Utrecht University

Tilburg University
Law School

Real (Computational) Argumentation in Practice

Floris Bex

Professor Data Science & Judiciary – Tilburg University

Associate Professor AI – Utrecht University

Scientific director Police Lab AI – Utrecht University

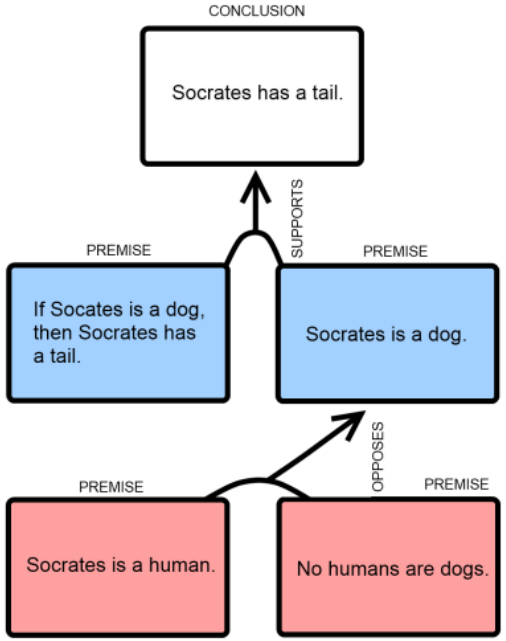
Real argumentation in practice

- What is “real” argumentation?
  - And how does it connect to computational argumentation?
- How can real argumentation be used in practical applications?

Is computational argumentation really argumentation?

- Dung-style semantics – abstract argumentation
  - What's in the arguments?
  - "Calculus of opposition" (abstract argumentation is not argumentation)
- (Defeasible) logics – structured argumentation
  - Mathematical proof theory
  - Does real argumentation follow the rules of logic?
- Dialogical argumentation
  - Process of argumentation – what are the claims and arguments themselves?

What is a “real” argument anyway?



- [Monty Python’s take on what an argument is]

Elements of argumentation

- Ethos: the credibility, expertise and charisma of the speaker
- Pathos: the emotions or values of the audience
- Logos: the facts or reasons that support the speakers claim

Argumentation

- Providing reasons for claims, giving counterarguments to claims
- Different senses of argument:
  - structures of reasons for conclusions
    - we regulate AI *because* it poses a risk
    - we should not regulate AI *because* it would stifle innovation
  - dialogue between agents
    - EU: “We should regulate AI.”
    - Tech industry: “Why?”
    - EU: “because it poses a risk.”
    - Tech industry: ”No, we shouldn’t; regulation is bad for innovation.”

Structured argumentation

- Premises (statements) lead to a conclusion (statement)
  - (premise) All men are mortal
  - (premise) Socrates is a man
  - (conclusion) Socrates is mortal

- The conclusion is inferred from the premises
- The premises serve as reasons for the conclusion

Structured (logical) argumentation

- Premises (statements) lead to a conclusion (statement)
  - (premise) $\forall x. man(x) \rightarrow mortal(x)$
  - (premise) $man(Socrates)$
  - (conclusion) $mortal(Socrates)$

- The conclusion is inferred from the premises
- The premises serve as reasons for the conclusion

Building arguments – examples from evidential reasoning

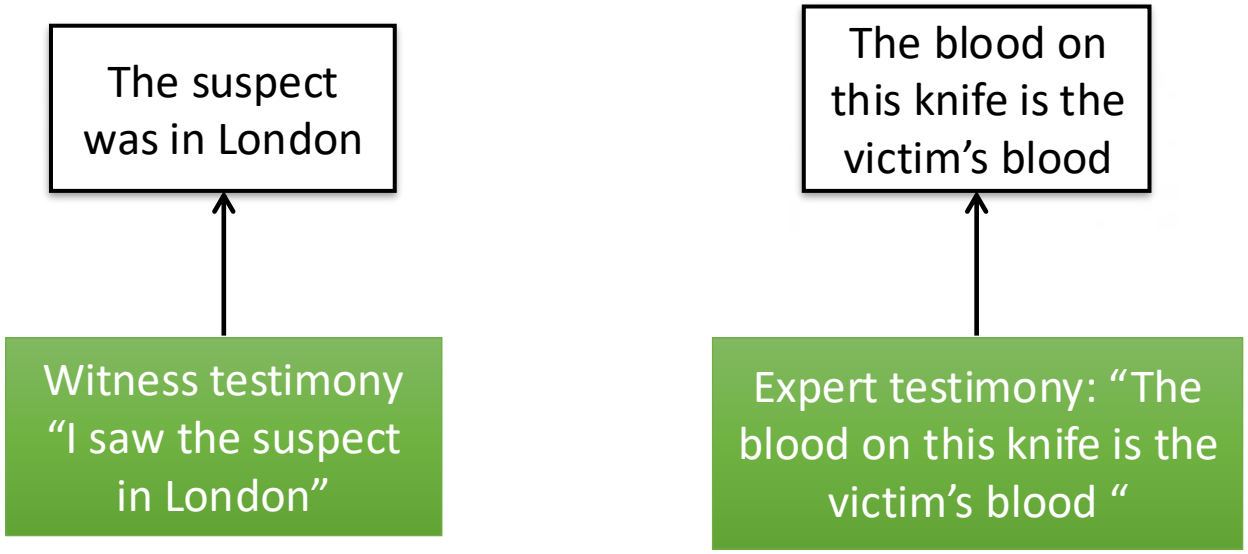
- Given the premises (the evidence in a case)…

Witness testimony "I saw the suspect in London"

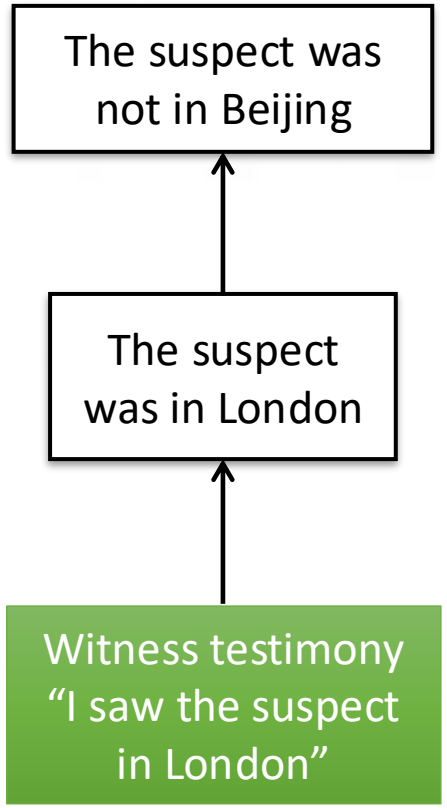
Expert testimony: "The blood on this knife is the victim's blood "

Building arguments – examples from evidential reasoning

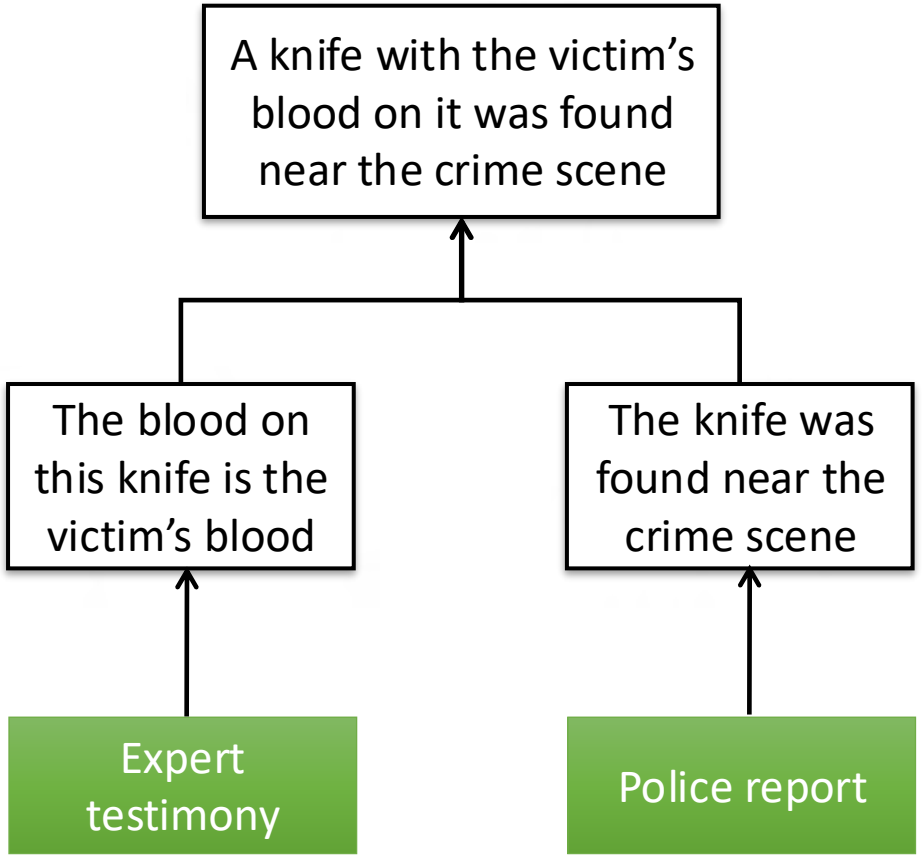
- …we can infer conclusions



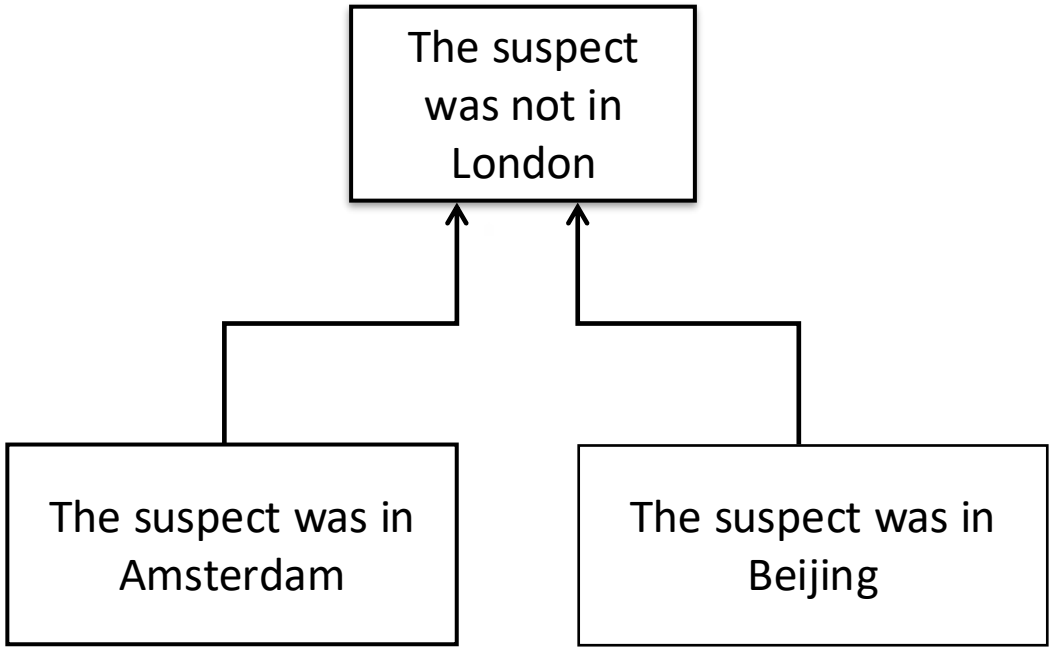
Conclusions can serve as premises for a new inference



