Abstract argumentation semantics: from limits to perspectives

Pietro Baroni
DII - Dip. di Ingegneria dell'Informazione
University of Brescia (Italy)

Roadmap

- Introduction and review

Dung's framework is (almost) nothing

Definition 2. An *argumentation framework* is a pair

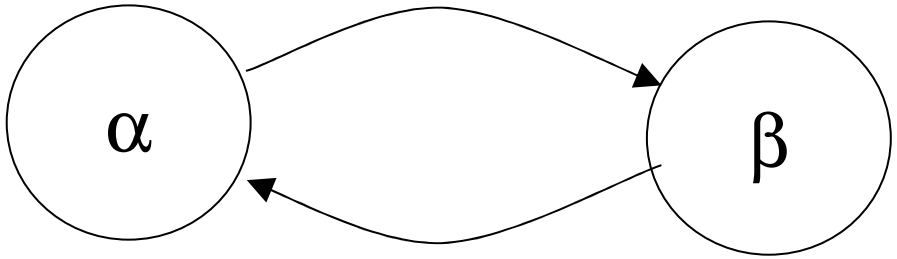
$$AF = \langle AR, attacks \rangle$$

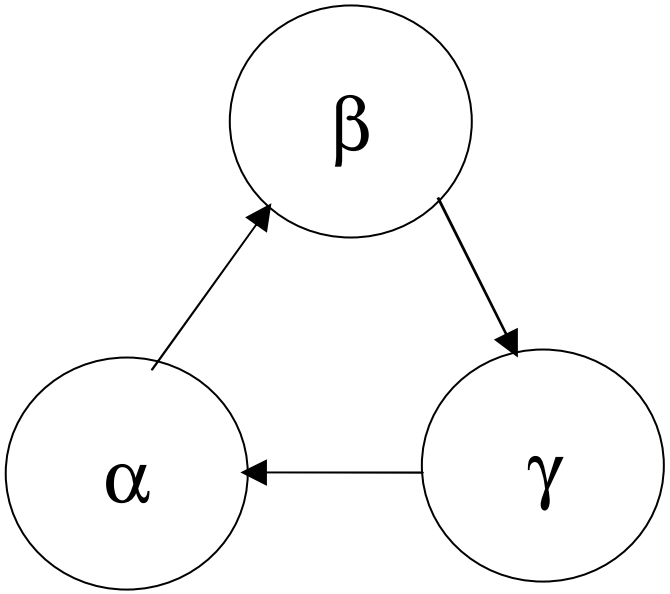
where AR is a set of arguments, and *attacks* is a binary relation on AR, i.e. $attacks \subseteq AR \times AR$.

- A directed graph (called *defeat graph*) where:
  - » arcs are interpreted as attacks
  - » nodes are called arguments "by chance" (let say historical reasons)

Here, an argument is an abstract entity
whose role is solely determined by its relations to other arguments. No special attention is paid to the internal structure of the arguments.

Dung's framework is (almost) nothing





Dung's framework is (almost) nothing

- Risk of rediscovering graph-theoretical results under new names and/or in specialized versions
- Too poor to be actually useful?
- Several extensions have been considered to enhance its expressiveness:
  - » Value-based argumentation frameworks
  - » Preference-based argumentation frameworks
  - » Bipolar argumentation frameworks

Dung's framework is (almost) everything

- Conflicts are everywhere
- Conflict management is a fundamental need with potential spectacular/miserable failures both in real life and in formal contexts (e.g. in classical logic)
- A general abstract framework centered on conflicts has a wide range of potential applications

Dung's framework is (almost) everything

- The pervasiveness of Dung's framework and semantics is witnessed by the correspondences drawn in the original paper with a variety of other formal contexts:
  - » default logic
  - » logic programming with negation as failure
  - » defeasible reasoning
  - » N-person games
  - » stable-marriage problem
- Many extensions and variations of Dung's framework allow a translation procedure back to the original framework to exploit its basic features

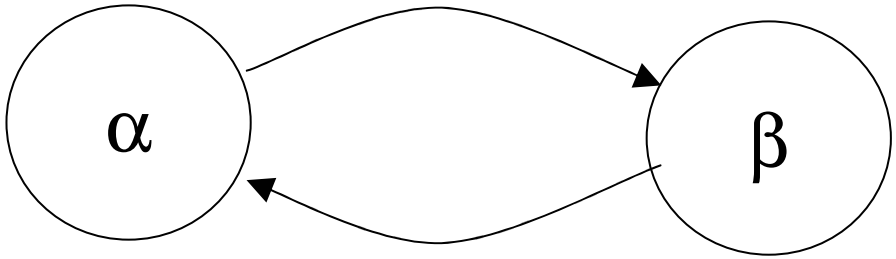
Abstract argumentation semantics

- A way to identify sets of arguments “surviving the conflict together” given the conflict relation only
- In general, several choices of sets of “surviving arguments” are possible
- The conflict-free principle (and no other one) is somehow embedded in the underlying intuition
- Two main styles for semantics definition: extension-based and labelling-based

Extension-based semantics

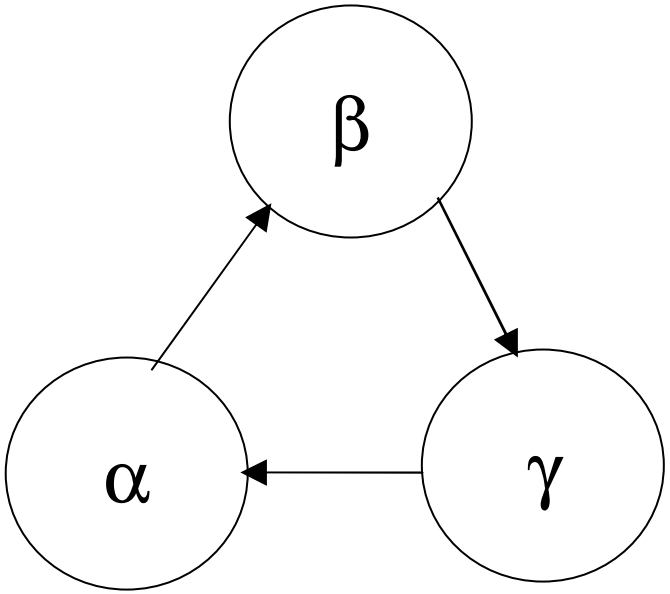
- A set of extensions is identified
- Each extension is a set of arguments which can “survive together” or are “collectively acceptable” i.e. represent a reasonable viewpoint
- The justification status of each argument can be defined on the basis of its extension membership

