Defeasible Reasoning & Argumentation

Joint International Doctoral (Ph.D) degree in Law, Science and Technology

Slides: `https://apice.unibo.it/xwiki/bin/view/Courses/LogicPhd22`

Roberta Calegari

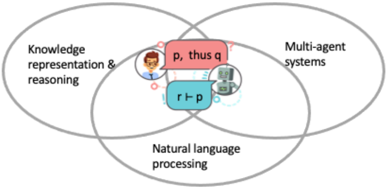
roberta.calegari@unibo.it

Alma Mater Studiorum – Università di Bologna

Academic Year 2022/2023

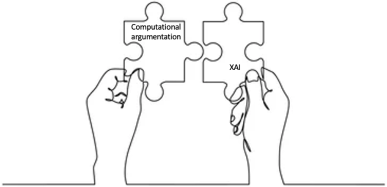
Next in Line...

Context I



Computational argumentation

- supporting the vision of Computable law and XAI
- argumentative explanation



Computational argumentation & defeasibility

Computational argumentation & defeasibility

- enabler for conversational capabilities and explanation
- closer to human thinking
- defeasibility for *defeasible inference* [Loui, 1986] (instead of LP SLD-resolution)
- argumentation to support defeasible logic (and non-monotonic reasoning in general) for *converse* reaching an agreement [Calegari et al., 2021]

Next in Line. . .

Computational argumentation

Computational argumentation

- identifying argumentation framework (AF)
  - abstraction of debate and existing conflicts
- identifying their semantics
  - evaluation of detected debates
- studying properties of this combination
  - goodness of semantics in general and in the case at hand

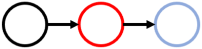
Argumentation Framework *(abstraction of debate)* + Argumentation Semantics *(evaluation of debate)* = Properties *(goodness of semantics)*

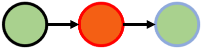
Computational argumentation: abstract argumentation I

Abstract argumentation [Dung, 1995a]

- AF is a graph (abstraction of debate focussing on disagreement)
  - Nodes are arguments
  - Arcs conflicts ⇒ relationship between arguments (a.k.a. attacks)
- semantics
  - allows to draw conclusions ⇒ recipes for detecting winning arguments
- studying properties of this combination
  - depending on the application

Argumentation Framework *(abstraction of debate)* + Argumentation Semantics *(evaluation of debate)* = Properties *(goodness of semantics)*





e.g. existence: yes
e.g. uniqueness: no

Computational argumentation: abstract argumentation II

Automobile Insurance Eligibility

Michigan law requires every vehicle owner to purchase automobile insurance and guarantees that automobile insurance will be available to all eligible Michigan citizens. An eligible person for automobile insurance is a person who is an owner or registrant of a car registered or to be registered in Michigan or has a valid Michigan driver's license based on Michigan law (MCL 500.2103).

There are circumstances when a company can refuse to insure you.

The Michigan no-fault system was adopted in 1973 to increase the level of benefits paid to injured persons, make sure such payments are made promptly, and reduce the proportion of premium dollars paid out for legal and administrative costs.

Reasons for Denial

You can be turned down for automobile insurance if:

- ✓ You are not required by law to have insurance because you do not plan to live or drive in Michigan for 30 or more days.
- ✓ Your driver's license is suspended or revoked.
- ✓ Within the past five years you have been convicted of trying to defraud an insurance company, or have been denied payment of a claim over $1,000 because there is evidence of fraud on your part.
- ✓ Within the past three years you have been found guilty of a felony with a motor vehicle, driving under the influence of alcohol or drugs, failing to stop at the scene of an accident, or reckless driving.
- ✓ The car you want to insure does not meet Michigan safety requirements.
- ✓ Within the past two years your automobile insurance has been cancelled because of non-payment of premium. This can be waived if you pay the entire premium on the policy you are buying in advance.
- ✓ The insurance you want to buy requires you to be a member of a group, club, or organization and you do not join the organization.
- ✓ Your driving record causes you to have more than the allowable number of insurance eligibility points (see the box below).

Insurance Eligibility Points

Insurance companies assign insurance eligibility points for certain traffic violations.

Insurance eligibility points are not the same as points on your official driving record. They are simply a guide to help the company decide whether you are eligible for automobile insurance. You can be turned down for automobile insurance if you have more than six eligibility points from violations within the past three years.

How insurance companies assign eligibility points:

- ✓ Driving more than 15 mph over the speed limit (careless driving) - four points
- ✓ Driving 11-15 mph over the speed limit – three points
- ✓ Driving 15 mph or fewer over the speed limit on freeways that used to have a maximum speed limit of 70 mph - two points
- ✓ Other moving violations - two points
- ✓ The first accident in which you are more than 50% at fault - three points
- ✓ The second and all following accidents in which you are more than 50% at fault – four points

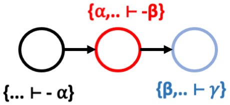
Example: *modelling reasoning with policies*:
Mary, account holder traveling in friend's car that breaks down.
$\rightarrow$ Is she covered by this policy?

Computational argumentation: abstract argumentation III

Abstract argumentation [Dung, 1995a]

- arguments can be (logic) rules

Argumentation Framework *(abstraction of debate)*	+	Argumentation Semantics *(evaluation of debate)*	=	Properties *(goodness of semantics)*



Mary (m): account holder traveling in friend's car (c); car breaks down

rules $\mathcal{R}$:

$$cov(m,c) \leftarrow ah(m), tr(m,c), pr(c), not\ \neg cov(m,c)$$
$$\neg cov(m,c) \leftarrow \neg reg(c,m), not\ cov'(m,c)$$
$$cov'(m,c) \leftarrow in(m,c)$$
$$ah(m) \leftarrow \qquad tr(m,c) \leftarrow \qquad pr(c) \leftarrow$$
$$\neg reg(c,m) \leftarrow \qquad in(m,c) \leftarrow$$

language $\mathcal{L}$: Herbrand base of $\mathcal{R}$ plus (all) NAF literals

assumptions $\mathcal{A}$: (all) NAF literals

contrary $\overline{not\ x} = x$ for all x in the Herbrand base of $\mathcal{R}$

*example from "Explainable AI via Computational Argumentation, F.Toni, 2021"

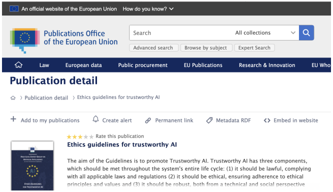
Computational argumentation: Hybrid approaches

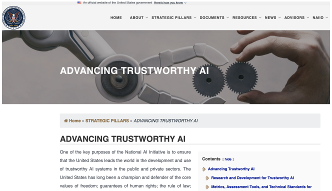
Argumentation for XAI

- ML techniques exploited for extracting arguments and relation (*argument mining* [Lawrence and Reed, 2020])
- formal argumentation for framework definition and model design [Dung, 1995a, Besnard et al., 2014]

Next in Line...

