Semantic Instantiation of Structured Argumentation

Giuseppe Pisano[1] Giovanni Sartor[1]

[1]ALMA-AI CIRSFID, University of Bologna, Italy

AI^3 2026, Perugia, 7 October 2026

Ministero dell'Università e della Ricerca | FIS | GenAI4Law

This work was supported by the Italian Ministry of University and Research (MUR) under the FIS project GenAI4Law, Grant No. FIS-2024-07323.

Outline

1. Motivation
2. Background
3. Semantic ASPIC$^+$
4. Requirements on the Semantic Relation
5. Running Example
6. Optional Properties
7. Towards the Rationality Postulates
8. Conclusions

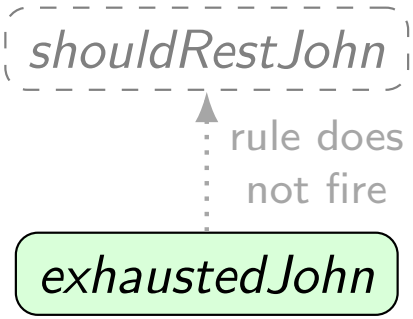
The problem: exact matching

In ASPIC$^+$ [7, 6]:

- a rule applies only if its antecedents are **exactly** the conclusions of sub-arguments
- two conclusions conflict only if one is the **negation** of the other

Natural-language knowledge does not work this way:

- "John is *exhausted*" should trigger the rule
  $tiredJohn \Rightarrow shouldRestJohn$
- "John should rest" and "John must keep working" conflict, with no negation in sight



Our proposal

Semantic ASPIC$^+$

Replace the two syntactic tests of ASPIC$^+$ by graded semantic tests:

$\Rightarrow$ rule applicability: **semantic matching** $\approx_{\theta_s}$ instead of identity

$\Rightarrow$ attacks: **semantic conflict** $\bowtie_{\theta_c}$ instead of negation

- Argument structure, defeat, preferences and semantics are **unchanged**
- The semantic relation is **abstract**: LLMs, textual entailment, sentence embeddings [2, 9]
- Design choice: the matched premise **stays** in the argument, the rule antecedent is **never added**
- We identify the **requirements** on the semantic relation that keep the framework well-behaved

ASPIC$^+$: argumentation systems

An **argumentation system** is a triple $(\mathcal{L}, \mathcal{R}, n)$:

- $\mathcal{L}$: logical language with negation $\neg$
- notation: $\psi = -\varphi$ iff $\psi = \neg\varphi$ or $\varphi = \neg\psi$
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$: strict $(\varphi_1, \ldots, \varphi_n \rightarrow \psi)$ and defeasible $(\varphi_1, \ldots, \varphi_n \Rightarrow \psi)$ rules
- $n : \mathcal{R}_d \rightarrow \mathcal{L}$: $n(r)$ is the **name** of rule r in $\mathcal{L}$, so that an argument for $-n(r)$ can undercut applications of r

A **knowledge base** $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p$: axioms and ordinary premises.

An **argumentation theory** is a pair $(AS, \mathcal{K})$.

ASPIC$^+$: arguments

Arguments chain rules over the knowledge base:

1. φ if $\varphi \in \mathcal{K}$, with $\texttt{Prem}(A) = \{\varphi\}$, $\texttt{Conc}(A) = \varphi$
2. $A_1, \ldots, A_n \rightsquigarrow \psi$ if $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightsquigarrow \psi \in \mathcal{R}$, with $\texttt{Prem}(A) = \bigcup_i \texttt{Prem}(A_i)$, $\texttt{Conc}(A) = \psi$, $\texttt{Sub}(A) = \bigcup_i \texttt{Sub}(A_i) \cup \{A\}$

Exact matching

Step 2 requires the conclusions of the sub-arguments to be **syntactically identical** to the rule antecedents.

ASPIC$^+$: attack and defeat

A **attacks** B iff A:

- **undercuts** B: $\texttt{Conc}(A) = -n(r)$, r the defeasible top rule of some $B' \in \texttt{Sub}(B)$
- **rebuts** B: $\texttt{Conc}(A) = -\varphi$, φ the conclusion of some $B' \in \texttt{Sub}(B)$ with **defeasible** top rule
- **undermines** B: $\texttt{Conc}(A) = -\varphi$ for some ordinary premise $\varphi \in \texttt{Prem}(B) \cap \mathcal{K}_p$

Defeat: undercuts always succeed; rebuts and undermines succeed unless the attacker is strictly weaker than the attacked sub-argument.

