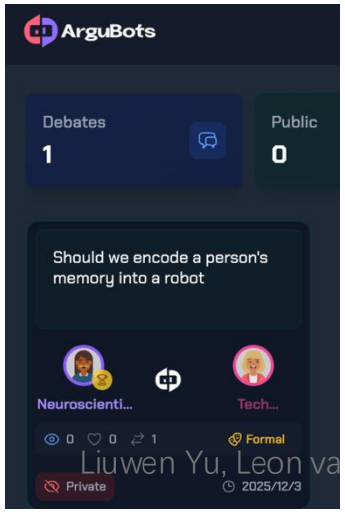
A bipolar argumentation framework

Applications of bipolar argumentation





Chatbot name	*ArguBot*
Authors	*Bistarelli, Santini, Taticchi*
Chatbot purpose	*Conversation (either supporting or challenging a specific topic)*
Chatbot type	*Retrieval-based (closed domain) chatbot*
Chatbot main features	*Developed using Google **DialogFlow**, ArguBot employs **ASPARTIX** to compute arguments from an underlying Bipolar AF to support or challenge a specific topic.* *The conversational capabilities of ArguBot are, however, restricted by the arguments stored in the BAF as its knowledge base, limiting its dialectical potential only to specific fully-developed interactions.*



Introduction: Bipolar argumentation

1. Extend Dung's theory with support
2. Three kinds of semantics
   - ➢ Defense-based
   - • Define new notion of defense using both support and attack
   - ➢ Selection-based
   - • Support is used only to select some of the extensions provided in Dung's semantics
   - ➢ Reduction-based
   - • Introduce indirect attacks based on the interpretation of support

Yu, Liuwen, et al. "A principle-based analysis of bipolar argumentation semantics." *European Conference on Logics in Artificial Intelligence*. Cham: Springer Nature Switzerland, 2023.

New semantics 1: Defense-based semantics

- adapt Dung's defense

1. argument a is defended by a set E if for every argument b attacking a, there is an argument c in E both attacks b and supports a

New semantics 1: Defense-based semantics

- adapt Dung's defense

1. argument a is defended by a set E if for every argument b attacking a, there is an argument c in E both attacks b and supports a

Example (Funding for renewable energy):

Argument a: "We should increase funding for renewable energy."

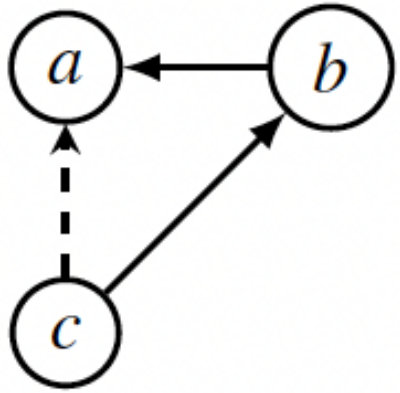
Argument b: "Renewable energy is too costly and inefficient."

Argument c: "Renewable energy technology is rapidly advancing and becoming more cost-effective."

New semantics 1: Defense-based semantics

- adapt Dung's defense

1. argument a is defended by a set E if for every argument b attacking a, there is an argument c in E both attacks b and supports a



{c} defends a by attacking b and supporting a

New semantics 1: Defense-based semantics

- adapt Dung's defense

2. argument a is defended by a set E if for every argument b attacking a, there is argument c and d in E such that c attacks b and d supports c

New semantics 1: Defense-based semantics

- adapt Dung's defense

2. argument a is defended by a set E if for every argument b attacking a, there is argument c and d in E such that c attacks b and d supports c

Example (Public health policy):

Argument a: "Vaccinations should be mandatory for all children."

Argument b: "Vaccinations can have severe side effects."

Argument c: "Severe side effects are extremely rare."

Argument d: "Extensive research and data support the rarity of severe side effects from vaccinations."

New semantics 1: Defense-based semantics

- adapt Dung's defense

2. argument a is defended by a set E if for every argument b attacking a, there is argument c and d in E such that c attacks b and d supports c

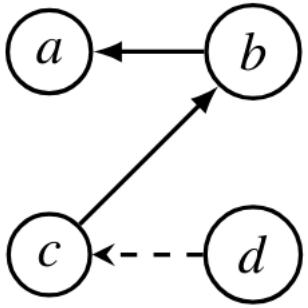
Example (Public health policy):

Argument a: "Vaccinations should be mandatory for all children."

Argument b: "Vaccinations can have severe side effects."

Argument c: "Severe side effects are extremely rare."

Argument d: "Extensive research and data support the rarity of severe side effects from vaccinations."



{c,d} defends a by c attacking b and d supporting c

New semantics 1: Defense-based semantics

- adapt Dung's defense

3. argument a is defended by a set E if for every argument b attacking a, there is argument c and e in E such that c attacks b and e attacks the supporters of b.

New semantics 1: Defense-based semantics

- adapt Dung's defense

3. argument a is defended by a set E if for every argument b attacking a, there is argument c and e in E such that c attacks b and e attacks the supporters of b.

Example (Corporate litigation):

Argument a: "The corporation engaged in illegal dumping of toxic waste."

Argument b: "The substances dumped were non-toxic and approved for disposal."

Argument c: "Scientific analysis shows that the substances are toxic and harmful to environment."

Argument d: "An environmental agency had categorized the substances as non-hazardous."

Argument e: "Recent findings indicate that the environmental agency's classifications were based on outdated or manipulated data."

New semantics 1: Defense-based semantics

- adapt Dung's defense

3. argument a is defended by a set E if for every argument b attacking a, there is argument c and e in E such that c attacks b and e attacks the supporters of b.

Example (Corporate litigation):

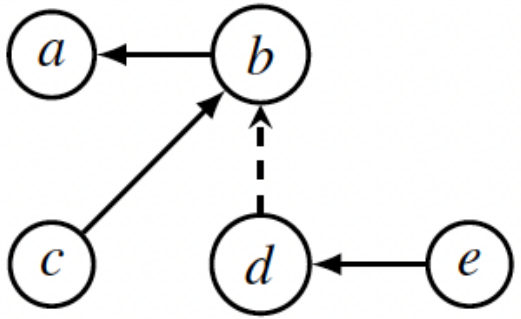
Argument a: "The corporation engaged in illegal dumping of toxic waste."

Argument b: "The substances dumped were non-toxic and approved for disposal."

Argument c: "Scientific analysis shows that the substances are toxic and harmful to environment."

Argument d: "An environmental agency had categorized the substances as non-hazardous."

Argument e: "Recent findings indicate that the environmental agency's classifications were based on outdated or manipulated data."



{c,e} defends a by attacking b and attacking d which is supporting b

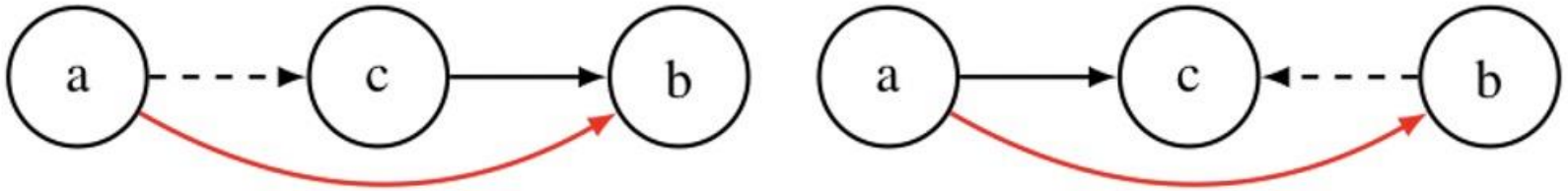
New semantics 2: Reduction-based semantics

- is based on different interpretation of support
  1. Various indirect attacks are introduced based on different interpretations

New semantics 2: Reduction-based semantics

- is based on different interpretation of support
  1. Various indirect attacks are introduced based on different interpretations

Deductive support: if a supports b, then the acceptance of a implies the acceptance of b;

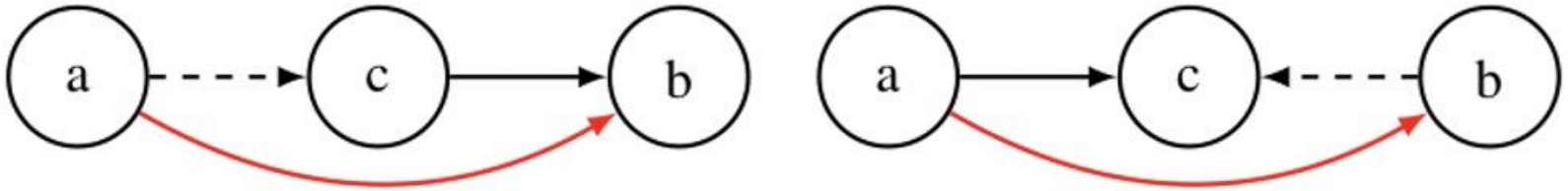


(a) Supported attack (b) Mediated attack

New semantics 2: Reduction-based semantics

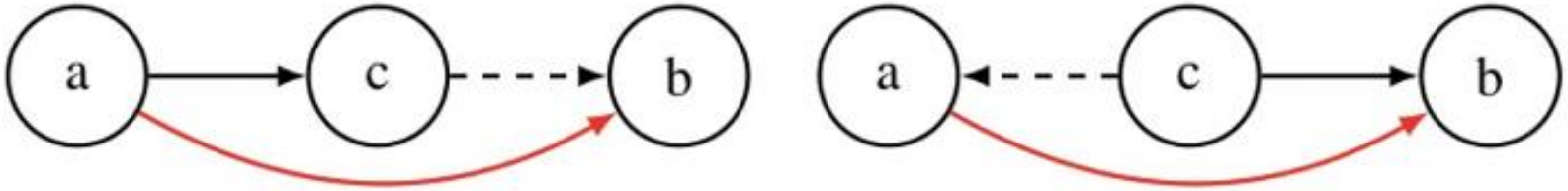
- is based on different interpretation of support
  1. Various indirect attacks are introduced based on different interpretations

Deductive support: if a supports b, then the acceptance of a implies the acceptance of b;



(a) Supported attack (b) Mediated attack

Necessary support: if a supports b, then the acceptance of a is necessary to get the acceptance of b;