Explainable AI: why?





Explainable AI: standard solutions

Attribution based explanation

- e.g., card denied because client is credit unworthy, despite good salary (LIME [Ribeiro et al., 2016])

Rule based explanation

- e.g., if the client is credit unworthy, then the card is denied (ANCHORS [Ribeiro et al., 2018])

Counterfactual explanation

- e.g., had the client had a good credit score, then the card would have been granted ([Lucic et al., 2022])

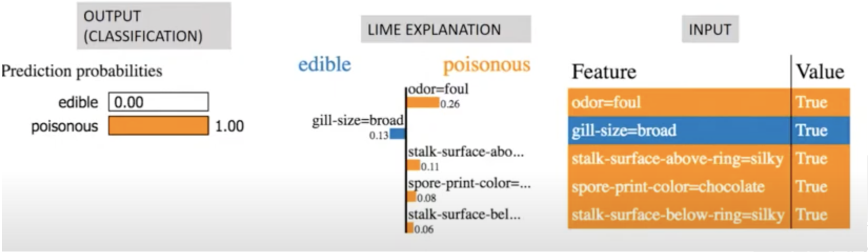
Machine- vs human-oriented XAI

Looking at how humans explain to each other can serve as a useful starting point for explanation in AI [Miller, 2019]

Argumentation based explanation

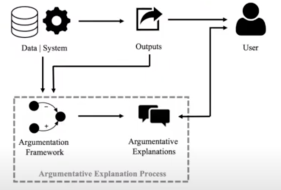
- what humans like, appreciate, buy for explanation?
- not only ML, but also social sciences → *holistic* combined approach
- argumentation identified by social science as a good abstraction for explanation → how people communicate

Argumentation based XAI I



Lime example `https://github.com/marcotcr/lime`

Argumentation based XAI II



System properties, like...

- *strict* dialectical monotonicity
  - adding an attacker → decreases the strength of an argument
  - increasing the strength of an attacker → decreases the strength of an argument
  - adding an supporter → increase the strength of an argument
  - increasing the strength of a supporter → increase the strength of an argument

Review of existing approaches [Čyras et al., 2021]

Next in Line...

Defeasibility

Reasonable results with minimum effort

philosophical logic kind of reasoning rationally compelling, though not deductively valid.

→ a rule is given, but there may be specific exceptions to the rule, or subclasses that are subject to a different rule

defeasibility is found in literatures that are concerned with argument and the process of argument

Defeasible reasoning: fullest expression

- jurisprudence
- ethics and moral philosophy, epistemology, pragmatics
- conversational conventions, constructivist decision theories
- knowledge representation and planning in AI

→ presumptive reasoning (i.e., reasoning on the "face" of evidence)

→ ceteris paribus (default) reasoning

Some schools of philosophy

All reasoning is at most defeasible, and there is no such thing as absolutely certain deductive reasoning, since it is impossible to be absolutely certain of all the facts (and know with certainty that nothing is unknown). Thus all deductive reasoning is in reality contingent and defeasible.

Conclusive vs Defeasible Reasoning I

Conclusive Reasoning Schema: definition

A reasoning schema R is conclusive if one can always adopt R's conclusions while endorsing R's premises (and one should never reject R's conclusions while endorsing R's premises).

- cognitive transition operate regardless of any further information the agent possesses, as long as the reasoner instantiates the pre-conditions in the schema
- *monotonic reasoning* $\rightarrow$ conclusions always grow

Conclusive vs Defeasible Reasoning II

Defeasible reasoning Schema: definition

A reasoning schema R is defeasible if one should, under certain conditions, refrain from adopting its conclusions though endorsing its premises.

- may be defeated by further information to the contrary
- *non-monotonic reasoning* [McCarthy, 1980] $\rightarrow$ conclusions may need to be abandoned when new information is available

Conclusive vs Defeasible Reasoning III

Reasoning schema: Conclusive syllogism

```
(1) all Y's are Z's; AND
(2) x is Y
IS A CONCLUSIVE REASON FOR
(3) x is Z
```

Reasoning schema: Defeasible syllogism

```
(1) all Y's are normally Z's; AND
(2) x is Y
IS A DEFEASIBLE REASON FOR
(3) x is Z
```

Conclusive vs Defeasible Reasoning IV

Reasoning schema: Conclusive syllogism

```
(1) all bachelors are unmarried; AND
(2) John is a bachelor
IS A CONCLUSIVE REASON FOR
(3) John is unmarried
```

Reasoning schema: Defeasible syllogism

```
(1) pet dogs are normally unaggressive; AND
(2) Fido is a pet dog
IS A DEFEASIBLE REASON FOR
(3) Fido is unaggressive
```

Conclusive vs Defeasible Reasoning: adding knowledge I

Adding knowledge

The difference between conclusive and defeasible reasoning emerges most clearly when one acquires beliefs that contradict the conclusions of one's previous inferences.

