Spring School on Artificial Intelligence and Law

Invited graduate course
at Central South University, Changsha

中南大學
CENTRAL SOUTH UNIVERSITY

Floris Bex, Enrico Francesconi, Bart Verheij, April 2019

Organisers: Juan Li, Minghui Xiong

Day 1 Sunday April 14

8:30	Introduction to AI & Law (Bart Verheij)
10:00	*Break*
10:30	AI & Law - Semantic Web, Open Data and AI in the Legal Domain (Enrico Francesconi)
12:00	*Break*
14:30	Evidence, Stories and Arguments in AI & Law (Floris Bex)
16:00	*Break*
16:30	Discussion
17:30	

http://www.ai.rug.nl/~verheij/ssail2019/


HUMANS NEED NOT APPLY
A GUIDE TO WEALTH & WORK IN THE AGE OF ARTIFICIAL INTELLIGENCE
BY JERRY KAPLAN


BAN KILLER ROBOTS
OUR CALL TO ACTION »

Published this week

Human agency and oversight
Technical robustness and safety
Privacy and data governance
Transparency
Diversity, non-discrimination and fairness
Societal and environmental well-being
Accountability

Artificial intelligence

Specialized artificial intelligence

Exists and is often in use.

Tax administration, photo classification

General artificial intelligence

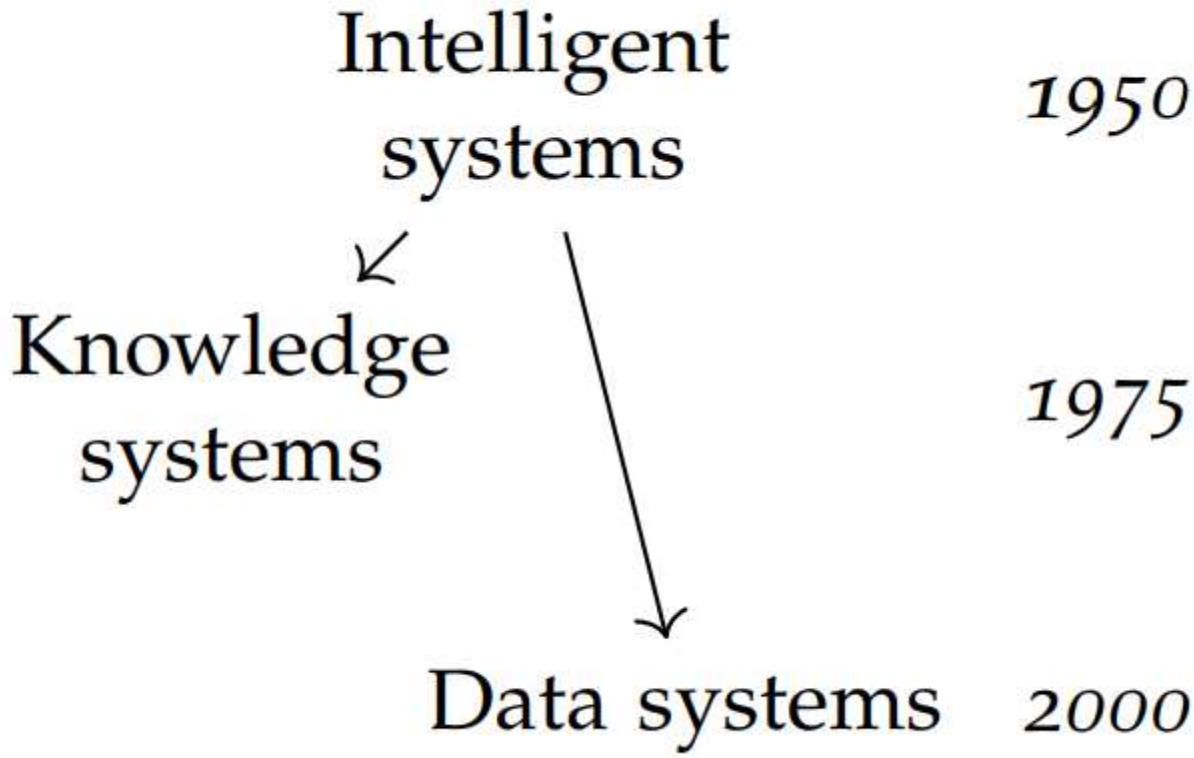
Does not exist. There is a natural variant of general intelligence.

Understand books, biking in a busy street

Superior artificial intelligence

Does not exist. By definition there is no natural variant.

Speculative: Automatic invention, robot uprise


Intelligent systems
1950
Knowledge systems
1975
Data systems
2000

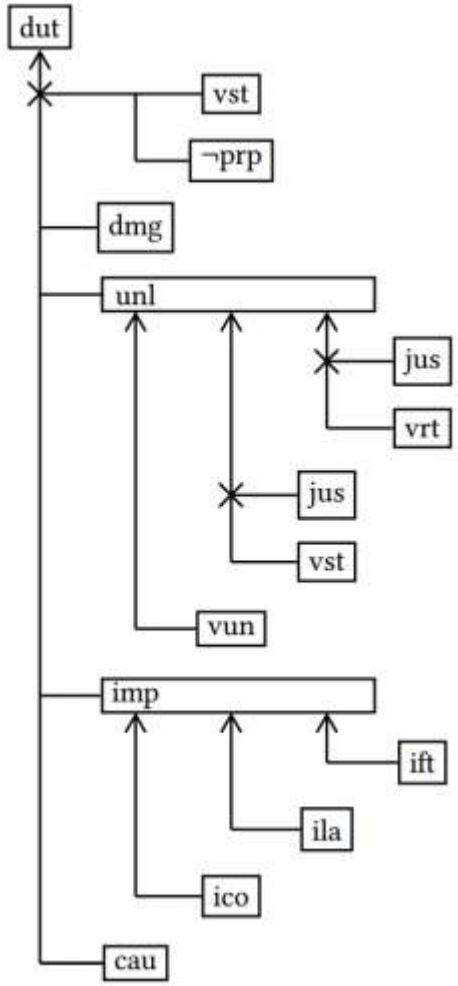
Knowledge systems

Art. 6:162.1 BW (Dutch civil code)

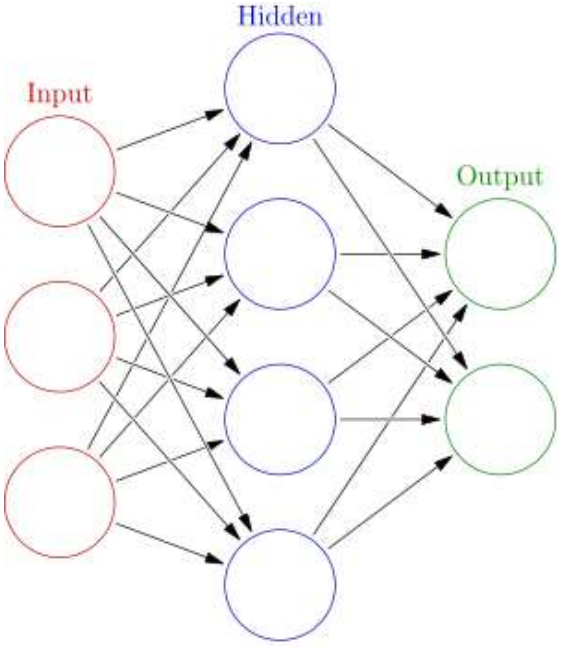
A person who commits an unlawful act toward another which can be imputed to him, must repair the damage which the other person suffers as a consequence thereof.

```
IF damages
   AND unlawful
   AND imputable
   AND causal-connection
   THEN duty-to-repair
```

$$\mathrm{dmg} \wedge \mathrm{unl} \wedge \mathrm{imp} \wedge \mathrm{cau} \rightsquigarrow \mathrm{dut}$$



Data systems



The two faces of Artificial Intelligence

Expert systems	Adaptive systems
Business rules	Machine learning
Open data	Big data
IBM's Deep Blue	IBM's Watson
Complex structure	Adaptive structure
Knowledge tech	**Data tech**
Foundation:	Foundation:
logic	probability theory
Explainability	Scalability

Realizing the **dreams** and
countering the **concerns**
connected to AI require the same innovation:

the development of **argumentation technology**

The law leads the way

Argumentation systems are systems that can conduct a critical discussion in which hypotheses can be constructed, tested and evaluated on the basis of reasonable arguments.

The two faces of Artificial Intelligence

Expert systems
Business rules
Open data
IBM's Deep Blue
Complex structure

Adaptive system
Machine le
Big d
son
aptive structure

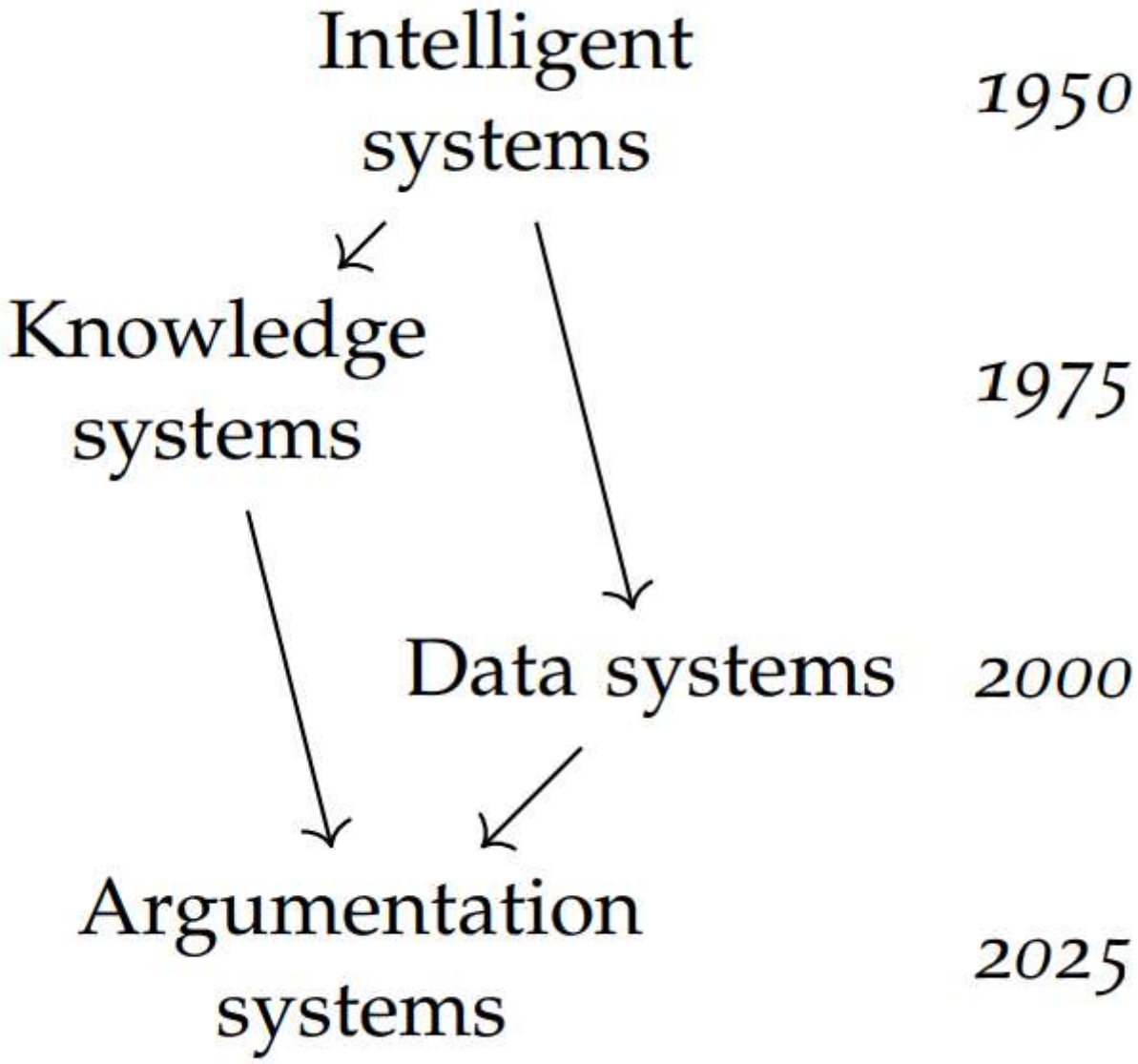
Knowled
Fou

Data tech
Foundation:
probability theory

Explainability

Scalability

Argumentation technology


Intelligent systems
1950
Knowledge systems
1975
Data systems
2000
Argumentation systems
2025

The law can be enhanced by artificial intelligence

Access to justice, efficient justice

The law can be enhanced by artificial intelligence

Access to justice, efficient justice

Artificial intelligence can be enhanced by the law

Ethical AI, explanatory AI

Artificial intelligence and Law

Legal artificial intelligence

Artificial intelligence and Law

ICAIL conferences since 1987 (biennially)

Next edition June 2019 Montreal

`iaail.org`

JURIX conferences since 1988 (annually)

Next edition December 2019 Madrid

`jurix.nl`

Artificial Intelligence and Law journal since 1992

Springer

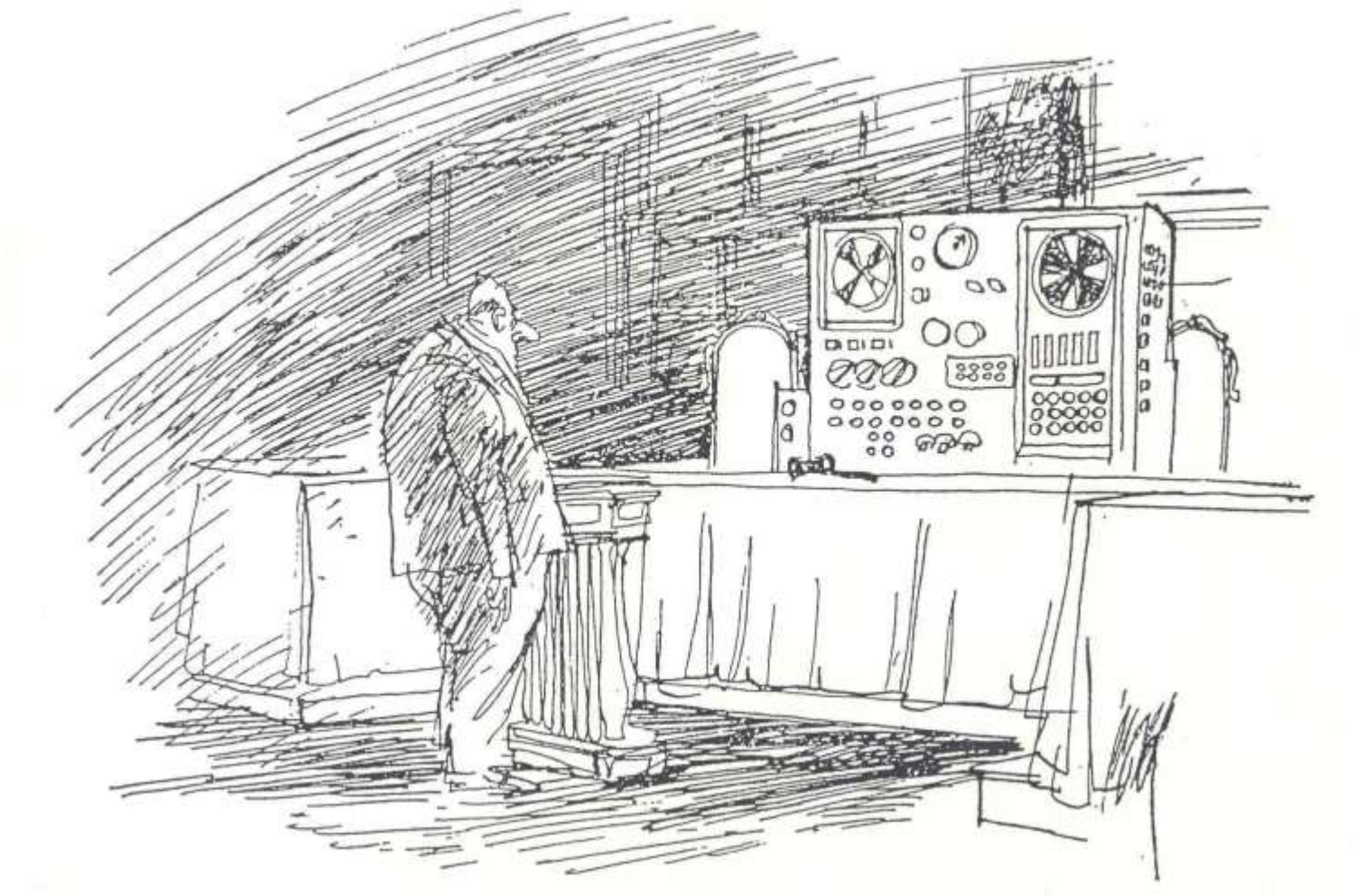
`link.springer.com/journal/10506`

Machines can decide legal cases (?)