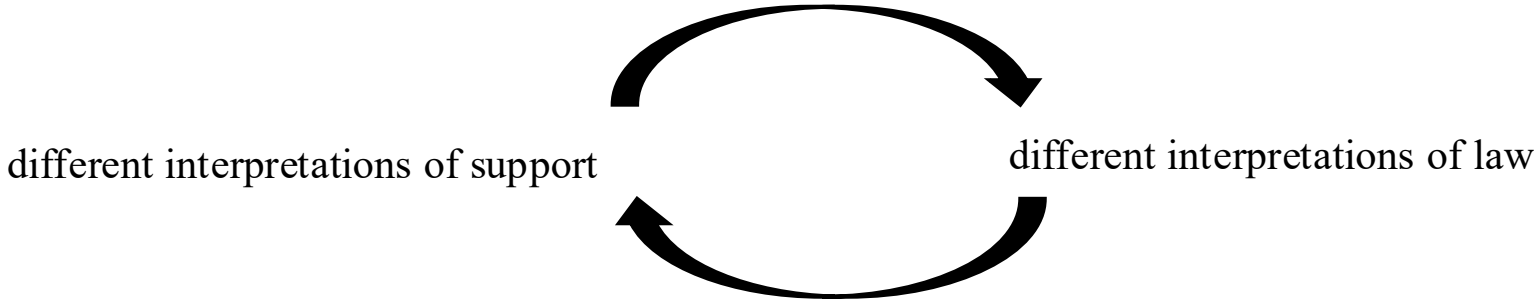


(c) Secondary attack (d) Extended attack

New semantics 2: Reduction-based semantics

- is based on different interpretation of support
  1. Various indirect attacks are introduced based on different interpretations

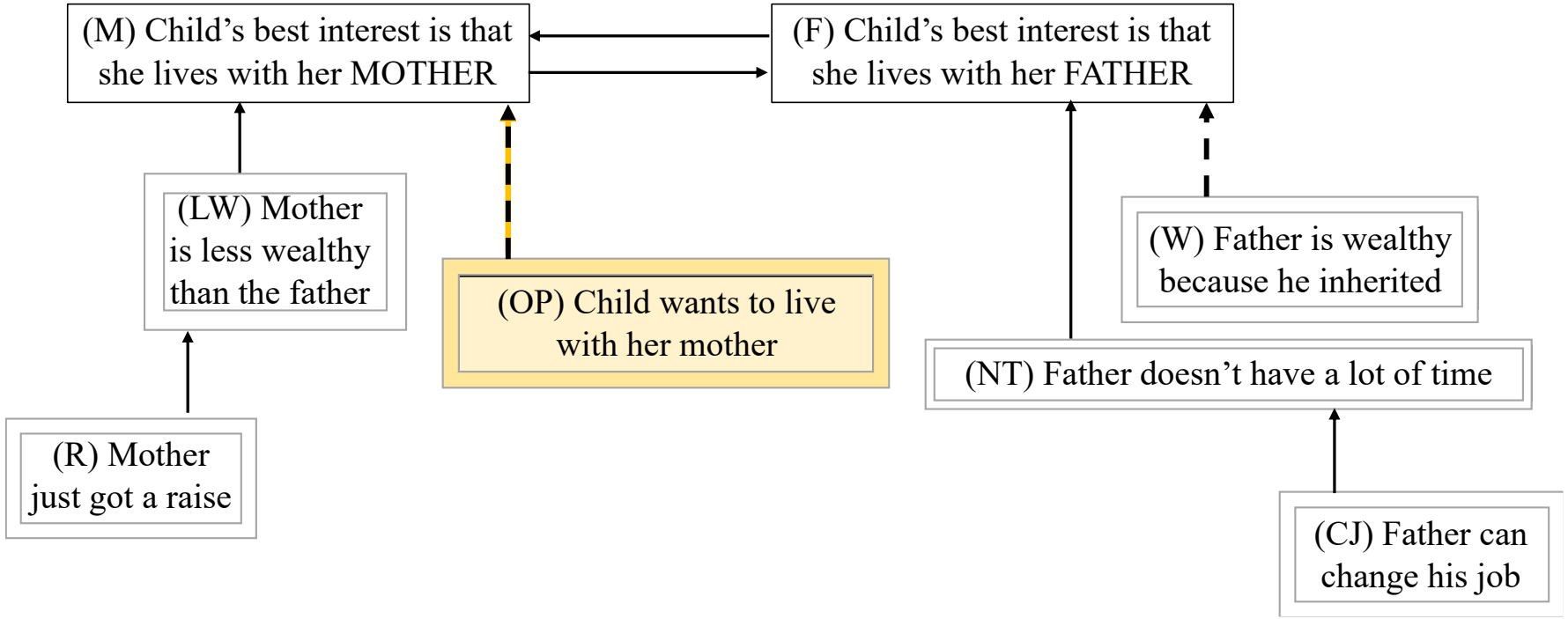
Deductive support: if a supports b, then the acceptance of a implies the acceptance of b;
Necessary support: if a supports b, then the acceptance of a is necessary to get the acceptance of b;



New semantics 2: Reduction-based semantics

- is based on different interpretation of support
  1. Various indirect attacks are introduced based on different interpretations

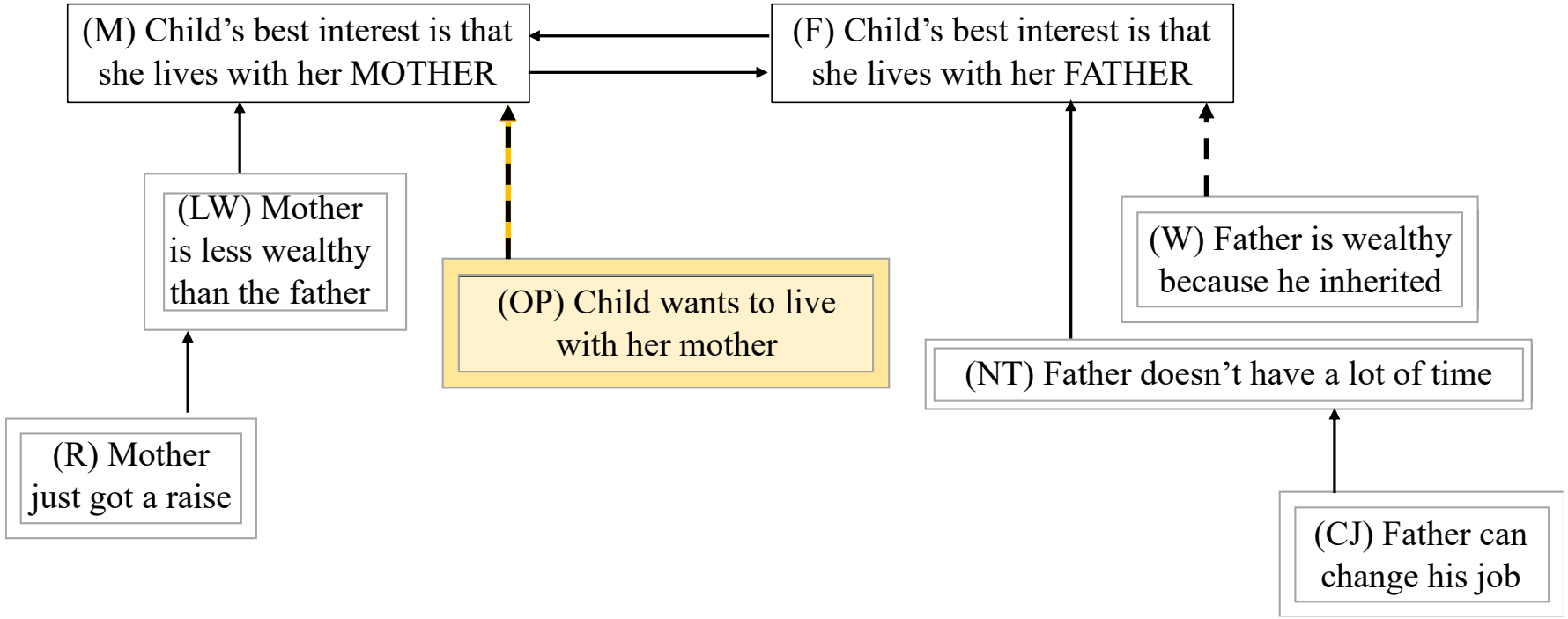
Deductive support: if a supports b, then the acceptance of a implies the acceptance of b;
Necessary support: if a supports b, then the acceptance of a is necessary to get the acceptance of b;



New semantics 2: Reduction-based semantics

- is based on different interpretation of support
  1. Various indirect attacks are introduced based on different interpretations

Deductive support: if a supports b, then the acceptance of a implies the acceptance of b;
Necessary support: if a supports b, then the acceptance of a is necessary to get the acceptance of b;



Civil Code:

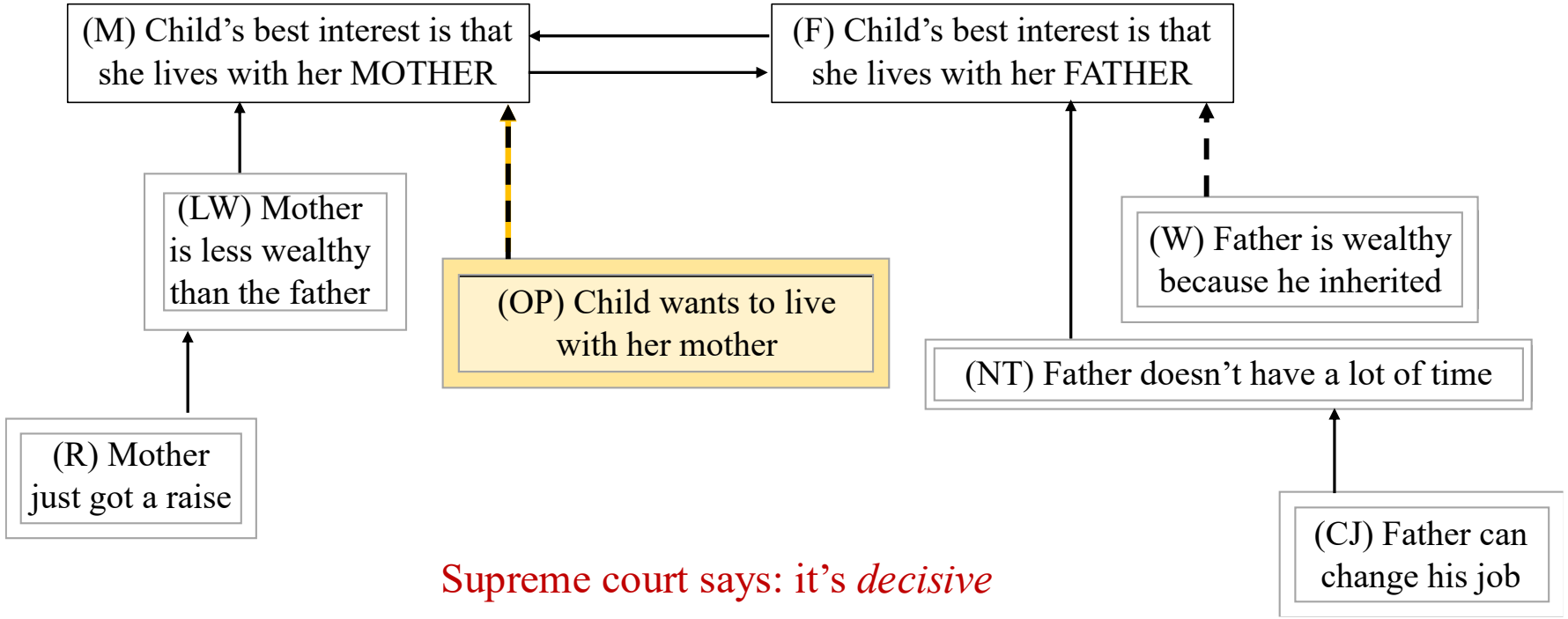
the judge has to take the child's opinion into consideration when deciding about custody

What does it mean in terms of interpreting the support relation?

