AI^3 2024

Spurious preferences in structured argumentation: a preliminary analysis

Pietro Baroni, Federico Cerutti, Massimiliano Giacomin
DII - University of Brescia (Italy)

Outline

- Preferences and spurious preferences in structured argumentation
- An example of spurious preferences in ASPIC+
- A formal requirement of spurious preference avoidance
- ASPIC+ revisited satisfies spurious preference avoidance
- Conclusions

Preferences in argumentation

- Given the arguments produced by an argumentation mechanism, a preference relation can be defined over them according to a variety of criteria (values, source, credibility of rules, …)
- The preference relation typically affects the relation of attack between arguments (e.g. attacks from less preferred to more preferred arguments are ignored)
- Some arguments receive a "better" treatment than others due to preferences

Spurious preferences

- Suppose that in some cases some arguments receive a "better" treatment than others without any preference justifying this disparity
- We refer to this situation as a case of spurious preferences: the argumentation system appears to follow some preferences which do not actually exist

ASPIC+ in a nutshell (1)

- A rule-based argumentation formalism
- Arguments are built by chaining rules starting from some premises
- Premises can be certain (axioms) or not (ordinary)
- Rules can be certain (strict) or not (defeasible)
- The set of strict rules needs to be closed under transposition in order to satisfy some rationality postulates

ASPIC+ in a nutshell (2)

- Three types of attacks between arguments are defined based on a contrariness relation
- Some attacks are preference dependent
- Given the constructed arguments and the attacks between them, a Dung's framework is derived
- An argumentation semantics can be applied to the derived framework to evaluate the acceptability of arguments in terms of sets of extensions

A simple reasoning example

- Evidences:
  - » three uncertain evidences about the birthdate (b), birthplace (p) and domicile (d) of a person, all equally preferred
- Inferences:
  - » From the birthdate, it can be derived with certainty that the person is over 18 (m)
  - » From the birthplace it can be derived with certainty that the person is a citizen of a given country (c) (assuming *ius soli* in the country)
  - » Finally, from the domicile, age majority, and citizenship, it can be derived that the person must be included in the taxpayers' list (ω).
- Certain fact:
  - » the person is not included in the taxpayers' list ($\neg\, \omega$).

Formalization in ASPIC+

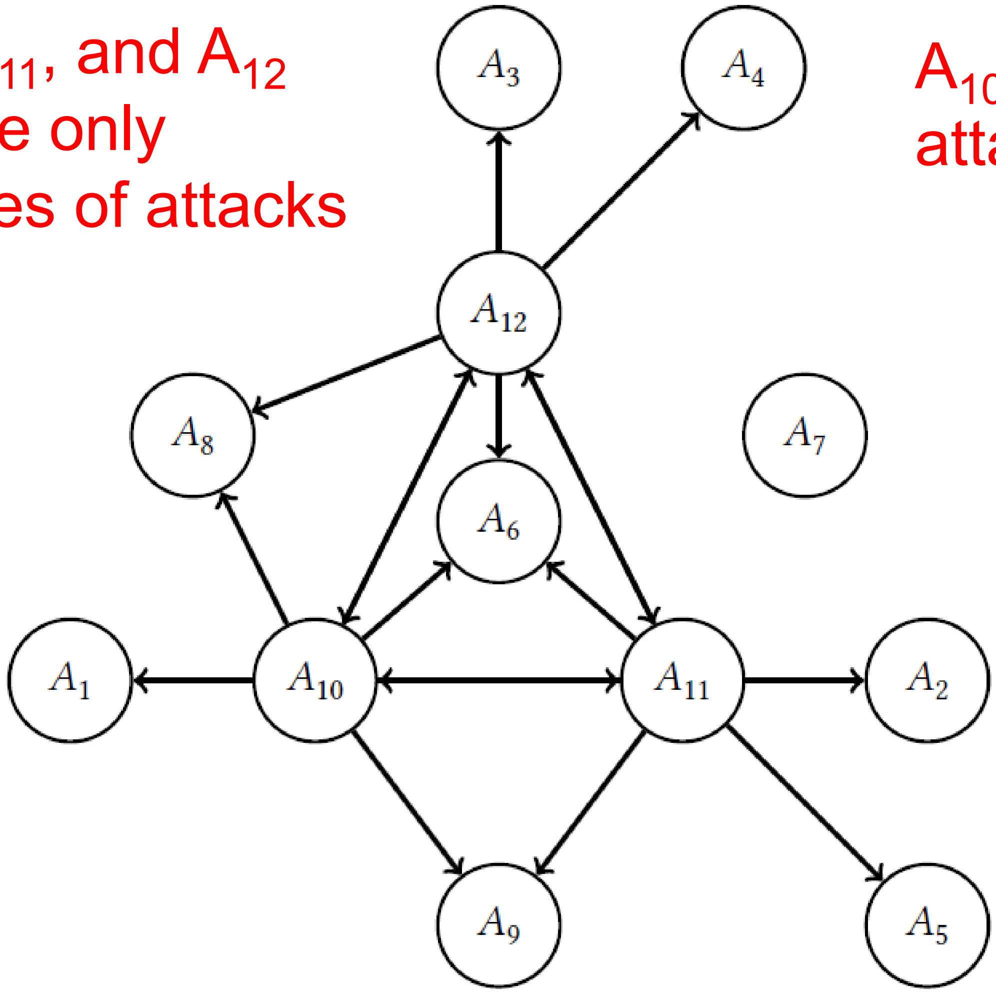
- Ordinary premises: $\mathcal{K}_p$ = {d, b, p}
- Axioms: $\mathcal{K}_n$ = {$\neg\omega$}
- Strict rules: $\mathcal{R}_{Sb}$ = {$b \rightarrow m$; $p \rightarrow c$; d, c, $m \rightarrow \omega$}
- Closure of strict rules: $\mathcal{R}_{St}$ = {$\neg m \rightarrow \neg b$; $\neg c \rightarrow \neg p$; d, c, $\neg\omega \rightarrow \neg m$; d, m, $\neg\omega \rightarrow \neg c$; c, m, $\neg\omega \rightarrow \neg d$}