Conclusive vs Defeasible Reasoning: adding knowledge II

Conclusive Syllogism: Let as assume, for example

- Mary knows that
  - `bachelors are unmarried` and
  - `husbands are married`
- after meeting John at a dinner party comes to believe
  - `John is a bachelor`
  - $\rightarrow$ `John is unmarried` (leads Mary to believe)
- However, the day after a friend tells Mary that
  - `John is Lisa's husband`
  - $\rightarrow$ `John is married` (leads Mary to believe)
- belief contradiction!!! $\rightarrow$ abandon the premises of one of these
  - withdraw `John is Lisa's husband`
  - or withdraw `John is a bachelor`

Conclusive vs Defeasible Reasoning: adding knowledge III

Defeasible Syllogism: Let as assume, for example

- one endorses all of the following propositions
  - `Fido is a pet dog`
  - `pet dogs are normally unaggressive`
  - `Fido is a Doberman`
  - `Dobermans are normally aggressive`
- premises licences both of the defeasible inferences
  - `Fido is unaggressive`
  - `Fido is aggressive`
- assume that the second inference is stronger then the first one
- can maintain the premises of both inferences

→ refrain from deriving the conclusion that
  - `Fido is unaggressive`

Defeasible Reasoning I

Defeasible Reasoning

- not *truth-preserving*
- failure to satisfy truth preservation does not entail logical faultiness
- epistemology has come to identify various kinds of sound defeasible inference [Pollock, 1995]

Defeasible Reasoning II

Perceptual inference

Perceptual contents are a defeasible reason to believe in the existence of corresponding external objects. More generally, having a percept with content ϕ is a defeasible reason to believe ϕ.

For instance, having an image of a red book at the centre of my visual field is a defeasible reason to believe that there is a red book in front of me. This conclusion is defeated if I come to know that there are circumstances which do not ensure the reliability of my perceptions (e.g., I am watching a hologram).

Memory inference

Recalling ϕ is a defeasible reason to believe ϕ .

For instance, my memory that yesterday I had a faculty meeting is a defeasible reason for me to believe that indeed there was such meeting. This inference is defeated if I come to know that the supposed memory was on outcome of my imagination.

Defeasible Reasoning III

Enumerative induction

Observing samples of Φ, all of which have the feature ψ, is a defeasible reason to believe that all $\phi \in \Phi$ have the feature ψ.

For instance, believing that all crows I have ever seen are black is a defeasible reason to believe that all crows are black. This inference is defeated if I perceive one white crow.

Statistical syllogism

Believing that most $\phi \in \Phi$ are $\psi \in \Psi$ and $a \in \Phi$ is a defeasible reason to believe that $a \in \Psi$.

For instance my beliefs that most printed books have even-numbered pages on their left side and that the volume on the top of my table is a printed book are defeasible reason for me to believe that this volume has even-numbered pages on its left side. This inference is defeated if I discover that this volume was wrongly printed with even-numbered pages on its right side.

Defeasible Reasoning IV

Temporal persistence

Believing that ϕ is the case at time t_1 is a defeasible reason to believe that ϕ is still the case at a later time t_2.

For instance, my belief that my computer was on the top of my table yesterday evening (when I last saw it) is a defeasible reason for me to believe that my computer is still there. This inference is defeated if I come to know that the computer was moved from the table after yesterday evening.

Similar defeasible reasoning schemata may be identified also for *practical*, and in particular *legal reasoning*.

Defeasible Reasoning: Legal Reasoning [Sartor, 1995]

- assume that propositions only concern the *majority of cases*, or the normal case
- the exception serves the rule, or at least, does not damage the rule

→ deal appropriately with the anomalous case: not necessary the case to abandon the rule or to change its formulation

→ assume that the operation of the rule is limited on grounds different from the grounds supporting the rule itself

  - exception, in a prevailing principles to the contrary
  - conflicting rules coupled with criteria for solving rule-conflicts (for instance, the traditional principles, according to which a higher, more special or later law norm over a lower, more general or earlier one)

Defeasible Reasoning: Collision and Defeat I

Assume that I believe that swans are normally white, and that I am told that there is a swan in the park

This enables the following defeasible inference (*statistical syllogism*)

```
(1) most swans are white;
(2) the bird in the park is a swan
--------------------------------------------
(3) the bird in the park is white
```

Defeasible Reasoning: Collision and Defeat II

When I look out of the window, I see that the bird in the park, although being unmistakably a swan, looks kind of pinkish

This prompts the following *perceptual inference*

```
(1) I am having a pink image of the bird in the park
-------------------------------------------
(2) the bird in the park is pink
```

Defeasible Reasoning: Collision and Defeat III

- conflicting conclusions, supported by competing defeasible inferences
  - according to the first inference, the bird is white, according to the second one it is pink
- being a moderate empiricist, I consider the perceptual inference to be stronger than the statistical one
  - I should abandon my *pro-tanto* belief that the swan is white and accept (though provisionally) that it is pink
- we say that the inference concluding that the swan is white is defeated by the perceptual inference according to which it is pink

Defeasible Reasoning: Collision and Defeat IV

Cognitive processes like this can be explained by introducing two notions

- collision
- defeat

Collision [Pollock and Cruz, 1999]

Let M be a reason for adopting Q and M^* be a reason for adopting Q^*. We say that that there is a collision between M and M^*, when the combined cognitive state which consists in endorsing both of M and M^* does not support adopting both of Q and Q^*

Defeasible Reasoning: Collision and Defeat V

When one finds oneself in a collision, one is prevented from performing both colliding inferences (just one or both)

- inferences which are prevented by the collision $\rightarrow$ to be defeated
- the reason that prevents deriving a conclusion $\rightarrow$ defeater

Defeat [Pollock and Cruz, 1999]

Let M be a reason for adopting Q. Premise M^* defeats (is a defeater for) M, iff the combined state consisting in endorsing both of M and M^* does not support adopting Q

Rebutting collision I

Rebutting collision

There is a rebutting collision between reasons M and M^* when

- M is a reason to adopt Q
- M^* is a reason to adopt Q^*
- Q is incompatible with Q^*

Rebutting collision: Inference $\mathcal{A}$ collides with inference $\mathcal{B}$

inference $\mathcal{A}$

```
(1) most swans are white;
(2) the bird in the park is
a swan
----------------------
(3) the bird in the park is
white
```

inference $\mathcal{B}$

```
(1) I am having a pink
image of the bird in the park

----------------------
(2) the bird in the park
is pink
```

Undercutting collision I

In many other cases, to assess the incompatibility of propositions one needs to consider further information, concerning meaning connections, causal links, or further facts.