The resulting abstract framework $(\mathcal{A}, \mathcal{D})$ [4] is evaluated with the usual semantics.

The semantic relation

Semantic relation

$R : \mathcal{L} \times \mathcal{L} \to [0,1]^3, \quad R(\phi,\psi) = \big(P_E(\phi,\psi),\ P_C(\phi,\psi),\ P_N(\phi,\psi)\big),$
$P_E + P_C + P_N = 1$

- P_E: degree to which ϕ **matches** ψ (can play its role)
- P_C: degree to which ϕ **conflicts** with ψ
- P_N: degree of **neutrality**

Thresholds

$\phi \approx_{\theta_s} \psi$ iff $P_E(\phi,\psi) \geq \theta_s$ $\qquad$ $\phi \bowtie_{\theta_c} \psi$ iff $P_C(\phi,\psi) \geq \theta_c$

R is **not** assumed symmetric: P_E may be directional (entailment), P_C may be one-sided (contrary).

Semantic arguments

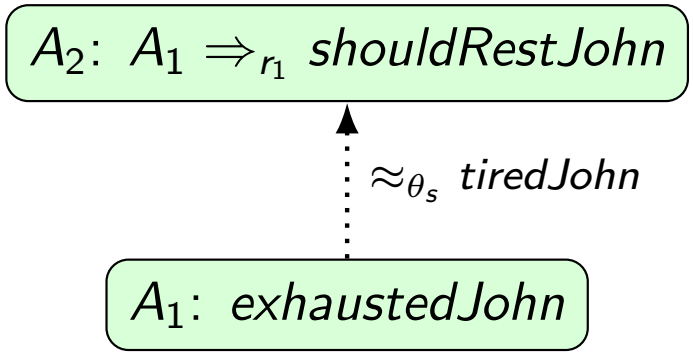
Rule application by matching

$A_1, \ldots, A_n \rightsquigarrow \psi$ is an argument if there is a rule $\phi_1, \ldots, \phi_n \rightsquigarrow \psi \in \mathcal{R}$ with

$$\texttt{Conc}(A_i) \approx_{\theta_s} \phi_i \qquad i = 1, \ldots, n$$

`Prem`, `Conc`, `Sub` and `TopRule` are defined exactly as in ASPIC$^+$.

- the antecedent ϕ_i is only an **applicability check**
- the argument keeps the **actual** formula $\texttt{Conc}(A_i)$
- attacks therefore target what is **really there**, never the antecedent



Semantic attacks

Attack by conflict

A semantically attacks B iff *A*:

- **undercuts** B: $\texttt{Conc}(A) \bowtie_{\theta_c} n(r)$, r the defeasible top rule of some $B' \in \texttt{Sub}(B)$
- **rebuts** B: $\texttt{Conc}(A) \bowtie_{\theta_c} \varphi$, φ the conclusion of some $B' \in \texttt{Sub}(B)$ with **defeasible** top rule
- **undermines** B: $\texttt{Conc}(A) \bowtie_{\theta_c} \varphi$ for some ordinary premise $\varphi \in \texttt{Prem}(B) \cap \mathcal{K}_p$

- same three attack types, same targets, same defeat rule
- the only change: $-\varphi$ is replaced by $\bowtie_{\theta_c} \varphi$
- since $\bowtie_{\theta_c}$ need not be symmetric, both directions are checked separately

Running example: the theory

$$\mathcal{K}_p = \{ \quad tiredJohn,\ urgentDeadlineJohn,\ -r_1 \quad \}$$

$$\mathcal{R}_d = \{ \quad r_1 : tiredJohn \Rightarrow shouldRestJohn,\ r_2 : busyJohn \Rightarrow keepWorkingJohn \quad \}$$

Thresholds $\theta_s = \theta_c = 0.8$ and semantic relation

$$R(urgentDeadlineJohn, busyJohn) = (0.89, 0.03, 0.08)$$
$$R(shouldRestJohn, keepWorkingJohn) = (0.03, 0.92, 0.05)$$
$$R(keepWorkingJohn, shouldRestJohn) = (0.02, 0.90, 0.08)$$

Premise arguments:

$$A_1 : tiredJohn \qquad A_2 : urgentDeadlineJohn \qquad A_3 : -r_1$$

Questions

Which arguments can be built? Which attacks hold?
The answers depend on what we may assume about R.