The constructed arguments

- Ordinary premises:
  $A_1 = d$; $A_2 = b$; $A_3 = p$;
- Inferences using strict rules:
  $A_4 = A_3 \rightarrow c$; $A_5 = A_2 \rightarrow m$; $A_6 = A_1, A_4, A_5 \rightarrow \omega$;
- Axiom:
  $A_7 = \neg\omega$;
- Inferences using transposed rules:
  $A_8 = A_1, A_4, A_7 \rightarrow \neg m$; $A_9 = A_1, A_5, A_7 \rightarrow \neg c$;
  $A_{10} = A_4, A_5, A_7 \rightarrow \neg d$; $A_{11} = A_8 \rightarrow \neg b$; $A_{12} = A_9 \rightarrow \neg p$.

The resulting argumentation framework

A_{10}, A_{11}, and A_{12} are the only sources of attacks



A_{10}, A_{11}, and A_{12} mutually attack each other

With preferred, stable and semi-stable semantics there are three extensions

Each extension corresponds to rejecting one of the uncertain premises

A variant of the reasoning example

- Consider an analogous case in a country where ius soli has been introduced "recently"
- Then the inference about citizenship requires both the birthplace and birthdate
- The only difference is that the strict rule $p \rightarrow c$; becomes $b, p \rightarrow c$;
- As a consequence the set of transposed rules is:
$\mathcal{R}'_{St}$ = $\{\neg m \rightarrow \neg b;\ p, \neg c \rightarrow \neg b;\ b, \neg c \rightarrow \neg p;$
$d, c, \neg \omega \rightarrow \neg m;\ d, m, \neg \omega \rightarrow \neg c;\ c, m, \neg \omega \rightarrow \neg d\}$

Comment on the variant of the example

- Only one strict rule has changed
- The three ordinary premises are still the only uncertain elements from which a contradiction with a certain fact is strictly derived
- Still there is no preference over them: they are equal candidates to be rejected
- However …