Deciding legal cases consists of applying the law.
The law consists of rules and cases.
Machines can apply rules and following cases.

THEREFORE:

Machines can decide legal cases.



Maar edelachtbare, u drinkt toch ook wel eens een glaasje?

But, Your Honour, you sometimes have a drink too, haven’t you?

Some hard questions

Deciding legal cases consists of applying the law.

-> Is applying the law sufficient for deciding cases?

-> How does one apply the law?

The law consists of rules and cases.

-> Does it?

-> Where are they?

Machines can apply rules and follow cases.

-> Can they?

THEREFORE:

Machines can decide legal cases.

-> Well, I don’t know!

AI & Law

Working hypothesis:

Deciding legal cases can be automated.

Research agenda:

Find out how!

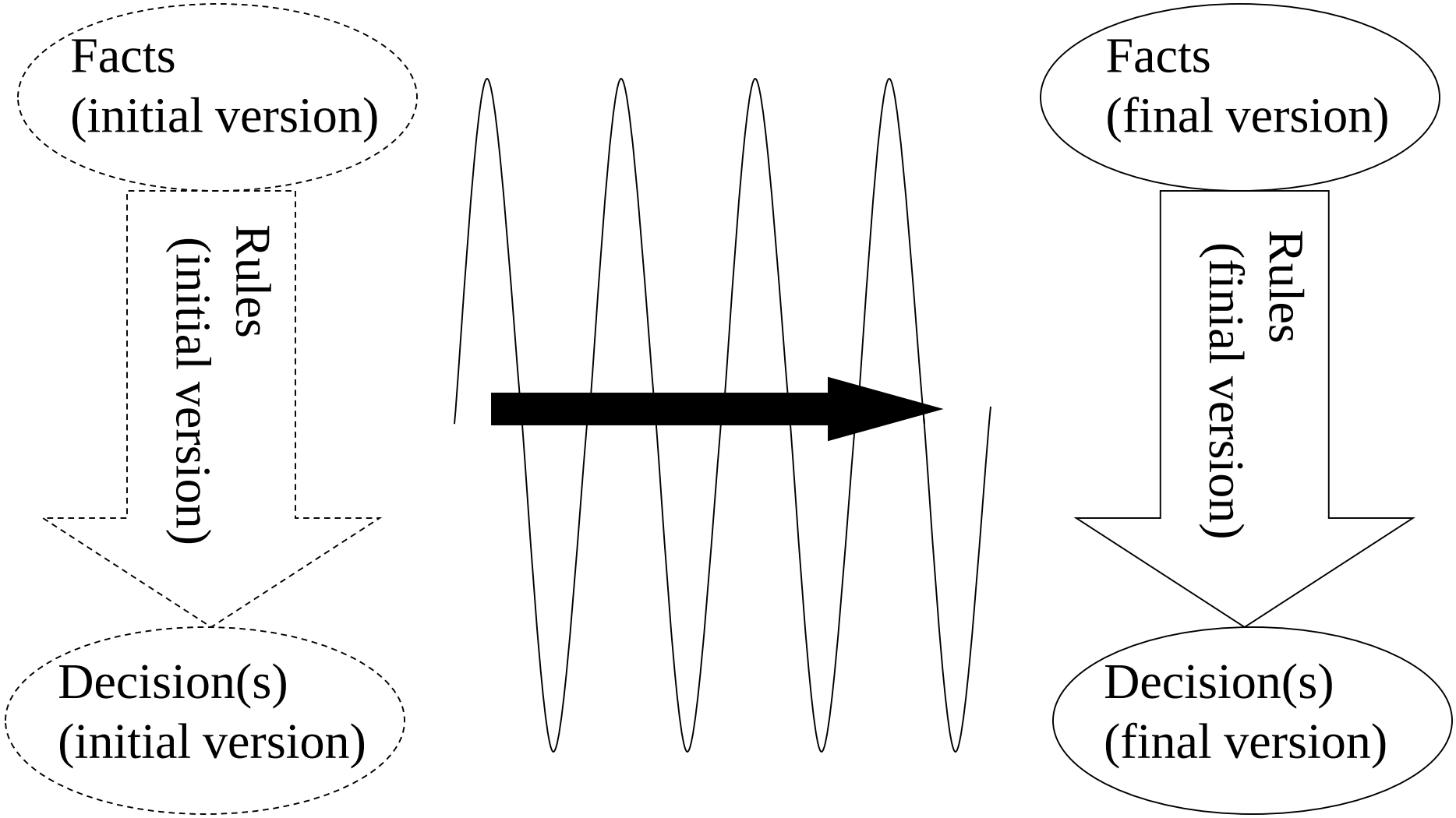
Law and artificial intelligence

The tension in the law between *legal security* on the one hand and *justice* on the other is related to the *gof-ai* vs. *new-ai* dichotomy.

The former are *top-down* and focus on *explicit knowledge* (rules, logic), the latter are *bottom-up* and use *implicit knowledge* (discretion, case analogy, learning, self-organisation).

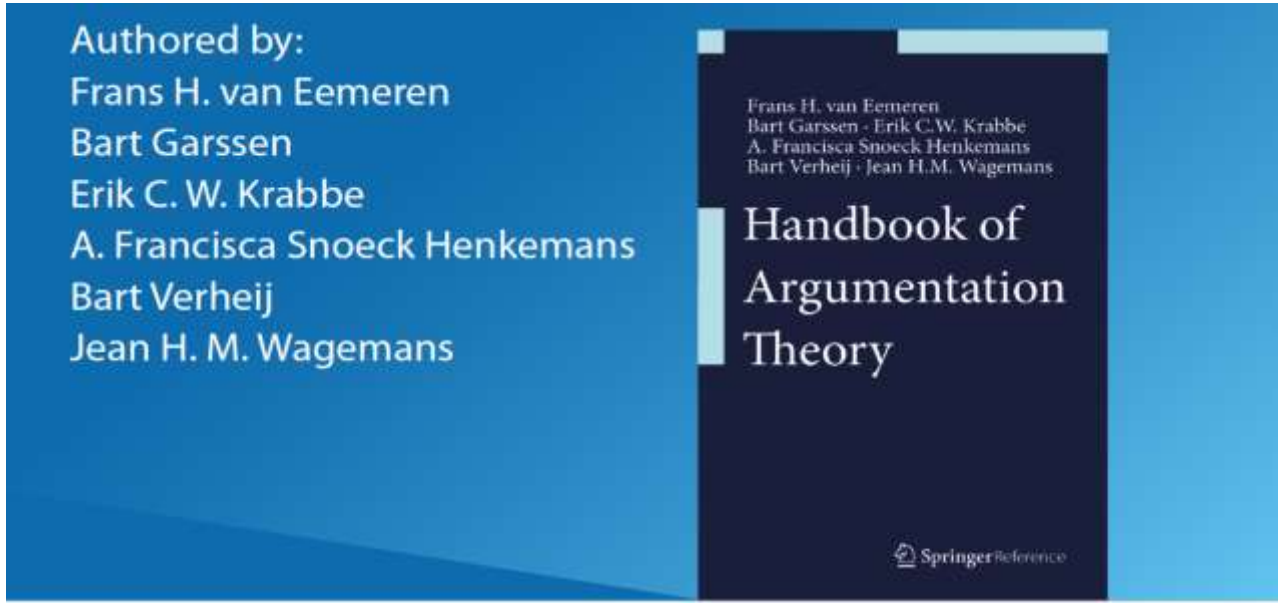
The law has a long history of struggling with this tension and developed pragmatic approaches.

Theory construction



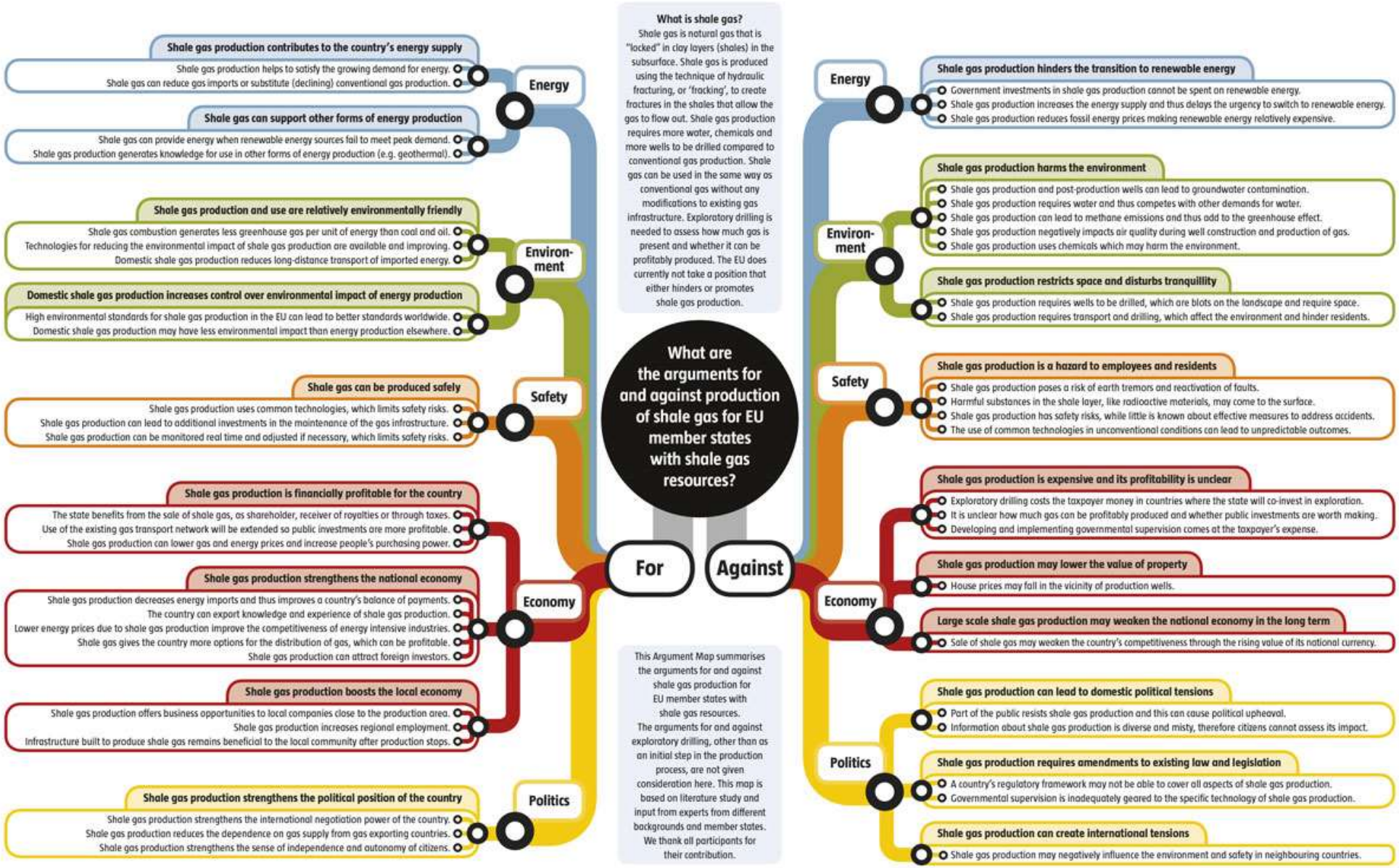
Argumentation

Argumentation
is an interactive social process aimed at the balancing of different positions and interests.