Sets of extensions



$E_1 = \{\{\alpha\},\{\beta\}\}$

$E_2 = \{\varnothing\}$



$E_1 = \{\{\alpha\},\{\beta\},\{\gamma\}\}$

$E_2 = \{\varnothing\}$

Labelling-based semantics

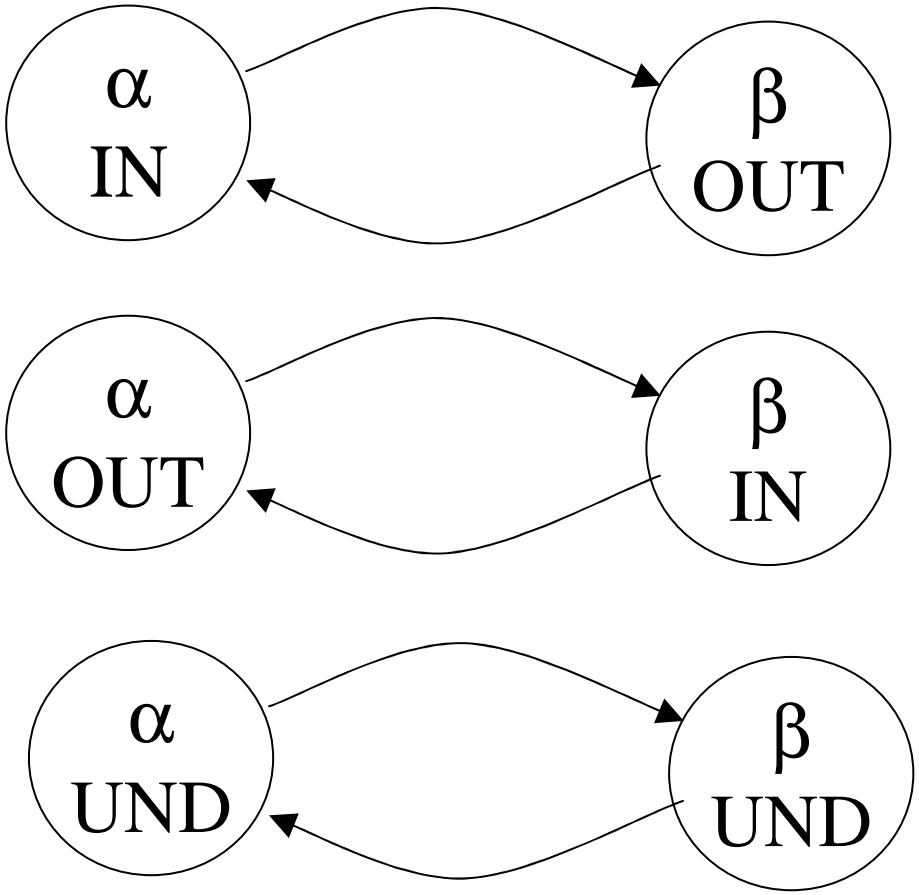
- A set of labels is defined (e.g. IN, OUT, UNDECIDED) and criteria for assigning labels to arguments are given
- Several alternative labellings are possible
- The justification status of each argument can be defined on the basis of its labels

Labelling-based semantics

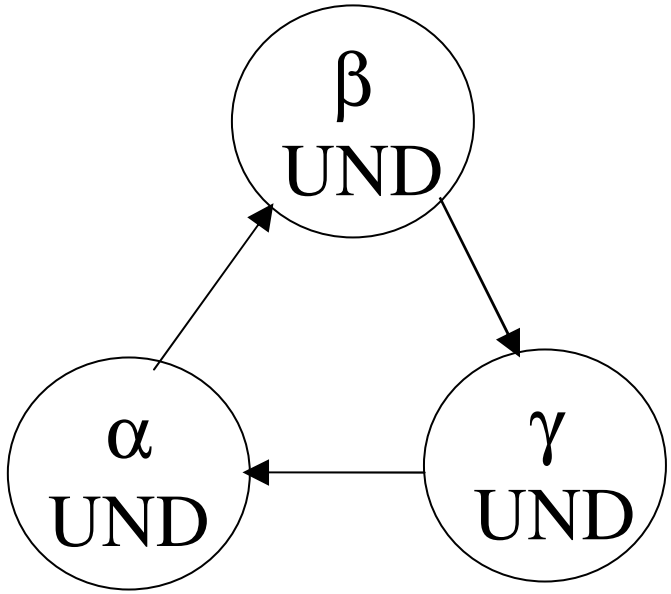
- A typical, but not the only conceivable, set of requirements on labellings consists of three simple rules

- If all attackers are OUT then the argument is IN
- If at least one attacker is IN then the argument is OUT
- If no attacker is IN and at least one attacker is UND then the argument is UND

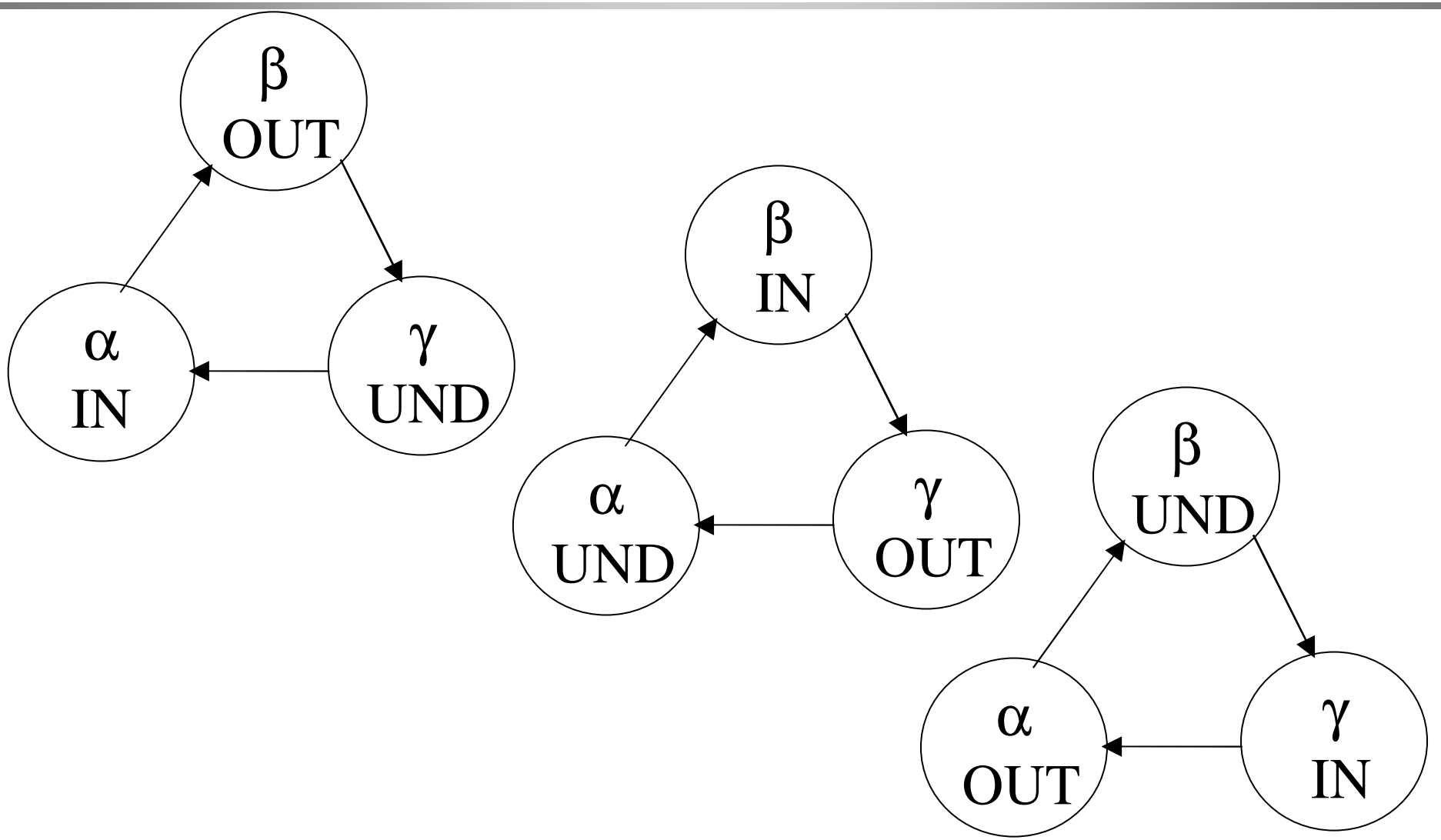
Labelling-based semantics



Labelling-based semantics



Labelling-based semantics



Labellings vs. extensions

- Labellings based on {IN, OUT, UNDEC} and extensions can be put in direct correspondence
- Given a labelling L, LabToExt(L) = in(L)
- Given an extension E, a labelling L=ExtToLab(E) can be defined as follows:
  in(L)=E
  out(L)=attacked(E)
  undec(L)=all other arguments

Dung's semantics

- Dung's original paper is focused on extension-based semantics
- Relatively simple intuitions underlying semantics definitions
- Dung's semantics are partly based on ideas in other pre-existing and less abstract formalisms and are related each other

Dung's "traditional" semantics

- Admissible set: defends (i.e. attacks the attackers of) its elements
- Complete extension: includes all arguments it defends
- Grounded extension: least complete extension (provably unique)
- Preferred extension: maximal admissible set ≡ maximal complete extension (in general not unique)
- Stable extension: conflict-free set attacking any other argument

