Argumentation in Artificial Intelligence,
With Applications in the Law

Course at the Institute of Logic and Cognition,
Sun Yat-Sen University

Ib Abstract Argumentation and Argument Structure

Bart Verheij
CodeX, Stanford University
Artificial Intelligence, University of Groningen
`www.stanford.edu/~bartv, www.ai.rug.nl/~verheij`

IA Introduction

Topics:
- Argumentation in Artificial Intelligence
- Historical Background

Goals:
- Get an overview of the course and its subject matter
- Acquire insight about the historical background

Literature:
- Van Eemeren et al. (in preparation). Sections 11.1-11-3.

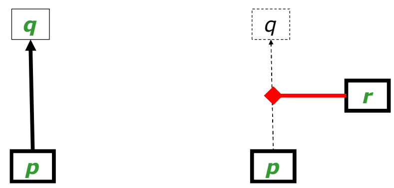
IB Abstract Argumentation, Argument Structure

Topics:
- Abstract Argumentation
- Argument Structure

Goals:
- Acquire knowledge of abstract argumentation and its semantics
- Acquire insight into the relation between argument structure and abstract argumentation

Literature:
- Van Eemeren et al. (in preparation). Sections 11.4-11.5.

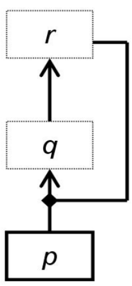
Pollock's undercutting defeat



q is warranted by the argument from *p*

q is not warranted by the argument from *p*

A philosophical puzzle (Pollock)



p	p
$p \Rightarrow q$	p is a prima facie reason for q
$q \Rightarrow r$	q is a prima facie reason for r
$r \Rightarrow \neg(p \Rightarrow q)$	r is an exception that undercuts the support of q by p

Is q warranted by the argument from p?

Pollock's research question

How is argumentative warrant determined by the structure of the available arguments and counterarguments?

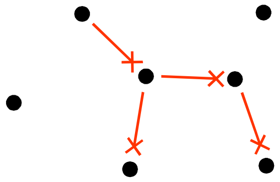
He produced a series of proposals, amongst other things driven by philosophical puzzles.

Dung 1995

‘On the acceptability of arguments
and its fundamental role
in non-monotonic reasoning,
logic programming
and n-person games’

Artificial Intelligence journal

The attack relation as a directed graph (Dung)



Pollock’s research question, revisited

How is argumentative warrant
determined by the structure
of the available arguments
and counterarguments?

Pollock’s research question, revisited

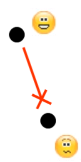
How is argumentative warrant
determined by the structure
~~of the available arguments~~
~~and counterarguments?~~
of the attack relation between arguments?

- Mathematically clean
- More abstract, so simpler structure

Dung’s basic principle of argument acceptability

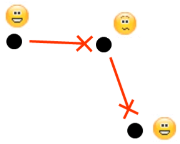
The one who has the last word laughs best.

Dung’s basic principle of argument acceptability



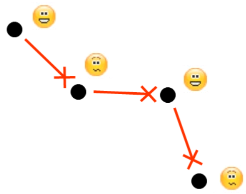
The one who has the last word laughs best.

Dung’s basic principle of argument acceptability



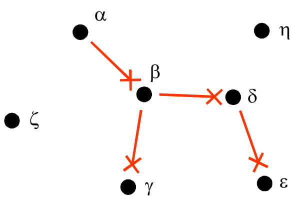
The one who has the last word laughs best.

Dung’s basic principle of argument acceptability



The one who has the last word laughs best.

Admissible sets



Admissible, e.g.: {α, γ}, {α, γ, δ, ζ, η}
Not admissible, e.g.: {α, β}, {γ}

Admissible sets

A set of arguments A is admissible if

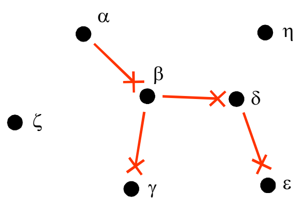
1. it is *conflict-free*: There are no arguments α and β in A, such that α attacks β.
2. the arguments in A are *acceptable* with respect to A: For all arguments α in A, such that there is an argument β that attacks α, there is an argument γ in A that attacks β.

Dung’s preferred and stable extensions

An admissible set of arguments is a *preferred extension* if it is an admissible set that is maximal with respect to set inclusion.