New semantics 2: Reduction-based semantics

- is based on different interpretation of support
  1. Various indirect attacks are introduced based on different interpretations

Deductive support: if a supports b, then the acceptance of a implies the acceptance of b;
Necessary support: if a supports b, then the acceptance of a is necessary to get the acceptance of b;



Supreme court says: it's *decisive*

Civil Code:

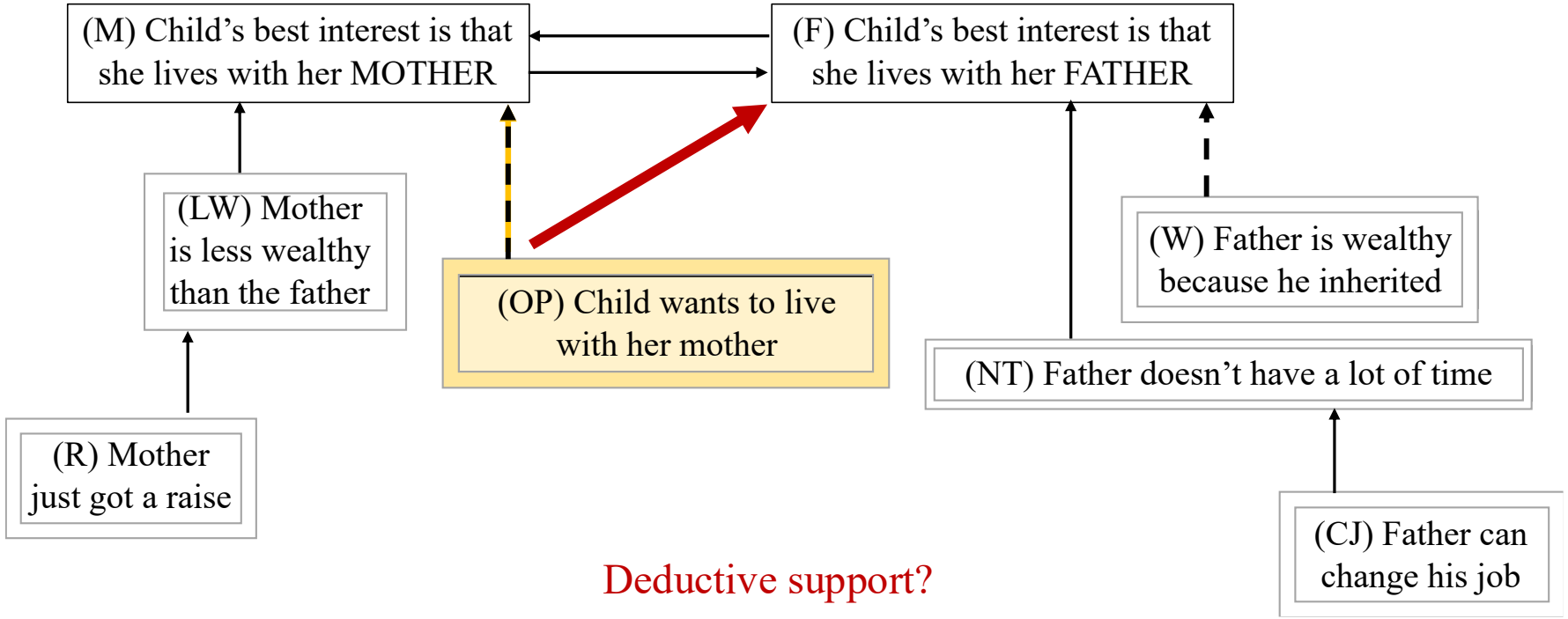
the judge has to take the child's opinion into consideration when deciding about custody

What does it mean in terms of interpreting the support relation?

New semantics 2: Reduction-based semantics

- is based on different interpretation of support
  1. Various indirect attacks are introduced based on different interpretations

Deductive support: if a supports b, then the acceptance of a implies the acceptance of b;
Necessary support: if a supports b, then the acceptance of a is necessary to get the acceptance of b;



Deductive support?

Civil Code:

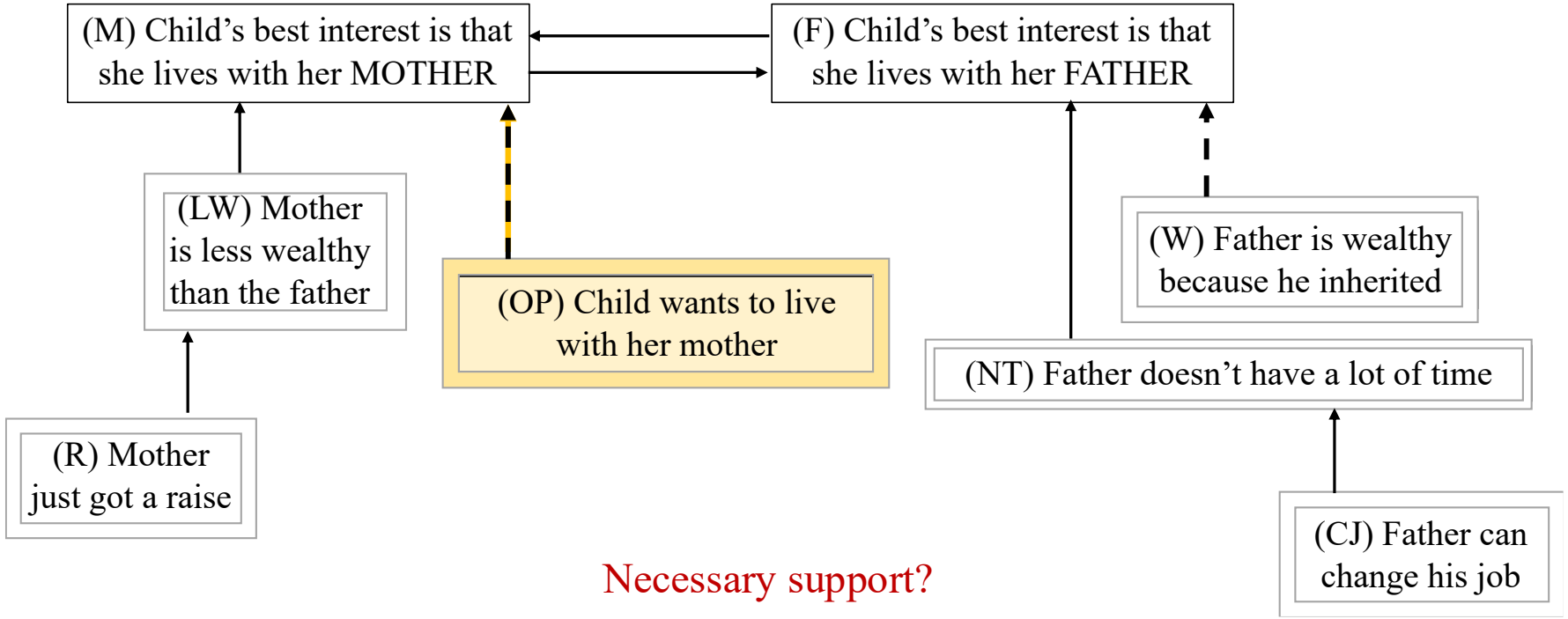
the judge has to take the child's opinion into consideration when deciding about custody

What does it mean in terms of interpreting the support relation?

New semantics 2: Reduction-based semantics

- is based on different interpretation of support
  1. Various indirect attacks are introduced based on different interpretations

Deductive support: if a supports b, then the acceptance of a implies the acceptance of b;
Necessary support: if a supports b, then the acceptance of a is necessary to get the acceptance of b;



Necessary support?

Civil Code:

the judge has to take the child's opinion into consideration when deciding about custody

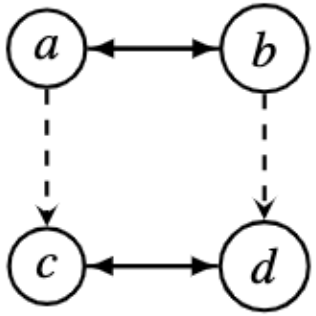
What does it mean in terms of interpreting the support relation?

New semantics 3: Selection-based semantics

- is based on coalition and voting
  1. select extensions based on internal support (number of support inside extensions)
  2. select extensions based on external support (number of support from outside extensions)

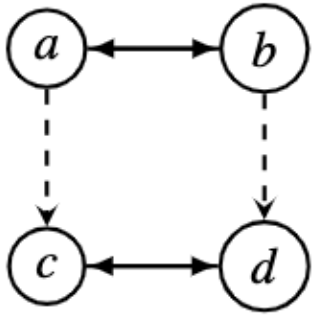
New semantics 3: Selection-based semantics

- is based on coalition and voting
  1. select extensions based on internal support (number of support inside extensions)
  2. select extensions based on external support (number of support from outside extensions)



New semantics 3: Selection-based semantics

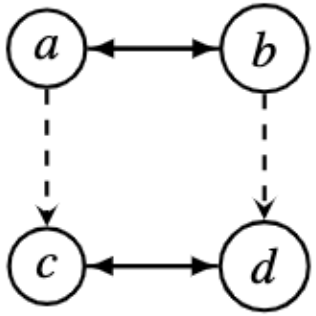
- is based on coalition and voting
  1. select extensions based on internal support (number of support inside extensions)
  2. select extensions based on external support (number of support from outside extensions)



There are two stable extensions {a, c}, {a, d}, {b, c}, {b, d} under Dung's theory
{a, c}, {b, d} are the extension of this BAF, because
{a, c}: 1 internal support
{a, d}: 0 internal support
{b, c}: 0 internal support
{b, d}: 1 internal support

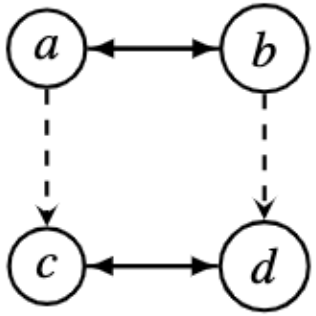
New semantics 3: Selection-based semantics

- is based on coalition and voting
  1. select extensions based on internal support (number of support inside extensions)
  2. select extensions based on external support (number of support from outside extensions)