Why constrain R?

An arbitrary R can block valid inferences, hide contradictions, or let the same pair both support and attack. Accordingly, we require:

Requirement	**Statement**	**Guarantees**
Reflexivity	$P_E(\phi, \psi) = 1 \iff \phi = \psi$	exact matches always work, and only they are perfect
Classical negation	$\psi = -\phi \iff P_C(\phi, \psi) = 1$	contradictions always conflict, and only they maximally
Exclusion	$\phi \approx_{\theta_s} \psi \Rightarrow \phi \not\bowtie_{\theta_c} \psi \wedge \psi \not\bowtie_{\theta_c} \phi$	a pair never both supports and attacks

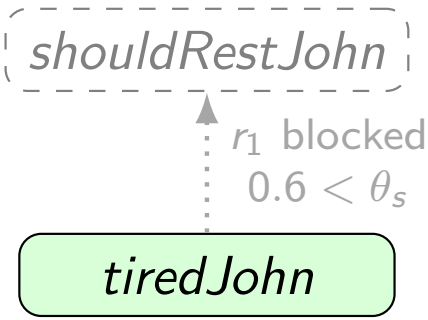
Optional: symmetry of P_E, conflict transmission. Not needed, and often false for real models.

Reflexivity

$$P_E(\phi, \psi) = 1 \iff \phi = \psi$$

Without $\Rightarrow$: self-match below 1

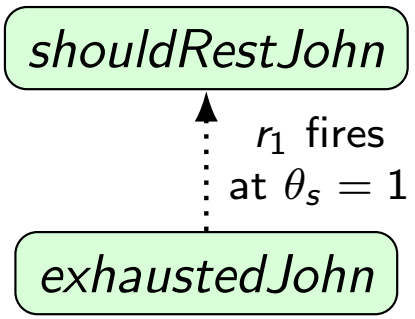
$P_E(\mathit{tiredJohn}, \mathit{tiredJohn}) = 0.6$, $\theta_s = 0.8$



a valid ASPIC$^+$ inference is lost

Without $\Leftarrow$: distinct formulas at 1

$P_E(\mathit{exhaustedJohn}, \mathit{tiredJohn}) = 1$,
$\theta_s = 1$



an argument ASPIC$^+$ does not have

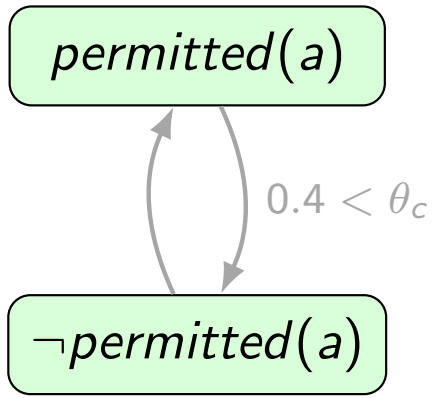
Instantiations: reserve 1 for the identity check, map distinct pairs into $[0, 1 - \epsilon]$.

Classical negation

$$(\psi = \neg\phi \vee \phi = \neg\psi) \iff P_C(\phi, \psi) = 1$$

Without $\Rightarrow$: a contradiction below θ_c

- $\mathcal{K}_p = \{permitted(a),\ \neg permitted(a)\}$
- $P_C(permitted(a), \neg permitted(a)) = 0.4$, $\theta_c = 0.8$
- no undermining in either direction
- grounded extension accepts **both**



Without $\Leftarrow$: a non-contradiction at 1 survives $\theta_c = 1$, so the attack relation is not the ASPIC$^+$ one.

Consequences of $P_E + P_C + P_N = 1$: $P_C(\phi, \phi) = 0$ and $P_E(\phi, \neg\phi) = 0$.