The constructed arguments

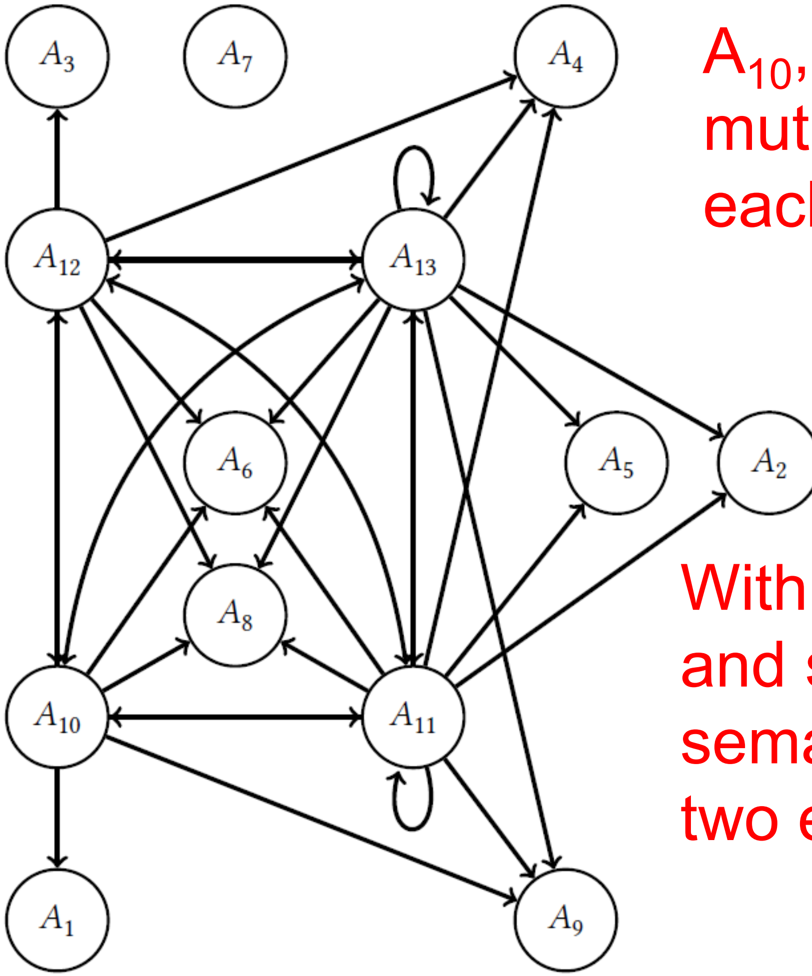
- Ordinary premises:
  $A_1 = d; A_2 = b; A_3 = p;$
- Inferences using strict rules:
  $A_4 = A_2, A_3 \rightarrow c; A_5 = A_2 \rightarrow m; A_6 = A_1, A_4, A_5 \rightarrow \omega;$
- Axiom:
  $A_7 = \neg\omega;$
- Inferences using transposed rules:
  $A_8 = A_1, A_4, A_7 \rightarrow \neg m; A_9 = A_1, A_5, A_7 \rightarrow \neg c;$
  $A_{10} = A_4, A_5, A_7 \rightarrow \neg d; A_{11} = A_8 \rightarrow \neg b;$
  $A_{12} = A_2, A_9 \rightarrow \neg p, A_{13} = A_3, A_9 \rightarrow \neg b$

The resulting argumentation framework

$A_{10}, A_{11}, A_{12}, A_{13}$ are the only sources of attacks

A_{11} and A_{13} are self-defeating

Each extension corresponds to rejecting one of the uncertain premises: the third one is always accepted



$A_{10}, A_{11}, A_{12}, A_{13}$ mutually attack each other

With preferred, stable and semi-stable semantics there are two extensions

A spurious preference

- One ordinary premise (the one corresponding to birthdate) is always accepted
- This amounts to ascribing to a sort of implicit preference with respect to the other premises
- This implicit preference can be regarded as an accidental side effect of the structure of the set of strict rules and can be considered ***spurious***

A requirement of spurious preference avoidance

- Not easy to formalize spurious preference in general
- There are many possible reasons to have different acceptance statuses for different arguments even in absence of preferences
- As a first step, we formalize a requirement of spurious preference avoidance in a specific family of reasoning cases

The SSDOP family

- SSDOP stands for Simple Strict Derivation from Ordinary Premises
- The idea is to focus on cases where one derives a contradiction with a certain fact and the only uncertain elements are ordinary premises