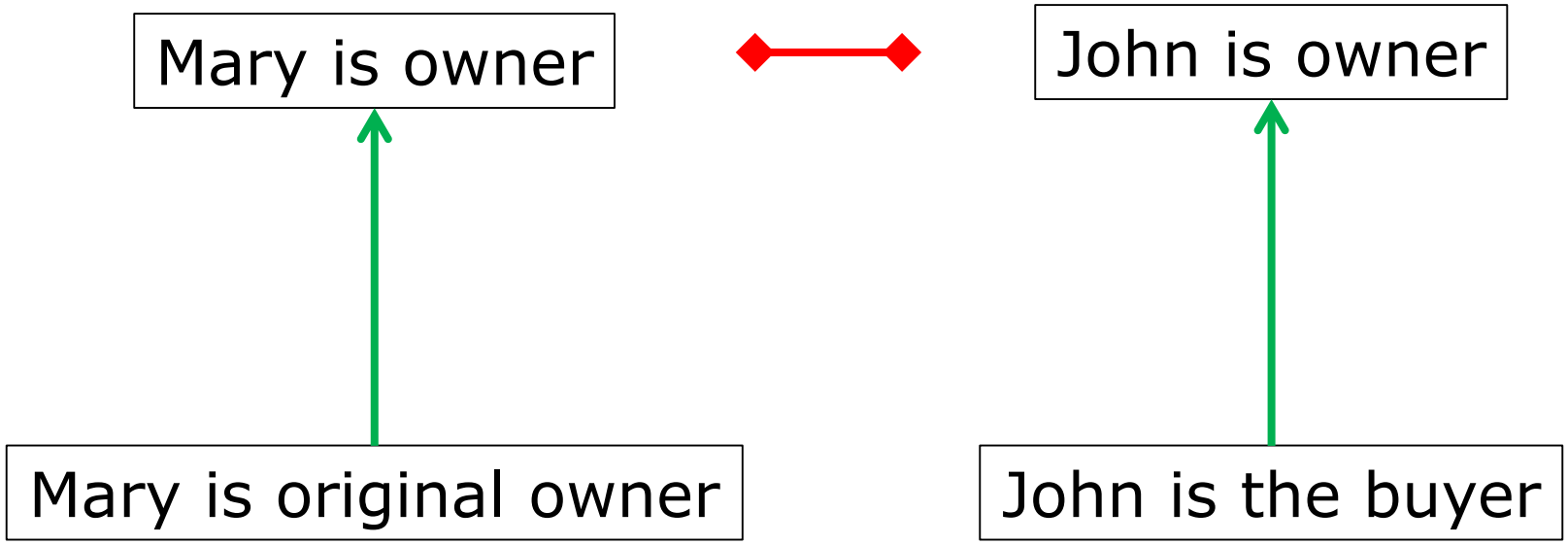
Chapter 11: Argumentation and Artificial Intelligence

ARGUMENT MAP SHALE GAS PRODUCTION IN EU MEMBER STATES



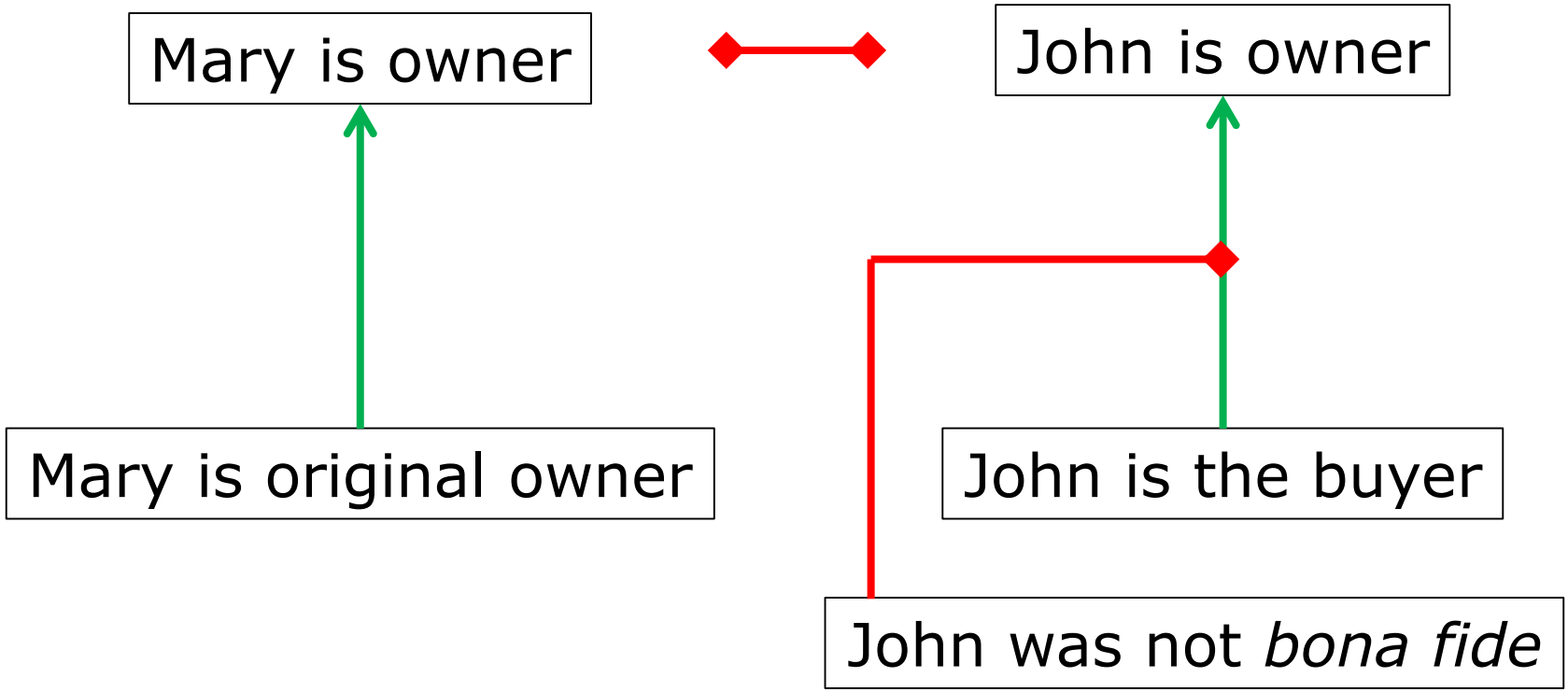
commissioned by TNO innovation for life

made by The Argumentation Factory Argument Map 2012


Mary is owner
John is owner
Mary is original owner
John is the buyer

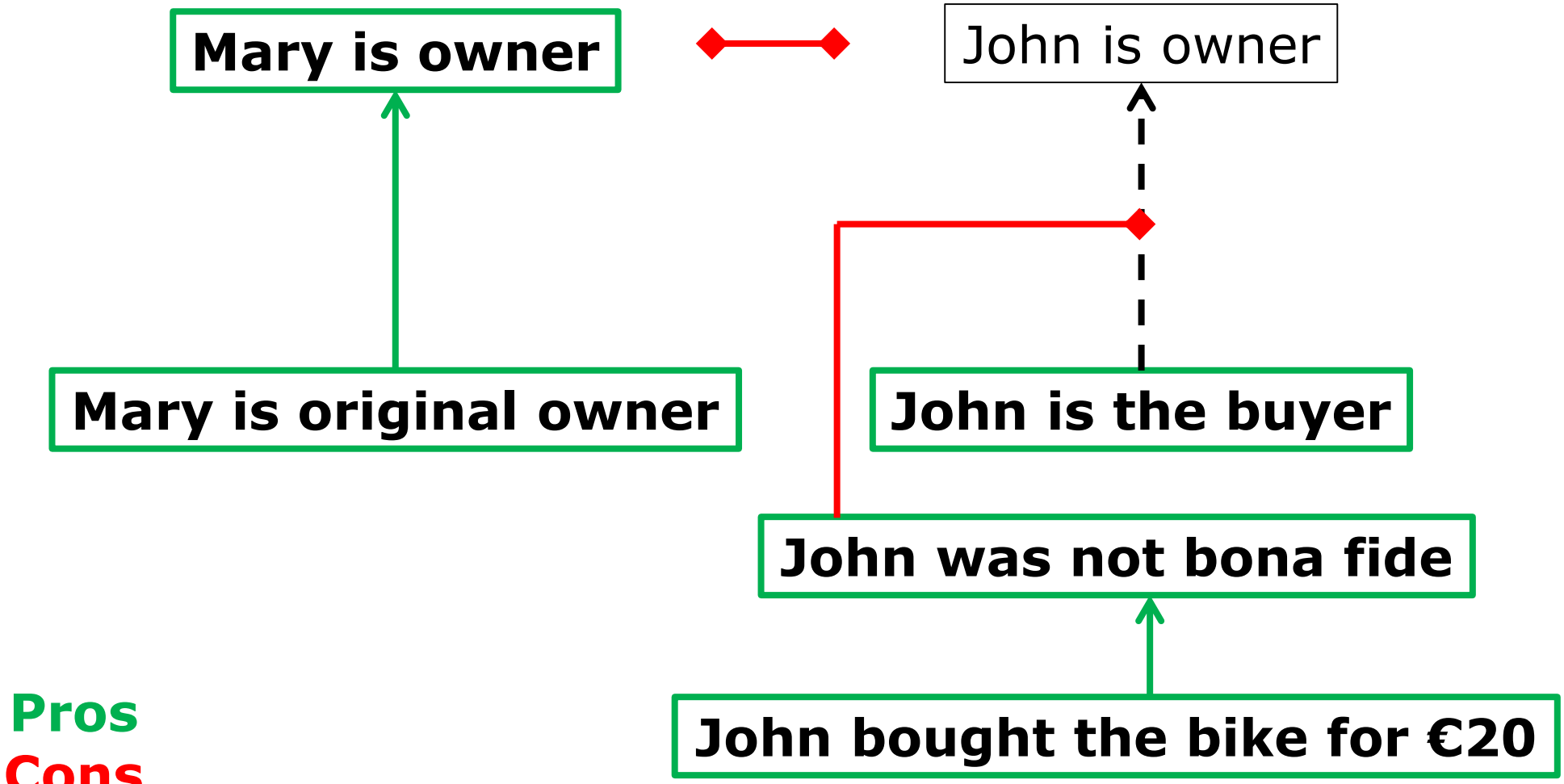
Pros

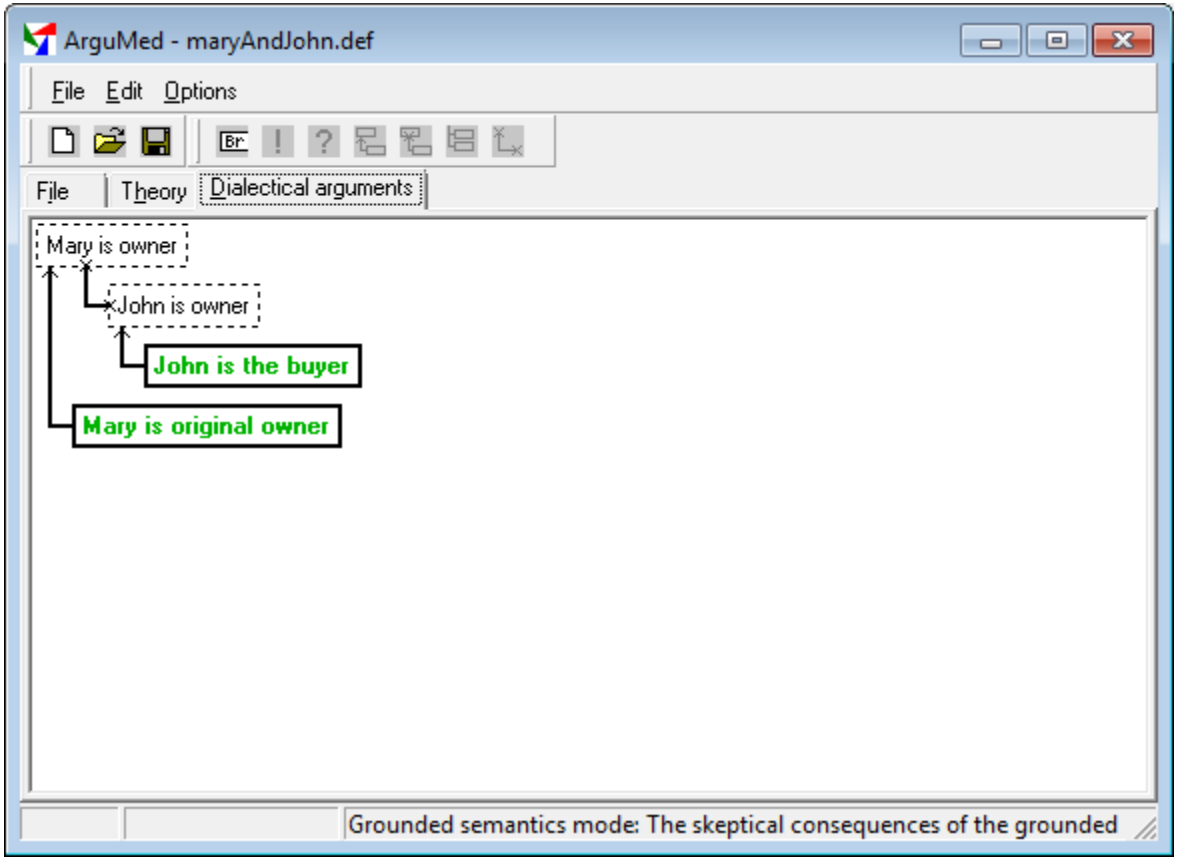
Cons


Mary is owner
John is owner
Mary is original owner
John is the buyer
John was not *bona fide*