Some “non-traditional” semantics

- Stage extension: conflict-free set with maximal range (union of arguments and attacked arguments)
- Semi-stable extension: complete extension with maximal range
- Ideal extension: maximal admissible set included in all preferred extensions (provably unique)
- CF2 and stage2 semantics: based on SCC decomposition of the defeat graph, can not be synthesized in a line
- Prudent semantics: variations of Dung’s traditional semantics based on the notion of “indirect conflict” (odd-length attack path)

Semantics principles: properties of extensions

- Conflict-free principle
- Admissibility and strong admissibility
- Reinstatement (with weak and CF versions)

Semantics principles: properties of sets of extensions

- I-maximality
- Directionality
- Skepticism-adequacy
- Resolution-adequacy

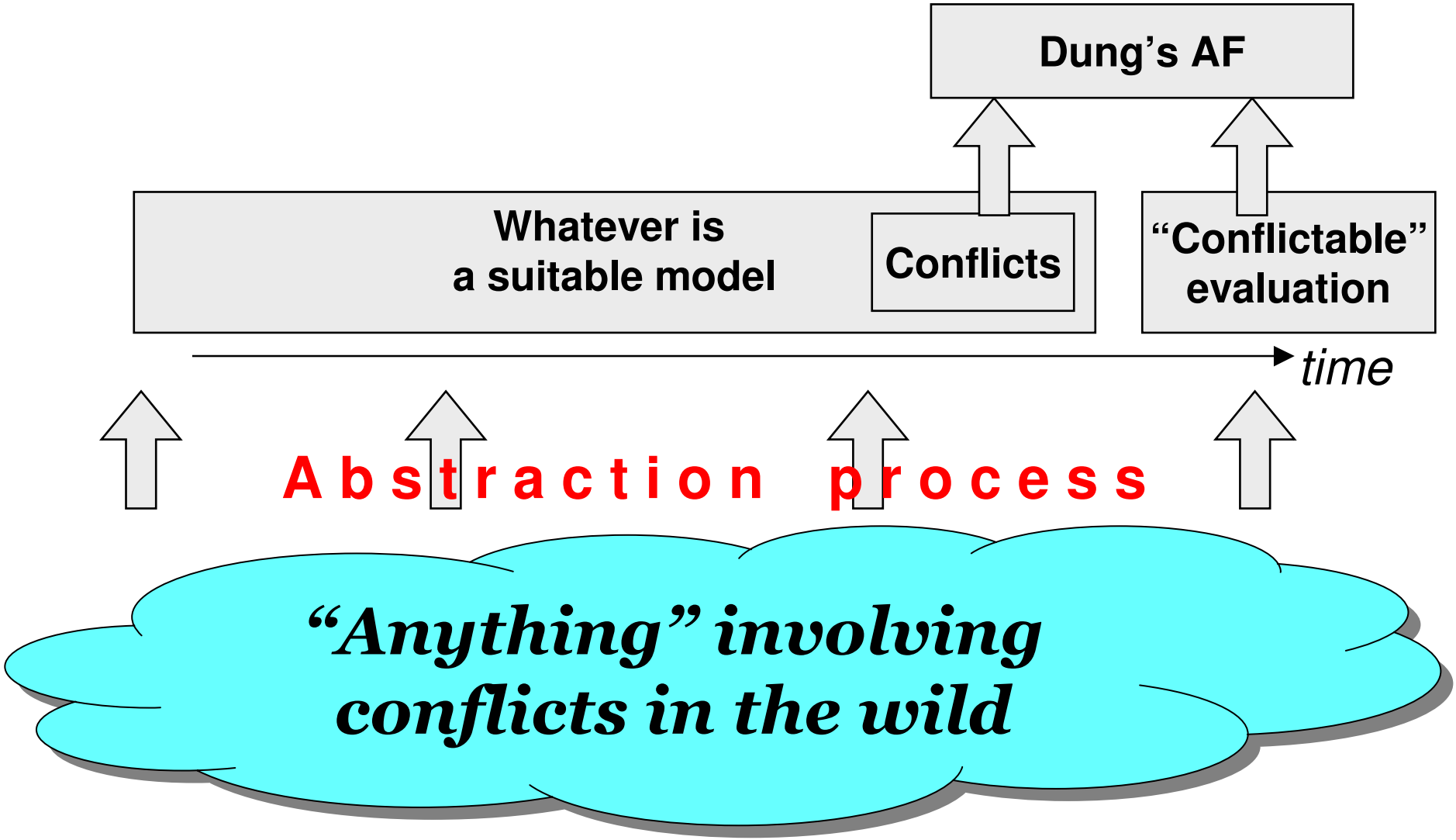
Semantics principles: properties wrt AF modifications

- Succinctness

Roadmap

- Introduction and review
- Too much (or too less) on conflicts?

Dung's AF: more and less



A logical bias?

- Many “instantiated argumentation” formalisms (ABA, DeLP, ASPIC+, …) assume an underlying logic and the derivation of arguments using some “inference rules”
- The emphasis on conflict might be related to the fact that, from a logical point of view, arguments *per se* are nothing really new, while having to cope with conflicts is
- Argument derivation is taken for granted and does not involve special relations between arguments

A logical bias?

- Argument construction is separated from argument evaluation (conflict management)
- “No reasoning” about the existence of conflicts
- Attacks come from other constructed arguments and are somehow related to the premises-rule-conclusion underlying structure
- Conflicts are binary
- Conflicts are all the same (at least in the evaluation)
- One or many (equally justified) attackers is the same
- Argument evaluation is rather crisp

Unbiasing

- Are there less biased (or differently biased) abstractions?
- Yes, both concerning argument structure and argument relations
- Less, as to my knowledge, on argument evaluation

Argumentation schemes

- Semi-formal model where arguments are instances of schemes, namely prototypical patterns of defeasible derivation of a conclusion from some premises
- A scheme is equipped with a set of critical questions, each stressing a specific aspect of the scheme (a sort of checklist)
- Direct relations with common-sense examples
- Sixty primary schemes (many with subschemes) in the Walton-Reed-Macagno 2008 book

Argumentation schemes

2. ARGUMENT FROM EXPERT OPINION

Major Premise: Source *E* is an expert in subject domain *S* containing proposition *A*.
Minor Premise: *E* asserts that proposition *A* is true (false).
Conclusion: *A* is true (false).

Critical Questions

CQ1: *Expertise Question*: How credible is *E* as an expert source?
CQ2: *Field Question*: Is *E* an expert in the field that *A* is in?
CQ3: *Opinion Question*: What did *E* assert that implies *A*?
CQ4: *Trustworthiness Question*: Is *E* personally reliable as a source?
CQ5: *Consistency Question*: Is *A* consistent with what other experts assert?
CQ6: *Backup Evidence Question*: Is *E*'s assertion based on evidence?

Argumentation schemes

3. ARGUMENT FROM WITNESS TESTIMONY

Position to Know Premise: Witness *W* is in a position to know whether *A* is true or not.
Truth Telling Premise: Witness *W* is telling the truth (as *W* knows it).
Statement Premise: Witness *W* states that *A* is true (false).
Conclusion: *A* may be plausibly taken to be true (false).

Critical Questions

CQ1: Is what the witness said internally consistent?
CQ2: Is what the witness said consistent with the known facts of the case (based on evidence apart from what the witness testified to)?
CQ3: Is what the witness said consistent with what other witnesses have (independently) testified to?
CQ4: Is there some kind of bias that can be attributed to the account given by the witness?
CQ5: How plausible is the statement *A* asserted by the witness?

Argumentation schemes

- Can be regarded as a sort of defeasible rule, but ...
- Is filling a scheme an inferential process?
- Just posing a critical question may affect an argument
- You don’t need to construct another argument to affect/attack an already existing one
- The idea of a non-just-logical prototypical and defeasible scheme is applicable also to other parts of the argumentation process

Argumentation schemes

- A chapter of the book is entitled “Attack, Rebuttal and Refutation”
- Detailed analysis and discussion of different types of conflicts
- More questions than answers
- Leaves you wondering whether all conflicts are (to be treated) the same
- Do we need “attack schemes”?