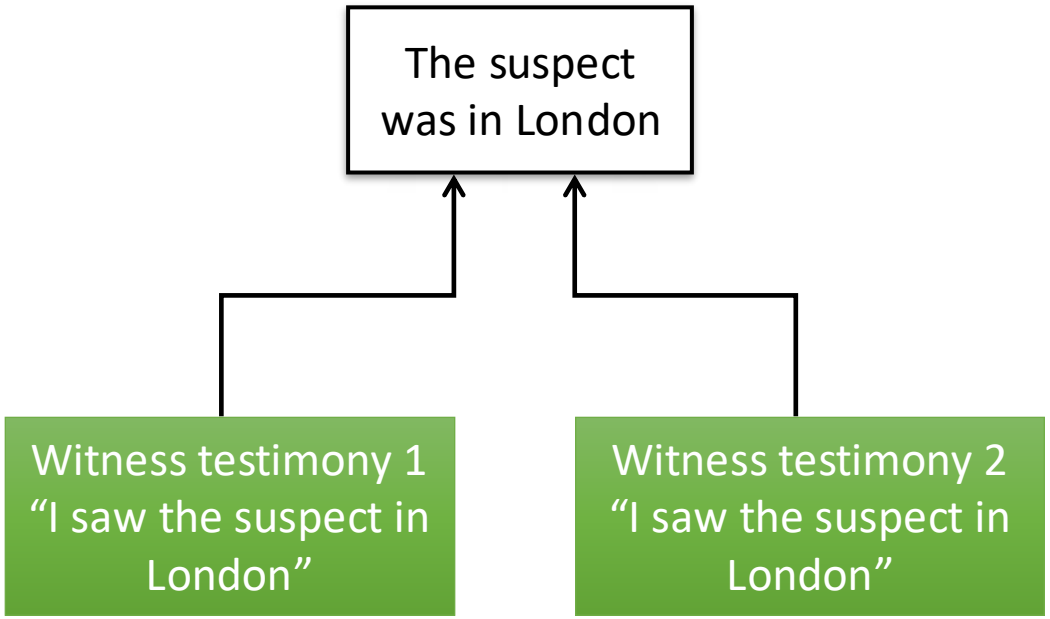
Linked arguments: both premises are needed (conjunction)



Convergent arguments: the pieces of evidence are separate reasons for the conclusion



Accrual: the pieces of evidence are separate reasons (and strengthen the conclusion)



Arguments built on commonsense knowledge

- General knowledge or knowledge from personal experience
  - Christmas is the 25th of December
  - Floris Bex does not have red hair
- Generalizations: statements about how we think the world around us works
  - the impact of a hammer can break a person's skull
  - witnesses under oath usually speak the truth
  - Chinese people are smaller than Dutch people
  - women are worse drivers than men

Commonsense knowledge and generalizations

- Generalizations are not always true!
  - Exceptions
- Qualify generalizations with words such as *usually*, *sometimes*
- Generalizations can appeal to prejudice and may depend on the community
  - Chinese basketball players are not shorter than Dutch people.
  - Women are not worse drivers than men.
- When arguing, make the generalizations used explicit

Commonsense knowledge in argumentation

- As premises or claims

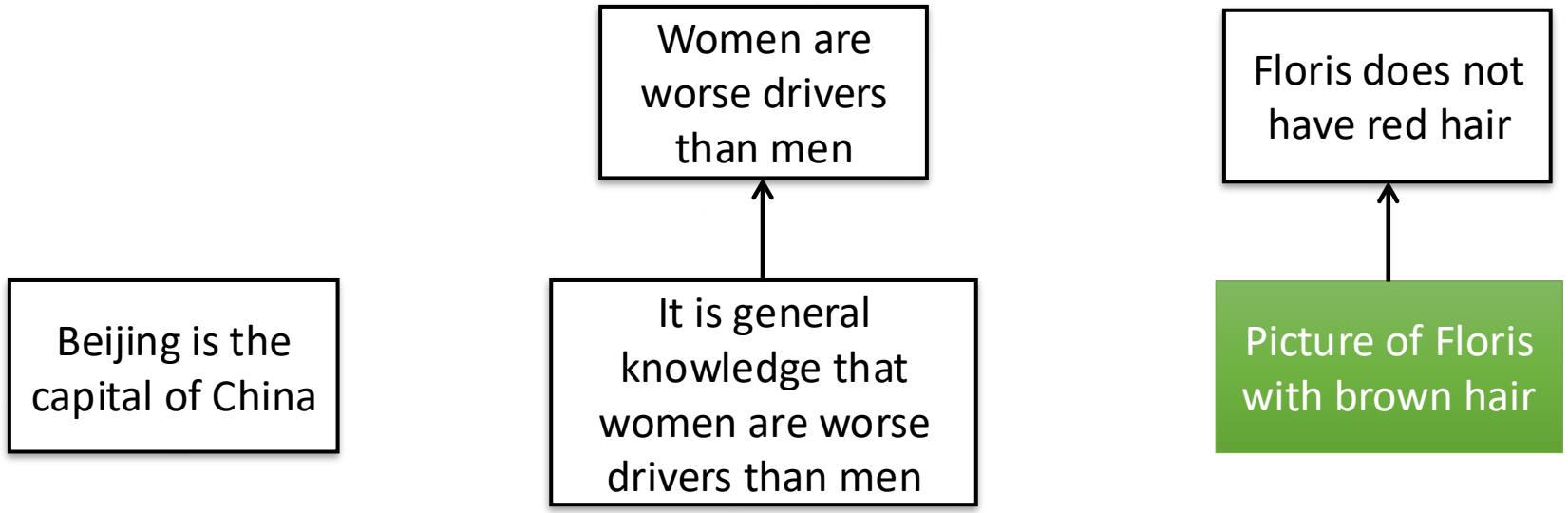
Beijing is the capital of China

Women are worse drivers than men

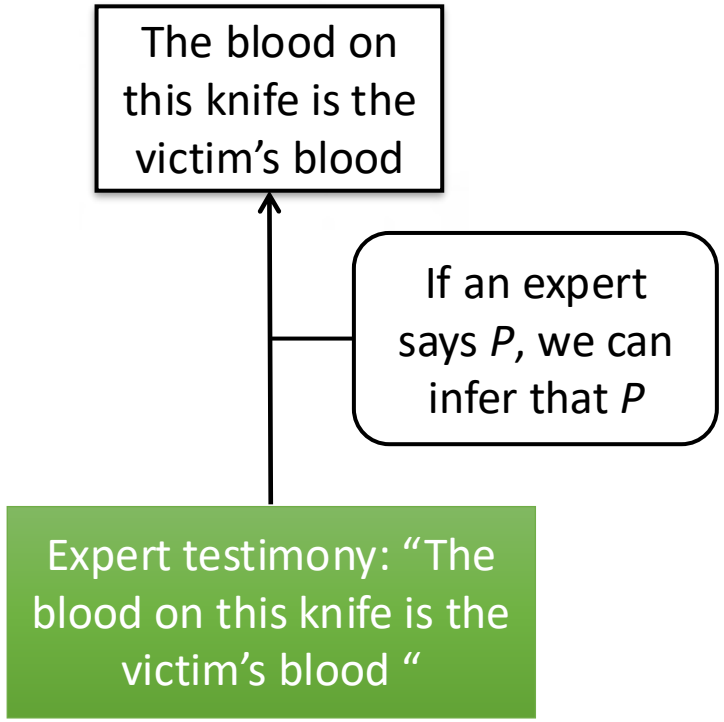
Floris does not have red hair

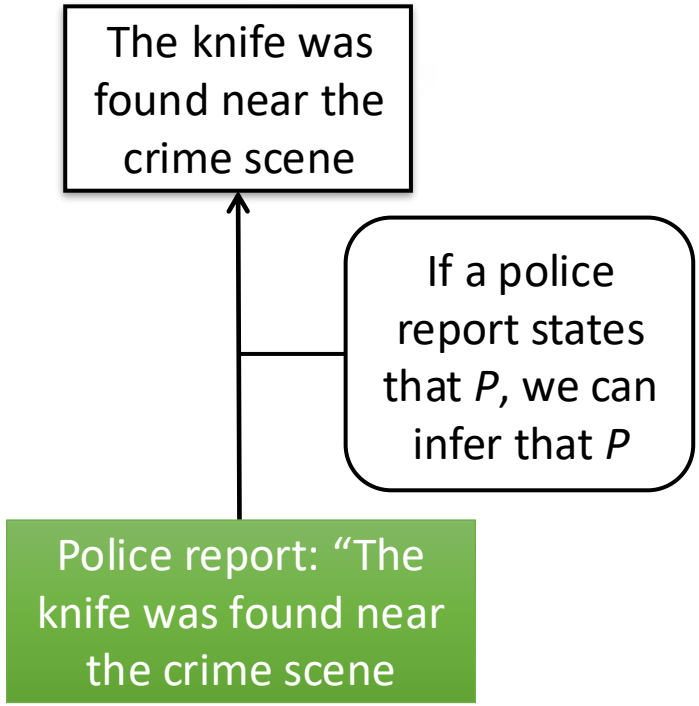
Commonsense knowledge in argumentation

- Claims can be further supported
- Different types of premises: necessary (cannot be denied), defeasible (can be denied)