Examples

- to assess that being pink and being white are incompatible states of one same object (our swan) one must know that an object cannot have two colours at the same time
- to know that there is a conflict between the fact that others detain and process one's personal data and one's free self-determination (as affirmed by privacy supporters), much psychological and sociological background-knowledge is to be assumed

Undercutting collision II

- incompatibility of two conclusions cannot be immediately detected by the reasoner
- but is rather the result of finding relevant information and of bringing it to bear through reasoning processes [Perelman and Olbrechts-Tyteca, 1997]

Undercutting collision III

Undercutting collision

There is an undercutting collision between reasons M and M^* when

- M is a reason to adopt Q
- M^* is a reason for believing that M does not support Q

Under these conditions we also say that M^* undercuts M

Undercutting collision IV

Undercutting collision: inference $\mathcal{A}$ collides with inference $\mathcal{B}$

inference $\mathcal{A}$

```
(1) I perceive a pink bird
in the park

----------------------
(2) the bird in the park is
pink
```

inference $\mathcal{B}$

```
(1) There is red light outside

(2) Perceiving a pink object
under red light does not
warrant that it is pink

----------------------
(3) My perceiving a pink bird
does not warrant that it is
pink
```

Preference-Based Reasoning I

When one endorses two colliding reasons:

- If one reason R_1 prevails over the other R_2, we may reject R_2 and endorse R_1 (so that only R_2 is defeated)
- If, on the contrary, none of the two reasons prevails, both reasons are defeated

Preference-Based Reasoning: rebutting collision

A reason R_1 defeats a reason R_2, whenever R_1 collides R_2, and R_1 does not prevail over R_1.
In addition, R_1 prevails over R_2 then R_1 strictly defeats R_2 (defeats R_2, but is not defeated by it).

Preference-Based Reasoning II

Health-care legal issue

Consider a recent case to be addressed by the Italian privacy authority:

- case of a woman who requested for health reasons to access data concerning the DNA of her father (the data was stored in the database of a hospital)
- the father did not give his consent

$\rightarrow$ the authority needed to balance

- privacy-based inference (the father's data could not be provided since sensitive data cannot be communicated without the consent of the data subject) and
- inference based upon the right to health (one has a right to obtain what is needed for one's health: access to the father's DNA).

$\rightarrow$ health-based inference was considered to be preferable to the privacy-based inference

Preference-Based Reasoning III

Preference-Based Reasoning: undercutting collision

In case of undercutting collisions, the undercutter prevails on the contrary: it strictly defeats the attacked reason.

Assume that I find in a law text two rules:

- a rule R_1, and
- a rule R_2 saying that R_1 is inapplicable under certain circumstances

$\rightarrow$ If these circumstances are satisfied in the present case I should conclude that R_1 is indeed inapplicable, and refrain from deriving its conclusion

Preference-Based Reasoning IV

Preference-Based Reasoning: undercutting collision

when facing a collision, we should reason as follows:

- in rebutting collisions, we should compare the strength of the conflicting reasons, and assume that any reason that is not stronger than its competitor is defeated (only stronger reasons prevail)
- in undercutting collisions, we should assume that the undercutter always prevails

Preference-Based Reasoning V

Legal issue

Assume that I have heard two different accounts of the same event, from two friends, John and Mary:

- John tells me A,
- Mary tells me $NOT\ A$ (A is not the case), and
- I consider both of them to be sufficiently reliable, under normal circumstances

I should view the statements of Mary and John as defeating one another, and refrain from forming any belief on the matter, unless I can assume that one of the two conclusions more reliable then the other

$\rightarrow$ In this case, I should provisionally (*defeasibly* or *pro-tanto*) believe the content of the more reliable statement, and reject the other

Preference-Based Reasoning VI

Legal issue

Assume that I come to know both of the following facts:

- John intentionally caused harm to Mary's property, driving into her fence, and,
- he acted in a state of necessity, to avoid hitting a child who was crossing the street

→ conclude that John has no duty to compensate Mary's damage, since the conclusion that he is not liable (having acted for the necessity of saving another's life) prevails over the conclusion that he is liable (having intentionally damaged another's property)

Preference-Based Reasoning VII

- this assumes that one has a way of determining what conclusion (if any) is more strongly supported. In some cases, one may do that by adopting mathematical methods
- one may use *probability* calculus, and assume that the strength of each conclusions corresponds to its probability, which can be computed by combining the probabilities of its pre-conditions
- other numerical calculations of the comparative strength of beliefs have also been proposed

Preference-Based Reasoning VIII

Legal scenario

Let us just observe, in general, that there are certain specific legal issues (in particular in the domain of evidence) where numerical calculi can provide appropriate answers.
However, these calculi do not provide a generally applicable solution for dealing with incompatible conclusions in moral and legal reasoning. In the law, it is rather the case that one needs to engage in priority reasoning, that is, one needs to bring to bear further information, and to decide accordingly which inference is to prevail. Though this information rarely is numerical, it enables in most cases a sufficiently precise assessment.

Reinstatement I

Reinstatement

When a defeater is strictly defeated by a further inference, then the inference originally attacked by the defeat may recover its capacity to establish its conclusion

Reinstatement II

Pink-swan example

- the bird of which I am having a pink vision is a swan, so that I would conclude for its whiteness, unless this conclusion was defeated by my perception of its pinkness
- I realise that the sun is setting, and that its light makes all white things look pink
- consequence of the perceptual inference being undercut, the inference for whiteness is reinstated

$\rightarrow$ I again *pro-tanto* believe that the bird is white

Reinstatement III

Criminal example: defeat by rebutting

- a detective is investigating the violent death of John, of which Mary, John's inconsolable girlfriend, is accused
- the detective believes that Mary loved John, but has evidence that her clothes were stained with John's blood
- him to build two inferences, which rebut one another

inference $\mathcal{A}$

```
(1) Mary loved John

(2) people normally do not
kill their loved ones

-----------------------
(3) Mary has not killed John
```

inference $\mathcal{B}$

```
(1) Mary's clothes were stained
with John's blood

(2) if one's clothes are stained
with the victim's blood, then
normally one has killed the
victim
-----------------------
(3) Mary has killed John
```

Reinstatement IV

- assume that the detective also believes that inference $\mathcal{B}$ is preferable to inference $\mathcal{A}$ (he gives more credit to chemistry then to psychology)