Exclusion

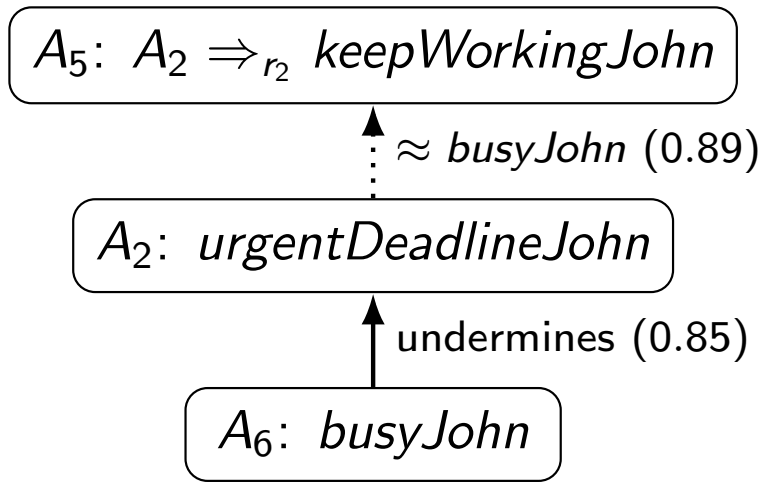
$$\phi \approx_{\theta_s} \psi \implies \phi \not\bowtie_{\theta_c} \psi \ \wedge \ \psi \not\bowtie_{\theta_c} \phi$$

- **same order**: follows from $P_E + P_C + P_N = 1$ whenever $\theta_s + \theta_c > 1$
- **reversed order**: R is not symmetric, so it must be **required** of R, e.g. $P_E(\phi, \psi) + P_C(\psi, \phi) \leq 1$ (automatic if P_E is symmetric)

Without the reversed order, $\theta_s = \theta_c = 0.8$:

$$R(\mathit{urgentDeadlineJohn}, \mathit{busyJohn}) = (0.89, 0.03, 0.08)$$
$$R(\mathit{busyJohn}, \mathit{urgentDeadlineJohn}) = (0.05, 0.85, 0.10)$$

busyJohn **licenses** r_2 for A_5 and **undermines** its premise: the same formula supports and attacks the same argument.



Conservative generalisation

With $\theta_s = \theta_c = 1$ we obtain exactly ASPIC$^+$

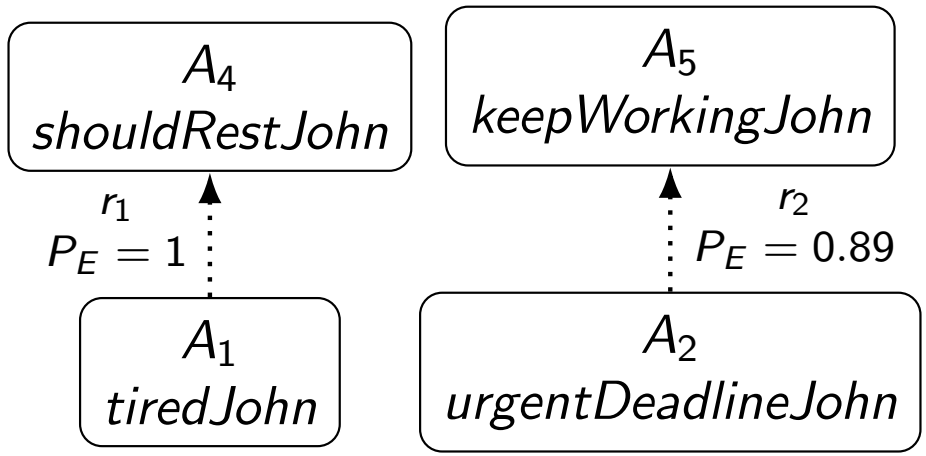
- reflexivity ($\Leftarrow$): $\approx_1$ is syntactic identity
- classical negation ($\Leftarrow$): $\bowtie_1$ is $-$

Arguments and attacks coincide with the standard definitions: no argument or attack is lost, none is added.

- The $\Rightarrow$ directions give **containment** (nothing lost); the $\Leftarrow$ directions give **exactness** (nothing added)
- Lowering the thresholds adds arguments and attacks **monotonically**