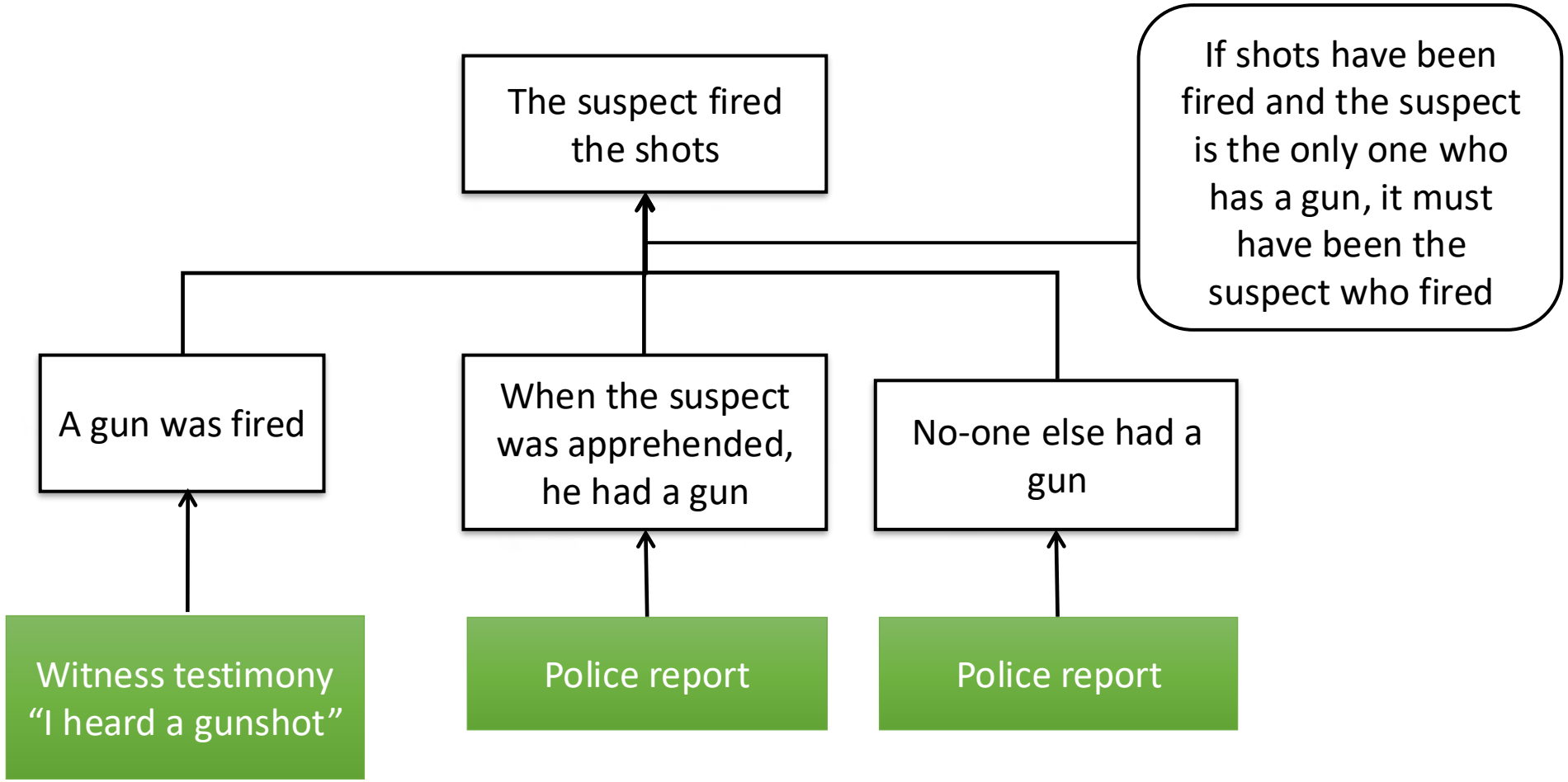
Commonsense knowledge as inference rules



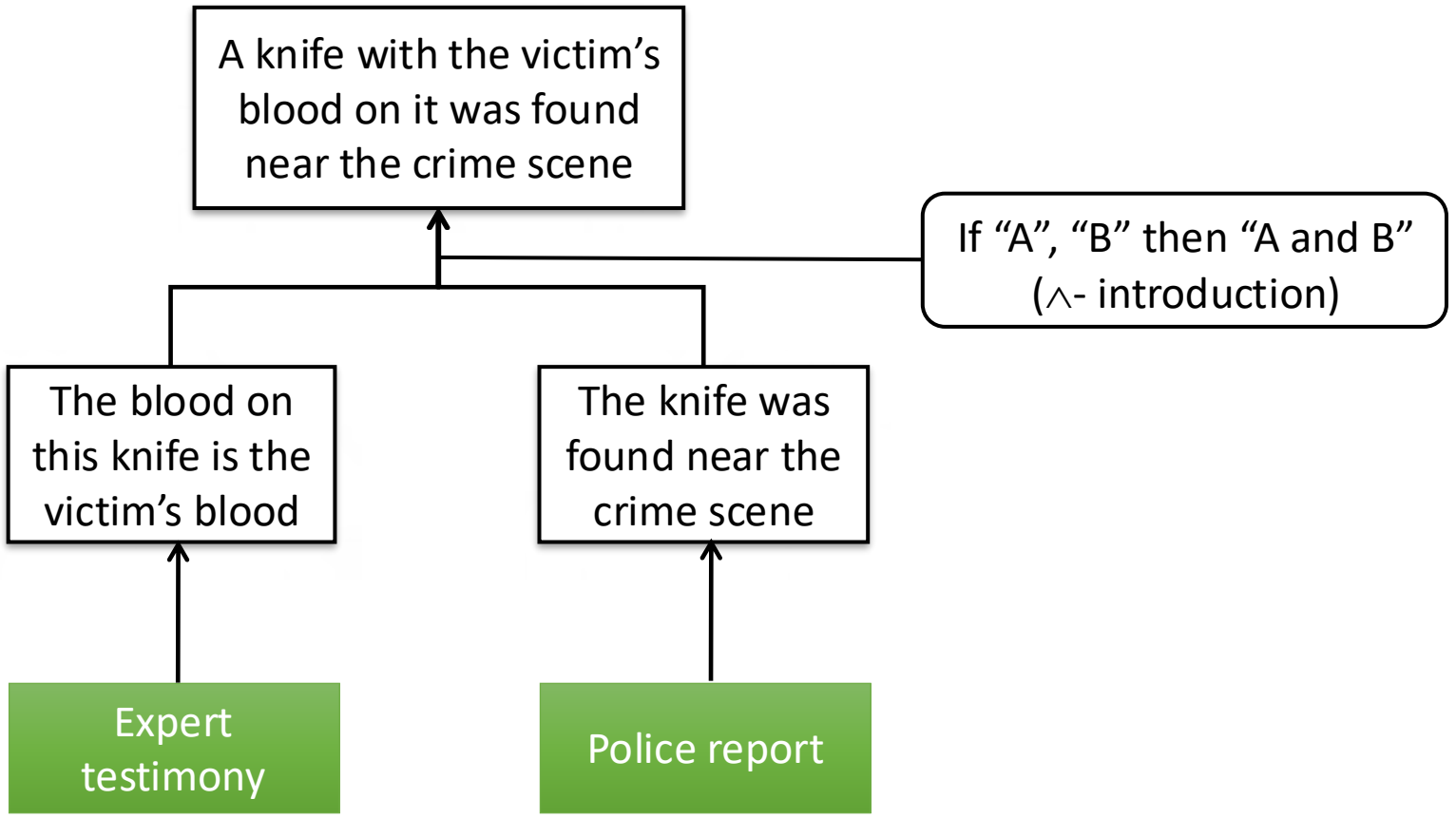


Commonsense knowledge as inference rules

Commonsense knowledge as inference rules



Logical inference rules



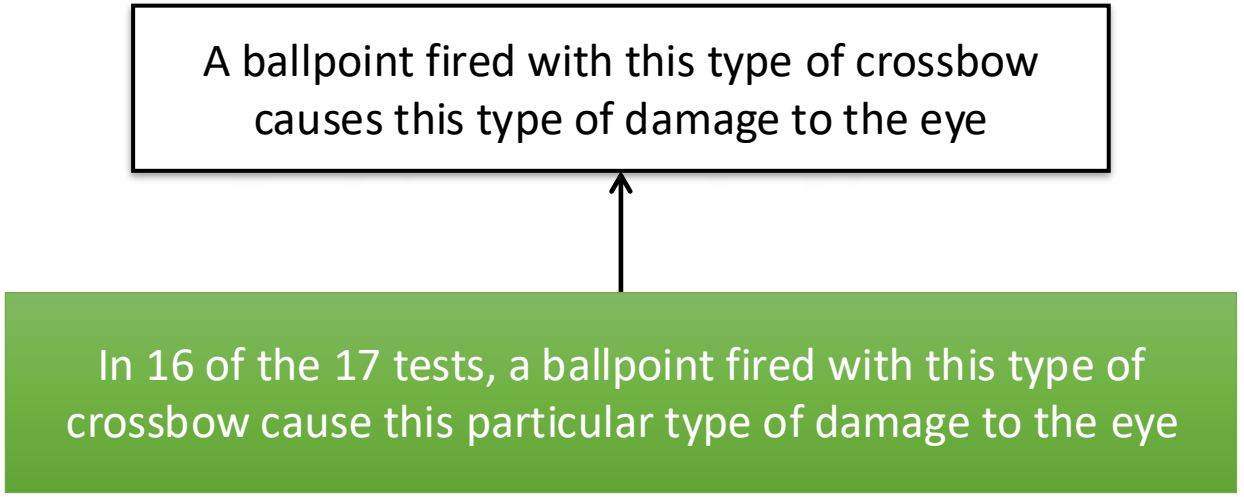
Where do the inference rules come from?

- There are many (possible) inference rules
  - Logical inference rules
  - Legal rules
  - Rules of rational argumentation
  - Rules following from scientific research
  - General knowledge
  - Prejudice

- Many rules or schemes have critical questions, which may be used to critically analyse the inference

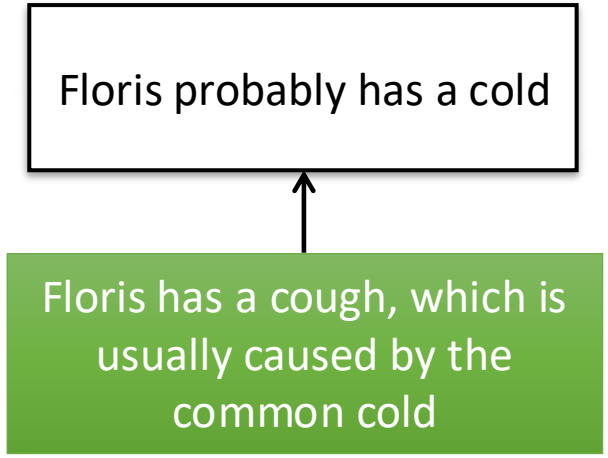
Inference rule for inductive reasoning

- Most/all observed P's were Q's *therefore* if P then usually Q
  - Was the sample big enough?
  - Was the sample randomly selected?