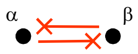
A conflict-free set of arguments is a *stable extension* if all arguments that are not in the set are attacked by an argument in the set.

Admissible sets



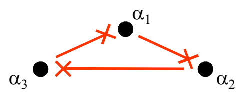
Preferred and stable extension: {α, γ, δ, ζ, η}

Even-length attack cycles



Preferred and stable extensions: {α}, {β}

Odd-length attack cycles



Preferred extensions: ∅ (the empty set)
Stable extensions: none

Basic properties of Dung's extensions

- A stable extension is a preferred extension, but not the other way around.
- An attack relation always has a preferred extension. Not all attack relations have a stable extension.
- An attack relation can have more than one preferred/stable extension.
- A well-founded attack relation has a unique stable extension.

Dung's grounded and complete extensions

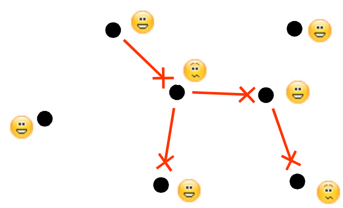
A set of arguments is a *complete* extension if it is an admissible set that contains all arguments of which all attackers are attacked by the set.

A set of arguments is a (the) *grounded* extension if it is a minimal complete extension.

Dung's four semantics

Preferred
Stable
Complete
Grounded

Labelings



Labelings

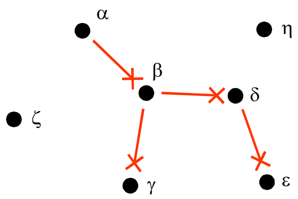
Stable labeling:

An argument α is labelled "Defeated"

if and only if

There is an argument β that attacks α and that is labelled "Justified."

Stages



Stages, e.g.: β (γ), α (β) γ, α (β) γ δ (ε) ζ η
Non-stages, e.g.: β γ, β (δ ε)

Labelings

1. Using labelings instead of sets simplifies the formal analysis and increases its transparency.

2. Labelings allow a new natural idea of maximal interpretation: maximize the set of labeled nodes.
   → Stage extensions

3. *Some preferred extensions are better than others.*
   → Semi-stable extensions

Verheij (1996). Two Approaches to Dialectical Argumentation: Admissible Sets and Argumentation Stages.

Semi-stable semantics

A set of arguments is a *semi-stable extension* if it is an admissible set, for which the union of the set with the set of arguments attacked by it is maximal.

Notion introduced by Verheij (1996)
Term coined by Caminada (2006)

Properties

1. Stable extensions are semi-stable.
2. Semi-stable extensions are preferred.
3. Preferred extensions are not always semi-stable.
4. Semi-stable extensions are not always stable.

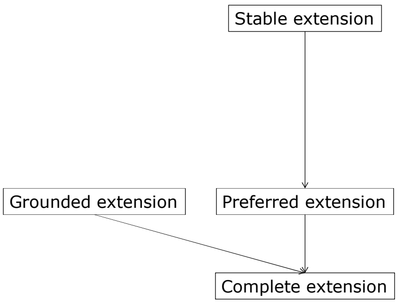
Preferred extensions always exist, but stable extensions do not.

Do all attack graphs have a semi-stable extension?
Answered negatively by Verheij (2000, 2003)

Properties

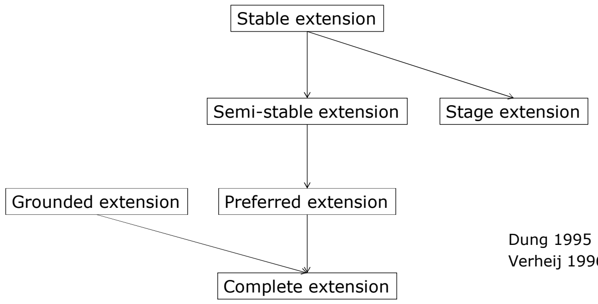
1. There exist attack graphs without a semi-stable extension.
2. Finite attack graphs always have a semi-stable extension.
3. An attack graph with a finite number of preferred extensions has a semi-stable extension.
4. An attack graph with a stable extension has a semi-stable extension.
5. If an attack graph has no semi-stable extension, then there is an infinite sequence of preferred extensions with strictly increasing ranges.

Abstract argumentation semantics (1995)



Dung 1995

Abstract argumentation semantics (1996)



Dung 1995
Verheij 1996

Pollock's research question, revisited

How is argumentative warrant
determined by the structure
~~of the available arguments~~
~~and counterarguments?~~
of the attack relation between arguments?