The SSDOP family

Definition 12. *Let $AS = (\mathcal{L}, \bar{\ }, \mathcal{R}, n)$ be an argumentation system and $\mathcal{K}$ a knowledge base in AS. An argumentation theory $AT = (AS, \mathcal{K})$ is said to be an instance of the SSDOP family if the following conditions hold:*

- *the language $\mathcal{L}$ consists of the closure of a given set Σ of symbols and their negation, namely $\mathcal{L} = \Sigma \cup \{\neg s \mid s \in \Sigma\}$;*
- *the contrariness function coincides with the classical notion of negation: for every $s \in \Sigma$, $\bar{s} = \{\neg s\}$ and $\overline{\neg s} = \{s\}$;*
- *$\mathcal{R} = (\mathcal{R}_{Sb}, \emptyset)$, namely the set of defeasible rules is empty;*
- *$\nexists r, r' \in \mathcal{R}_{Sb} : cons(r) \in \overline{cons(r')}$, namely no contradiction can be derived using the strict rules only;*
- *$\mathcal{K}_n = \{\neg\omega\}$ for some $\omega \in \Sigma$ that will be called* contradiction focus*;*
- *$\forall r \in \mathcal{R}_{Sb}, ant(r) \cap \{\omega, \neg\omega\} = \emptyset$;*
- *the set of ordinary premises $\mathcal{K}_p$ satisfies the following conditions*
  - *$|\mathcal{K}_p| \geq 2$;*
  - *$\mathcal{K}_p \cap \{\omega, \neg\omega\} = \emptyset$;*
  - *$\nexists p, p' \in \mathcal{K}_p : p \in \overline{p'}$;*
  - *$\nexists r \in \mathcal{R}_{Sb} : cons(r) \in \mathcal{K}_p \cup \overline{\mathcal{K}_p}$*
  - *there is an argument α such that $Prem(\alpha) = \mathcal{K}_p$, $Conc(\alpha) = \omega$, and there is no argument α' such that $Prem(\alpha') \subsetneq \mathcal{K}_p$, $Conc(\alpha') = \omega$;*
  - *for every $p_1, p_2 \in \mathcal{K}_p$, $p_1 \simeq p_2$.*

A simple language

No defeasible rules

Contradiction only with the unique axiom

The axiom is not used by any rule

The ordinary premises do not "directly interphere" between them or with other elements

The ordinary premises together lead to a contradition with the axiom

The ordinary premises are equally preferred

A P-addition

- Given a SSDOP instance, a P-addition consists in just adding a premise to the antecedents of a strict rule

Definition 14. *Given an argumentation system* $AS = (\mathcal{L}, \bar{\ }, \mathcal{R}, n)$*, where* $\mathcal{R} = (\mathcal{R}_{Sb}, \emptyset)$ *and a knowledge base* $\mathcal{K}$ *such that* $AT = (AS, \mathcal{K})$ *belongs to the SSDOP family, we say that* $\mathcal{R}'_{Sb}$ *is a P-addition of* $\mathcal{R}_{Sb}$ *iff* $\exists r \in \mathcal{R}_{Sb}$ *such that* $\mathcal{R}'_{Sb} = (\mathcal{R}_{Sb} \setminus \{r\}) \cup \{r'\}$ *where* $cons(r') = cons(r)$ *and* $ant(r') = ant(r) \cup \{p\}$ *for some* $p \in \mathcal{K}_p$.

Basic spurious preference avoidance

- An argument evaluation mechanism which treats equally (credulous acceptance) the premises of a SSDOP instance should treat equally the premise in every P-addition of the instance

Definition 15. *An argumentation theory* $AT = (AS, \mathscr{K})$ *which belongs to the SSDOP family is* Cr-premise-fair *with respect to an evaluation mechanism* E *iff for every ordinary premise* $p \in \mathscr{K}_p$, $E_{AT}(p) = Cr$.