Inference rule for inference to the best explanation

- Usually, A causes B. B has been observed, so A must have occurred
  - Is A a common cause of B?
  - Is there another better explanation for B?
  - Have all possible explanations of B been considered?



Inference rule for witness testimony

- Witness W says that P *therefore* P
  - Is W speaking the truth?
  - Is W's memory good?
  - Is W's perception good?
  - What do other witnesses say?

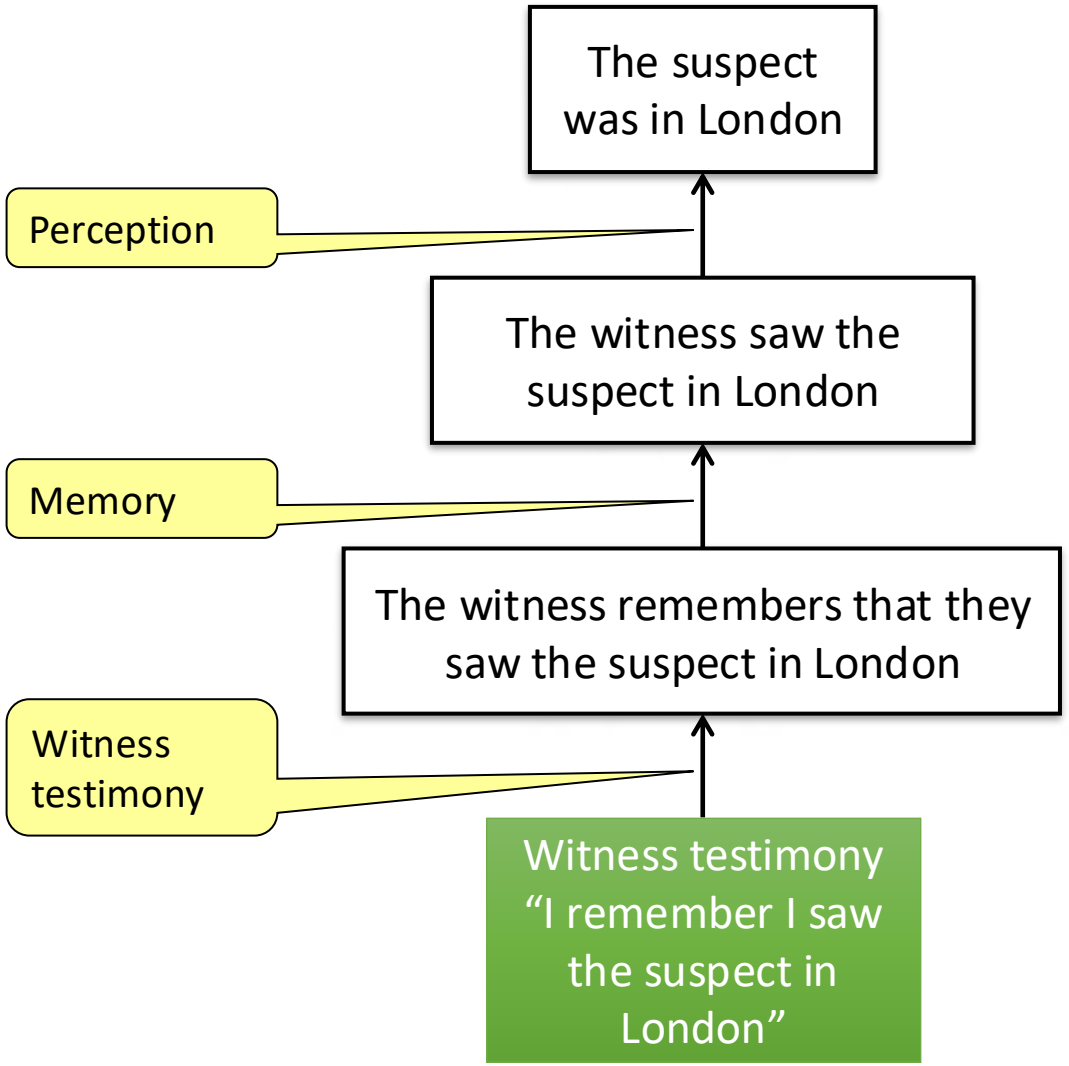
Inference rule for perception

- X saw that P *therefore* P
  - Were the circumstances such that X could see P?
  - Does X have good sight?

Inference rule for memory

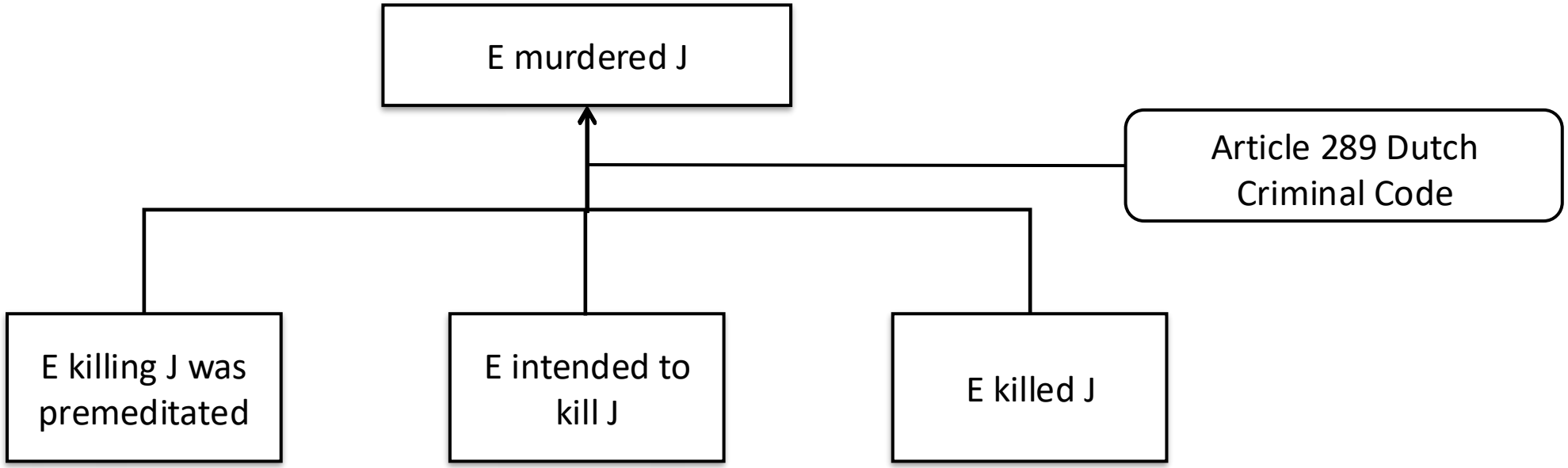
- X remembers that P *therefore* P
  - Can it be that X's memory has been tainted?
  - Does X have good memory?

Inference rule for witness testimony unpacked



Inference rules based on legal rules

- Anyone who deliberately and with malice aforethought takes the life of another shall be considered guilty of murder



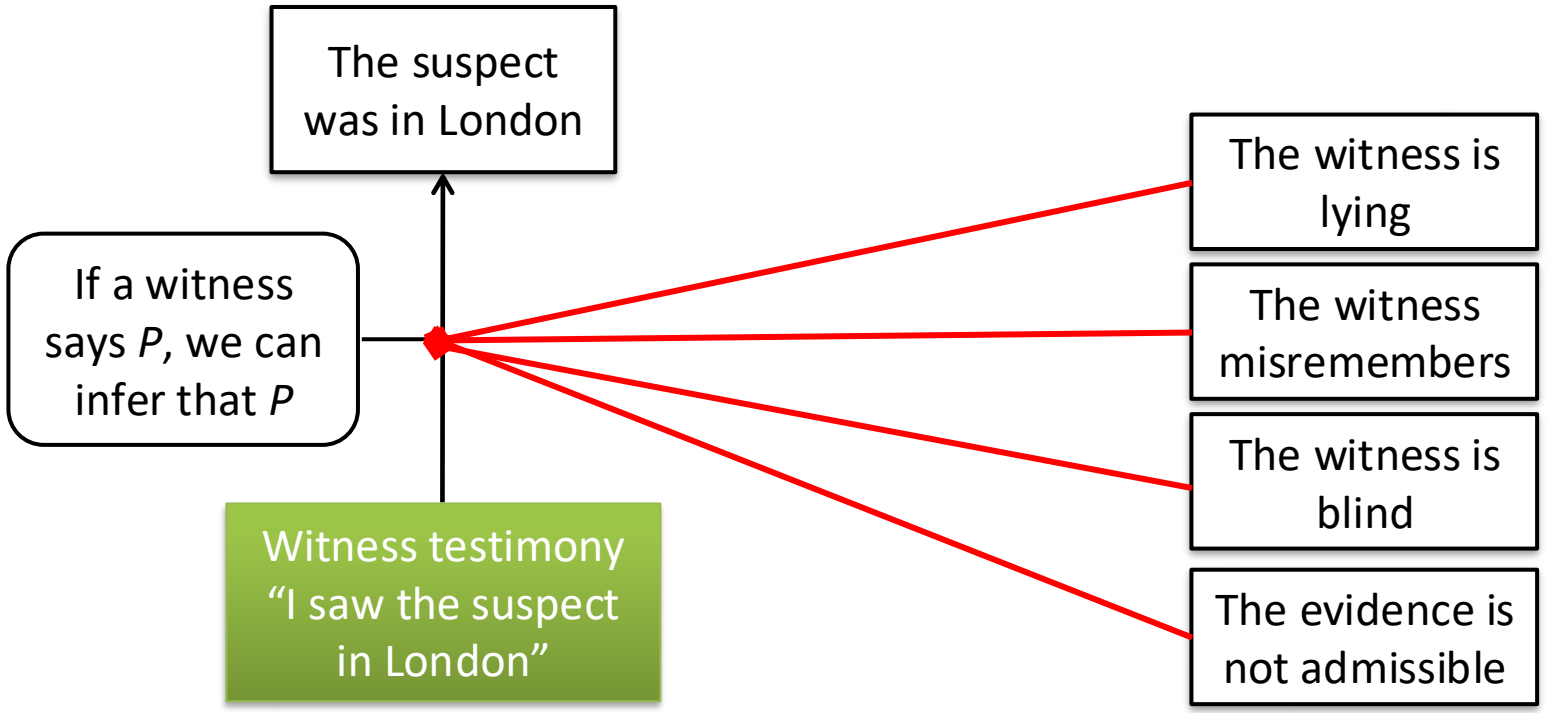
Counterarguments

- Arguments may be attacked on each of their elements.
  - Underminer: attack a (non-necessary) premise
  - Rebuttal: attack a (sub-)conclusion
  - Undercutter: Attack the application of an inference rule by arguing for an exception
    - Note: inference rules themselves cannot be attacked
- Critical questions point to possible attackers of argument based on a specific rule

Undermining: premise attack

Rebutting: conclusion attack

Undercutting: attacking the application of the inference rule



Realistic arguments & argumentation

- Reasons for claims, and counterarguments against claims or inferences
  - Structured (logic, diagrams, bullet lists, outlines)
  - As pertaining to a realistic context, involving explicit or implicit (commonsense) knowledge

So how much of COMMA is about “real” arguments?