Roadmap

- Introduction and review
- Too much (or too less) on conflicts?
- An asset or a plethora?

An asset or a plethora?

- Motivating a new semantics with examples built directly jumping from a natural language description to abstract representation is a very risky game
- Many ambiguities and adhoceries may be hidden in this “too long step”
- Motivating a new semantics with general principles is (probably) a less risky game, but also principles may be questionable and may have no direct relationships with applications

A theory/application interplay?

- Identify an application area where conflict resolution plays a key role
- Define an abstraction procedure from application problem instances to Dung's framework
- Define a "counter-abstraction" procedure to map Dung's extensions into problem solutions
- Try different semantics and check:
  - » do the corresponding solutions make sense?
  - » do alternative semantics give an insight on novel solution strategies in the original problem?

What can be learned?

- Some semantics may not fit some applications
- An application $\leftrightarrow$ semantics map is badly needed

- Different semantics may not make any difference: under some topological conditions many (or all) semantics agree
- An application $\leftrightarrow$ topology map is badly needed

- Different semantics correspond to different flavors of the application problems

What can be learned?

- Different semantics correspond to different flavors of the original application problems
- Example of maps between abstract semantics principles and application-related principles (or intuitions) in some contexts would be very useful

- Benchmark problems are more than badly needed to stimulate the discussion within and outside the community and to provide some guidelines to an otherwise anarchic (but, in a sense, very creative) research development

Roadmap

- Introduction and review
- Too much (or too less) on conflicts?
- An asset or a plethora?
- Abstracting even more

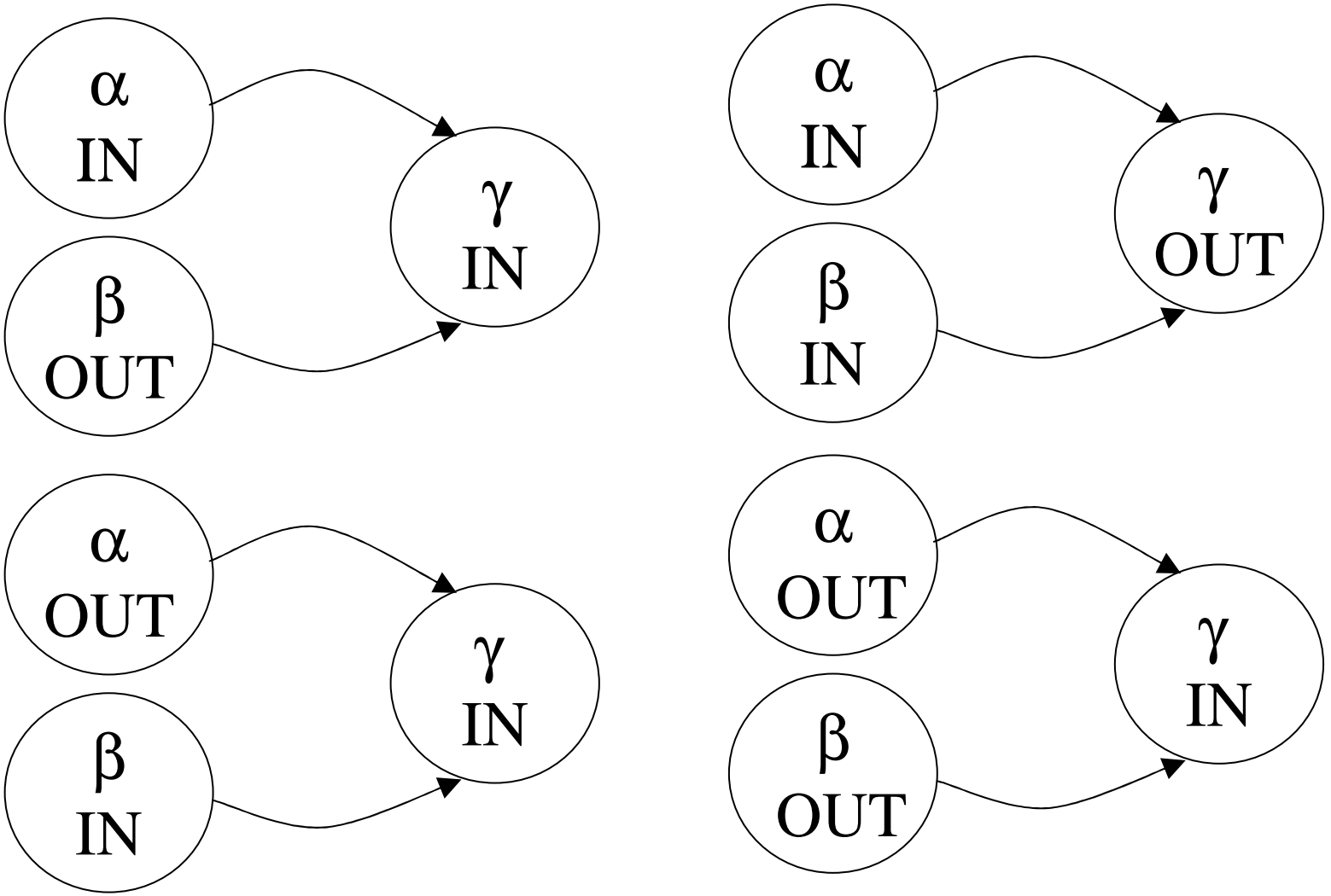
Even more abstract: abstract dialectical frameworks

Definition 5. An *abstract dialectical framework* is a tuple $D = (S, L, C)$ where

- S is a set of statements,
- $L \subseteq S \times S$ is a set of links,
- $C = \{C_s\}_{s \in S}$ is a set of total functions $C_s : 2^{par(s)} \rightarrow \{in, out\}$, one for each statement s. C_s is called acceptance condition of s.

- Even the nature of the relation between “arguments” is not specified: links of different nature all belong to the relation L
- All the meaning is embedded into the acceptance conditions (one for each node: heterogeneous situations may occur)

A non-Dung semantics: "unanimity of attacks"



More than a plethora

- ADFs represent an alternative perspective where the only embedded principle seems the one of directionality (rather than conflict-free)
- Large variety of “semantics”, actually of acceptance functions, even inside the same framework
- Semantics evaluation principles and skepticism comparisons to be revisited/redefined in this more general formal context
- A new unexplored universe for lovers of abstract argumentation semantics

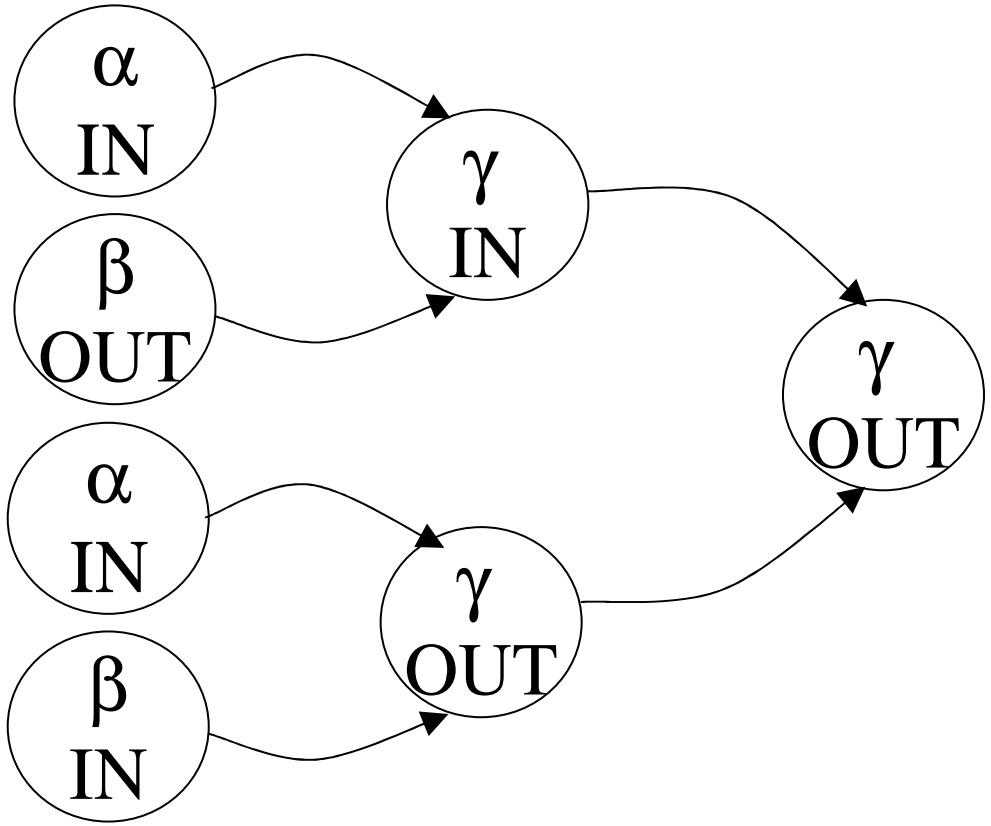
Revisiting principles: conflict-freeness