$\rightarrow$ this allows the detective to endorse the conclusion of inference $\mathcal{B}$

$\rightarrow$ at this stage of the inquiry he forms the belief that Mary killed John (inference $\mathcal{B}$ defeats inference $\mathcal{A}$, but inference $\mathcal{A}$ does not defeat $\mathcal{B}$)

Reinstatement V

Criminal example: reinstatement through undercutting

- assume the detective discovers that Lisa (John's previous girlfriend) tried to frame Mary, by staining Mary's dress with John's blood
- this alone would not be a reason to believe that Mary did not kill John, but rather a reason for considering that the inference $\mathcal{B}$ (from Mary's having blood-stained clothes to her being the murderer) is unreliable, and therefore to view this inference as being undercut

→ inference $\mathcal{C}$ by undercutting inference $\mathcal{B}$ results in reinstating inference $\mathcal{A}$

Reinstatement VI

Criminal example: reinstatement through undercutting

inference $\mathcal{A}$	inference $\mathcal{B}$	inference $\mathcal{C}$
(1) Mary loved John (2) people normally do not kill their loved ones ----------------- (3) Mary has not killed John	(1) Mary's clothes were stained with John's blood (2) if one's clothes are stained with the victim's blood, then normally one has killed the victim ----------------- (3) Mary has killed John	(1) Lisa stained John's clothes with blood ----------------- (2) Mary's clothes being stained with John's blood does not warrant that Mary killed John

Reinstatement VII

Presumption: typical undercutting in legal reasoning [Moore, 1987, Sartor, 1995]

- whenever a rule makes a certain legal effect dependent upon the fact that certain facts are not shown to hold
- reasoning with legal presumptions

inference $\mathcal{A}$

```
(1) If I do not believe that
one is guilty, then I
presume that one is innocent


(2) I do not believe that
Mary is guilty

----------------------
(3) Mary is innocent
```

inference $\mathcal{B}$

```
(1) I have evidence proving
that Mary is guilty




----------------------
(2) Mary is guilty
```

Summing up

Basic ingredients of defeasible logic

- strict rules (rules in the classical sense: whenever the premises are indisputable then so is the conclusion)
- defeasible rules
- defeaters
- superiority relation

Next in Line...

Argumentation systems

Argumentation systems: basic elements

- underlying logical language
- definitions of argument
- definitions of conflict
- superiority relation
- status of arguments

Argumentation systems with defeasible logic I

Argumentation systems: basic elements

- underlying logical language $\rightarrow$ defeasible logic
- definitions of argument $\rightarrow$ proof trees (or monotonic derivations)
  - $\rightarrow$ distinction between unrefuted chain of reasoning and a refuted chain
- definitions of conflict
  - $\rightarrow$ An argument A *attacks* a defeasible argument B if a conclusion of A is the complement of a conclusion of B, and that conclusion of B is not part of a strict subargument of B.
  - $\rightarrow$ A defeasible argument A is *supported* by a set of arguments S if every proper subargument of A is in S.
  - $\rightarrow$ A defeasible argument A is *undercut* by a set of arguments S if S supports an argument B attacking a proper non-strict subargument of A. That an argument A is undercut by S means that we can show that some premises of A cannot be proved if we accept the arguments in S.

Argumentation systems with defeasible logic II

Argumentation systems: basic elements

- superiority relation
- status of arguments
  - $\rightarrow$ notion of an acceptable argument: based on this define justified arguments and justified conclusions, conclusions
  - $\rightarrow$ intuitively, an argument A is acceptable w.r.t. a set of arguments S if, once we accept S as valid arguments, we feel compelled to accept A as valid $\rightarrow$ depending on the specific *semantics*

Mapping: formal accounts

Formal accounts

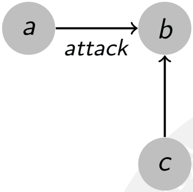
- Argumentation semantics for defeasible logic [Governatori et al., 2004]
- An argumentation-theoretic characterization of defeasible logic [Governatori and Maher, 2000]
- Defeasible logic programming: An argumentative approach [García and Simari, 2004]

Next in Line...

Abstract argumentation [Dung, 1995b]

An argumentation system consists of a couple (A, R), where A is a set of elements (arguments) and R a binary relation representing attack relation between arguments

- represented by a directed graph
- each node represents an argument
- each arc denotes an attack by one argument on another



Acceptability Criteria (defined by specific semantic)
→ analyse the graph to determine which arguments are acceptable according to some general criteria

Justification state of arguments: Dialectical Justification

Arguments accepted under a given semantics → *argument evaluation*
[Baroni and Giacomin, 2009]

Most common approaches

- Extension-based approach: semantics specification concerns generation of a set of extensions (“collective acceptable" arguments)
  - Determine conflict-free sets
  - Determine extensions (naive, admissible, preferred, complete, stable,...)
- Labelling-based approach: semantics specification concerns generation of a set of labellings (e.g. states of an argument)

→ any extension-based can be equivalently expressed in a simple labelling-based, adopting a set of two labels (let say $L = \{\text{in,out}\}$)

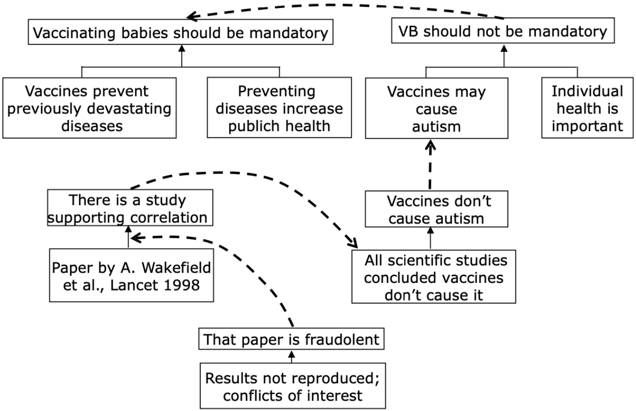
→ on the other hand, an arbitrary labelling can not in general be formulated in terms of extensions