- COMMA 2022
  - Realistic (ML & AM): 2
  - Realistic (formal logic): 3
  - Toy examples: 3
  - No realistic argumentation: 20
- COMMA 2024
  - Realistic (ML & AM): 5
  - Realistic (formal logic): 8
  - Toy examples: 4
  - No realistic argumentation: 11

Constructing real arguments yourself

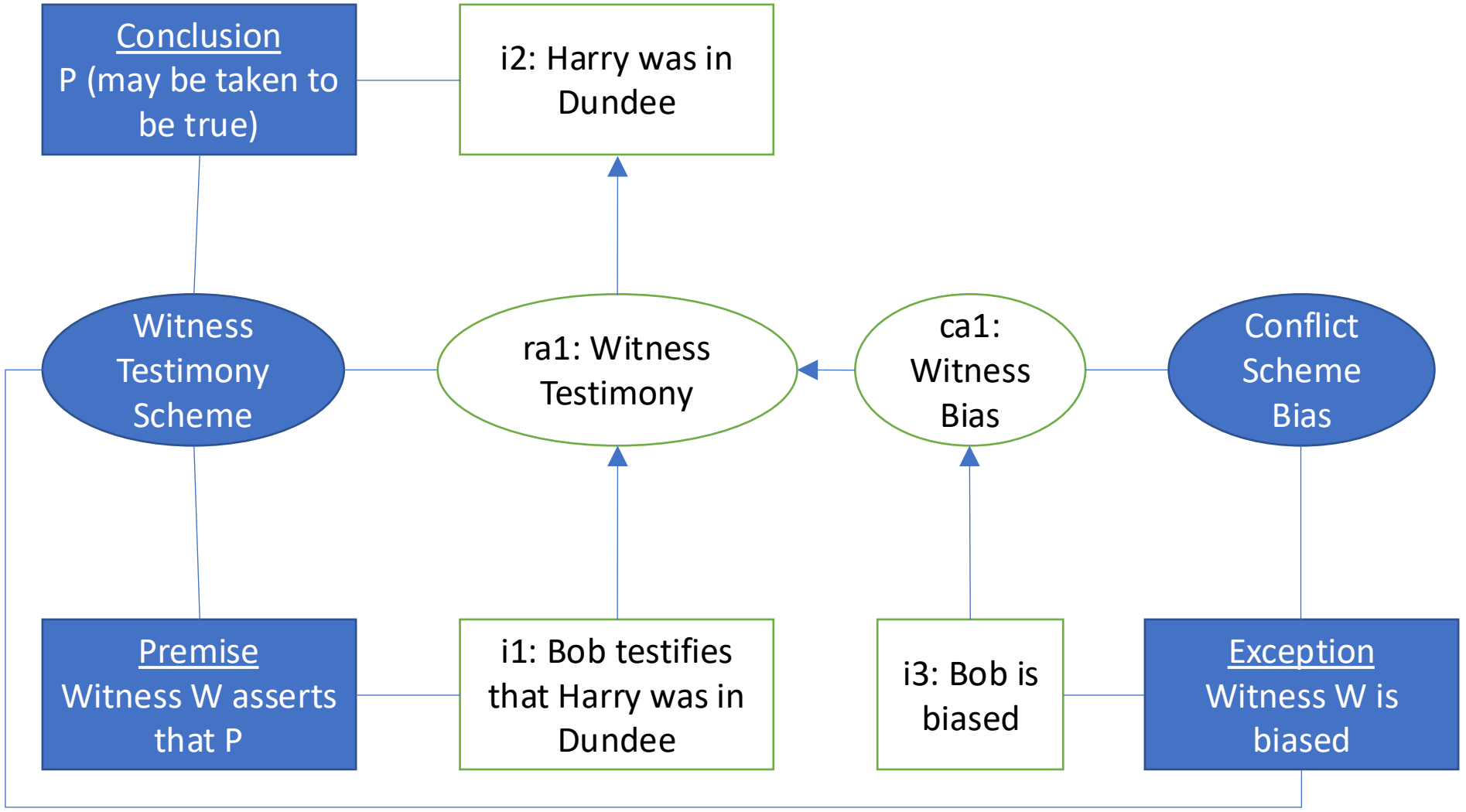
- Open the below URL in OVA (http://ova.arg-tech.org/)
  - https://www.florisbex.com/SaccoVanzetti.htm

- Construct an argument with counterarguments

Linked argument data – the AIF

- AIF – Argument Interchange Format – is an ontology for argumentation
  - Linked data, knowledge graphs
- I-nodes: information
  - S-nodes: relations between information
    - RA-nodes (inference), CA-nodes (conflict), PA-nodes (preference)
- Forms ontology for representing concepts
  - Argumentation Schemes
  - Conflict Schemes

Arguments in the AIF



Formal (logical) argumentation

- How do argumentation diagrams relate to formal (logical) argumentation?

- AIF <-> ASPIC+ framework for structured argumentation
  - ASPIC+ arguments for a Dung-argumentation-framework
  - ASPIC+ has clearly defined links to e.g. ABA, DELP, logical argumentation

F.J. Bex, S. Modgil, H. Prakken & C. Reed (2013) On Logical Specifications of the Argument Interchange Format. *Journal of Logic and Computation*, 23 (5): 951-989

ASPIC+ (without preferences)

- Arguments are trees where
  - Nodes are wff of a logical language $\mathcal{L}$
  - Links are applications of inference rules
    - $\mathcal{R}_s$ = Strict rules (ϕ_1, ..., $\phi_1 \rightarrow \phi$); or
    - $\mathcal{R}_d$= Defeasible rules (ϕ_1, ..., $\phi_1 \Rightarrow \phi$)
  - Constructed from a knowledge base $\mathcal{K} \subseteq \mathcal{L}$
    - $\mathcal{K}_n$ = (necessary) axioms
    - $\mathcal{K}_p$ = ordinary premises

Attack and defeat in ASPIC+

- Negation:
  - generalised to arbitrary contrary relation between formulas (cf. ABA)

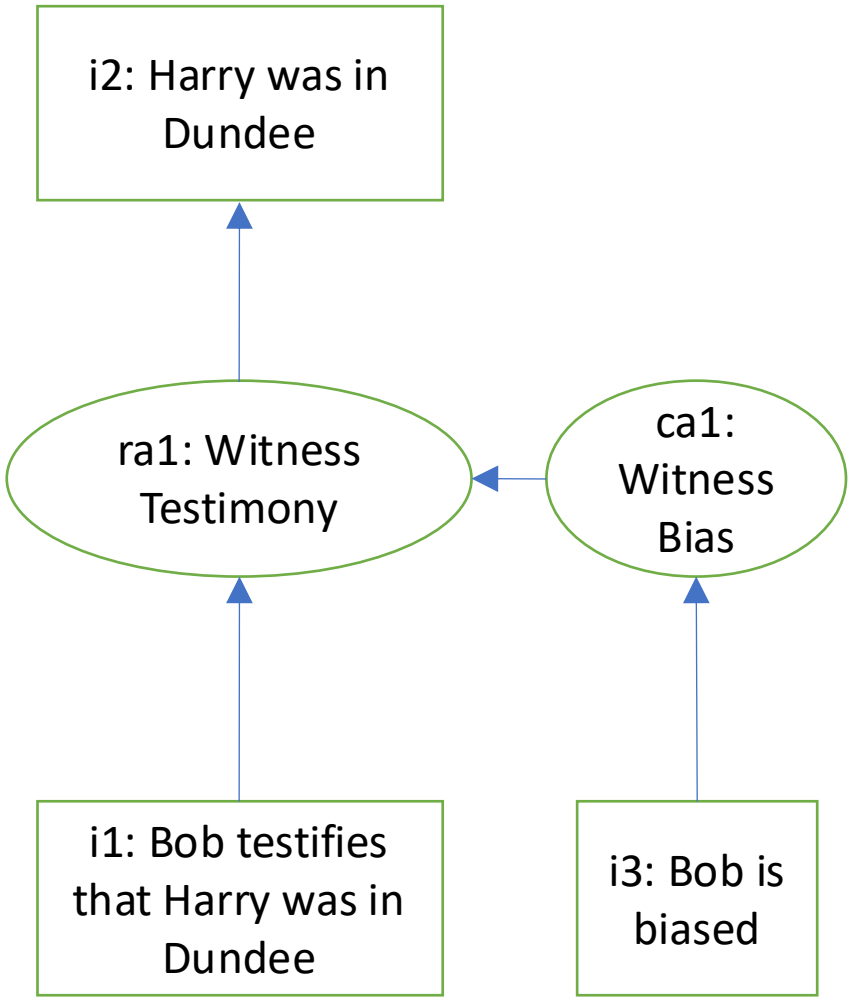
- Attack:
  - on premise (undermining),
  - on conclusion of defeasible inference (rebutting),
  - on defeasible inference (undercutting)

- A defeats B iff for some sub-argument B′ of B, A attacks B′

ASPIC+ argumentation theory

- An argumentation theory is a triple $AT = (AS, KB)$ where:
  - AS is an argumentation system
    - Logical language $\mathcal{L}$ with contrary relation, strict and defeasible inference rules $\mathcal{R}_s$ and $\mathcal{R}_d$
  - KB is a knowledge base in AS
- Since we have defined a binary defeat relation on $Args_{AT}$ this instantiates Dung's (1995) abstract argumentation frameworks