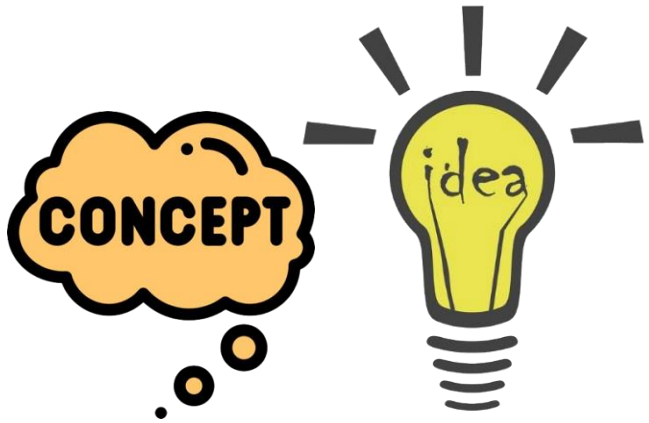
New semantics 3: Selection-based semantics

- is based on coalition and voting
  1. select extensions based on internal support (number of support inside extensions)
  2. select extensions based on external support (number of support from outside extensions)

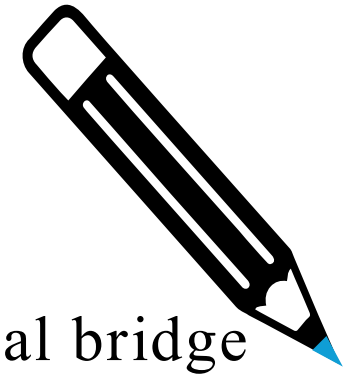


There are two stable extensions {a, c}, {a, d}, {b, c}, {b, d} under Dung's theory
{a, d}, {b, c} are the extension of this BAF, because
{a, c}: 0 internal support
{a, d}: 1 internal support
{b, c}: 1 internal support
{b, d}: 0 internal support

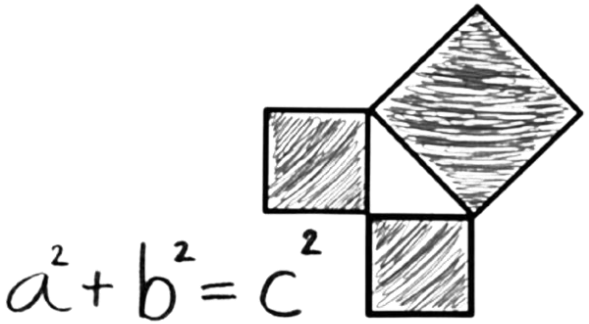
Knowledge representation and reasoning



Philosophy



Methodological bridge



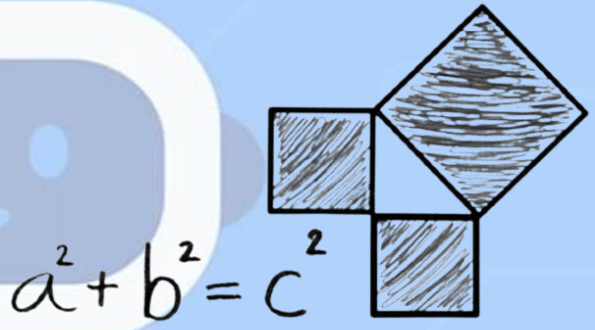
theorem

Mathematics

Knowledge representation and reasoning

Philosophy

Methodological bridge



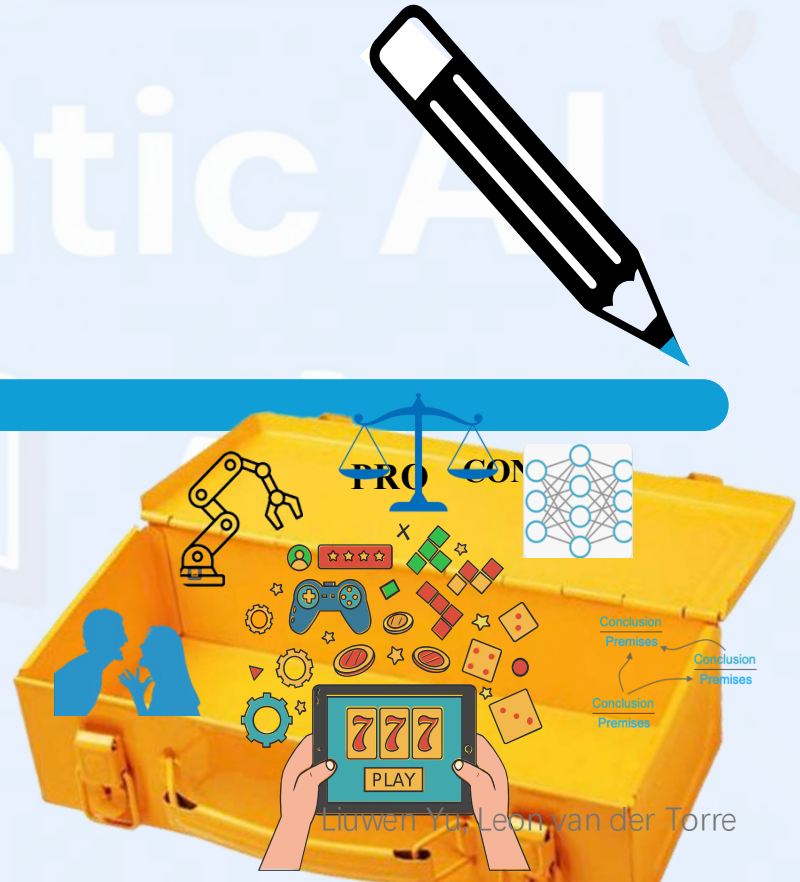
theorem

Mathematics

Reasons and arguments

Reasons
Pro and con
Weight of reasons
(im)Permissibility
obligation
…

Reason-based
normative reasoning



Arguments and counter-arguments
Strength of arguments
Attack
Support
Acceptability of arguments
…

Formal argumentation

Reasons and arguments

Reasons
Pro and con
Weight of reasons
(im)Permissibility
obligation
…

Arguments and counter-arguments
Strength of arguments
Attack
Support
Acceptability of arguments
…

Balancing for Agent Decision Making through Argumentation

Authors: Liuwen Yu, Chenyang Cai, Leon Van der Torre, Réka Markovich | Authors Info & Claims

AAMAS 2026: Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems • Pages 2876 - 2885
https://doi.org/10.65109/VBZK6091

Published: 24 May 2026 Publication History | Check for updates

Reason-based
normative reasoning

Formal argumentation

Simplified normative reasoning in philosophy

Simplified normative reasoning in philosophy

Reason first theory

Simplified normative reasoning in philosophy

Reason first theory

Reasons as defaults
(John F. Horty 2012)

Simplified normative reasoning in philosophy

Reason first theory

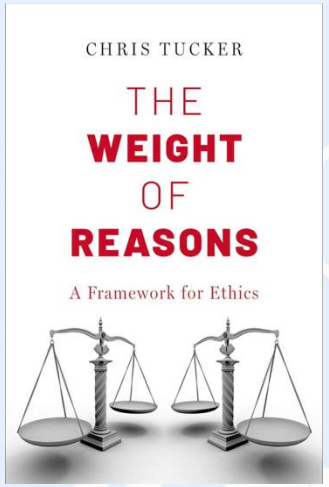
Reasons as defaults
(John F. Horty 2012)

The weight of reasons
(Chris Tucker 2025)

Different output types

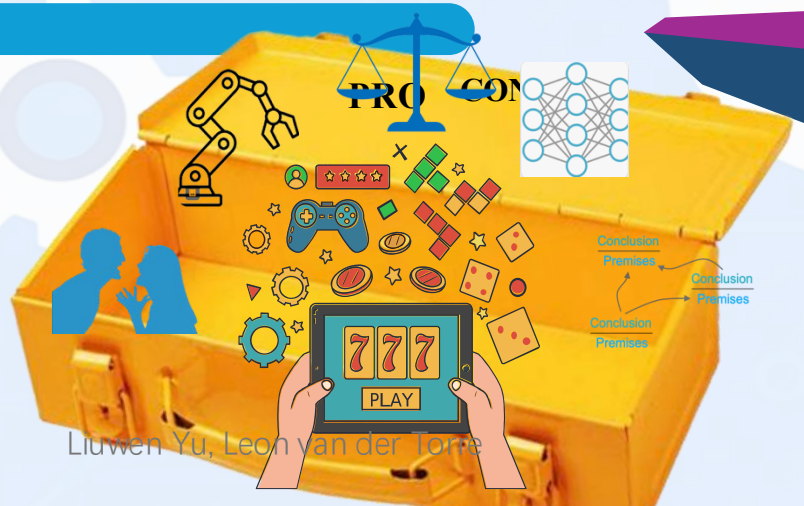
Output type	Example
Deontic status	permitted / impermissible / required
Ranking	better than / worse than
Choice set	selected options
Argument status	accepted / rejected /rankings/ weight

Formalization of Tucker's theory

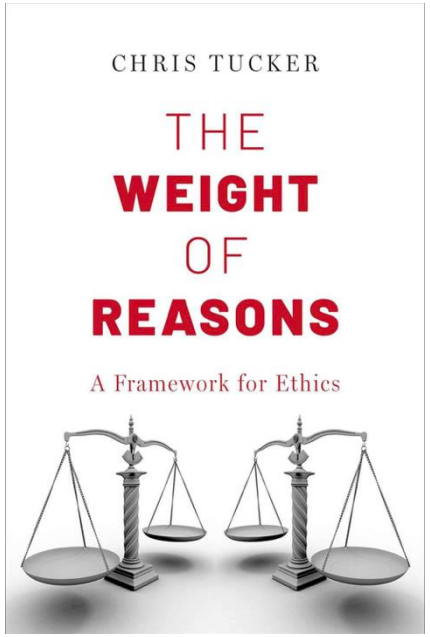


Reason-based normative reasoning

Formal argumentation



Tucker's Metaphor: **Tournament** and Dual Scale



Tournament: ϕ is permissible by winning a tournament, a pairwise competition with each alternative

Contrastivism: "for some possible ϕ, the reason for ϕ can vary as you vary the alternative.

Contrastivism

For some possible ϕ, the reason for ϕ can vary as you vary the alternative.