Definition 16. *An evaluation mechanism* E *satisfies the requirement of basic spurious preference avoidance if given any argumentation theory* AT *which is Cr-premise-fair with respect to* E*, it holds that every P-addition of* AT *is Cr-premise-fair too.*

ASPIC+ revisited (ASPIC^R)

- ASPIC^R has been introduced in 2018 to deal with another problem of ASPIC+ related to the presence of multiple contraries
- No time to present it in detail, only a couple of basic ideas

ASPIC+ revisited (ASPIC^R)

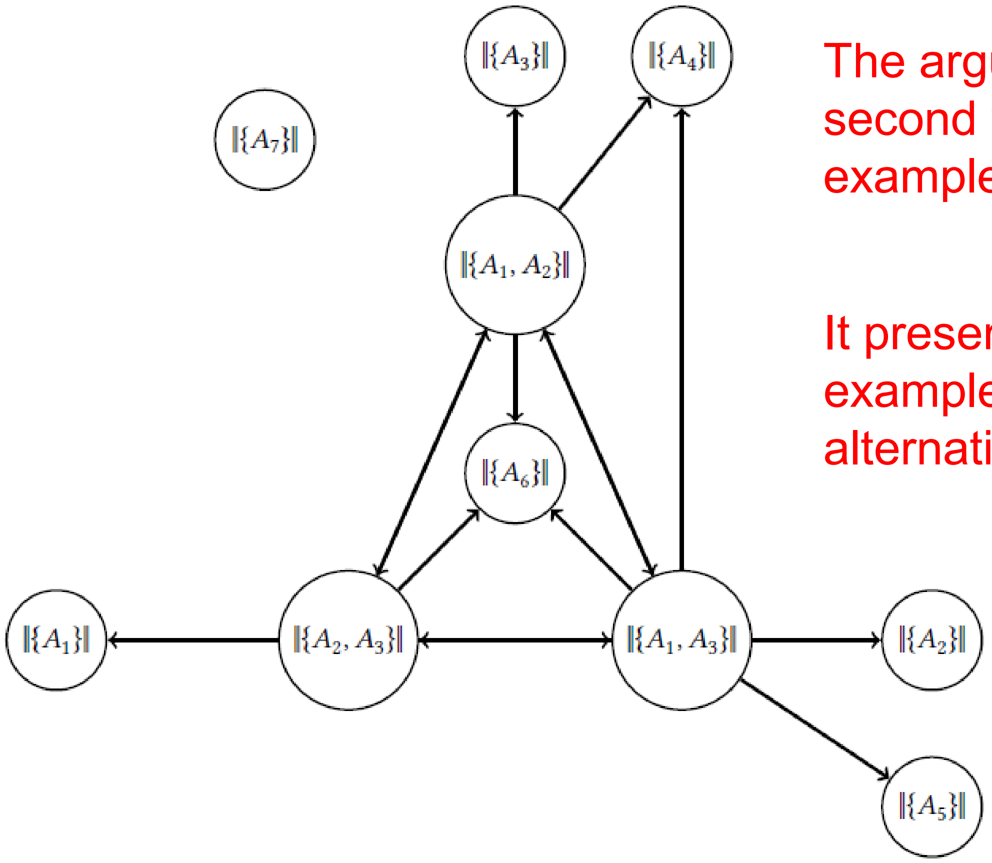
- Basic idea 1: satisfaction of rationality postulates is not achieved through closure by transposition of strict rules but through a closure of the contrariness relation (taking into account strict rules) at the level of sets of language elements
- Basic idea 2: conflicts occur between sets of arguments. Each node of the generated argumentation framework represents a set of arguments (including singletons)

Good news for ASPIC^R

Theorem 1. *The evaluation mechanism provided by* ASPIC^R *under the choice of preferred, stable, and semi-stable semantics satisfies the basic spurious preference avoidance requirement.*

- ASPIC^R behaves "natively well" in SSDOP instances, though it has been conceived to address a rather different issue

ASPIC^R at work



The argumentation framework for the second version of the taxpayers' list example is simpler than in ASPIC+

It preserves the structure of the first example with three mutually exclusive alternatives

Conclusions and future work

- Identified peculiar phenomenon in ASPIC+
- Shown that it is avoided (in a specific context) by ASPIC^R
- Many possible future developments
  - » Spurious preferences in other formalisms?
  - » More general characterization
  - » Which are the causes? Is any feature of ASPIC+ to blame?
  - » Language dependence of the phenomenon: $b, p \rightarrow c$ versus $(b \wedge p) \rightarrow c$