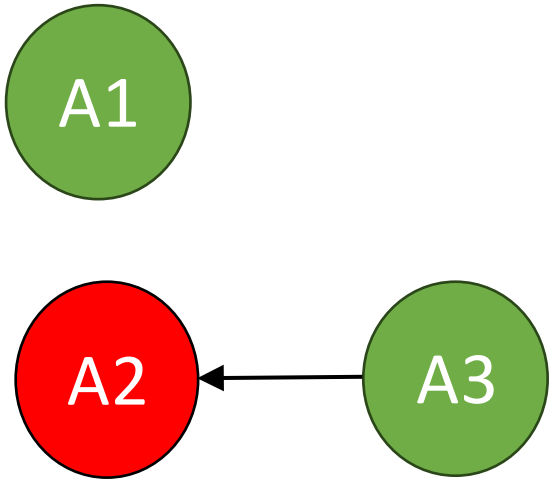
AIF <-> ASPIC+



- $\mathcal{K}$ = {i1, i3}
- $\mathcal{R}_d$ = {*ra1*: i1 ⇒ i2}
- –*ra1* = {i3}

- A1: i1
- A2: A1 ⇒ i2
- A3: i3

- A3 undercuts A2



Back to Sacco & Vanzetti

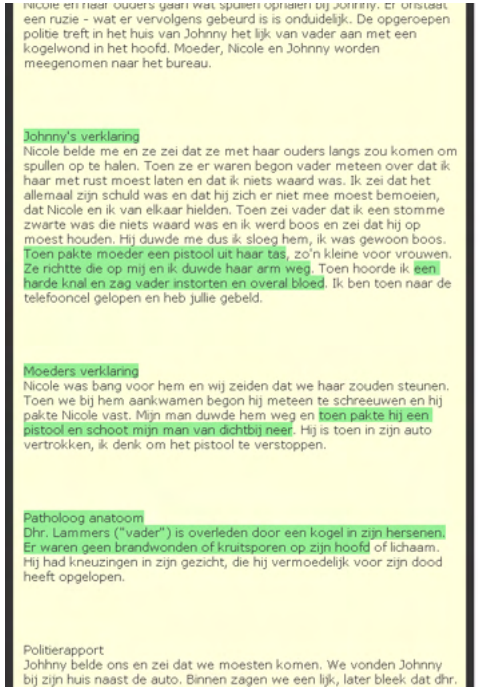
- Which extensions do you get from your argumentation theory about Sacco & Vanzetti?

Applications of “real” computational argumentation

- Argument diagramming
  - Sensemaking & critical thinking
- Automated reasoning
  - Interactive dialogues
  - Decision-making and -support

Argument diagramming

- Explicitly (manually) diagramming arguments in a (semi-) formal structure

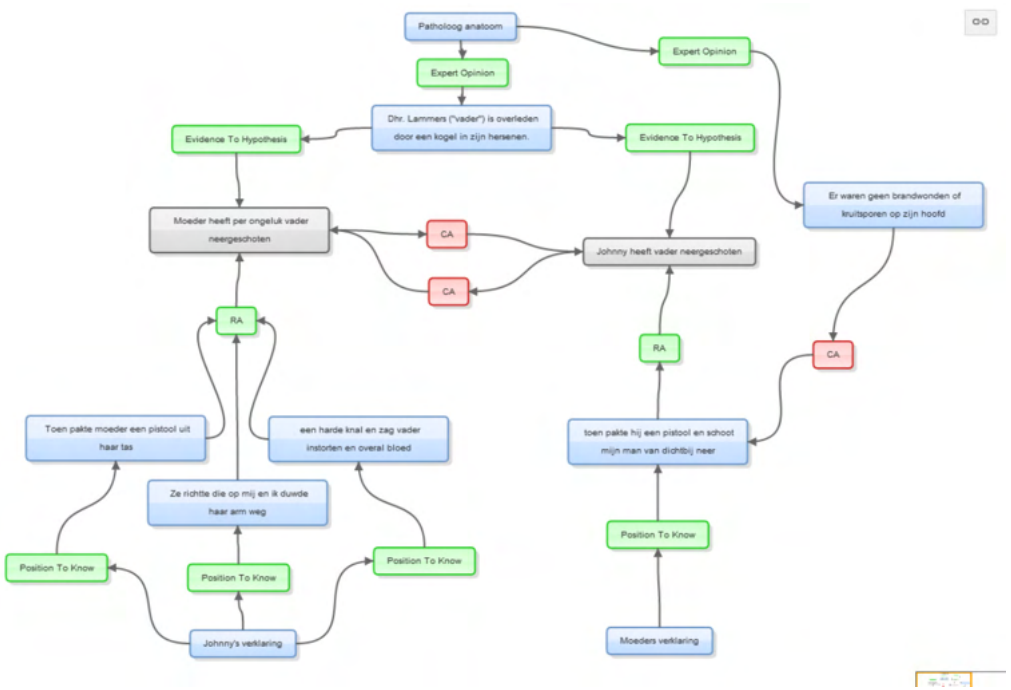

een ruzie - wat er vervolgens gebeurd is is onduidelijk. De opgeroepen politie treft in het huis van Johnny het lijk van vader aan met een kogelwond in het hoofd. Moeder, Nicole en Johnny worden meegenomen naar het bureau.

Johnny's verklaring
Nicole belde me en ze zei dat ze met haar ouders langs zou komen om spullen op te halen. Toen ze er waren begon vader meteen over dat ik haar met rust moest laten en dat ik niets waard was. Ik zei dat het allemaal zijn schuld was en dat hij zich er niet mee moest bemoeien, dat Nicole en ik van elkaar hielden. Toen zei vader dat ik een stomme zwarte was die niets waard was en ik werd boos en zei dat hij op moest houden. Hij duwde me dus ik sloeg hem, ik was gewoon boos. Toen pakte moeder een pistool uit haar tas, zo'n kleine voor vrouwen. Ze richtte die op mij en ik duwde haar arm weg. Toen hoorde ik een harde knal en zag vader instorten en overal bloed. Ik ben toen naar de telefooncel gelopen en heb jullie gebeld.

Moeders verklaring
Nicole was bang voor hem en wij zeiden dat we haar zouden steunen. Toen we bij hem aankwamen begon hij meteen te schreeuwen en hij pakte Nicole vast. Mijn man duwde hem weg en toen pakte hij een pistool en schoot mijn man van dichtbij neer. Hij is toen in zijn auto vertrokken, ik denk om het pistool te verstoppen.

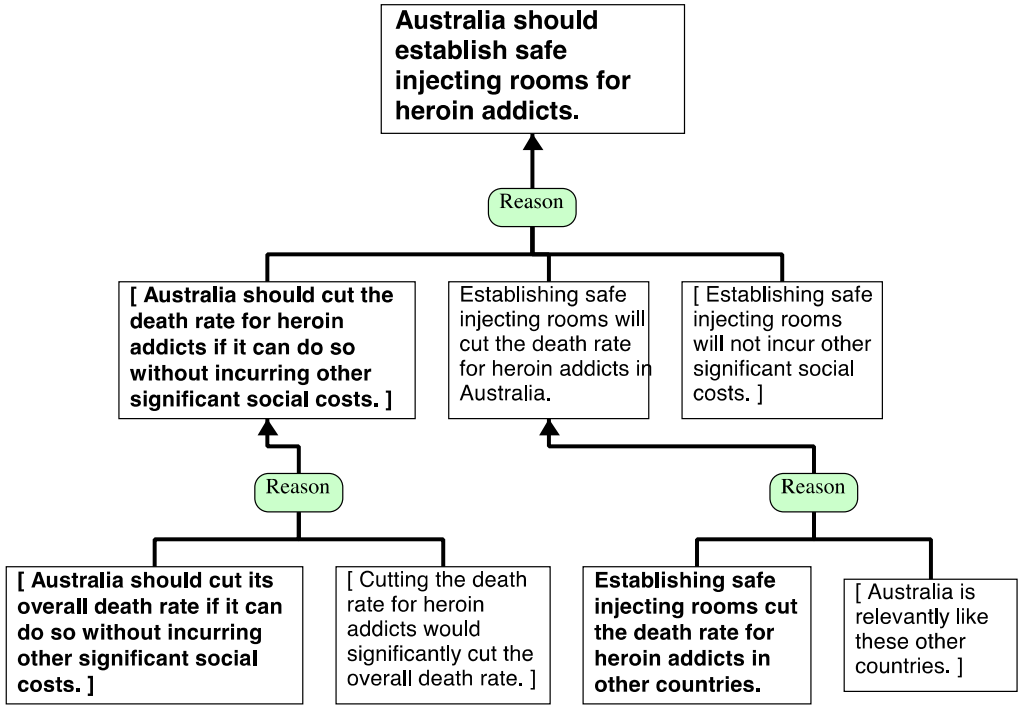
Patholoog anatoom
Dhr. Lammers ("vader") is overleden door een kogel in zijn hersenen. Er waren geen brandwonden of kruitsporen op zijn hoofd of lichaam. Hij had kneuzingen in zijn gezicht, die hij vermoedelijk voor zijn dood heeft opgelopen.

Politierapport
Johhny belde ons en zei dat we moesten komen. We vonden Johnny bij zijn huis naast de auto. Binnen zagen we een lijk, later bleek dat dhr.