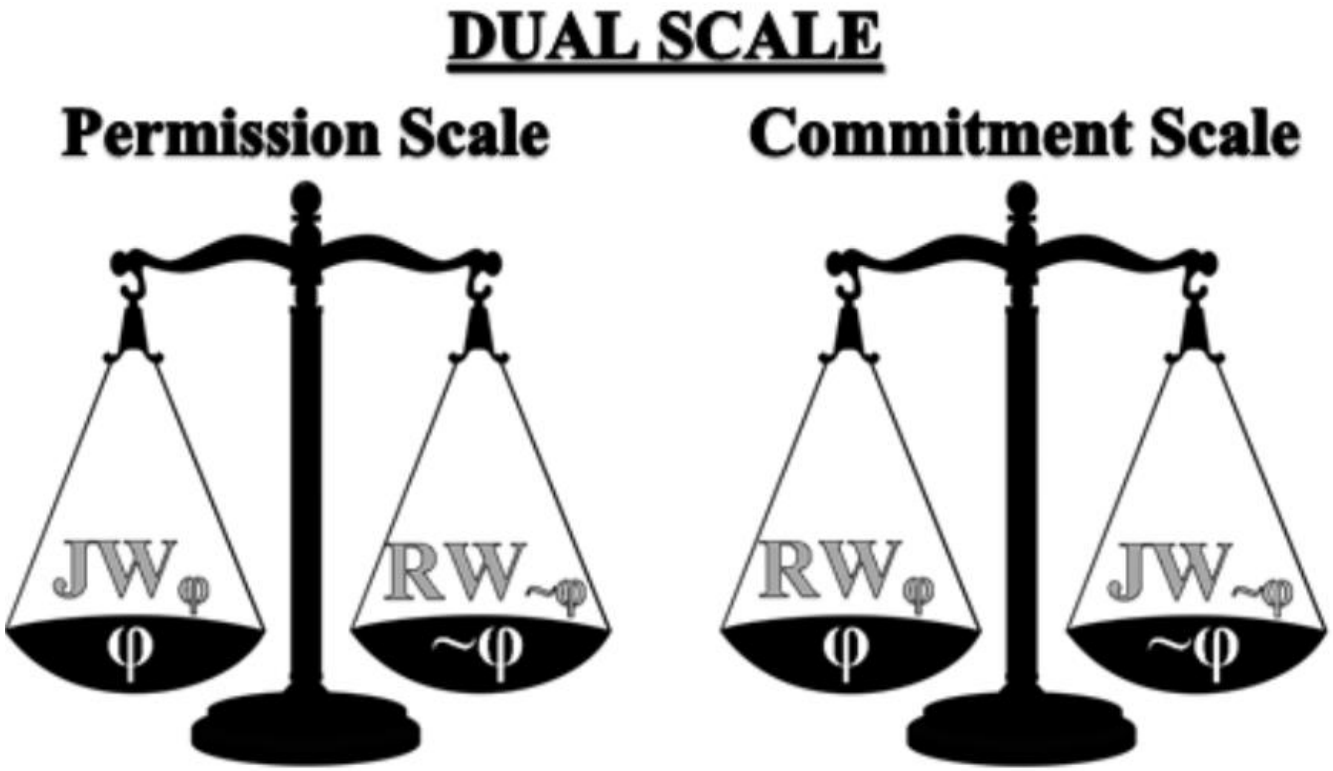
Definition (Comparative reasons)

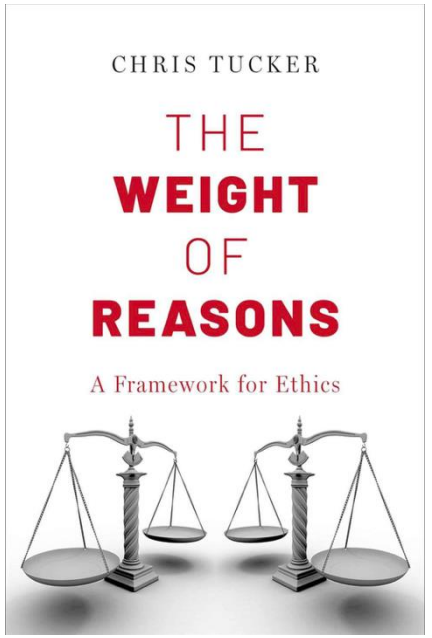
Let G and O be two sets called grounds and options, respectively. The set of reasons R_{cr} is a three-place relation

$$R_{cr} \subseteq G \times O \times O,$$

where $(g, o_1, o_2) \in R_{cr}$ implies that $o_1 \neq o_2$. We say that the ground g is a reason for option o_1 compared to option o_2.

Tucker's Metaphor: Tournament and **Dual Scale**





justifying weight: force that pushes an option toward permissibility
requiring weight: force that pushes the alternative toward impermissibility

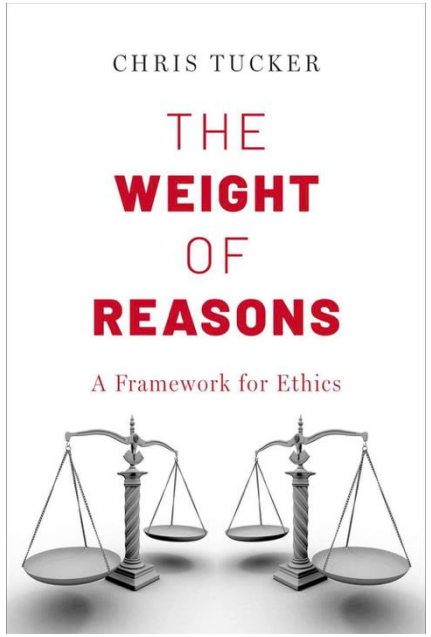
$$JW(a,b)\ is\ pro(a), RW(a,b)\ is\ \ con(b).$$

All or Nothing Example

Option 1 (Bystander): bystander and avoid getting burned

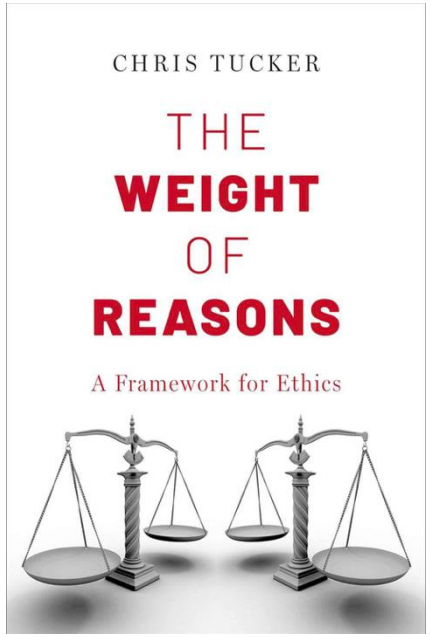
Option 2 (Save 1): save one child and get severe burns

Option 3 (Save 2): save two children and get severe burns

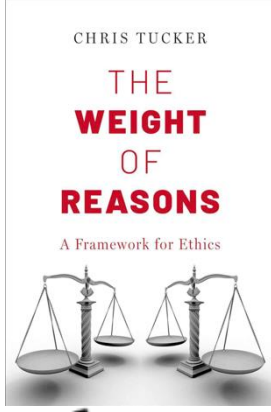


All or Nothing Example

Option 1 (Bystander): bystander and avoid getting burned
Option 2 (Save 1): save one child and get severe burns
Option 3 (Save 2): save two children and get severe burns

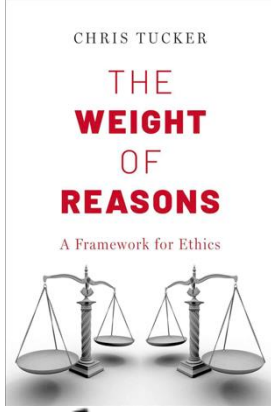


All or Nothing Example



In my view, it is a huge cost to deny Right > Wrong. I say instead that the All or Nothing Problem is a ranking reversal. In the two-option case, Save1 is morally better than Bystander. In the three-option case, Bystander is morally better than Save1. The addition of Save2 somehow causes a ranking reversal.

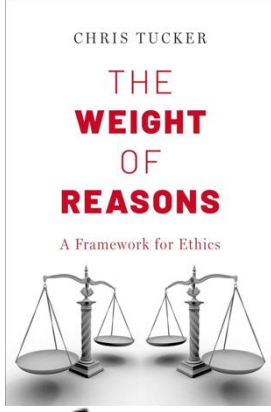
All or Nothing Example



In my view, it is a huge cost to deny Right > Wrong. I say instead that the All or Nothing Problem is a ranking reversal. In the two-option case, Save1 is morally better than Bystander. In the three-option case, Bystander is morally better than Save1. The addition of Save2 somehow causes a ranking reversal.

We use binary relation which we call "attack", and then define reinstatement

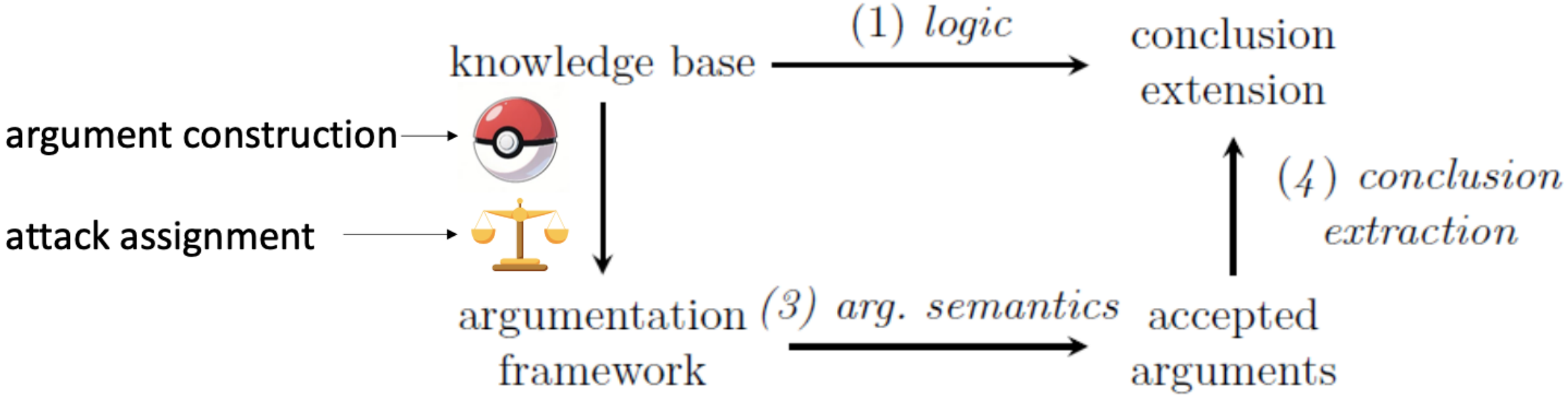
All or Nothing Example



In my view, it is a huge cost to deny Right > Wrong. I say instead that the All or Nothing Problem is a ranking reversal. In the two-option case, Save1 is morally better than Bystander. In the three-option case, Bystander is morally better than Save1. The addition of Save2 somehow causes a ranking reversal.