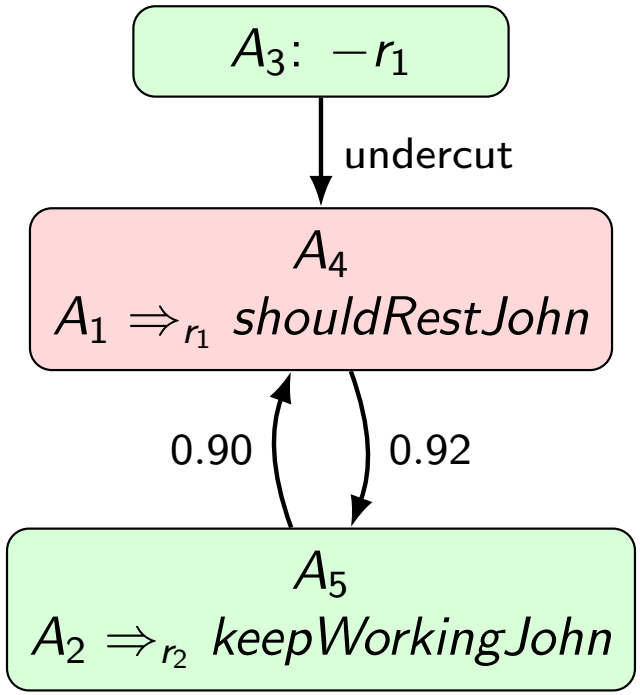
Running example: construction

- $A_4 = A_1 \Rightarrow_{r_1}$ *shouldRestJohn*: **exact** match. **Reflexivity** gives $P_E(\mathit{tiredJohn}, \mathit{tiredJohn}) = 1 \geq \theta_s$
- $A_5 = A_2 \Rightarrow_{r_2}$ *keepWorkingJohn*: **semantic** match, $P_E(\mathit{urgentDeadlineJohn}, \mathit{busyJohn}) = 0.89 \geq \theta_s$
- the premise of A_5 is still *urgentDeadlineJohn*; *busyJohn* was only the applicability check



Running example: attacks and evaluation

- A_3 **undercuts** A_4: **classical negation** gives $P_C = 1$ for the contradiction of r_1's applicability
- A_4 and A_5 **rebut** each other: $0.92, 0.90 \geq \theta_c$
- **exclusion**: the same pair is not also a match ($P_E = 0.03$)
- grounded extension: $\{A_1, A_2, A_3, A_5\}$



In words: the urgent deadline counts as being busy, so John keeps working; the exception to the rest rule removes the competing argument.

Optional: symmetry

$$P_E(\phi, \psi) = P_E(\psi, \phi)$$

- natural if P_E is read as **similarity**
- wrong if P_E is read as **entailment** or applicability, which is what rule application asks for:
  - "John has not slept for two days" can play the role of "John is tired"
  - "John is tired" cannot play the role of "John has not slept for two days"
- **not imposed**: the framework works with a directional P_E and a one-sided P_C

Optional: conflict transmission

Matching statements share their conflicts

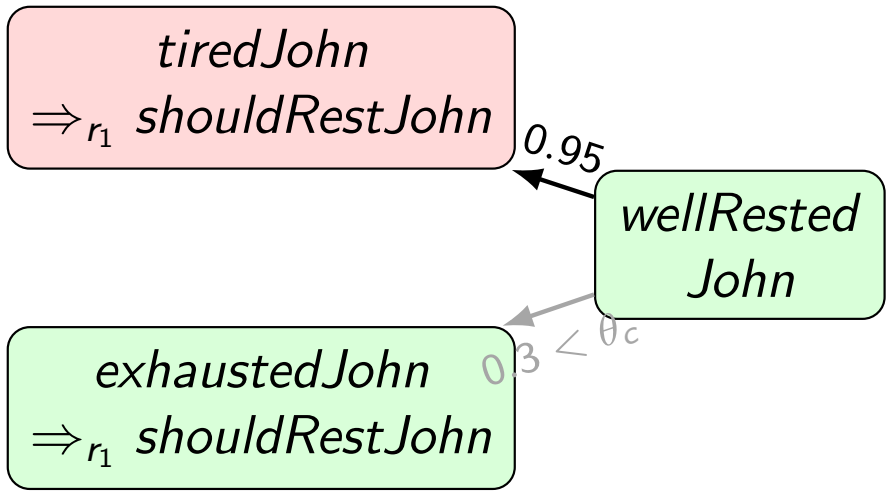
$$P_E(\phi,\psi) \geq \theta_s \implies \begin{cases} |P_C(\phi,\chi) - P_C(\psi,\chi)| \leq 1-\theta_s & \text{(outgoing)} \\ |P_C(\chi,\phi) - P_C(\chi,\psi)| \leq 1-\theta_s & \text{(incoming)} \end{cases} \quad \forall \chi$$

Without it, $\theta_s = \theta_c = 0.8$:

$P_E(\mathit{exhaustedJohn}, \mathit{tiredJohn}) = 0.9$
$P_C(\mathit{wellRestedJohn}, \mathit{tiredJohn}) = 0.95$
$P_C(\mathit{wellRestedJohn}, \mathit{exhaustedJohn}) = 0.3$

The **paraphrase** is immune to an attack that hits the **literal** premise.