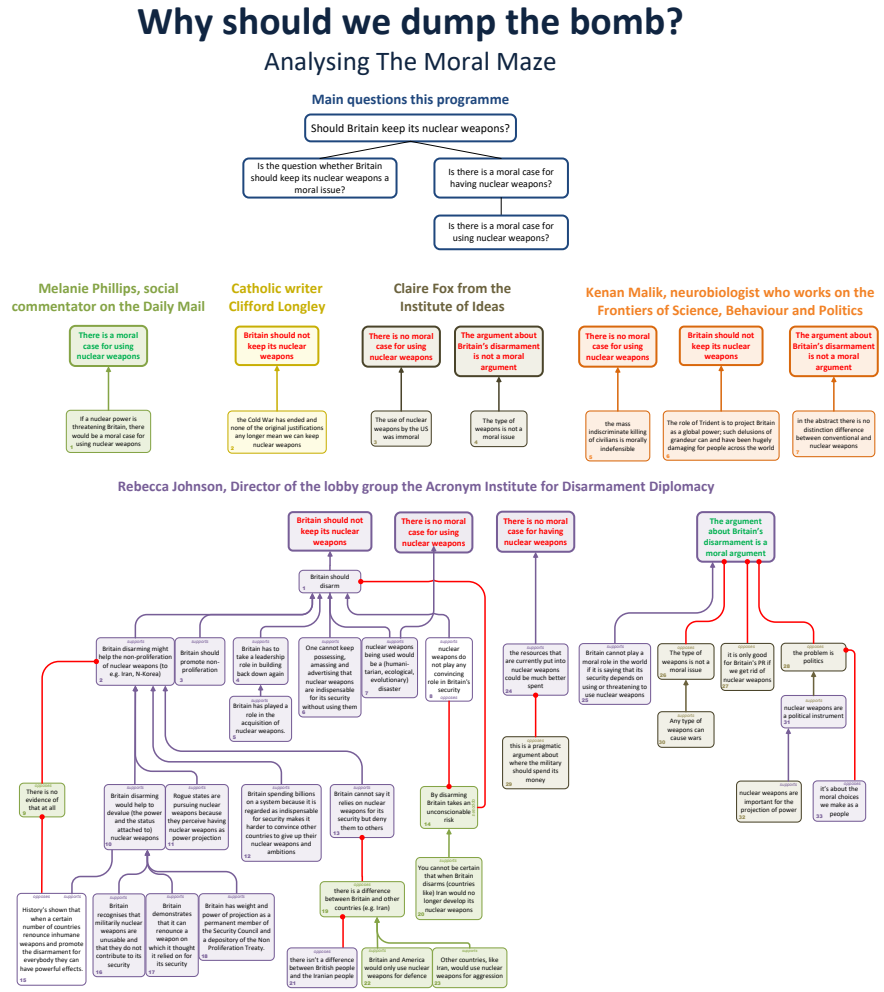
Argument diagramming – for critical thinking

- Explicitly mapping arguments makes implicit knowledge explicit, provides an overview of the structure of arguments



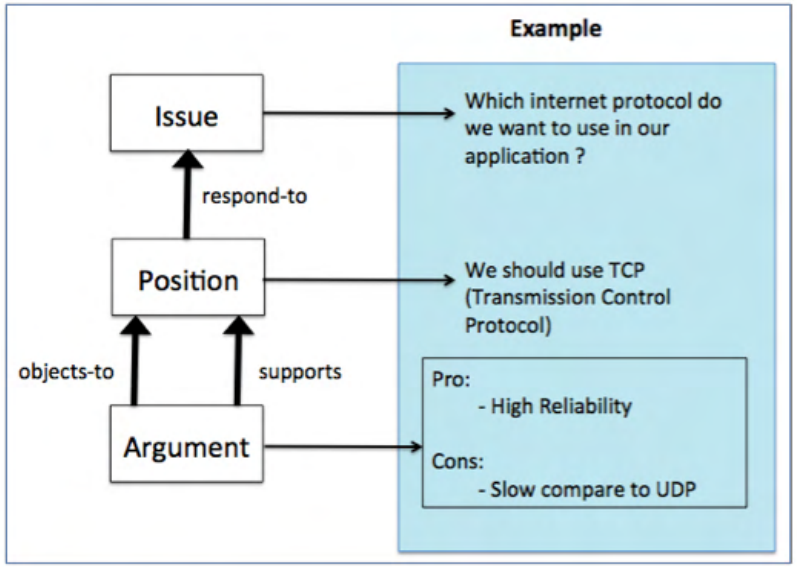
Argument diagramming – for mapping complex debates

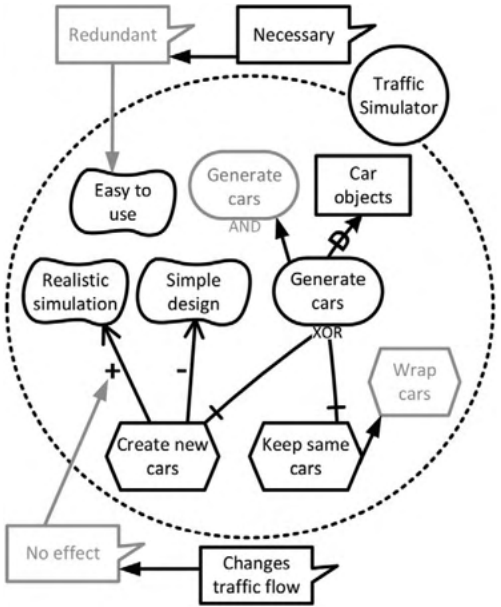
- Providing overviews of large and complex debates
  - Debategraph
  - Argument Web

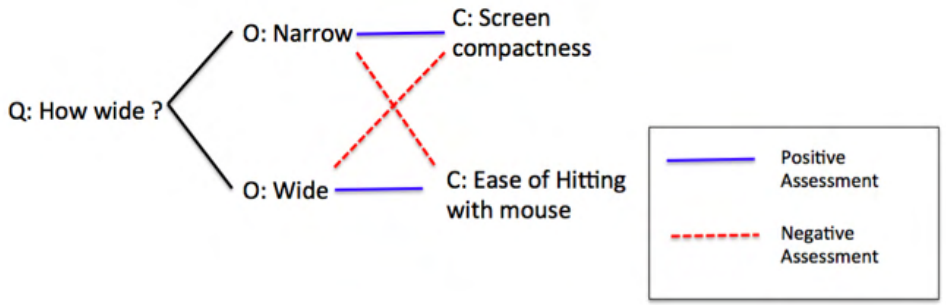


Argument diagramming – for design reasoning

- Reasoning about e.g. system or software design
  - Make choices explicit, documentation of the design
  - Correlation between rationalization & better designs



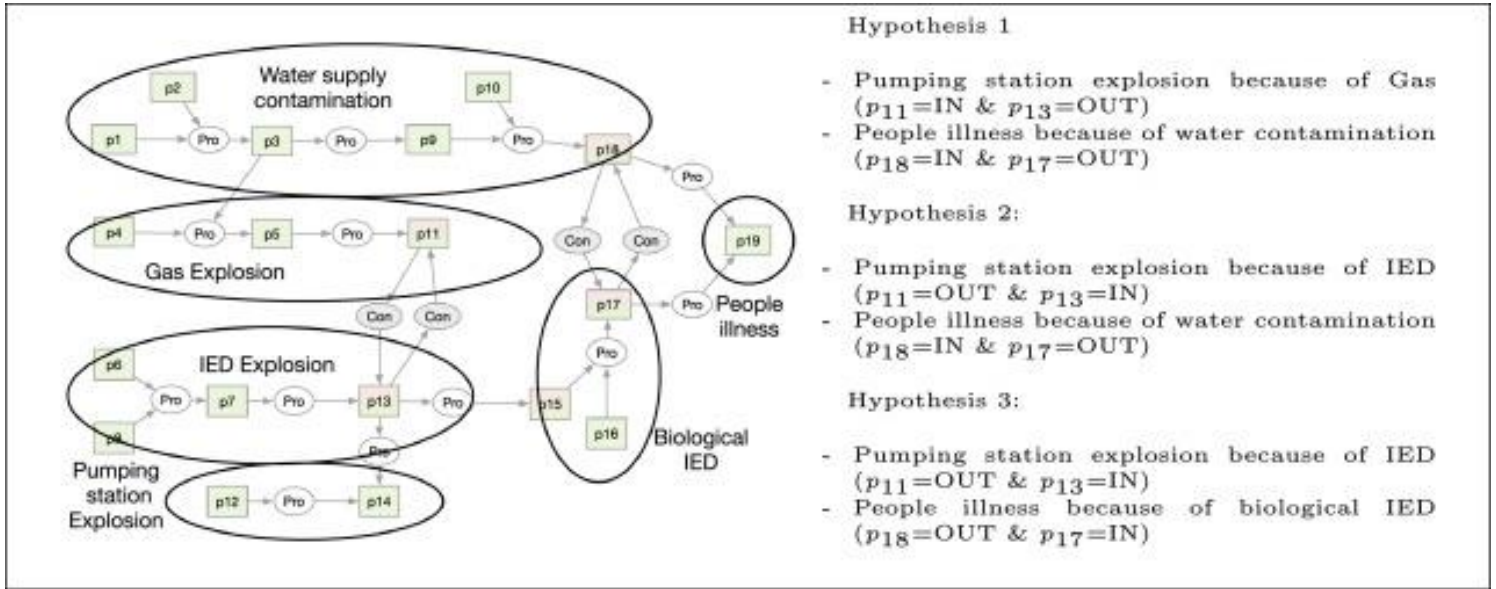




Argument diagramming – for collaborative sensemaking

Argument diagramming – for intelligence

- Mapping out different hypotheses, pro and counterarguments
  - Asking critical questions
- Can improve intelligence reports



Toniolo, A., Cerutti, F., Norman, T. J., Oren, N., Allen, J. A., Srivastava, M., & Sullivan, P. (2023). Human-machine collaboration in intelligence analysis: An expert evaluation. *Intelligent Systems with Applications*, *17*, 200151.

Kruger, A., Thorburn, L., & van Gelder, T. (2022). Using argument mapping to improve clarity and rigour in written intelligence products. *Intelligence and National Security*, *37*(5), 607-626.

Applications of argument diagramming

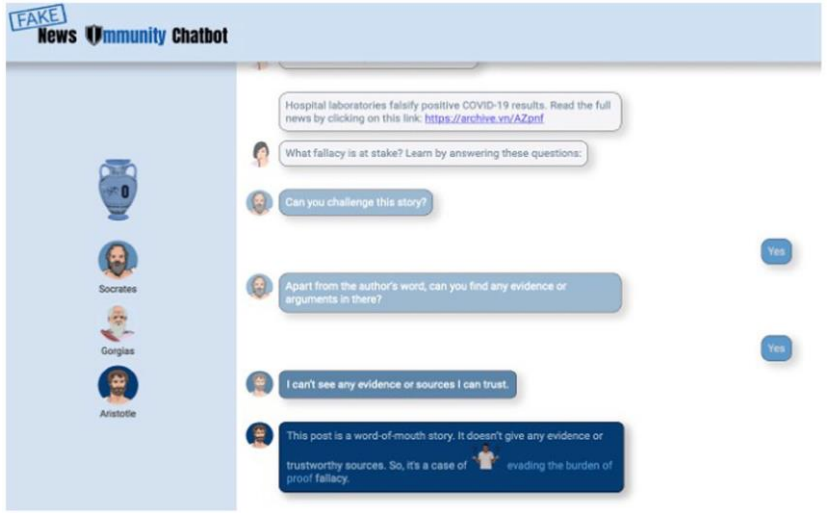
- Visualizing and mapping arguments, scenarios, hypotheses can help
  - Structuring, providing overviews
  - Mind maps etc.