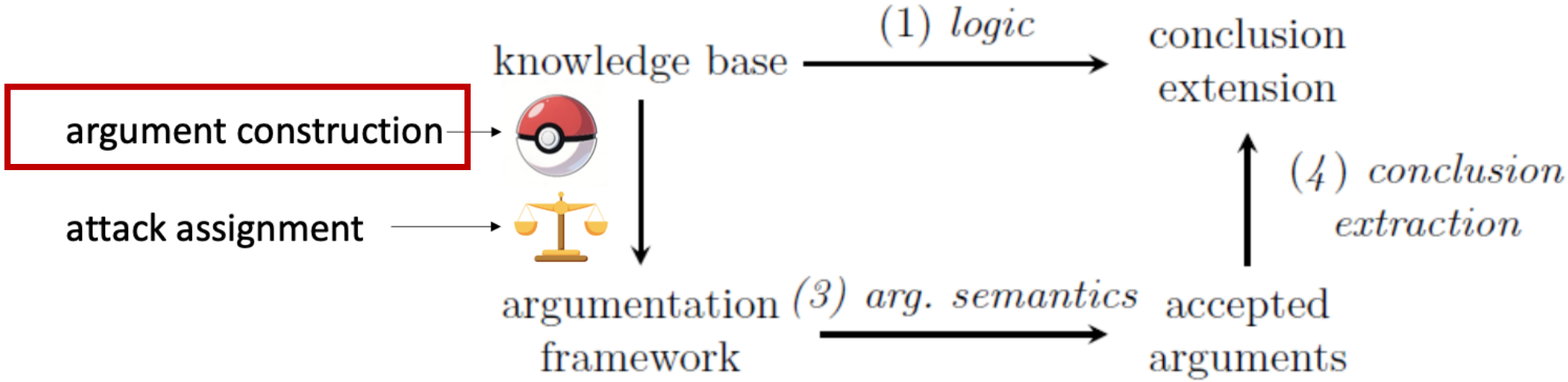
We use binary relation which we call "attack", and then define reinstatement

In our paper: reasoning alignment diagram



https://yuliuwen.github.io/assets/files/HOFA3_Chapter_Thirteen_Challenges.pdf
https://yuliuwen.github.io/assets/files/ifcolog00074.pdf

Example: All-or-Nothing as an Option Argument Framework



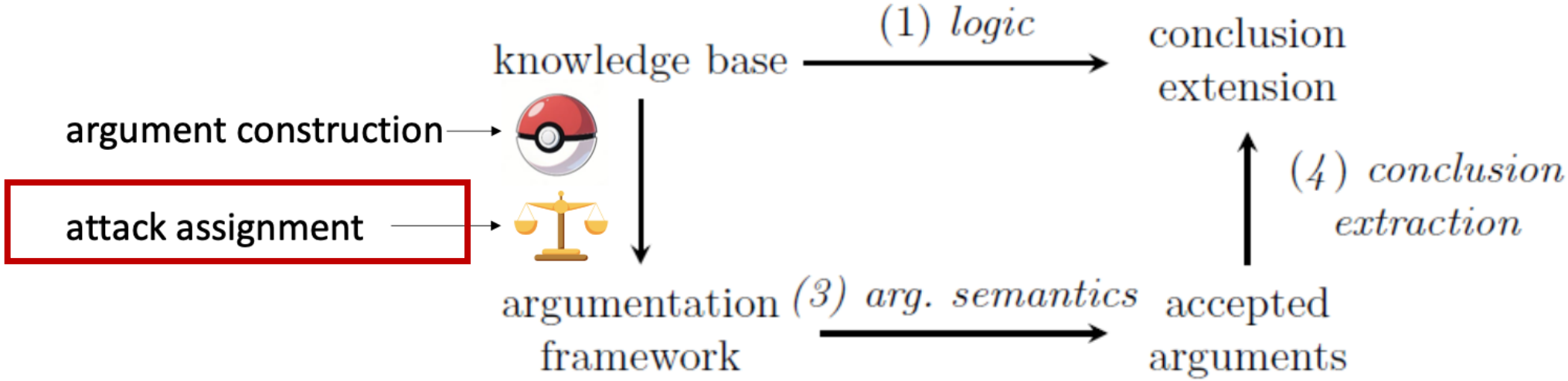
We have three options:

B = Bystander: save no child and avoid severe burns

S_1 = Save1: save one child and get severe burns

S_2 = Save2: save two children and get the same severe burns

Example: All or Nothing as an Option Argument Framework

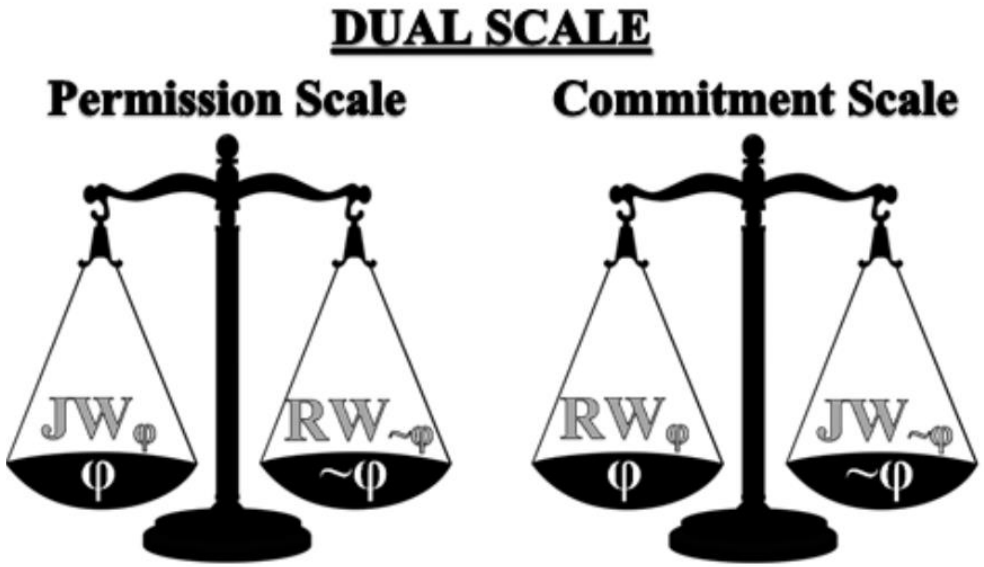


We formalize a reason in ○: (g, o_1, o_2)

$o_1\ attacks\ o_2$ iff

$\sum_{JW}(o_2, o_1) < \sum_{RW}(o_1, o_2)$

Choose from permitted options based on preference



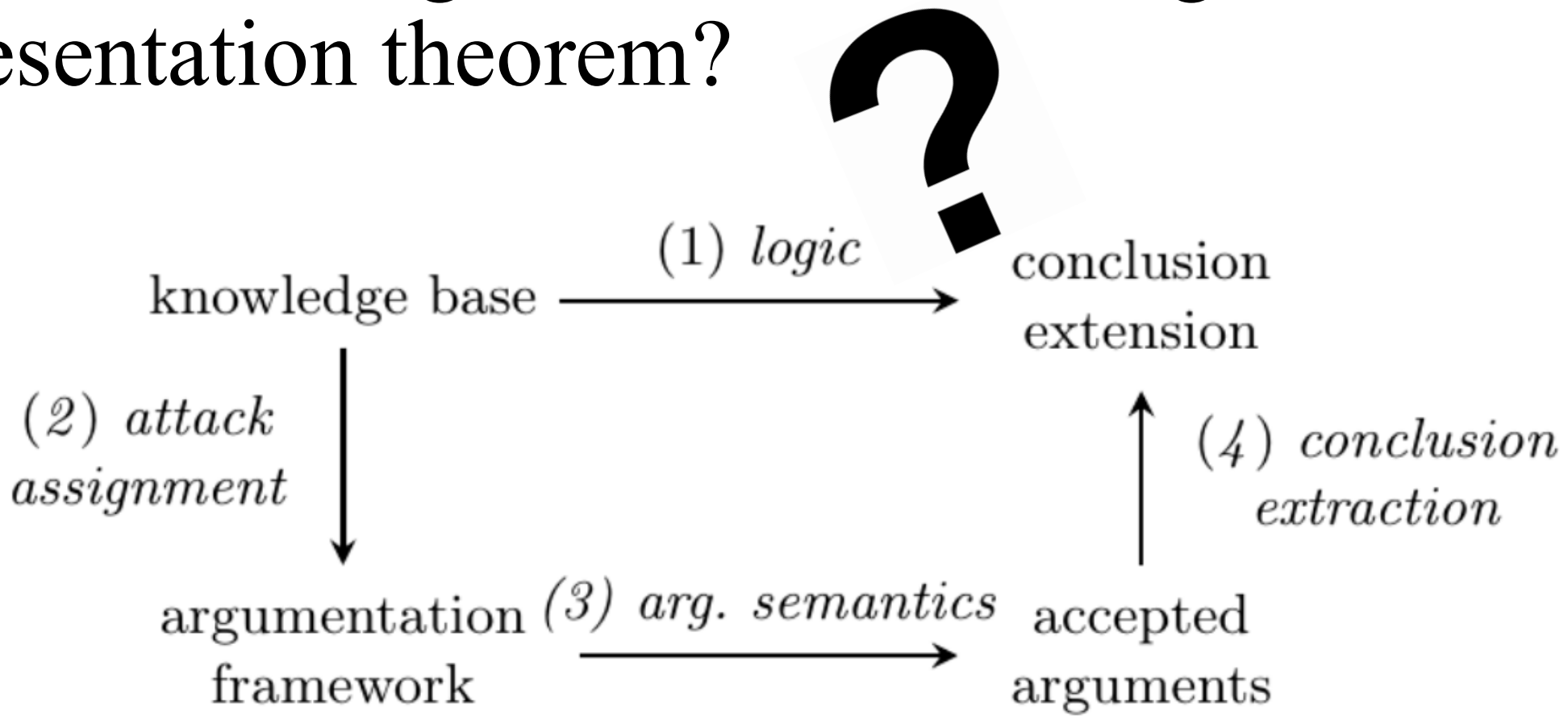
More questions…

Bipolar versus attack assignment, which is better?

	Bipolar representation	**Attack assignment**
Pro / con	preserved as support / attack	evaluated before abstraction
Abstract structure	BAF = (A, attack, support)	AF = (A, attack)
Where balancing happens	in bipolar semantics	in attack assignment
Dung semantics	extended / adapted	ordinary Dung semantics
Information retained	positive + negative structure remains visible	balance may be compiled into attack
Main design question	what does support do?	when does balance warrant attack?
Potential cost	richer semantics	possible information loss

Balancing at a different representational level

Alignment through commutative diagram Representation theorem?