- The “unanimity of attacks” violates the traditional conflict-free principle (assuming L represents attacks only)
- Weak conflict-freeness of the acceptance condition: C(par(s))=OUT
- A possible spectrum of conflict free properties

- Dually, weak reinstatement in the acceptance condition:
  C(∅)=IN

Mixing heterogeneous acceptance functions

- Are all kinds of acceptance functions freely mixable?



Mixing heterogeneous acceptance functions

- Given the properties of the individual acceptance functions, which properties can be derived for the global result?

- Are there principles/requirements on the global result driving/constraining the definition of the individual acceptance functions?

Roadmap

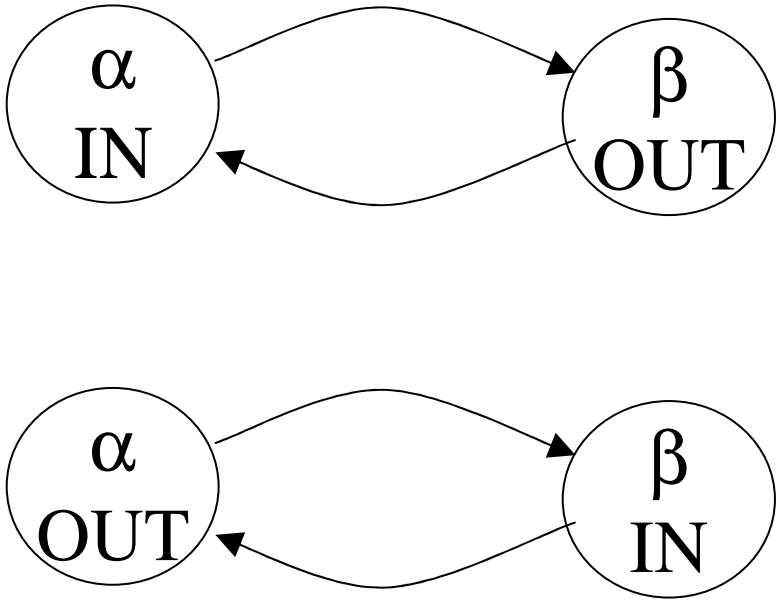
- Introduction and review
- Too much (or too less) on conflicts?
- An asset or a plethora?
- Abstracting even more
- A richer notion of justification status (beyond three labels)

Is three the perfect number?

- Most works on labellings in the literature adopt the so called “Caminada-labelling” with three possible labels: IN, OUT, UNDEC
- As we have already seen, one can freely move from 3-labellings to extensions and viceversa
- Accordingly, 3-labellings and extensions are alternative ways to express the same thing
- However labellings have an “unlimited” potential if one goes beyond the three “standard” labels

Justification states

- A semantics prescribes a set of labellings (extensions): an argument gets one or more different labels from a set of labellings



Justification states

- To summarize the justification state of an argument it seems “natural” to consider the set of labels the argument gets in the alternatives prescribed by a semantics
- Seven states
  {IN} : accepted in all alternatives
  {OUT} : rejected in all alternatives
  {UND} : undecided in all alternatives
  {IN,OUT} : “controversial” accepted or rejected
  {IN,UND} : not always accepted, never rejected
  {OUT,UND} : not always rejected, never accepted
  {IN,OUT,UND} : anything possible – who knows

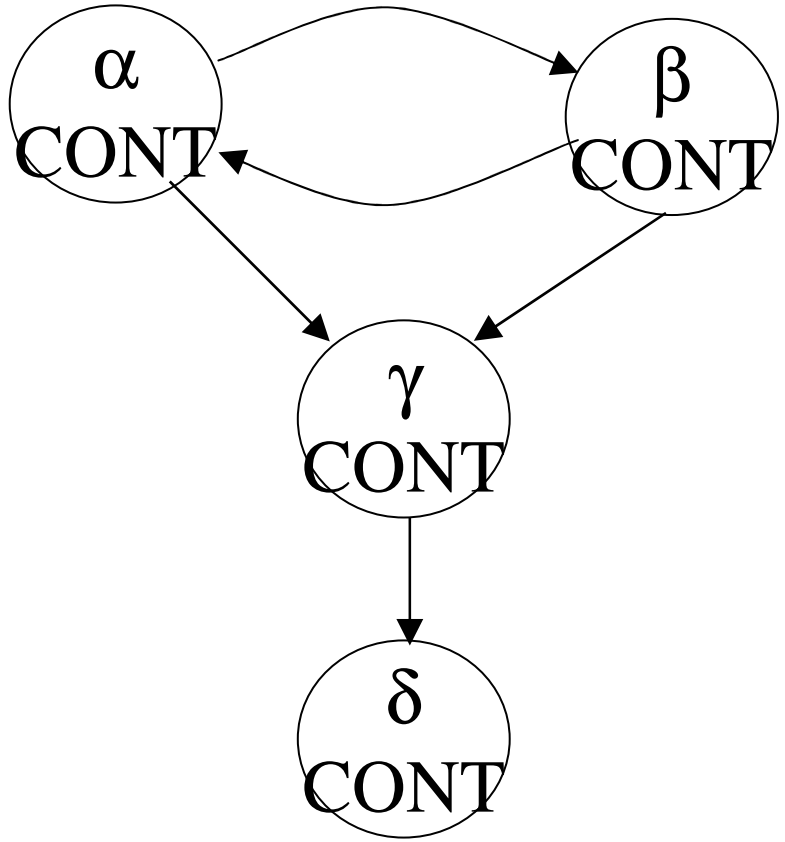
Is seven the perfect number?