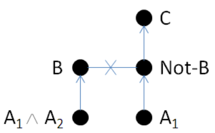
- Mathematically clean
- More abstract, so simpler structure
- ***Philosophically still complex***

What happens if we add structure?

Not just **attack**,
also **support**

Specificity

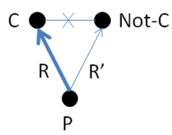
1. Conflict by inconsistency
2. Defeat by specificity



Simari & Loui 1992

Conclusive force

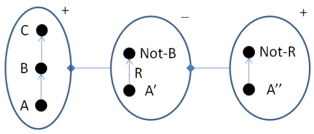
1. Conflict by inconsistency
2. Defeat by conclusive force



Vreeswijk 1997

Combining support and attack

Approach 1:
  Dung's abstract arguments have internal structure

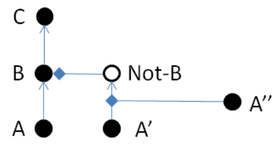


Abstract version:

ASPIC+ Prakken 2010

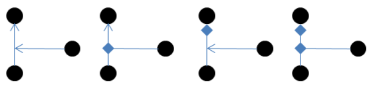
Combining support and attack

Approach 2:
  Arguments can attack or support



Nute 1994, DefLog Verheij 2003

Arguing about support and attack



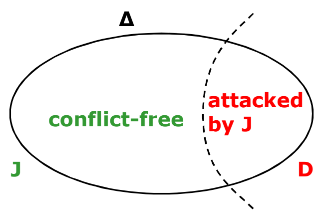
DefLog Verheij 2003
ArguMed software Verheij 2003

DefLog

A conditional ~> that validates Modus ponens
A connective × that expresses 'negation as defeat'
  (dialectical negation)

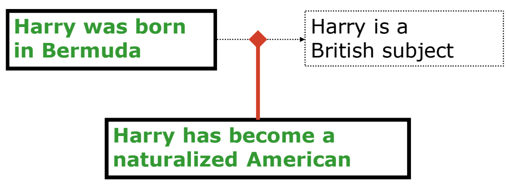
pro:	φ ~> ψ
con:	φ ~> $\times\psi$
warrant:	φ ~> (ψ ~> χ)
undercutter:	φ ~> $\times$(ψ ~> χ)
rebutter:	((φ ~> ψ) $\wedge$ φ) ~> $\times$(χ ~> not-ψ) or ψ ~> $\times$(χ ~> not-ψ)

DefLog

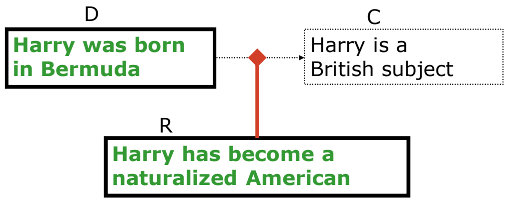


A **defeasible theory Δ** is divided in
a **justified part J** and a **defeated part D**.

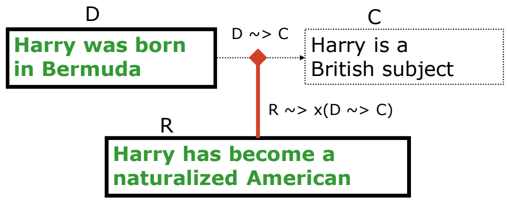
Attack I (no warrants)



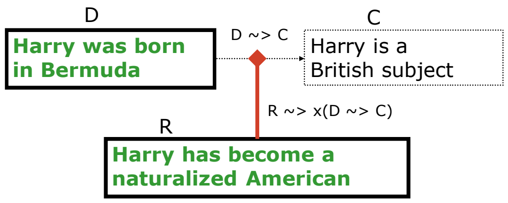
Attack I (no warrants)



Attack I (no warrants)

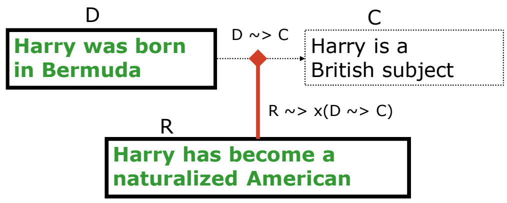


Attack I (no warrants)