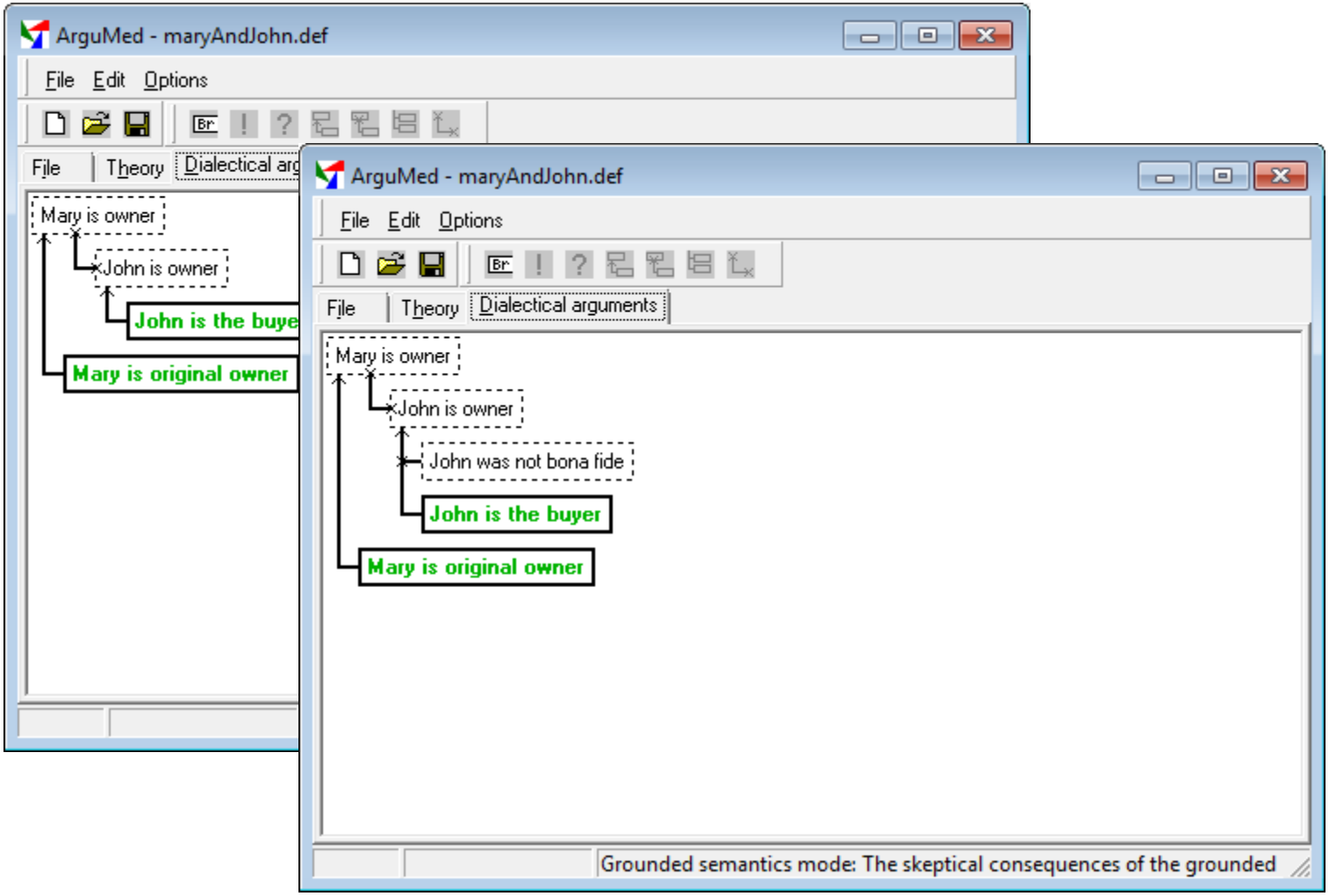
Pros

Cons

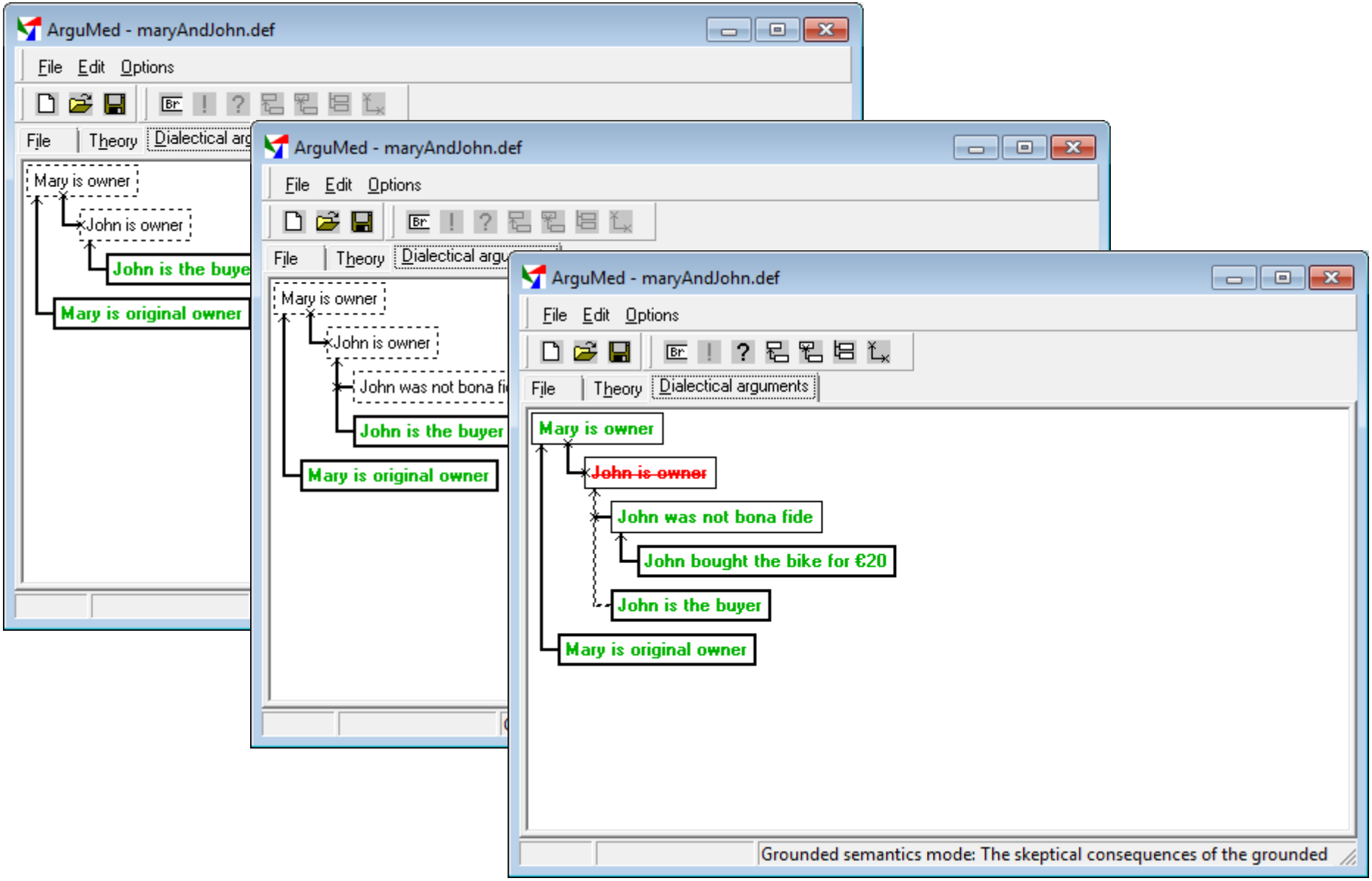

Mary is owner
John is owner
Mary is original owner
John is the buyer
John was not bona fide
John bought the bike for €20
Pros
Cons



Verheij, B. (2005). *Virtual Arguments. On the Design of Argument Assistants for Lawyers and Other Arguers.* T.M.C. Asser Press, The Hague.



Verheij, B. (2005). *Virtual Arguments. On the Design of Argument Assistants for Lawyers and Other Arguers.* T.M.C. Asser Press, The Hague.



Verheij, B. (2005). *Virtual Arguments. On the Design of Argument Assistants for Lawyers and Other Arguers.* T.M.C. Asser Press, The Hague.

现在我们着手目前案例中最后的信息片段，其仍未被整合进该论证中：第二个先例更切中要点，并且被一个更具体的规则所解释。[7]该规则解释了先例2，即伴有并发症的数处肋骨损伤构成严重身体伤害，所产生的作用是使先例1的规则（即数处肋骨损伤不构成严重身体伤害）为非击败的。先例2的规则之所以能做到这点是因为它更加具体。在图4.13中表示了此结论。最后，通过理由
69 “被告人对被害人造成严重身体伤害”证成结论“被告人将被处以最高8年的监禁”。

图4.13　对底切命题展开攻击

图4.14表示了基于先例推理的一个转化。它阐明先例2比先例1更切中要点。该论证可通过证成为何出现这种情形而得到进一步延伸：理由可能是，与先例1相比先例2与当前案例有更多共同因素，因为先例2涉及伴有并发症的肋骨损伤情景，而并发症是一个相关因素。

〔7〕 更多关于形式建模法律领域中基于案例推理的细节，读者可参考如阿什利（Ashley，1990）和罗斯（Roth，2003）的工作。另参见第3章的注释6。

图4.16　编辑论辩材料

71 新近添加了一个工具栏（图4.17）。通过点击其中某个按键可制定论证活动。该工具栏对语境敏感：某些按键允许活动关联于现行命题，仅这些按键能被点击。例如，当现行命题为争议时，则“设为争议”（Set as issue）按键不能被点击，而“设为假定”（Set as assumption）按键是可点击的。有按键用于添加基本命题、设置命题为假定或争议、支持或攻击命题和添加合取支。请注意ArguMed

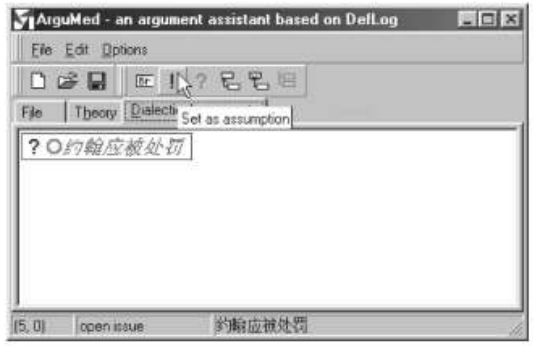
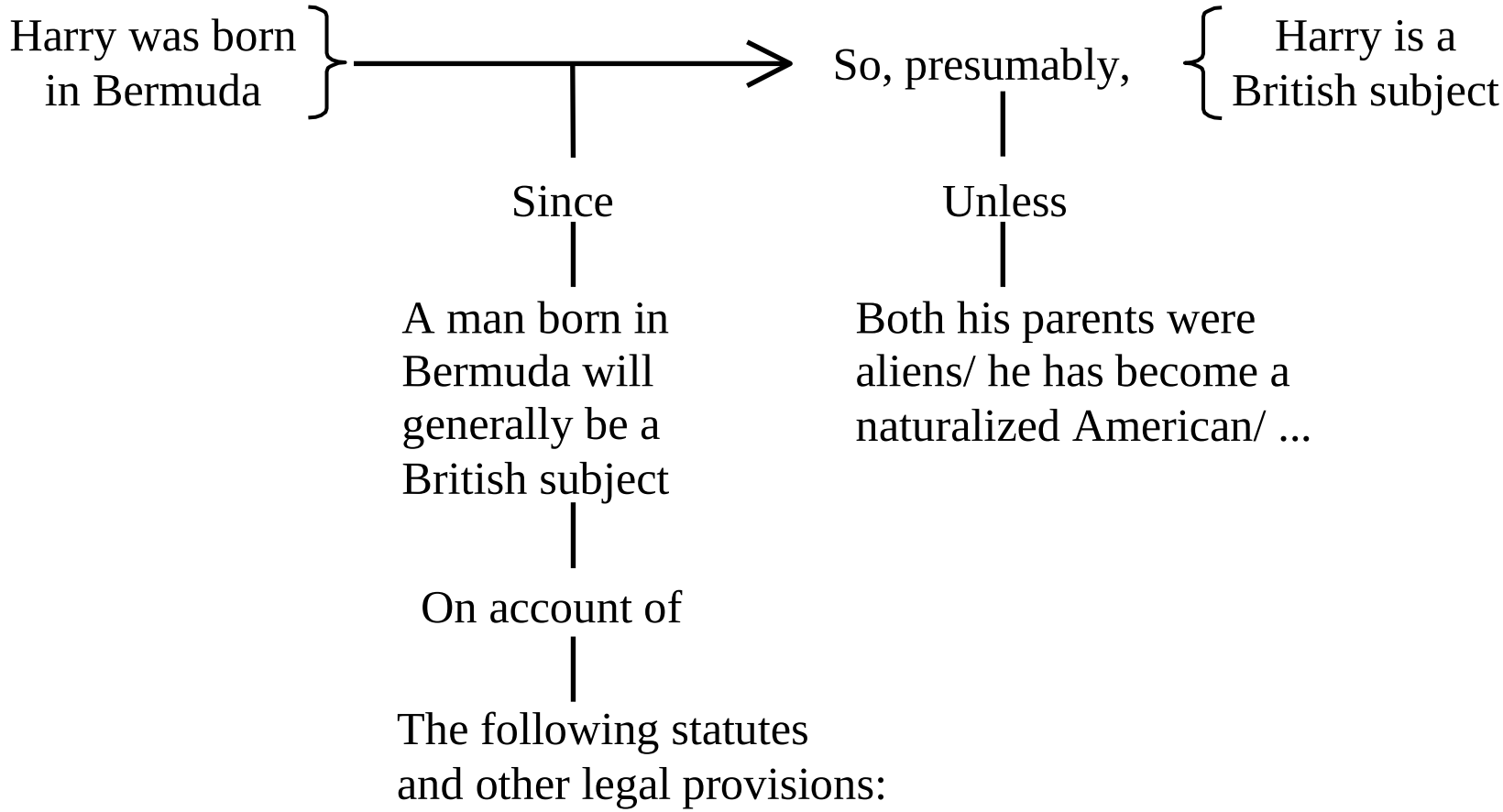


图4.17　使用工具栏

Toulmin's model



Reiter's logic for default reasoning

Birds fly

$$\mathrm{BIRD}(x) : \mathrm{M}\ \mathrm{FLY}(x) \,/\, \mathrm{FLY}(x)$$

A penguin does not fly

$$\mathrm{PENGUIN}(x) \rightarrow \neg \mathrm{FLY}(x)$$

FLY(t) **follows from** BIRD(t)

FLY(t) **does not follow from** BIRD(t), PENGUIN(t)

Defeasible reasoning

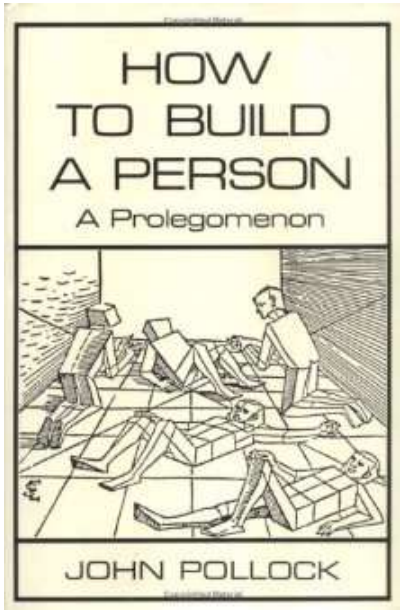
In 1987, John Pollock published the paper ‘Defeasible reasoning’ in the *Cognitive Science* journal.

What in AI is called “non-monotonic reasoning” coincides with the philosophical notion of “defeasible reasoning”.

Pollock on argument defeat

(2.2) P is a *prima facie reason* for S to believe Q if and only if P is a reason for S to believe Q and there is an R such that R is logically consistent with P but (P & R) is not a reason for S to believe Q.

(2.3) R is a *defeater* for P as a prima facie reason for Q if and only if P is a reason for S to believe Q and R is logically consistent with P but (P & R) is not a reason for S to believe Q.

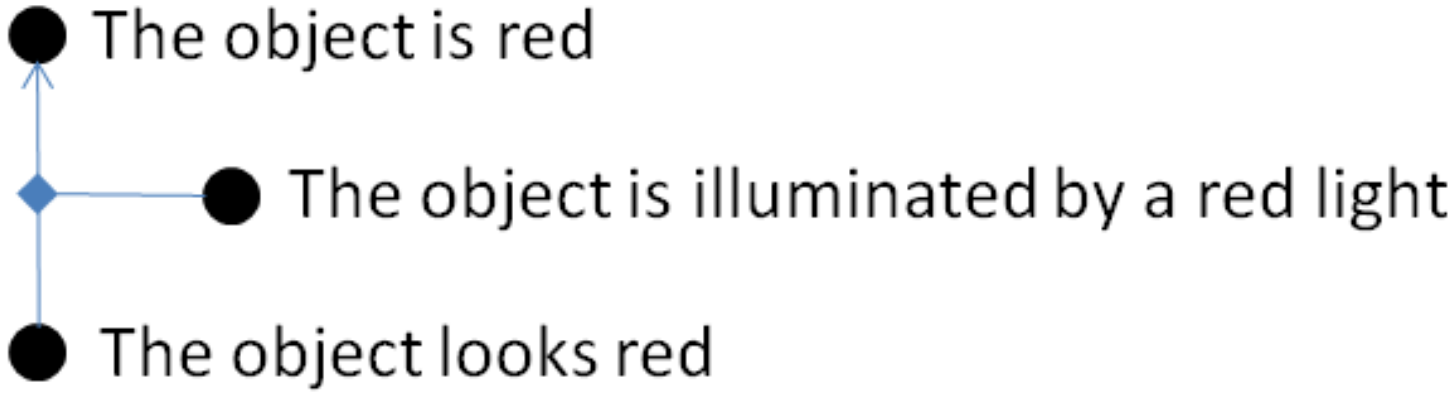


Pollock on argument defeat

(2.4) R is a *rebutting defeater* for P as a prima facie reason for Q if and only if R is a defeater and R is a reason for believing ~Q.