- One could adopt the seven justification states directly as labels rather than as a derived concept and define non-Dung semantics
- Full redefinition of labelling principles needed
  From:
  if an argument has an attacker IN then it should be OUT
  To:
  if an argument has an attacker CONTROVERSIAL then …
  it can not be IN

  if an argument has all attackers CONTROVERSIAL then …
  it should be CONTROVERSIAL

Using directly the seven labels…

- “Non standard” outcomes are possible



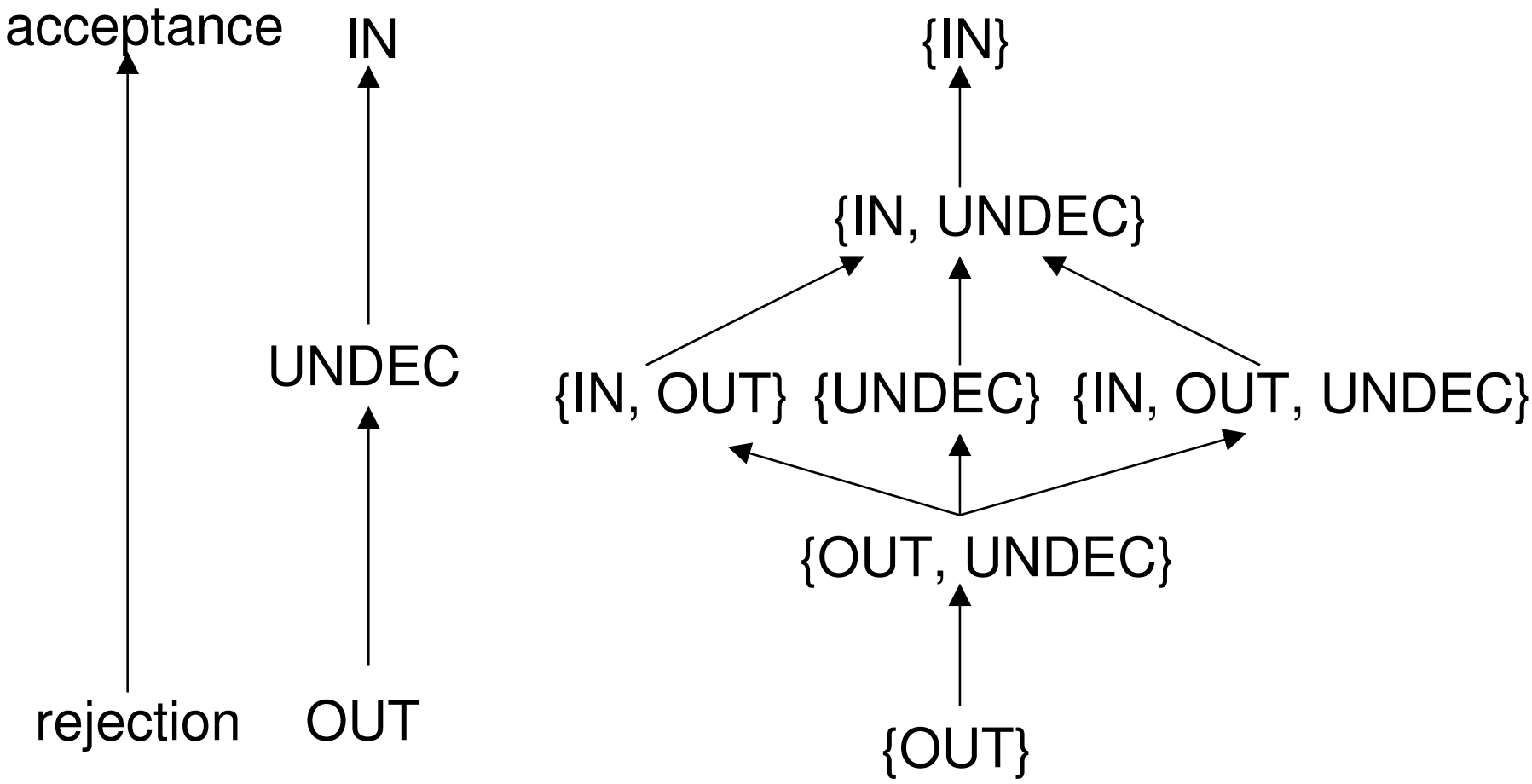
Using directly the seven labels…

- Hardly fits semantics notions like “maximal admissible set” implicitly based on the IN or OUT alternative
- Can be encompassed in directionality/topology centered approaches like the acceptance function of abstract dialectical frameworks or the SCC-recursive scheme

Why just seven?

- Human reasoning is rich of nuances and gradual evaluations
- What makes a set of labels suitable for argumentation labellings?
- Identifying at least the cases of definite acceptance, definite rejection and an intermediate case
- Ordering labels