Extension-based approaches

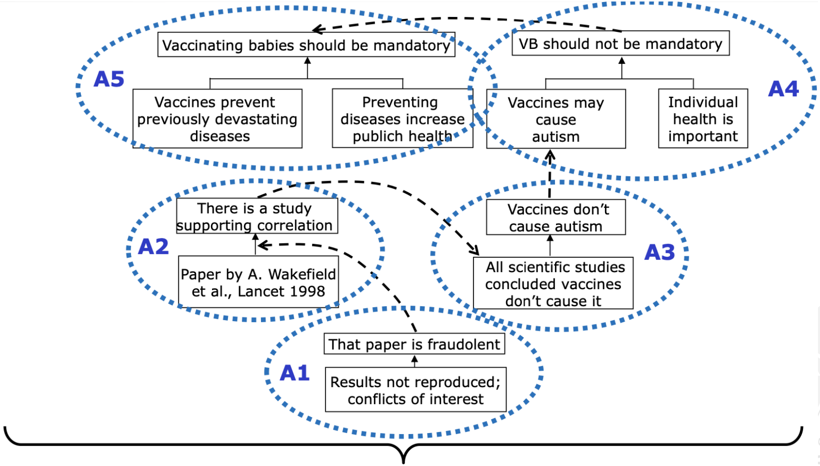
- four “traditional” semantics, considered in Dung’s original paper, namely semantics
  - *complete*: is a set which is able to defend itself and includes all arguments it defends
  - *grounded*: includes those and only those arguments whose defense is “rooted” in initial arguments (aka strong defense [Baroni and Giacomin, 2007])
  - *stable*: attack all arguments not included in it
  - *preferred*: The aggressive requirement that an extension must attack anything outside it may be relaxed by requiring that an extension is as large as possible and able to defend itself from attacks
- subsequent proposals: *stage*, *semi-stable*, *ideal*, *CF2*, and *prudent* semantics.

For a full review see [Baroni and Giacomin, 2009]

Argumentation example: Vaccination I



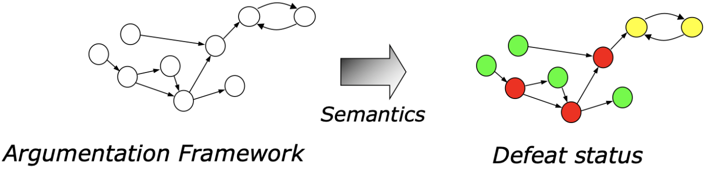
Argumentation example: Vaccination II

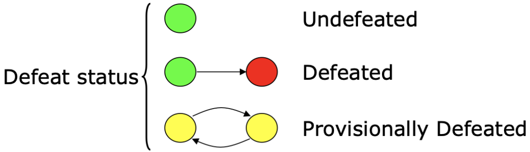


SOME ARGS OMITTED!

Argumentation semantics

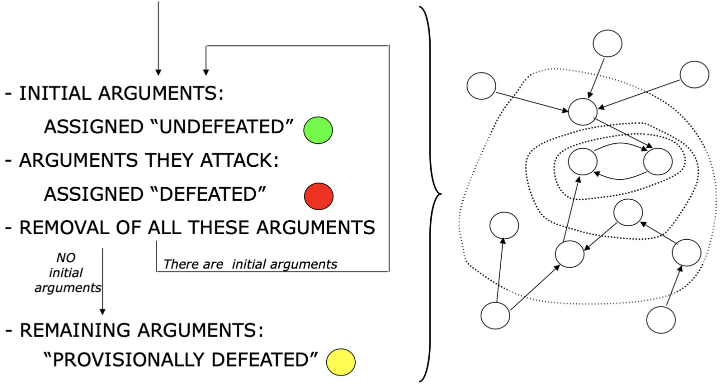
Specification of a method/criteria for argument evaluation, i.e. to determine, given a set of arguments, "their defeat status"





The Grounded Semantics: a unique status approach I

Grounded extension: includes the set of undefeated arguments



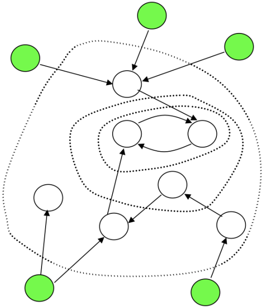
The Grounded Semantics: a unique status approach II

- INITIAL ARGUMENTS:
  ASSIGNED “UNDEFEATED”
- ARGUMENTS THEY ATTACK:
  ASSIGNED “DEFEATED”
- REMOVAL OF ALL THESE ARGUMENTS

NO initial arguments

There are initial arguments

- REMAINING ARGUMENTS:
  “PROVISIONALLY DEFEATED”



The Grounded Semantics: a unique status approach III

- INITIAL ARGUMENTS:

  ASSIGNED "UNDEFEATED"

- ARGUMENTS THEY ATTACK:

  ASSIGNED "DEFEATED"

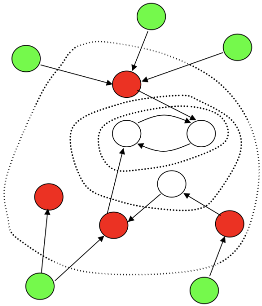
- REMOVAL OF ALL THESE ARGUMENTS

NO initial arguments

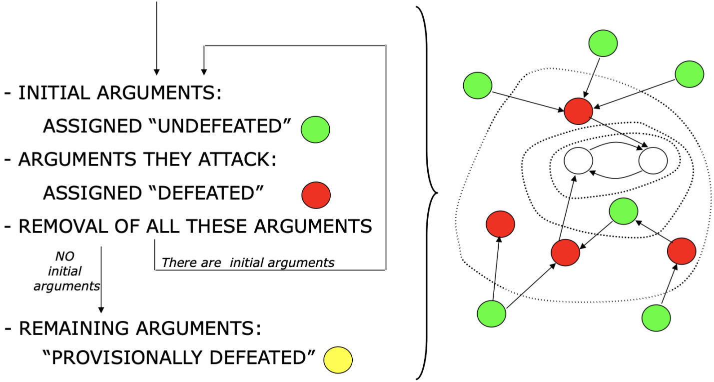
There are initial arguments

- REMAINING ARGUMENTS:

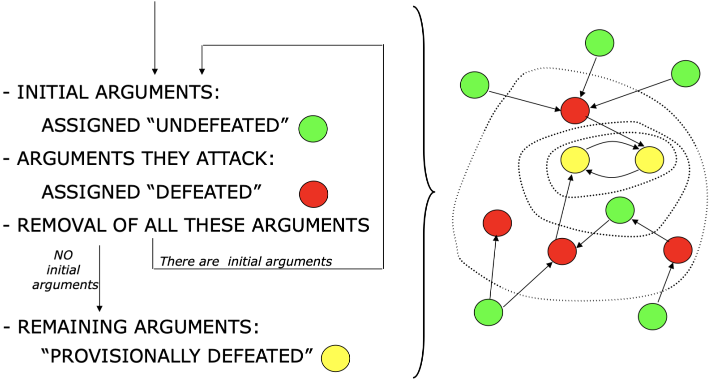
  "PROVISIONALLY DEFEATED"



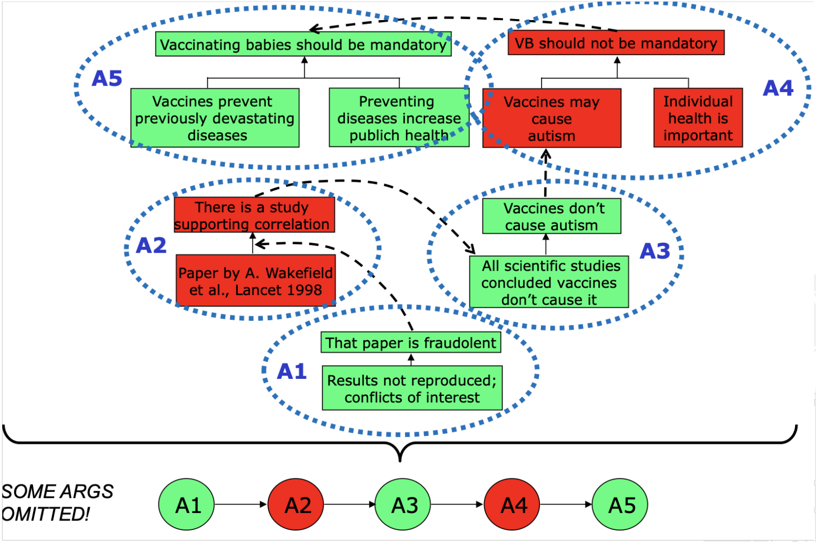
The Grounded Semantics: a unique status approach IV



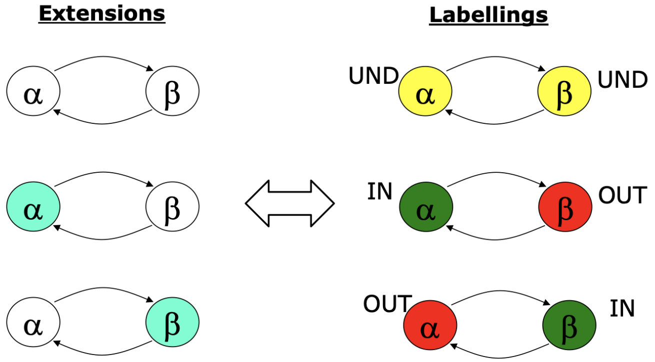
The Grounded Semantics: a unique status approach V



Argumentation example: Vaccination



Relationship between extensions and labellings



LABEL OF AN ARGUMENT

- IN = belongs to the extension
- OUT = attacked by the extension
- UNDEC= does not belong to + not attacked by the extension

Abstract vs Structured argumentation

Abstract argumentation: perspective for studying its nature

- each argument regarded as atomic
- no internal structure to an argument
- no specification of what is an argument or an attack

} assumed to be given

→ not cover needs for understanding argumentation or for building tools for supporting or undertaking argumentation

→ turn to structured argumentation: formal language for

i) representing knowledge,

ii) specifying how arguments and counterarguments can be constructed from that knowledge

Structured argument (normally)

- premises and claim of the argument are explicit
- relationship between the premises and claim is formally defined (for instance using logical entailment)

Structured argumentation

- provides a model of the structure and origin of arguments and attacks
- allows to construct/derive arguments (e.g. from a knowledge base using rules of inference)
- in principle, a structured argumentation system can be conceived as an 'instance' of abstract argumentation (not necessarily Dung-based)

Structured argumentation: main approaches

Four main approaches to structured argumentation [Besnard et al., 2014]

- ABA
- ASPIC+
- Defeasible Logic Programming (DeLP)
- Deductive argumentation

ABA Framework

Assumption-based argumentation (ABA)

A framework for logic-based instantiation of abstract argumentation which captures defeasibility through assumptions

- logic-based instantiation of abstract argumentation
- based on a deductive system $\rightarrow$ set of inference rules
- subset of the language $\rightarrow$ assumption
- argument $\rightarrow$ assumptions set that, with inference rules, proves a claim
- Attacks between arguments $\rightarrow$ notion of 'contrary' assumptions, specifying how to contradict them
- an argument attacks another if its claim contradicts (is the contrary of) some assumption in the other

ABA is equipped with a family of computational mechanisms for capturing specific argumentation semantics (abstract argumentation)

ASPIC+ Framework

ASPIC+ Framework

A general framework incorporating strict and defeasible rules, and handling preferences

- arguments are often resolved with explicit preferences,
- arguments are built with two kinds of inference rules: strict, or deductive rules, whose premises guarantee their conclusion, and defeasible rules, whose premises only create a presumption in favour of their conclusion

→ arguments can be attacked in three ways: on their uncertain premises or on their defeasible inferences, and the latter by either attacking their conclusion or the inference itself

DeLP system

DeLP system

Based on the language of logic programming, answers queries by constructing a dialectical tree

- includes strong negation and default negation
- argument, supporting a claim in the form of a literal, is considered a warrant for that claim if all the arguments that can be posed against it are defeated
- attacks can be directed to the claim, on internal points of an argument, or its premises if they are presumptions of any form
- construction of a *dialectical tree* with the original argument in the root and whose nodes are arguments that can be marked defeated or undefeated, thus determining if the root argument is a warrant or not