Bipolar argumentation/option argumentation

Liuwen Yu, Leendert van der Torre, and Réka Markovich, "Thirteen Challenges in Formal and Computational Argumentation", in Handbook of Formal Argumentation, Volume 3, College Publications, 2024, pp. 890–976.

Representation, reduction, or impossibility?

Three increasingly strong questions

1. Simulation	2. Characterisation	3. Impossibility
Given a bipolar model, can we construct an attack assignment with the same output?	Which principles on bipolar semantics make such a reduction possible?	Which balancing behaviours cannot be compiled into binary attack without loss?
constructive representation theorem	**principle-based representation theorem**	**information-loss / impossibility theorem**

The principle-based bipolar analysis becomes a tool for locating the boundary of reducibility.

Challenge: Tucker's contrastivism

A reason for an option can vary with the alternative against which it is compared.

Ordinary bipolar edge	**Tucker reason**	**Attack assignment**
r —support— o_1	**$r(g, o_1, o_2)$**	evaluate (o_1,o_2) then assign attack
Looks context-independent: r supports o_1.	The force of r for o_1 is indexed by contrast o_2.	Contrast can remain upstream of the binary AF.

Can a fixed bipolar framework preserve contrast-sensitive reasons?

If not: enrich bipolarity with contrast indices — or prove that attack assignment is strictly more expressive for Tucker-style balancing.

Outlook 1

Nonmonotonic reasoning descriptive and prescriptive

Generate sets and test
Explanation and construction
reasons in progress for PDL&Fatio/Libra

Outlook 2

Other AI problems:
Negotiation, decision making, practical reasoning..

Arguing facts ≠ arguing decisions

Reason graph & causal graphs

Outlook 3

Argumentation and causality

How to deal with time?

Outlook 4

A-BDI metamodel

The A-BDI Metamodel for Human-Level AI:

Argumentation as Balancing, Dialogue and Inference

Liuwen Yu[1][0000−0002−7200−6001] and Leendert van der Torre[1,2][0000−0003−4330−3717]

[1] University of Luxembourg, Luxembourg
[2] Zhejiang University, China

Abstract. In this paper, we introduce A-BDI, the first metamodel for formal and computational argumentation. It contains three models, conceptualizing argumentation as balancing, argumentation as dialogue, and argumentation as inference respectively. Each model looks at argumentation from a different perspective, addressing its own concerns and using its own formal and computational methods. Whereas balancing is inspired by the scale metaphor and uses quantitative techniques typically found in theories in economics and neural computing, dialogue is developed in multiagent communication and interaction and uses chatbot and Large Language Models (LLMs) technology, and inference is derived from theoretical investigations in knowledge representation and reasoning and uses techniques from symbolic reasoning. By bringing together new and traditional Artificial Intelligence (AI) approaches, the A-BDI metamodel provides a formal and computational framework for human-level, neuro-symbolic, and hybrid AI.

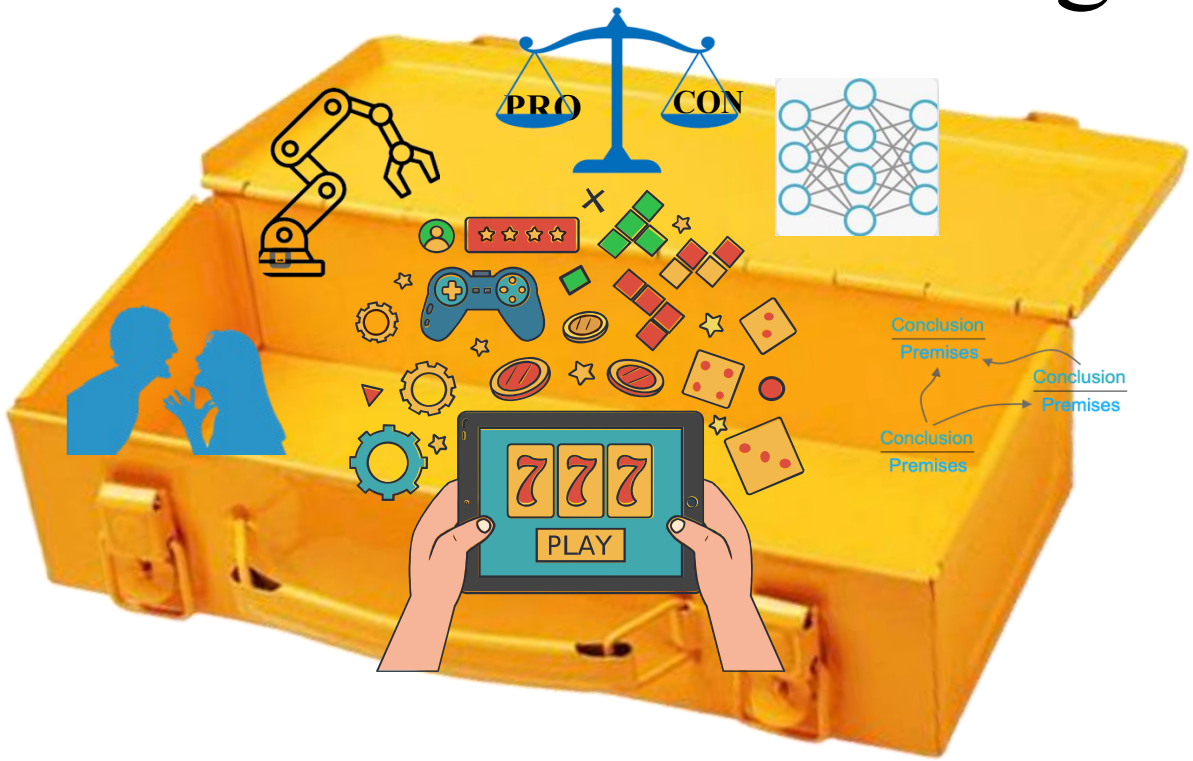
Keywords: Metamodel for Argumentation· Methodology of Argumentation · Knowledge Representation and Reasoning · Hybrid Artificial Intelligence

How to relate, integrate, and combine

Argumentation as balancing, dialogue and inference?

Outlook 5

Logic toolbox



Dov Gabbay

Example. *Mother reasons on her daughter*

A mother goes into her teenage daughter's bedroom. Her instant impression is that it is a big mess. There is stuff scattered everywhere. The mother's impression is that it is not characteristic of the girl to be like this.

Example from Dov Gabbay:
https://videolectures.net/godelfellowship2011_gabbay_pnd/

Example. *Mother reasons on her daughter*

A mother goes into her teenage daughter's bedroom. Her instant impression is that it is a big mess. There is stuff scattered everywhere. The mother's impression is that it is not characteristic of the girl to be like this.

What has happened?

Example from Dov Gabbay:
https://videolectures.net/godelfellowship2011_gabbay_pnd/

Example. *Mother reasons on her daughter*

A mother goes into her teenage daughter's bedroom. Her instant impression is that it is a big mess. There is stuff scattered everywhere. The mother's impression is that it is not characteristic of the girl to be like this.

What has happened?

Conjecture: The girl may be experiencing boyfriend issues.

Example from Dov Gabbay:
https://videolectures.net/godelfellowship2011_gabbay_pnd/

Example. *Mother reasons on her daughter*

A mother goes into her teenage daughter's bedroom. Her instant impression is that it is a big mess. There is stuff scattered everywhere. The mother's impression is that it is not characteristic of the girl to be like this.

What has happened?

Conjecture: The girl may be experiencing boyfriend issues.

Further Analysis: The mother notices a collapsed shelf and realizes that the disarray is due to the shelf's collapse under excessive weight, which, upon reflection, follows a logical (gravitational) pattern.

Example from Dov Gabbay:
https://videolectures.net/godelfellowship2011_gabbay_pnd/

Example. *Mother reasons on her daughter*

Several types of reasoning are illustrated through this scenario:

Example from Dov Gabbay:
https://videolectures.net/godelfellowship2011_gabbay_pnd/

Example. *Mother reasons on her daughter*

Several types of reasoning are illustrated through this scenario:

Neural Network Reasoning:

The mess is instantly recognized, similar to facial recognition by neural networks.

Example from Dov Gabbay:
https://videolectures.net/godelfellowship2011_gabbay_pnd/

Example. *Mother reasons on her daughter*

Several types of reasoning are illustrated through this scenario:

Neural Network Reasoning:

The mess is instantly recognized, similar to facial recognition by neural networks.

Nonmonotonic Deduction:

The mother deduces from context and knowledge: daughter doesn't typically live in disarray.

Example from Dov Gabbay:
https://videolectures.net/godelfellowship2011_gabbay_pnd/

Example. *Mother reasons on her daughter*

Several types of reasoning are illustrated through this scenario:

Neural Network Reasoning:

The mess is instantly recognized, similar to facial recognition by neural networks.

Nonmonotonic Deduction:

The mother deduces from context and knowledge: daughter doesn't typically live in disarray.

Abductive Reasoning:

She hypothesizes a plausible explanation that her daughter is facing social issues.

Example from Dov Gabbay:
https://videolectures.net/godelfellowship2011_gabbay_pnd/

Example. *Mother reasons on her daughter*

Several types of reasoning are illustrated through this scenario:

Neural Network Reasoning:

The mess is instantly recognized, similar to facial recognition by neural networks.

Nonmonotonic Deduction:

The mother deduces from context and knowledge: daughter doesn't typically live in disarray.

Abductive Reasoning:

She hypothesizes a plausible explanation that her daughter is facing social issues.

Database AI Deduction:

A re-evaluation: the mess is due to gravitational effects rather than disorganization.

Example from Dov Gabbay:
https://videolectures.net/godelfellowship2011_gabbay_pnd/

Example. *Mother reasons on her daughter*

Several types of reasoning are illustrated through this scenario:

Neural Network Reasoning:

The mess is instantly recognized, similar to facial recognition by neural networks.

Nonmonotonic Deduction:

The mother deduces from context and knowledge: daughter doesn't typically live in disarray.

Abductive Reasoning:

She hypothesizes a plausible explanation that her daughter is facing social issues.

Database AI Deduction:

A re-evaluation: the mess is due to gravitational effects rather than disorganization.

Pattern Recognition:

Someone accustomed to similar patterns may identify the cause as easily as recognizing a face.