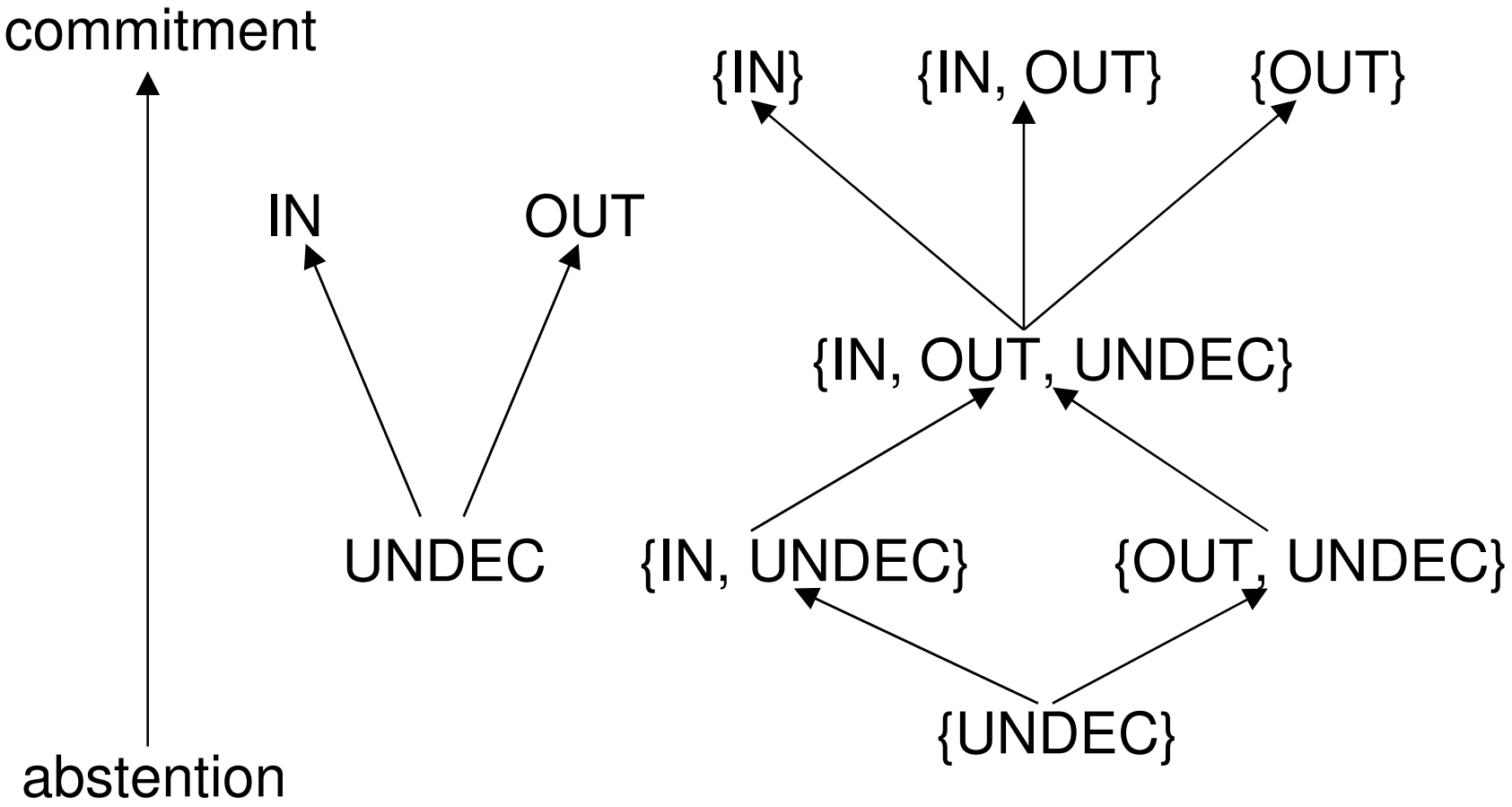
Ordering labels



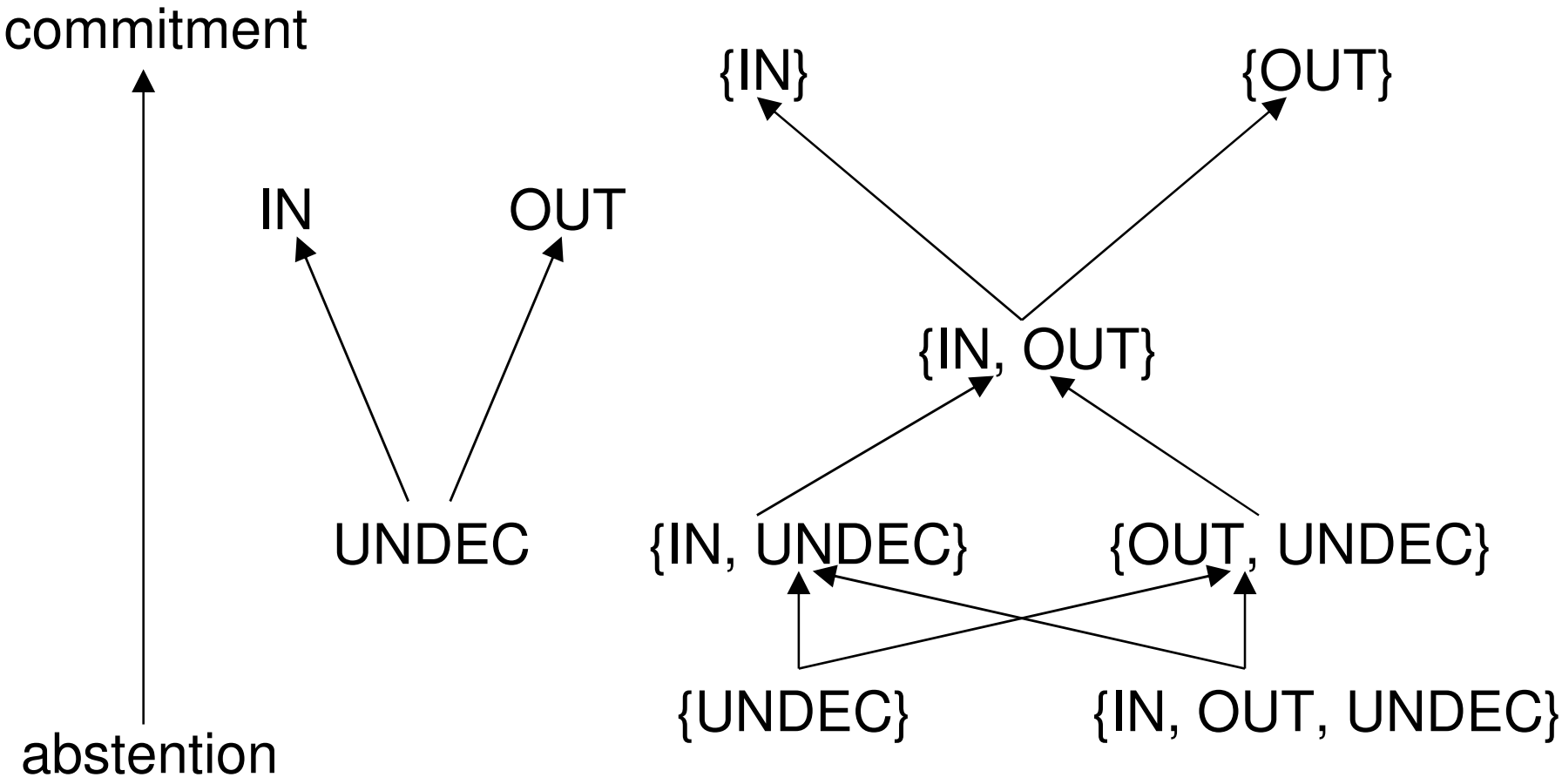
Commitment ordering

- Ordering according to “acceptance level” is not the only meaningful/useful one in a set of labels
- Different commitment levels can be identified: a label is more committed if it corresponds to a more clearcut choice
- Definite acceptance and definite rejection are equivalent according to commitment
- The commitment ordering may play a key role in defining principles for labellings and for skepticism comparison

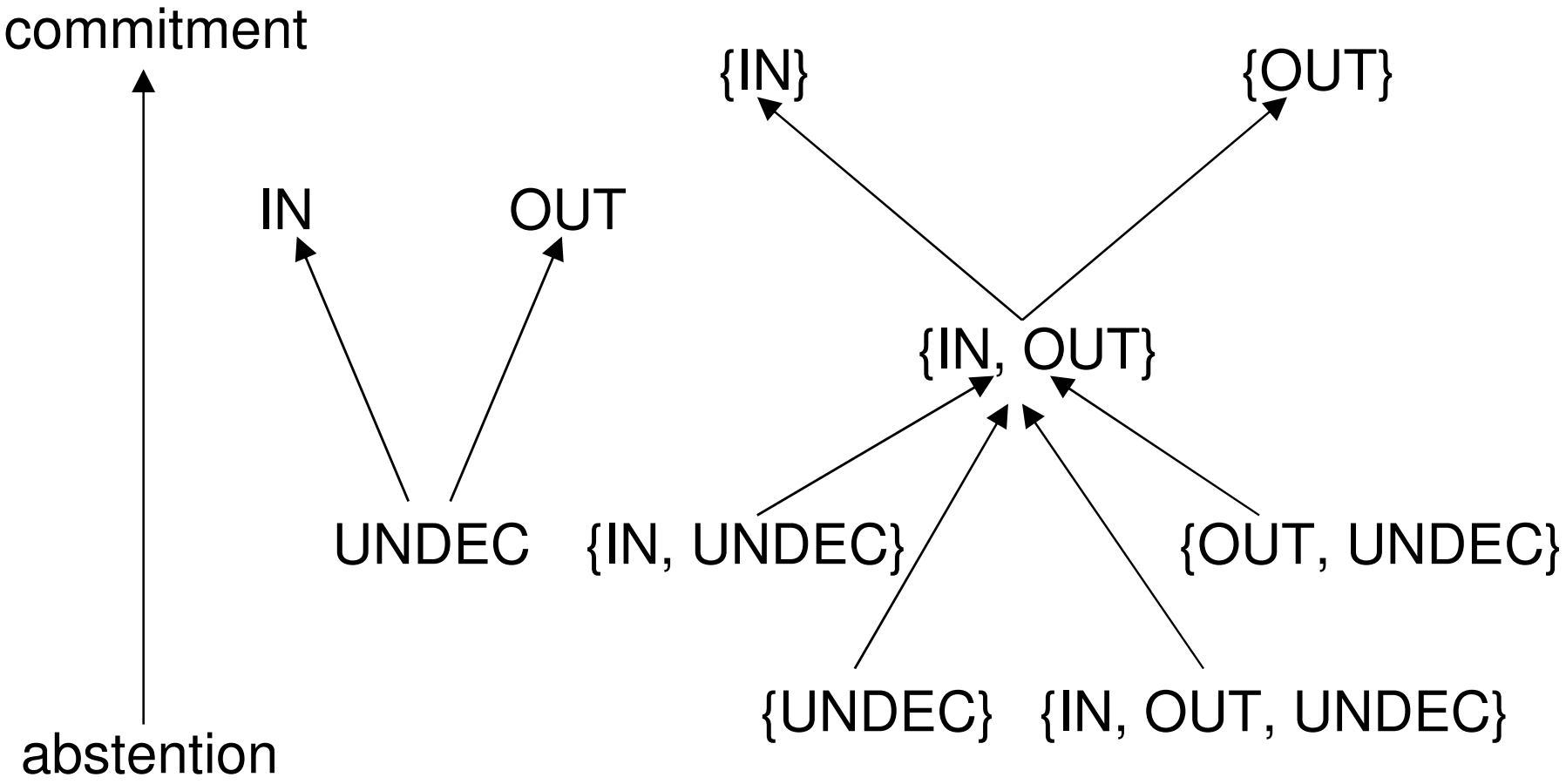
Commitment ordering



Another commitment ordering



A simpler commitment ordering



Roadmap

- Introduction and review
- Too much (or too less) on conflicts?
- An asset or a plethora?
- Abstracting even more
- A richer notion of justification status (beyond three labels)
- Collective attacks

Argument interactions

- Both attacks in Dung's framework and links in abstract dialectical frameworks are binary relations
- Arguments interact one-to-one
- "Simple" one-to-one interactions are the basis of a rich set of more articulated notions
- Are binary relations "too simple" and implicitly limiting the range of derivable notions?
- Collective attacks have been considered early in argumentation literature, but were shadowed by the prevailing Dung's wave

Collective attacks

- In the “semi-abstract” approach of Vreeswijk (AIJ 97 “Abstract argumentation systems”) a defeater of an argument A is a set S of arguments being altogether incompatible with A
- In the extension of Dung’s framework by Nielsen and Parsons attacks arise from sets of arguments

Definition 1 (Argumentation System*). *An* argumentation system *is a pair* $(\boldsymbol{A}, \triangleright)$*, where* $\boldsymbol{A}$ *is a set of arguments, and* $\triangleright \subseteq (\mathcal{P}(\boldsymbol{A}) \setminus \{\emptyset\}) \times \boldsymbol{A}$ *is an* attack relation.