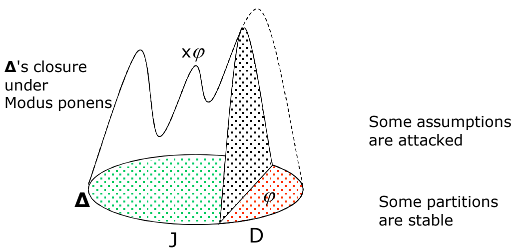
Δ = {D ~> C,
R ~> x(D ~> C),
R, D}

Attack I (no warrants)



Δ = {~~D ~> C,~~
R ~> x(D ~> C),
R, D}

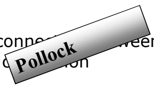
DefLog

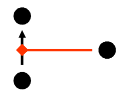


Some assumptions are attacked

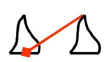
Some partitions are stable

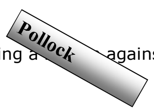
Undercutting & rebutting

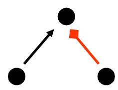
Undercutting-1:
Attacking the conneeen a reason and its c on



Undercutting-2:
Attacking an assumption of an argument



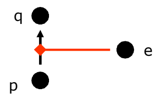
Rebutting:
Attacking by giving aagainst a conclusion



Undercutting-1 in DefLog

Attacking a conditional assumption

Δ:
- p ~> q $\in$ D
- e ~> x(p ~> q) $\in$ J
- p $\in$ J
- e $\in$ J



Undercutting-2 in DefLog

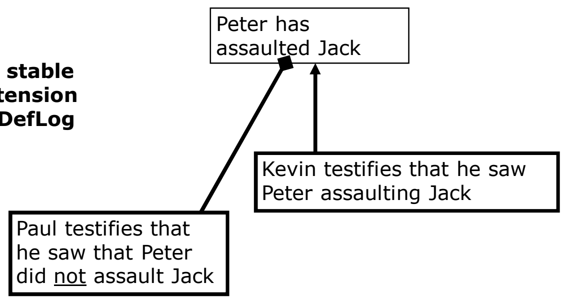
Passim

Rebutting in DefLog

Side-step:
- Conflicting reasons

Conflicting reasons

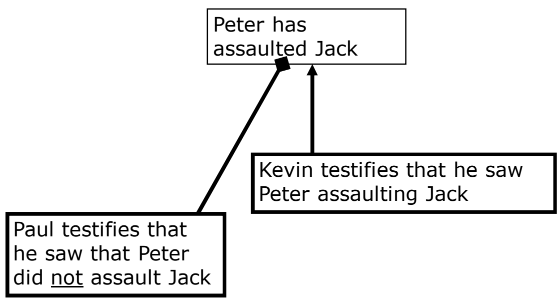
No stable extension in DefLog



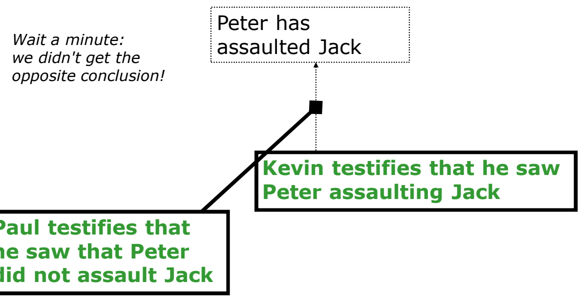
Rebutting in DefLog

Modelled using undercutting-1

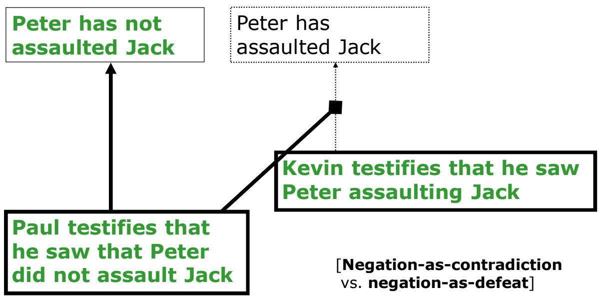
Conflicting reasons



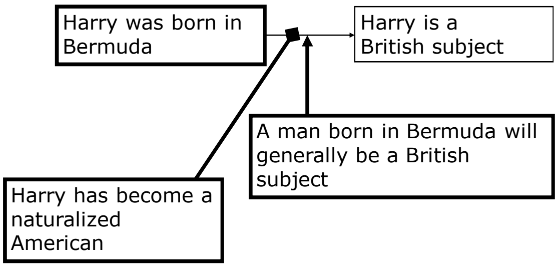
Conflicting reasons



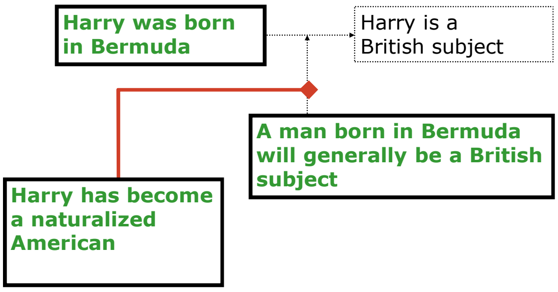