Deductive argumentation

Deductive argumentation

A logic-based approach to structured argumentation

- each argument is a *pair*(X, p) where X is a set of logical formulae, called the premises, that entails p , the claim, where entailment is specified by the choice of base logic (e.g. classical logic, modal logic, description logic, temporal logic, conditional logic, etc)
- options available for specifying when one argument attacks another

Argumentation Method: a Comparison I

- ABA and ASPIC+ explicitly within Dung's abstract approach to argumentation, deductive argumentation leaves various choices open, while DeLP deviates from Dung's approach
- The four approaches are at varying levels of abstraction. ABA and ASPIC+ (almost) fully abstract from the nature of the logical language and the inference rules, deductive argumentation is based on general definitions for any monotonic logic, while DeLP has a logic-programming like language with rules composed of propositional or first-order literals
- ASPIC+ and DeLPs explicitly distinguish between two kinds of inference rules called strict and defeasible rules, while ABA reduces defeasible rules to strict rules plus assumptions. In contrast, in deductive argumentation the inference rules are the inference rules of the base logic
- ASPIC+, DeLP, and deductive argumentation allow for the use of explicit preferences to resolve attacks, while ABA encodes preferences in assumption-premises of rules

Argumentation Method: a Comparison II

	ABA	ASPIC+
ARGUMENT STATUS	DUNG	DUNG
ATTACK (DEFEAT) DEFINITION	Attack on assumptions (no distintion wrt conflict definition)	Refinement of conflict taking into account preference order
CONFLICT DEFINITION	Attack on assumptions	Three forms of conflict
ARGUMENT DEFINITION	Strict rules, assumptions + axioms	Strict and defeasible rules, preference order, ordinary premises + axioms
LANGUAGE	Generic language Constrained Contrary relation on assumptions	Generic language General contrary relation

References I

[Baroni and Giacomin, 2007] Baroni, P. and Giacomin, M. (2007).
On principle-based evaluation of extension-based argumentation semantics.
Artificial Intelligence, 171(10):675–700.
Argumentation in Artificial Intelligence.

[Baroni and Giacomin, 2009] Baroni, P. and Giacomin, M. (2009).
Semantics of Abstract Argument Systems, pages 25–44.
Springer US, Boston, MA.

[Besnard et al., 2014] Besnard, P., Garcia, A., Hunter, A., Modgil, S., Prakken, H., Simari, G., and Toni, F. (2014).
Introduction to structured argumentation.
Argument & Computation, 5(1):1–4.

[Calegari et al., 2021] Calegari, R., Omicini, A., and Sartor, G. (2021).
Explainable and ethical AI: A perspective on argumentation and logic programming.
In *AIxIA 2020 – Advances in Artificial Intelligence*, volume 12414 of *Lecture Notes in Computer Science*, pages 19–36. Springer Nature.

[Čyras et al., 2021] Čyras, K., Rago, A., Albini, E., Baroni, P., and Toni, F. (2021).
Argumentative xai: a survey.
In *IJCAI*.

References II

[Dung, 1995a] Dung, P. M. (1995a).
An argumentation-theoretic foundation for logic programming.
The Journal of logic programming, 22(2):151–177.

[Dung, 1995b] Dung, P. M. (1995b).
On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games.
Artificial Intelligence, 77(2):321–357.

[García and Simari, 2004] García, A. J. and Simari, G. R. (2004).
Defeasible logic programming: An argumentative approach.
Theory and practice of logic programming, 4(1–2):95–138.

[Governatori and Maher, 2000] Governatori, G. and Maher, M. J. (2000).
An argumentation-theoretic characterization of defeasible logic.
14th European Conference on Artificial Intelligence.

[Governatori et al., 2004] Governatori, G., Maher, M. J., Antoniou, G., and Billington, D. (2004).
Argumentation semantics for defeasible logic.
Journal of Logic and Computation, 14(5):675–702.

References III

[Lawrence and Reed, 2020] Lawrence, J. and Reed, C. (2020).
Argument mining: A survey.
Computational Linguistics, 45(4):765–818.

[Loui, 1986] Loui, R. (1986).
Defeat among arguments: a system of defeasible inference.
Computational intelligence, 3(2):100–106.

[Lucic et al., 2022] Lucic, A., Ter Hoeve, M. A., Tolomei, G., De Rijke, M., and Silvestri, F. (2022).
Cf-gnnexplainer: Counterfactual explanations for graph neural networks.
In *International Conference on Artificial Intelligence and Statistics*, pages 4499–4511. PMLR.

[McCarthy, 1980] McCarthy, J. (1980).
Circumscription—a form of non-monotonic reasoning.
Artificial intelligence, 13(1–2):27–39.

[Miller, 2019] Miller, T. (2019).
Explanation in artificial intelligence: Insights from the social sciences.
Artificial intelligence, 267:1–38.

References IV

[Moore, 1987] Moore, R. (1987).
Possible-world semantics for autoepistemic logic.
In *Readings in Nonmonotonic Reasoning*, pages 137–142. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[Perelman and Olbrechts-Tyteca, 1997] Perelman, C. and Olbrechts-Tyteca, L. (1997).
Traité de l'argumentation. La nouvelle rhétorique.
Vrin.

[Pollock, 1995] Pollock, J. L. (1995).
Cognitive carpentry: A blueprint for how to build a person.
Mit Press.

[Pollock and Cruz, 1999] Pollock, J. L. and Cruz, J. (1999).
Contemporary theories of knowledge, volume 35.
Rowman & Littlefield.

[Ribeiro et al., 2016] Ribeiro, M. T., Singh, S., and Guestrin, C. (2016).
" why should i trust you?" explaining the predictions of any classifier.
In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144.

References V

[Ribeiro et al., 2018] Ribeiro, M. T., Singh, S., and Guestrin, C. (2018).
Anchors: High-precision model-agnostic explanations.
In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

[Sartor, 1995] Sartor, G. (1995).
Defeasibility in legal reasoning.
In *Informatics and the foundations of legal reasoning*, pages 119–157. Springer.

Defeasible Reasoning & Argumentation

Joint International Doctoral (Ph.D) degree in Law, Science and Technology

Slides: https://apice.unibo.it/xwiki/bin/view/Courses/LogicPhd22

Roberta Calegari

roberta.calegari@unibo.it

Alma Mater Studiorum – Università di Bologna

Academic Year 2022/2023