(2.5) R is an *undercutting defeater* for P as a prima facie reason for S to believe Q if and only if R is a defeater and R is a reason for denying that P wouldn't be true unless Q were true.

Pollock's red light example



Undercutting defeat

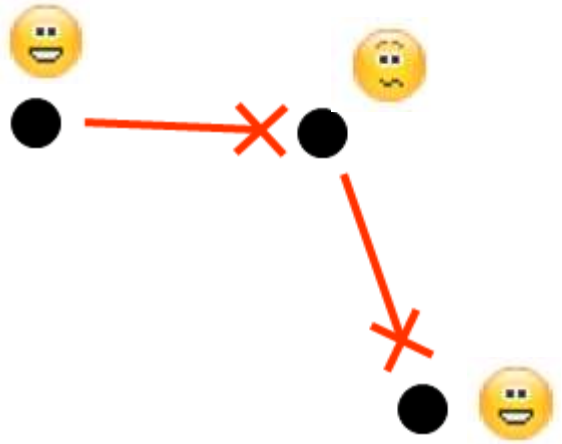
Dung's basic principle of argument acceptability

The one who has the last word laughs best.

Dung's basic principle of argument acceptability

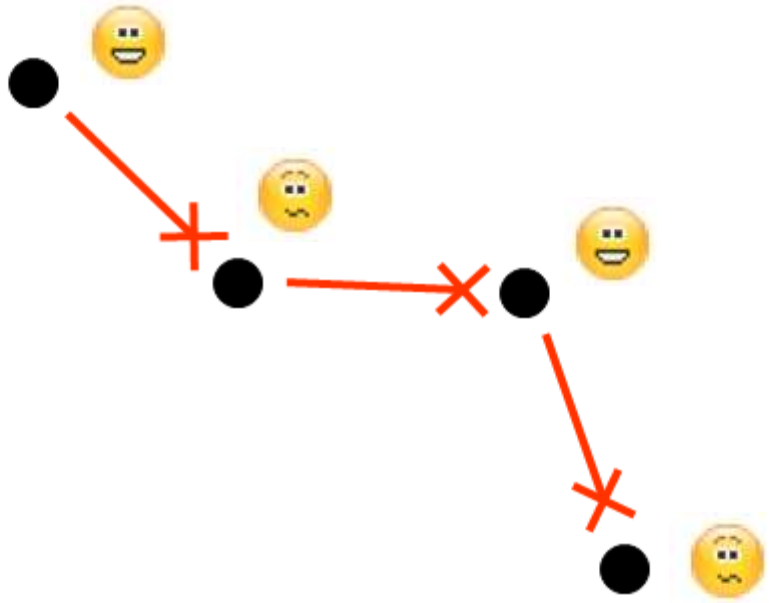
The one who has the last word laughs best.

Dung's basic principle of argument acceptability



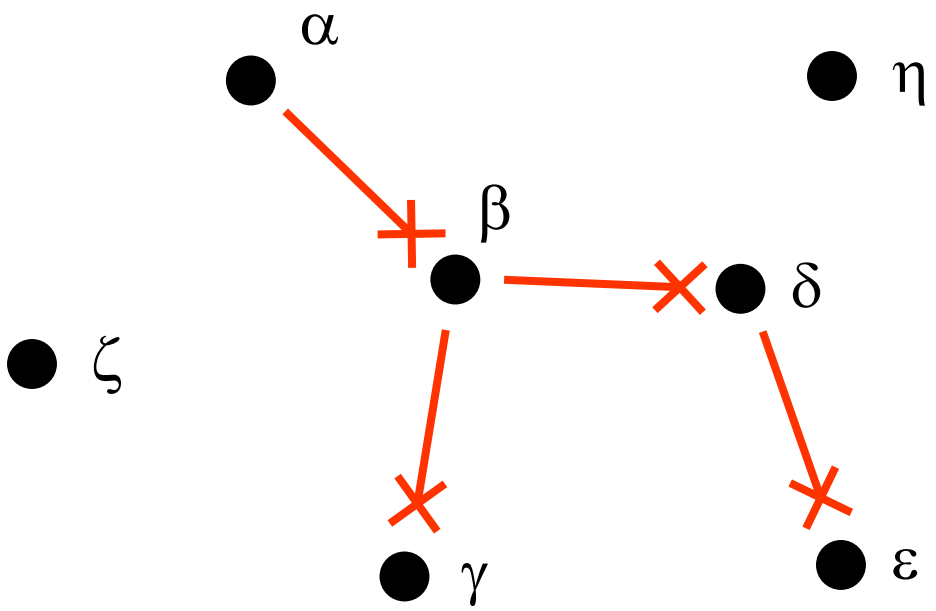
The one who has the last word laughs best.

Dung's basic principle of argument acceptability



The one who has the last word laughs best.

Dung's admissible sets



Admissible, e.g.: {α, γ}, {α, γ, δ, ζ, η}

Not admissible, e.g.: {α, β}, {γ}

Dung's admissible sets

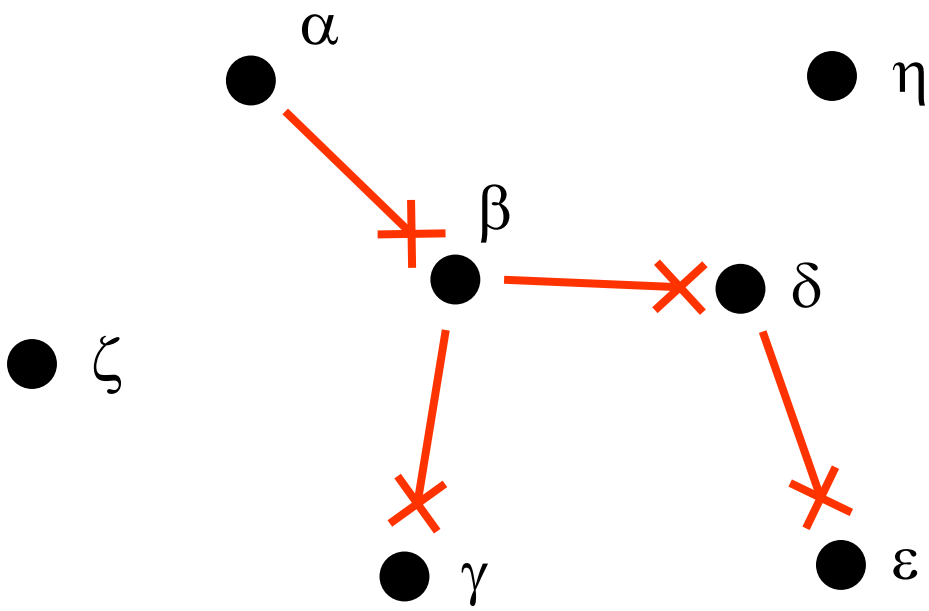
A set of arguments A is admissible if

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

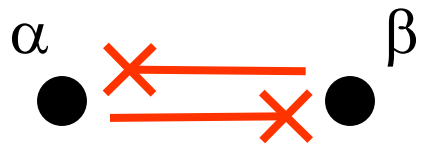
An admissible set of arguments is a *preferred extension* if it is an admissible set that is maximal with respect to set inclusion.

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.



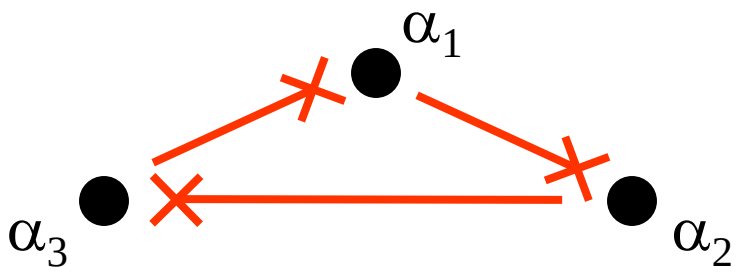
Preferred and stable extension: {α, γ, δ, ζ, η}

Even-length attack cycles



Preferred and stable extensions: $\{\alpha\}$, $\{\beta\}$

Odd-length attack cycles



Preferred extensions: $\varnothing$ (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.

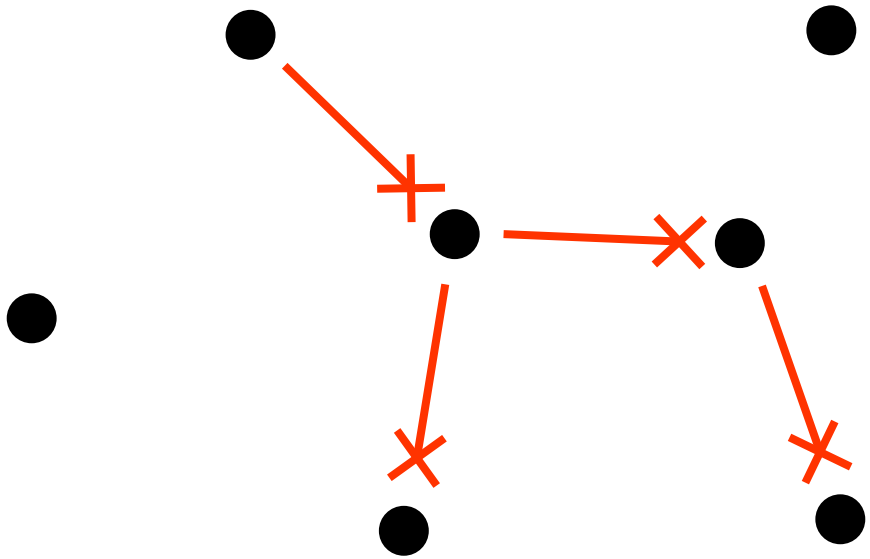
Computing a grounded extension

1. Label all nodes without attackers or with all attackers labeled out as in.
2. Label all nodes with an in attacker as out.
3. Go to 1 if changes were made; else stop.

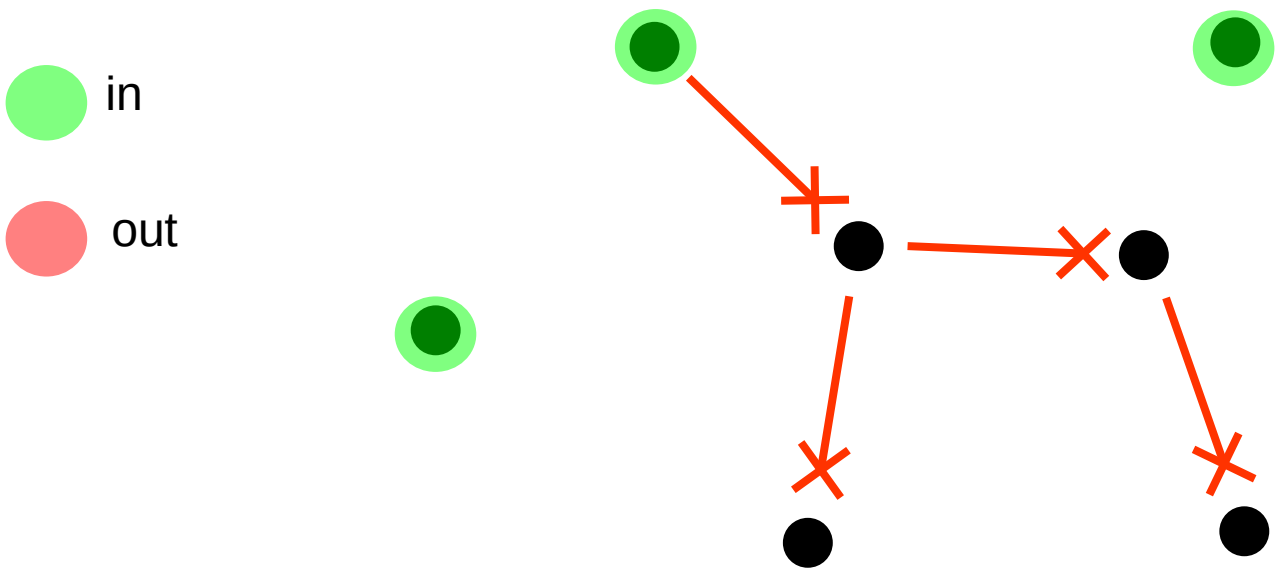
The attack relation as a directed graph (Dung)