Collective attacks

- Nielsen and Parsons provide a complete and "seamless" reformulation of Dung's theory and "traditional" extension-based semantics
- They use a "partial" notion of defense (it suffices to attack one of the members of an attacking set)

We say that a set of arguments $\boldsymbol{S}$ *attacks* an argument A, if there is $\boldsymbol{S'} \subseteq \boldsymbol{S}$ such that $\boldsymbol{S'} \triangleright A$. In that case we also say that A *is attacked by* $\boldsymbol{S}$. If there is no set $\boldsymbol{S''} \subsetneq \boldsymbol{S'}$ such that $\boldsymbol{S''}$ attacks A, then we say that $\boldsymbol{S'}$ is a *minimal* attack on A. Obviously, if there exists a set that attacks an argument A, then there must also exist a minimal attack on A. If for two sets of arguments $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$, there is an argument A in $\boldsymbol{S}_2$ that is attacked by $\boldsymbol{S}_1$, then we say that $\boldsymbol{S}_1$ attacks $\boldsymbol{S}_2$, and that $\boldsymbol{S}_2$ is attacked by $\boldsymbol{S}_1$.

Let $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ be sets of arguments. If $\boldsymbol{S}_2$ attacks an argument A, and $\boldsymbol{S}_1$ attacks $\boldsymbol{S}_2$, then we say that $\boldsymbol{S}_1$ is a *defense* of A from $\boldsymbol{S}_2$, and that $\boldsymbol{S}_1$ *defends* A from $\boldsymbol{S}_2$. Obviously, if $\boldsymbol{S}_3$ is a superset of $\boldsymbol{S}_1$, $\boldsymbol{S}_3$ is also a defense of A from $\boldsymbol{S}_2$.

A significant expressivity gain

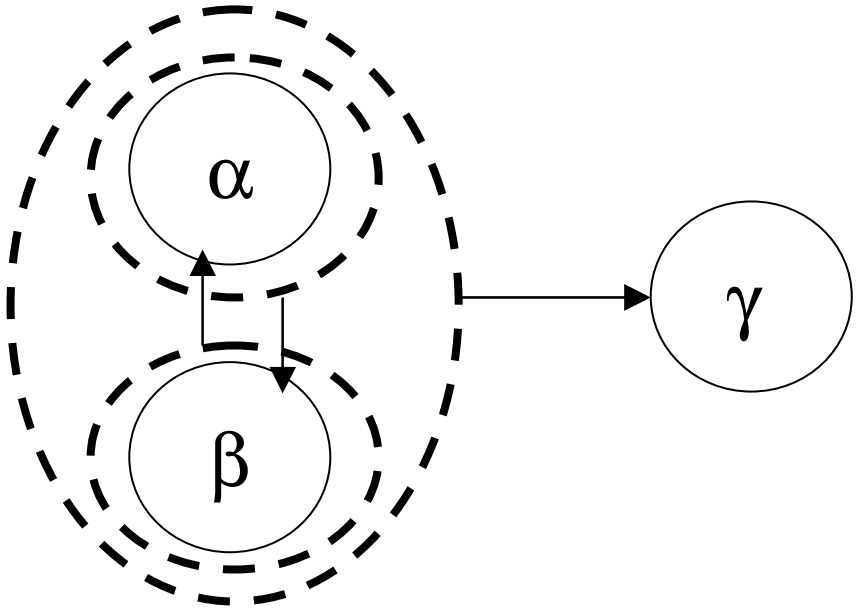
- Attacks are sometimes interpreted as arising from some form of incompatibility relation
- But incompatibility may be non-binary

- Alternative actions requiring bounded resources may not be pairwise incompatible but larger sets of actions can be unfeasible
- A logical contradiction may arise from a set of sentences which are not pairwise incompatible

A challenge for principles

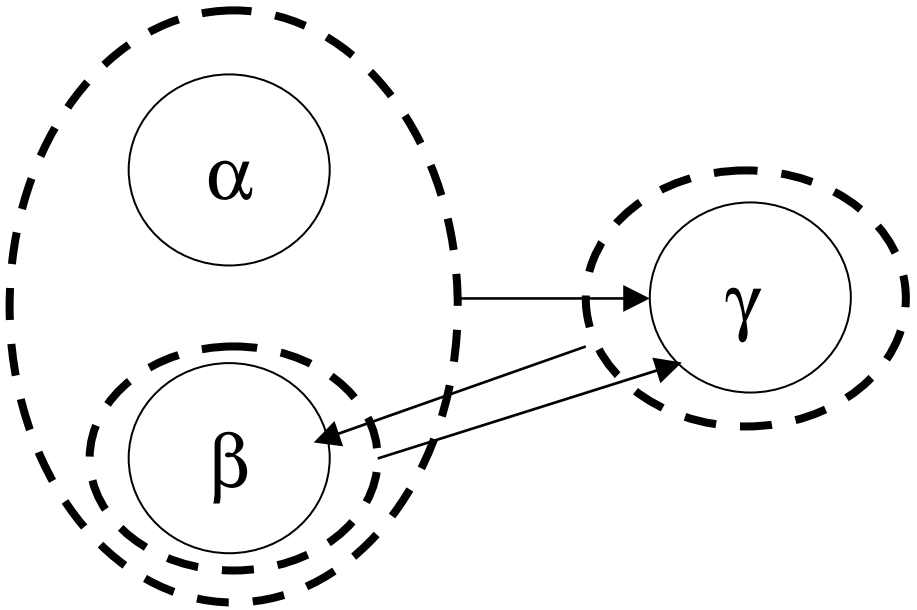
- Principles for standard extension-based semantics should be extended to the case of attacking sets (maybe not too difficult)
- Are “soundness” principles for the attack relation needed?
- Is there any advantage in imposing attacks to arise from conflict-free sets?
- Should there be a minimality requirement in the definition of $\triangleright$?

Sound $\triangleright$ relations?



γ is defended exactly by the set attacking it

Sound $\triangleright$ relations?



γ receives two distinct attacks but one counterattack is enough for defense

A challenge for labelling?

- The labelling-based approach to semantics definition appears more general and “flexible” than the extension-based one
- However, the extension-based approach can be “directly” upgraded to the case of attacking sets
- Upgrading the labelling-based approach seems less immediate: a labelling of sets of arguments is needed with “circular dependencies” wrt the labelling of individual arguments

A challenge not only for labelling?

- The definition of Nielsen and Parsons seems to be based on the intuition of “unanimity of attacks” for sets of arguments
- Other alternative intuitions are possible, e.g. ”survival of a single attack” or even “majority of attacks”

Conclusions on abstract argumentation …

- Practical applications may need rethinking (reasoning benchmarks)
- Semantics notions may need rethinking (abstract dialectical frameworks)
- The attack relation may need rethinking (collective attacks)
- The status of arguments may need rethinking (beyond three labels)

TUTTO SBAGLIATO
TUTTO DA RIFARE !


Thank you for
your attention!