Not needed for construction or postulates: attacks target the actual formulas.



Towards the rationality postulates

Postulates of [3, 6], consistency w.r.t. $\bowtie_{\theta_c}$. Preliminary intuitions:

Closure under sub-arguments and under strict rules — easy

Only the applicability test changes: attack targets and defeat are decided exactly as before. Standard proofs transfer.

Direct consistency — harder

Read $\bowtie_{\theta_c}$ as the ASPIC$^+$ **contrariness** relation: well-formedness and transposition transfer, applied to the **semantic instances** of strict rules (the rules as actually used):

$$\text{instance } \chi_1, \ldots, \chi_n \to \psi \;\; (\chi_j \approx_{\theta_s} \phi_j) \quad \rightsquigarrow \quad \chi_1, \ldots, -\psi, \ldots, \chi_n \to -\chi_i$$

The transposed argument concludes $-\chi_i$ and attacks the **matched** formula with degree 1.

Indirect consistency — easy

Follows from closure and direct consistency [3].

Extensions (sketched)

- **Confidence in arguments**: aggregate the matching degrees of a derivation, $w(A) = \min_i P_E$ (weakest link), mean, or $\max_i P_E$; use w in preferences or rankings [8]
- **Weighted attacks**: P_C as attack strength, connecting to weighted and gradual argumentation [5, 1]
- **Explicit alternative**: bridge rules $exhaustedJohn \Rightarrow tiredJohn$ keep ASPIC$^+$ untouched and make the link itself attackable.

Conclusions

- **Semantic ASPIC$^+$**: matching $\approx_{\theta_s}$ decides rule applicability, conflict $\bowtie_{\theta_c}$ decides attacks; structure, defeat and semantics unchanged
- **Three essential requirements**: reflexivity and classical negation as biconditionals, exclusion in both orders
- **Exact conservativity**: $\theta_s = \theta_c = 1$ gives ASPIC$^+$
- **Postulates**: closure transfers; consistency via ASPIC$^+$ well-formedness with $\bowtie_{\theta_c}$ as contrariness and transposition of semantic instances

Future work

- full proofs of the postulates; a notion of consistency relative to $\bowtie_{\theta_c}$
- instantiating R with NLI or LLM scorers wrapped to satisfy the requirements
- confidence-based preferences and weighted attacks

Thanks for the attention!

Semantic Instantiation of Structured Argumentation

Giuseppe Pisano[1] Giovanni Sartor[1]

[1]ALMA-AI CIRSFID, University of Bologna, Italy

AI^3 2026, Perugia, 7 October 2026

This work was supported by the Italian Ministry of University and Research (MUR) under the FIS project GenAI4Law, Grant No. FIS-2024-07323.

References I



Amgoud, L., Ben-Naim, J., Doder, D., and Vesic, S.
Acceptability semantics for weighted argumentation frameworks.
In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017* (2017), C. Sierra, Ed., ijcai.org, pp. 56–62.



Bowman, S. R., Angeli, G., Potts, C., and Manning, C. D.
A large annotated corpus for learning natural language inference.
In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015* (2015), L. Màrquez, C. Callison-Burch, J. Su, D. Pighin, and Y. Marton, Eds., The Association for Computational Linguistics, pp. 632–642.



Caminada, M., and Amgoud, L.
On the evaluation of argumentation formalisms.
Artif. Intell. 171, 5-6 (2007), 286–310.



Dung, P. M.
On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games.
Artif. Intell. 77, 2 (1995), 321–358.

References II



Dunne, P. E., Hunter, A., McBurney, P., Parsons, S., and Wooldridge, M. J.
Weighted argument systems: Basic definitions, algorithms, and complexity results.
Artif. Intell. 175, 2 (2011), 457–486.



Modgil, S., and Prakken, H.
The *ASPIC*$^+$ framework for structured argumentation: a tutorial.
Argument Comput. 5, 1 (2014), 31–62.



Prakken, H.
An abstract framework for argumentation with structured arguments.
Argument Comput. 1, 2 (2010), 93–124.



Prakken, H.
An abstract and structured account of dialectical argument strength.
Artif. Intell. 335 (2024), 104193.



Reimers, N., and Gurevych, I.
Sentence-bert: Sentence embeddings using siamese bert-networks.
In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019* (2019), K. Inui, J. Jiang, V. Ng, and X. Wan, Eds., Association for Computational Linguistics, pp. 3980–3990.