in

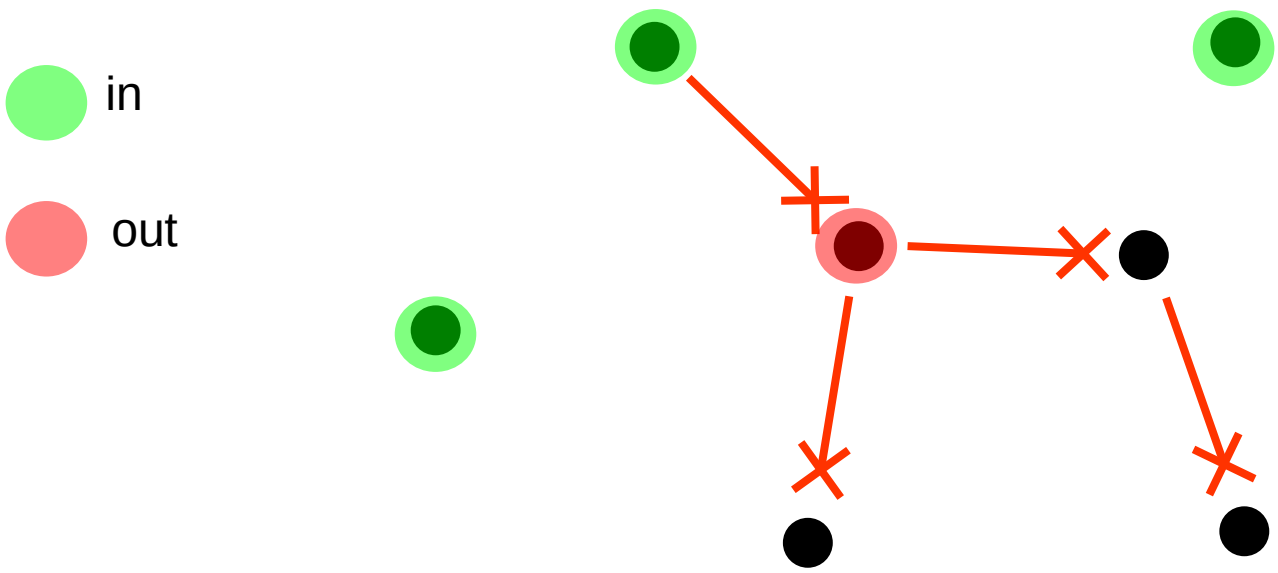
out



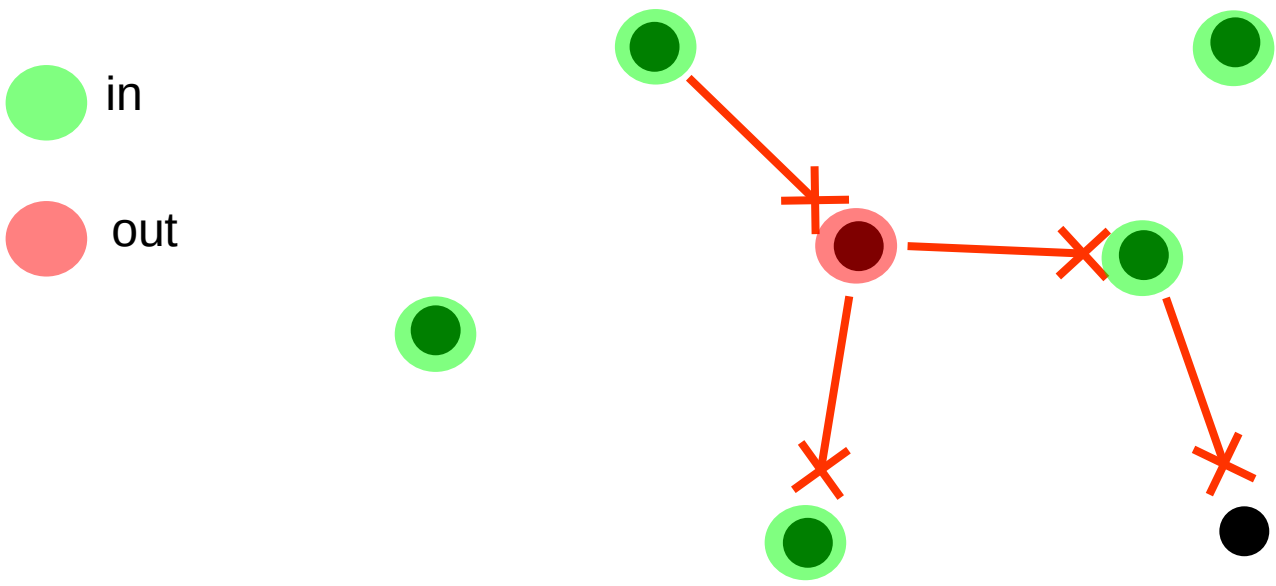
The attack relation as a directed graph (Dung)



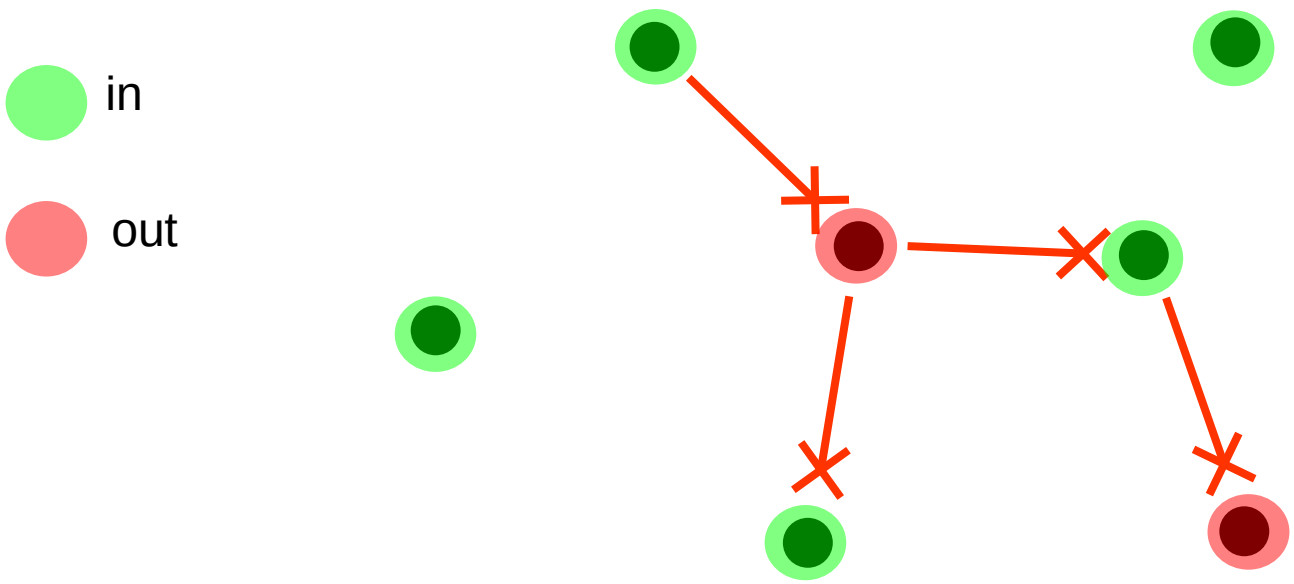
The attack relation as a directed graph (Dung)



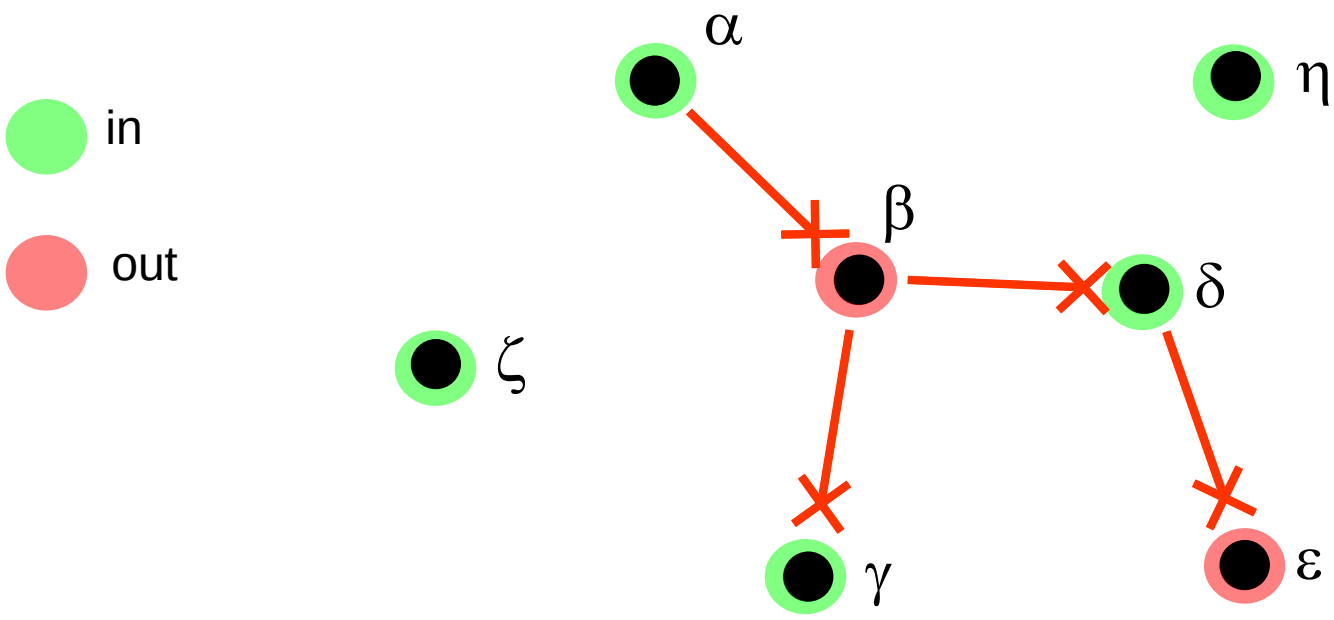
The attack relation as a directed graph (Dung)



The attack relation as a directed graph (Dung)



The attack relation as a directed graph (Dung)



Preferred, stable, grounded extension:

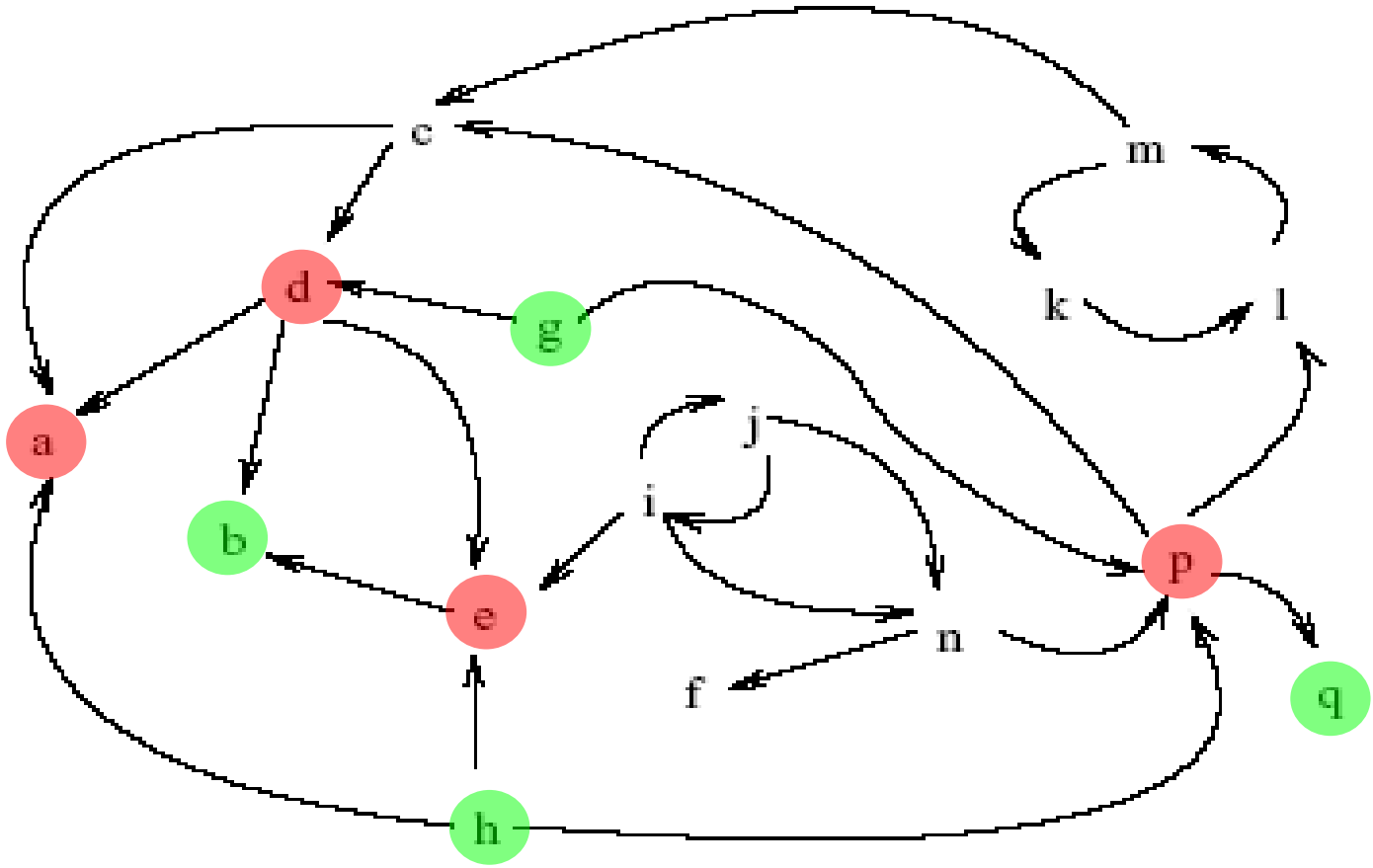
{α, γ, δ, ζ, η}

An Example Abstract Argument System

Note: arrows indicate attack

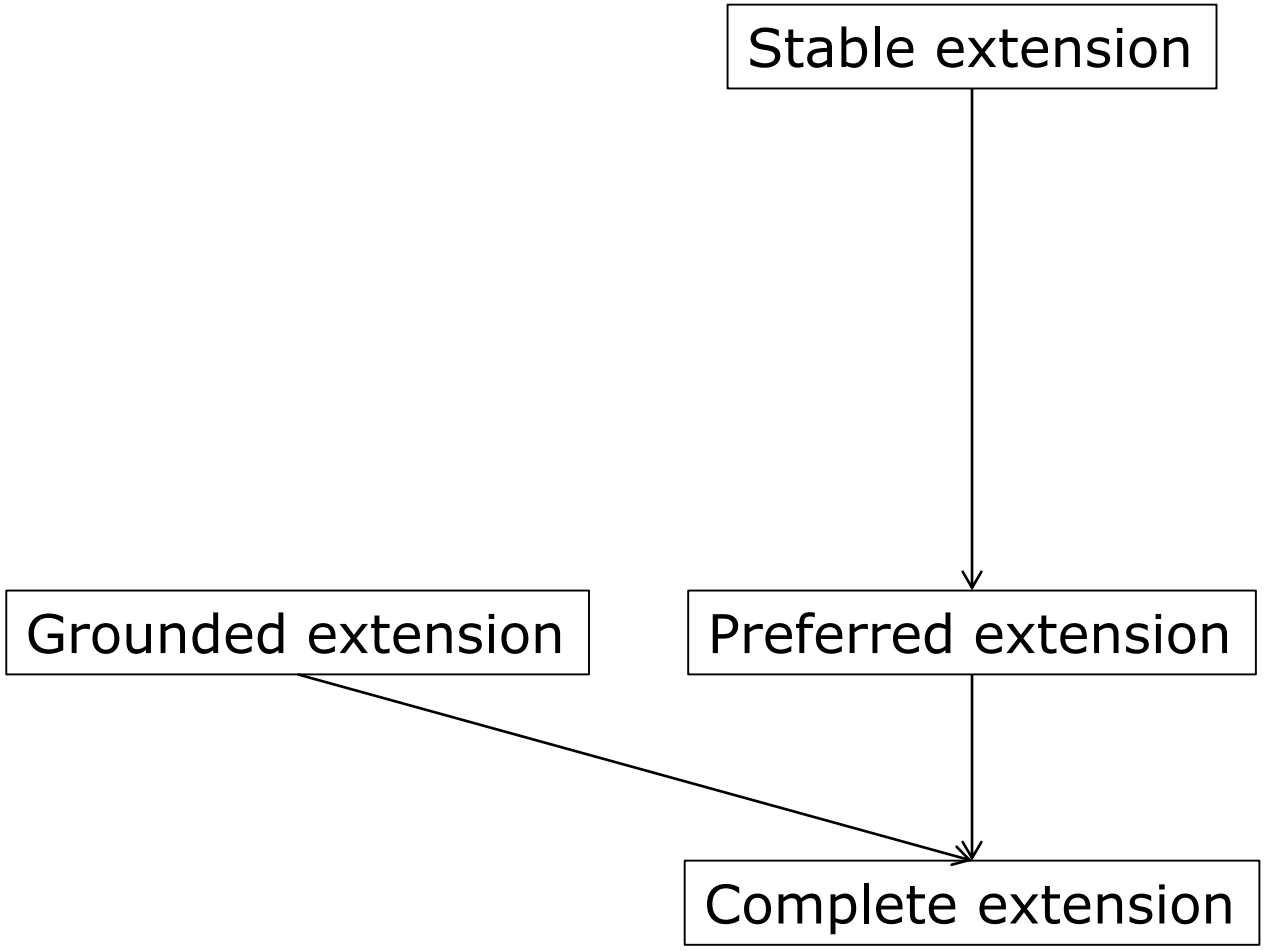
in

out



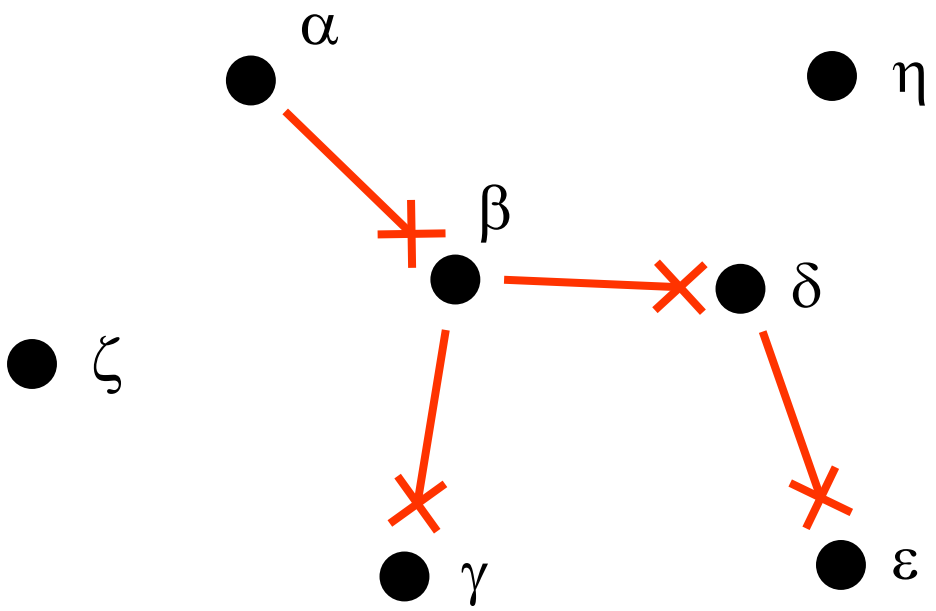
That's it! By the way: there is no stable extension.
(Why? And is there a preferred extension?)

Abstract argumentation semantics (1995)



Dung 1995

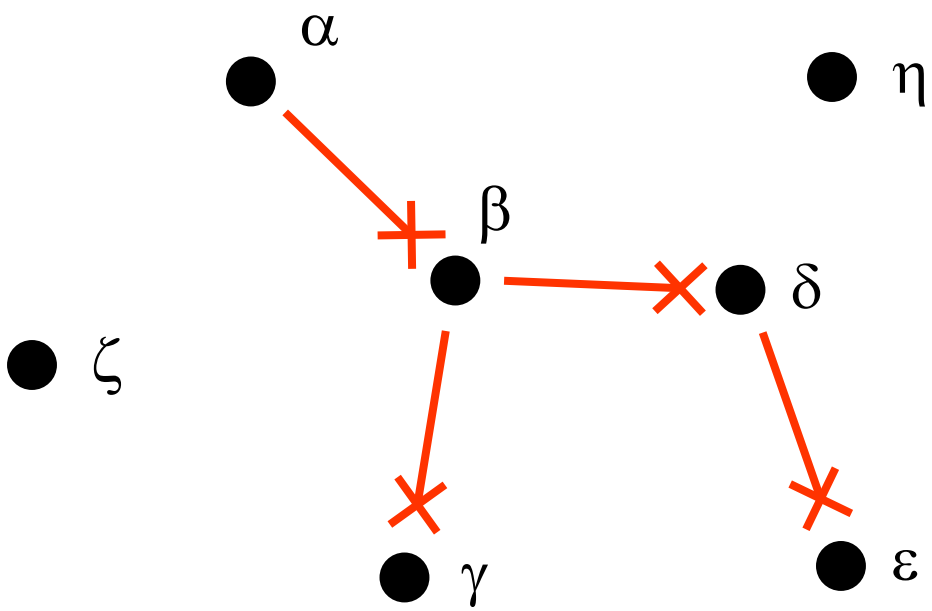
Dung's admissible sets



Admissible, e.g.: {α, γ}, {α, γ, δ, ζ, η}

Not admissible, e.g.: {α, β}, {γ}

Labelings



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. A labeling (*J*, *D*) has *justified defeat* if for all elements *Arg* of *D* there is an element in *J* that attacks *Arg*.
2. A labeling (*J*, *D*) is *closed* if all arguments that are attacked by an argument in *J* are in *D*.
3. A conflict-free labeling (*J*, *D*) is *attack-complete* if all attackers of arguments in *J* are in *D*.
4. A conflict-free labeling (*J*, *D*) is *defense-complete* if all arguments of which all attackers are in *D* are in *J*.

Some properties

Let *J* be a set of arguments and *D* be the set of arguments attacked by the arguments in *J*. Then the following properties obtain:

1. *J* is conflict-free if and only if (*J*, *D*) is a conflict-free labeling.
2. *J* is admissible if and only if (*J*, *D*) is an attack-complete stage.
3. *J* is a complete extension if and only if (*J*, *D*) is a complete stage.
4. *J* is a preferred extension if and only if (*J*, *D*) is an attack-complete stage with maximal set of justified arguments.
5. *J* is a stable extension if and only if (*J*, *D*) is a labeling with no unlabeled arguments.

Remarks on labelings

1. Using labelings can be used to define set-theoretic notions, but also inspire new ones.

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

3. *Some preferred extensions are better than others,* in the sense that they label more nodes.

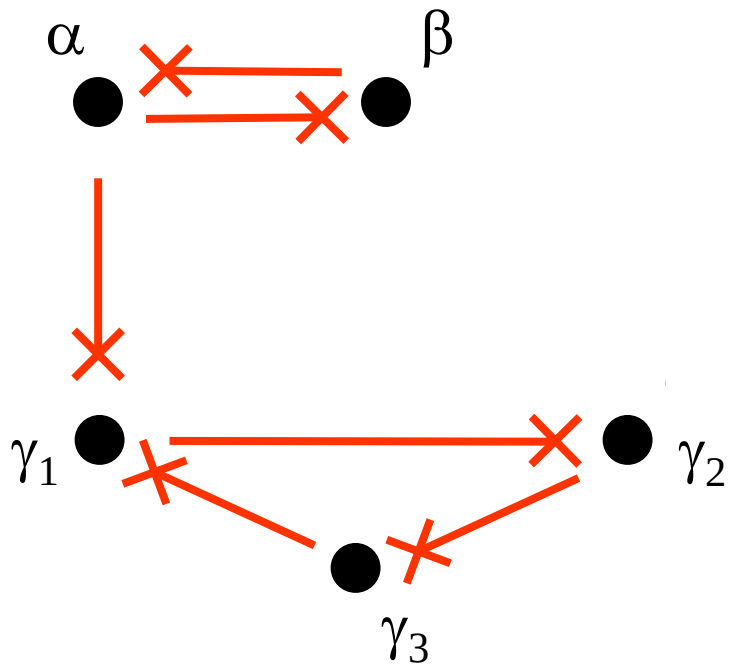
→ Semi-stable extensions

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)

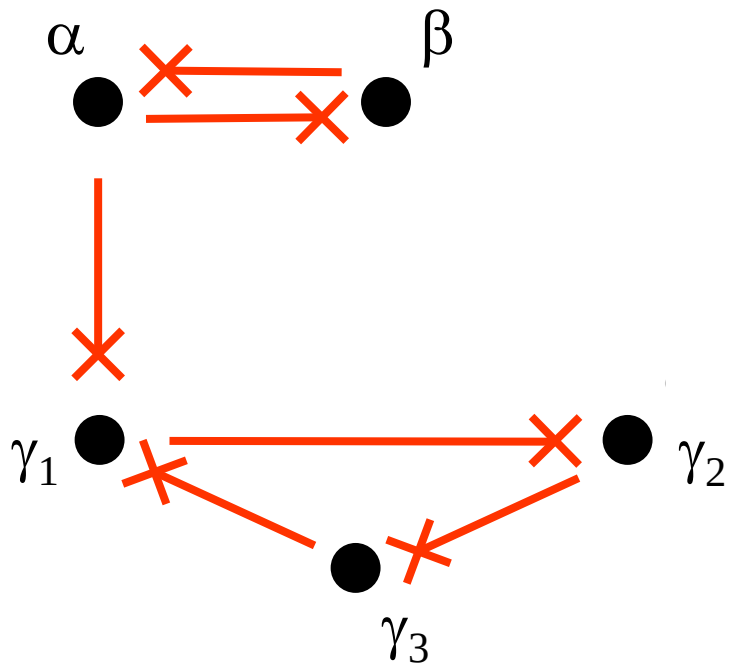
Term coined by Caminada (2006)



Preferred extensions: $\{\alpha, \gamma_2\}, \{\beta\}$

Semi-stable extension: $\{\alpha, \gamma_2\}$

Stable extension: $\{\alpha, \gamma_2\}$



Preferred labelings: $\alpha\ (\beta\ \gamma_1)\ \gamma_2\ (\gamma_3)$, $(\alpha)\ \beta$

Semi-stable labeling: $\alpha\ (\beta\ \gamma_1)\ \gamma_2\ (\gamma_3)$

Stable labeling: $\alpha\ (\beta\ \gamma_1)\ \gamma_2\ (\gamma_3)$

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.

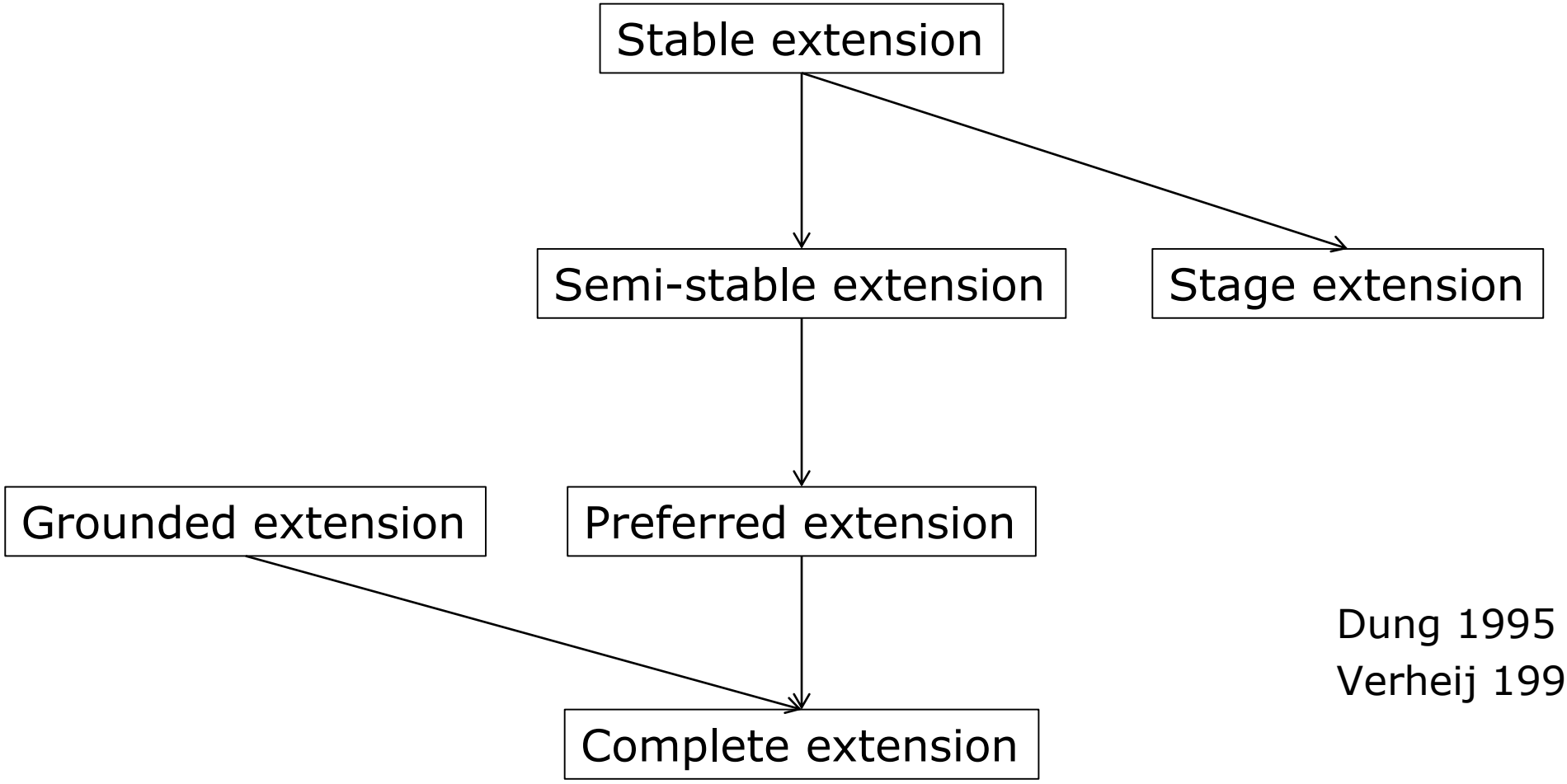
Do all attack graphs have a semi-stable extension?