Example from Dov Gabbay:
https://videolectures.net/godelfellowship2011_gabbay_pnd/

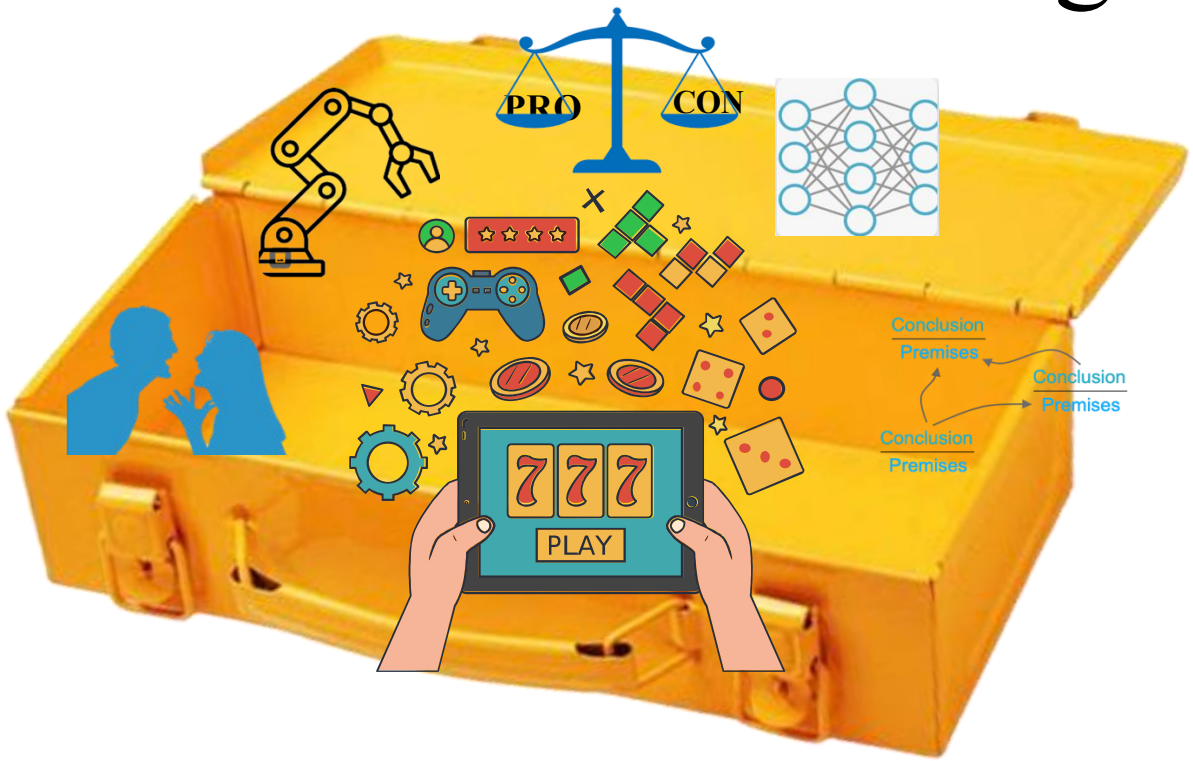
Outlook 5

Reasoning Alignment for Agentic AI: Argumentation, Belief Revision, and Dialogue

Tjitze Rienstra[1 [0000−0001−9971−3067]], Leendert van der Torre[2,3 [0000−0003−4330−3717]], and Liuwen Yu[2 [0000−0002−7200−6001]]

[1] Maastricht University, Maastricht, The Netherlands
t.rienstra@maastrichtuniversity.nl
[2] University of Luxembourg, Esch-sur-Alzette, Luxembourg
{leon.vandertorre,liuwen.yu}@uni.lu
[3] Zhejiang University, China

Logic toolbox



Dov Gabbay

Reasoning alignment diagram

Outlook 5

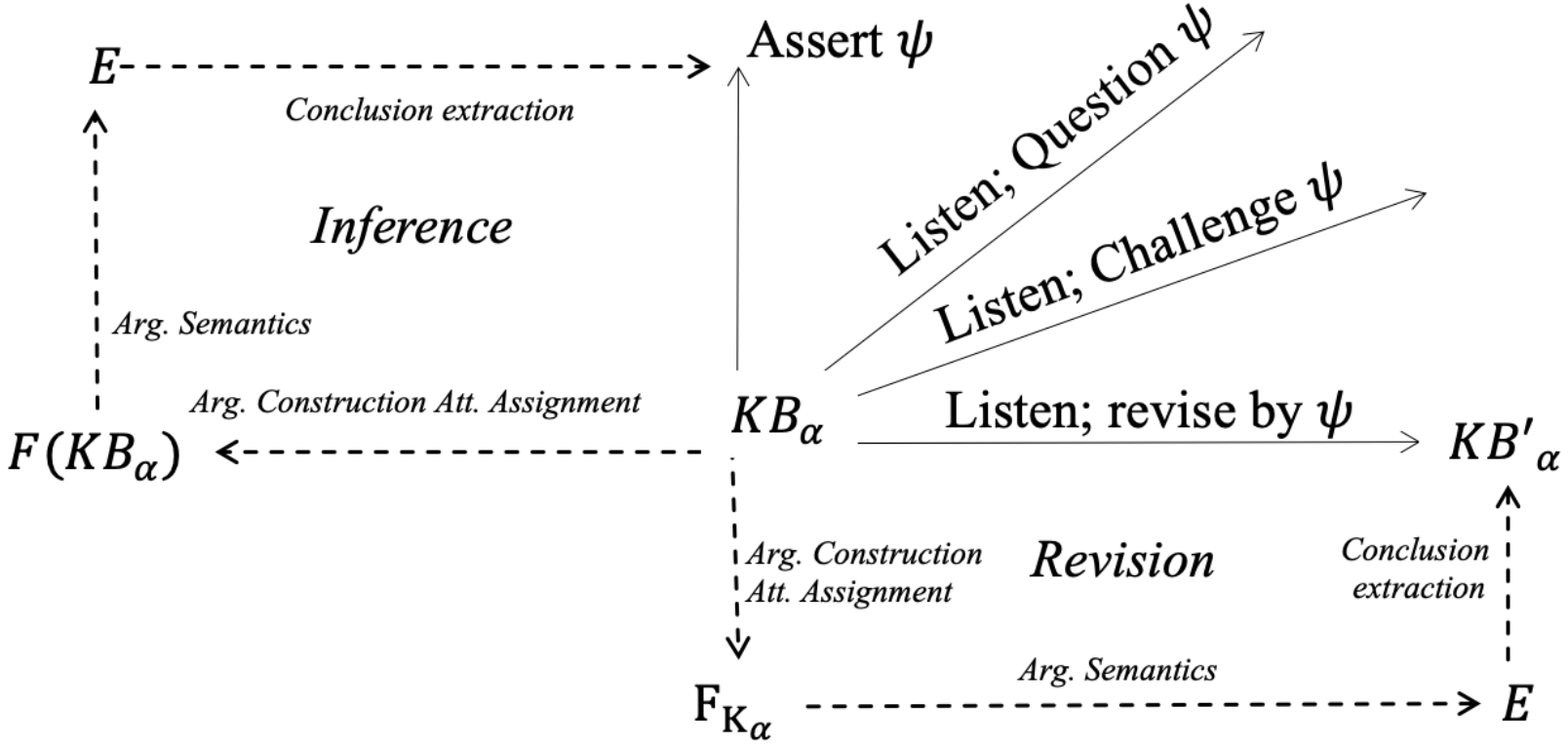
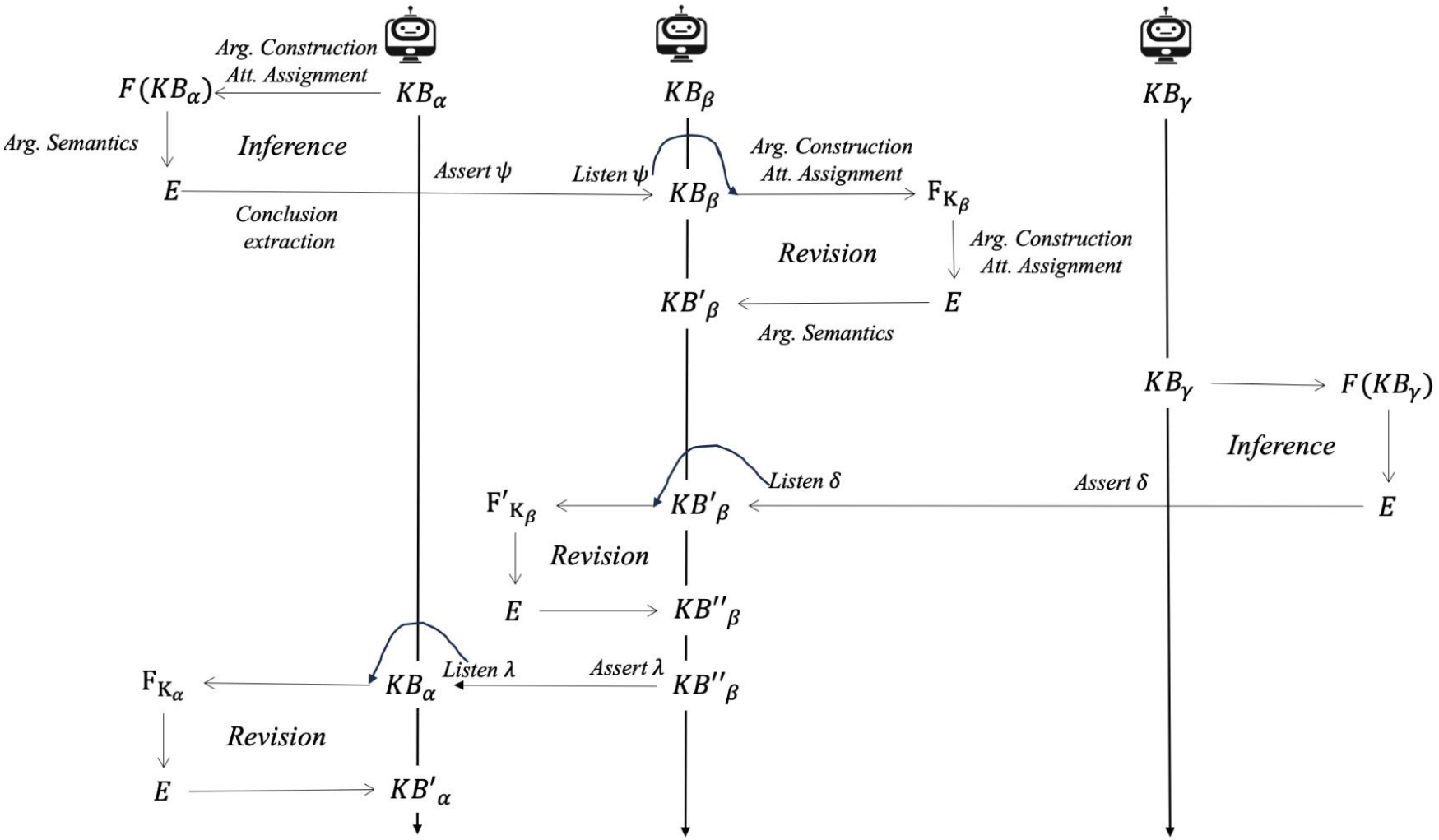


Fig. 6: Multi-RAD for individual agent decision

Reasoning alignment diagram

Outlook 5



Reasoning alignment diagram