Conflicting reasons



Conflicting reasons



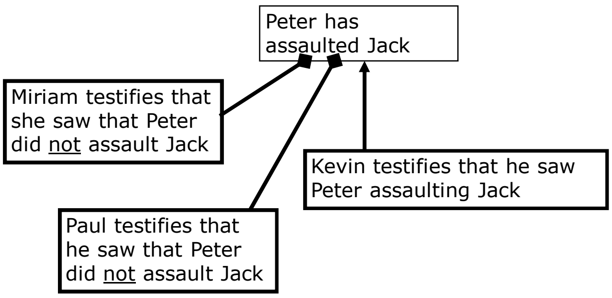
Attack II (with warrants)



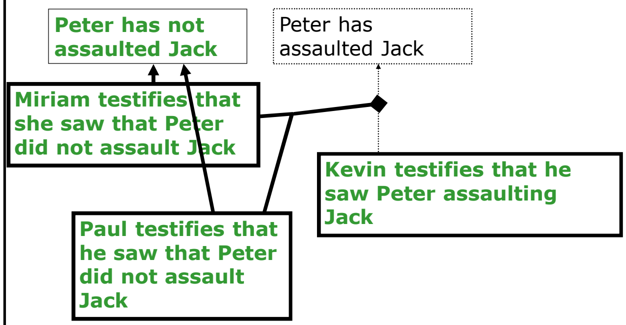
Rebutting in DefLog

Also the weighing of reasons (cf. Reason-Based Logic) can be modeled using undercutting-1.

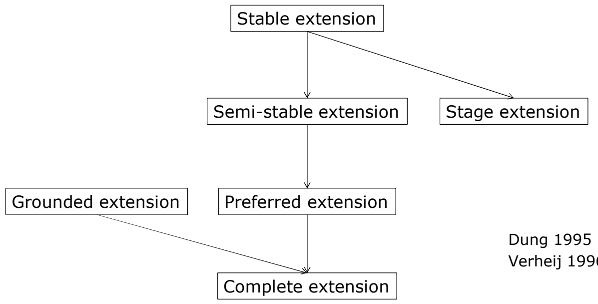
Conflicting reasons



Conflicting reasons

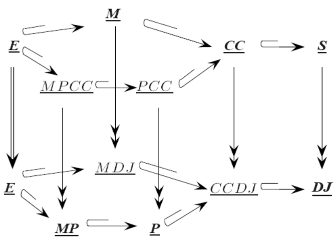


Abstract argumentation semantics (1996)



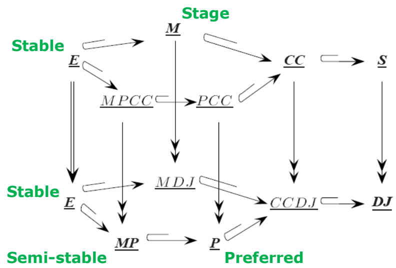
Dung 1995
Verheij 1996

Argumentation semantics (2003)



DefLog Verheij 2003

Argumentation semantics (2003)



DefLog Verheij 2003

Pollock's research question, revisited

We return to the original version (attack + support) :

How is argumentative warrant
determined by the structure
of the available arguments
and counterarguments?

- Mathematically clean (still)
- Less abstract, less simple structure
- Philosophically <u>very</u> complex

Argumentation in Artificial Intelligence, With Applications in the Law

Course at the Institute of Logic and Cognition, Sun Yat-Sen University

Ib Abstract Argumentation and Argument Structure

Bart Verheij
CodeX, Stanford University
Artificial Intelligence, University of Groningen
www.stanford.edu/~bartv, www.ai.rug.nl/~verheij