Answered negatively by Verheij (2000, 2003)

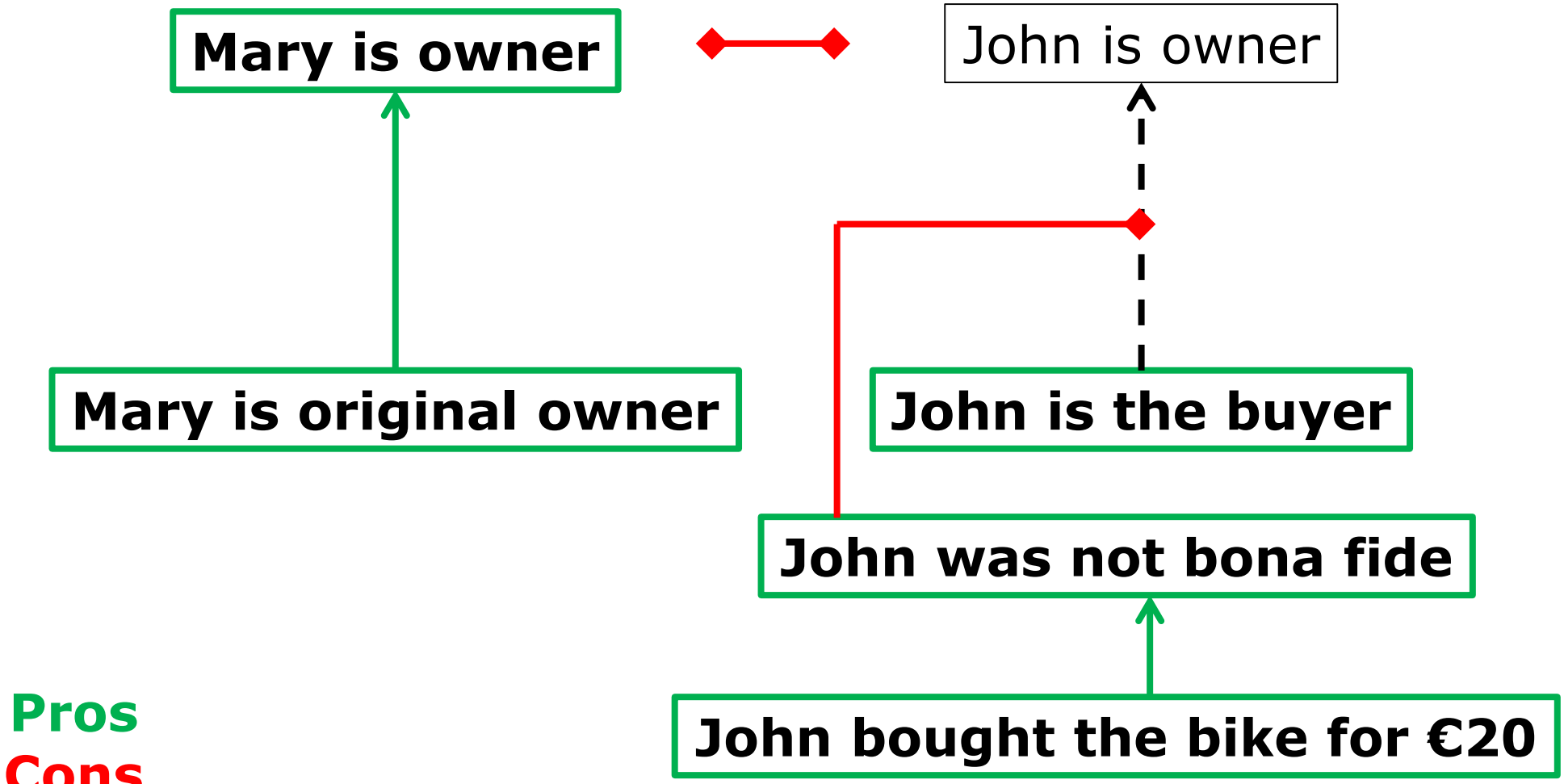
Properties

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.

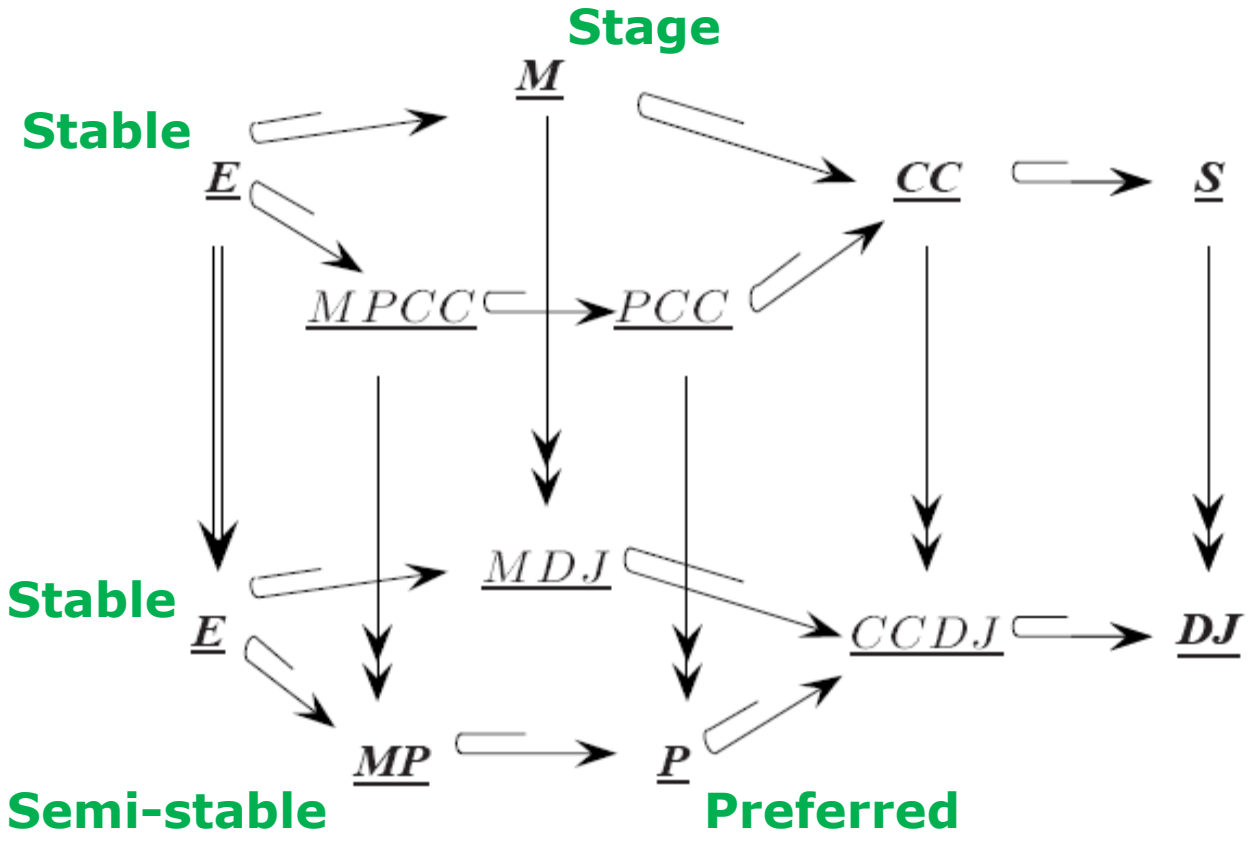
Abstract argumentation semantics (1996)



Dung 1995
Verheij 1996


Mary is owner
John is owner
Mary is original owner
John is the buyer
John was not bona fide
John bought the bike for €20
Pros
Cons

Argumentation semantics (2003)

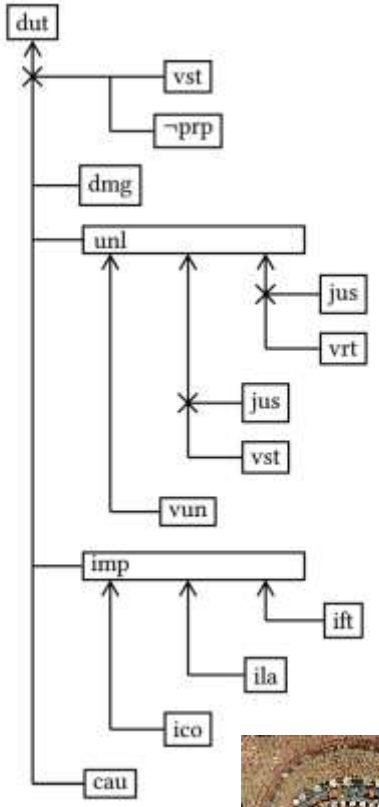


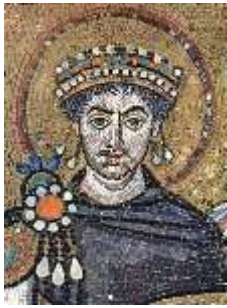
DefLog Verheij 2003

AI & Law

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	¬dut	¬dut	¬dut	dut	dut	dut	dut	dut	dut	dut	dut	dut	¬dut	¬dut	¬dut
¬dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg	dmg
	¬unl	unl	unl	unl	unl	unl	unl	unl	unl	unl	unl	unl	¬unl	¬unl	unl
		¬imp	imp	imp	imp	imp	imp	imp	imp	imp	imp	imp			imp
			¬cau	cau	cau	cau	cau	cau	cau	cau	cau	cau			cau
	¬vrt			vrt	vrt	vrt	¬vrt	¬vrt	¬vrt	¬vrt	¬vrt	¬vrt	vrt	¬vrt	
	¬vst			¬vst	¬vst	¬vst	vst	vst	vst	¬vst	¬vst	¬vst	¬vst	vst	vst
	¬vun			¬vun	¬vun	¬vun	¬vun	¬vun	¬vun	vun	vun	vun			
		¬ift		ift	¬ift	¬ift	ift	¬ift	¬ift	ift	¬ift	¬ift			
		¬ila		¬ila	ila	¬ila	¬ila	ila	¬ila	¬ila	ila	¬ila			
		¬ico		¬ico	¬ico	ico	¬ico	¬ico	ico	¬ico	¬ico	ico			
				¬jus	¬jus	¬jus	¬jus	¬jus	¬jus	¬jus	¬jus	¬jus	jus	jus	
				prp	prp	prp									¬prp

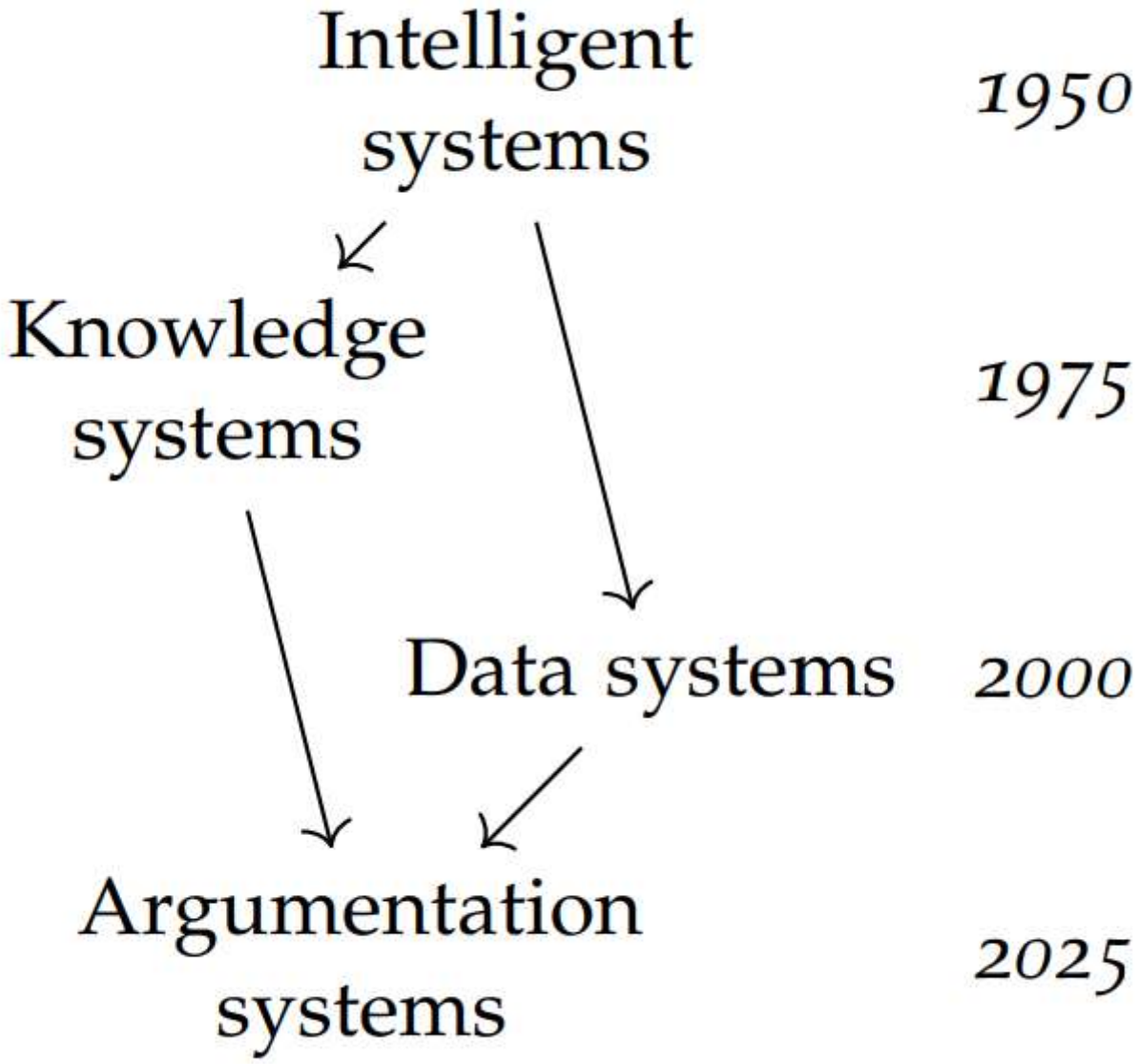
1 > 2 > 3 > 4 > 5 ~ 6 ~ 7 ~ 8 ~ 9 ~ 10 ~ 11 ~ 12 ~ 13 > 14 ~ 15 ~ 16





Data

Knowledge


Intelligent systems
1950
Knowledge systems
1975
Data systems
2000
Argumentation systems
2025

Readings

Verheij, B. (2018). *Arguments for Good Artificial Intelligence.* Groningen: University of Groningen. www.ai.rug.nl/~verheij/oratie/

van Eemeren, F.H., & Verheij, B. (2018). Argumentation Theory in Formal and Computational Perspective. *Handbook of Formal Argumentation* (eds. Baroni, P., Gabbay, D., Giacomin, M., & van der Torre, L.), 3-73. London: College Publications. www.ai.rug.nl/~verheij/publications/hofa2018ch1.htm

Verheij, B. (1996). Two Approaches to Dialectical Argumentation: Admissible Sets and Argumentation Stages. *NAIC'96. Proceedings of the Eighth Dutch Conference on Artificial Intelligence* (eds. Meyer, J.J., & Van der Gaag, L.C.), 357-368. Utrecht: Universiteit Utrecht. www.ai.rug.nl/~verheij/publications/cd96.htm

Verheij, B. (2007). A Labeling Approach to the Computation of Credulous Acceptance in Argumentation. *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007* (ed. Veloso, M.M.), 623-628. www.ai.rug.nl/~verheij/publications/ijcai2007.htm