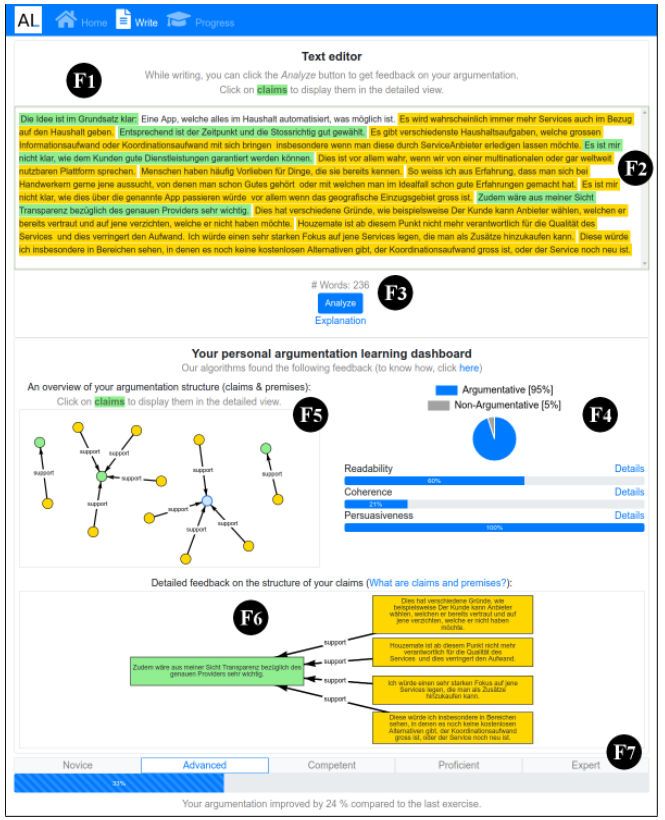
- Being constrained to diagrams does not help

- Manual work by the user
  - Is this “computational argumentation”?

Argumentative (dialogue) support systems

- For training argumentation skills
- For discussing fake news



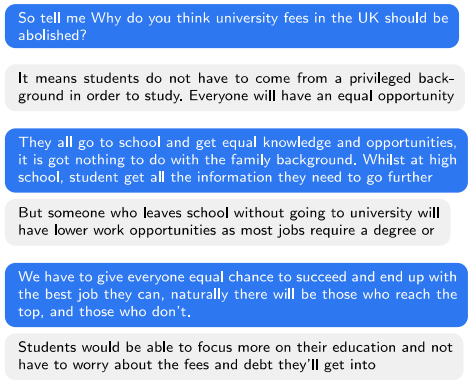


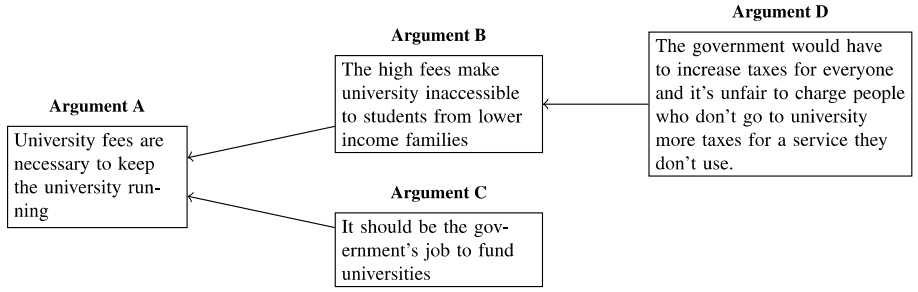
Musi, E., Carmi, E., O'Halloran, K., & Reed,C. (2023) "Developing misinformation immunity: how to reason-check fallacious news in a human computer interaction environment", *Social Media & Society, 9 (1).*

Wambsganss, Thiemo, et al. "AL: An adaptive learning support system for argumentation skills." *Proceedings of the 2020 CHI conference on human factors in computing systems*. 2020.

Argumentation & automated reasoning

- Argumentative “expert systems” performing automated reasoning
  - Medical, legal, systems
  - Persuasion, inquiry





Chalaguine, L. A., & Hunter, A. (2020). A persuasive chatbot using a crowd-sourced argument graph and concerns. In *Computational Models of Argument* (pp. 9-20). IOS Press.

Kakas, Antonis C., Moraitis, Pavlos, and Spanoudakis, Nikolaos I. ‘*GORGIAS*: Applying Argumentation’. 1 Jan. 2019 : 55 – 81.

Formal argumentation for citizen complaint/report intake

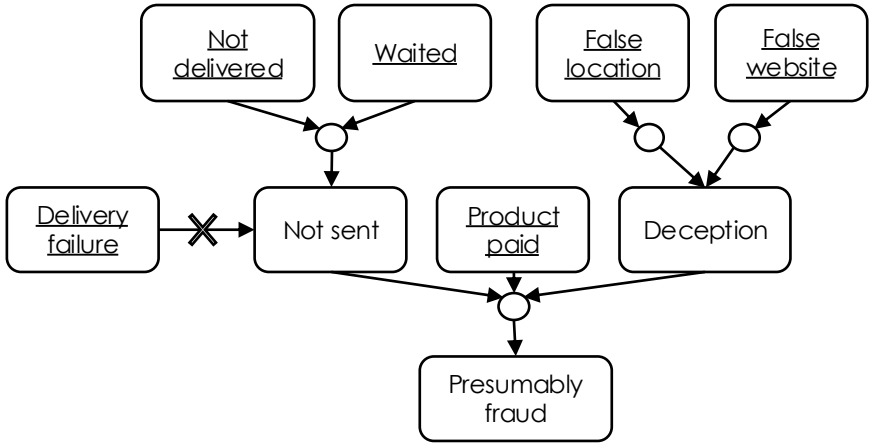
- Trade fraud: false webshops, malicious traders on Ebay
  - 60,000+ reports of alleged online fraud per year
  - Not all fraud: wrong product, not paid
  - Manually checked by case workers

- Automatically recommend to file report or not
  - Citizen fills in a form w. details & free text story
  - Possible fraud or not?

Schraagen, Testerink, Odekerken, Bex (2018) Argumentation-driven information extraction for online crime reports. *LeDAM 2018*

AI for intake – legal model

Legal model



Computational argumentation
ASPIC+ theory based on
Dutch Criminal Code & police policy rules

Schraagen, Bex (2019) Extraction of semantic relations in noisy user-generated law enforcement data, *IEEE Semantic Computing (ICSC)*.

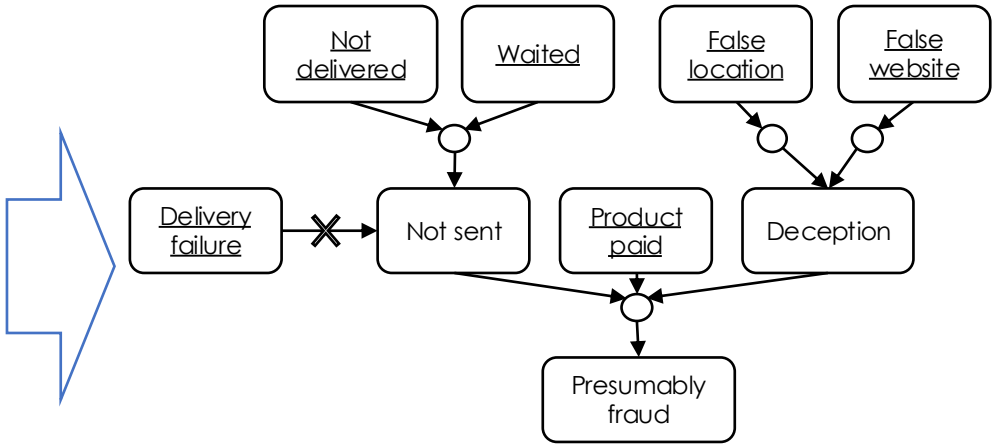
Schraagen, Brinkhuis & Bex (2017) Evaluation of Named Entity Recognition in Dutch Online Criminal Complaints. *DESI VII @ ICAIL 2017*

AI for intake – free text

Complaint form

Fictitious example report 1
I would like to report fraud. I recently saw a bicycle for sale on eBay and contacted the advertiser. He said he lived far away, so he would send me the bike. I paid him in good faith, but have still not received anything. I saw on Facebook he lives nearby.

Legal model



Computational argumentation
Rules w. exceptions based on DCC & police policy rules

D. Odekerken, F. Bex, A. Borg, B. Testerink (2022) Approximating Stability for Applied Argument-based Inquiry. *Intelligent Systems with Applications*.

AI for intake – combining data and knowledge

Extracting observations from complaint form

Paid

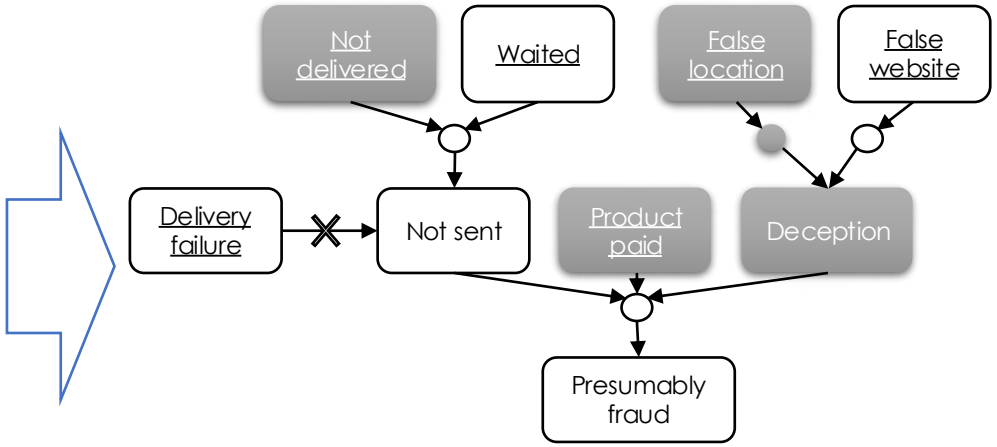
Fictitious example report 1
I would like to report fraud. I recently saw a bicycle for sale on eBay and contacted the advertiser. He said he lived far away, so he would send me the bike. I paid him in good faith, but have still not received anything. I saw on Facebook he lives nearby.

False location

Not delivered

Basic information extraction

Inferring possible fraud (or not)



Computational argumentation
Rules w. exceptions based on DCC & police policy rules

D. Odekerken, F. Bex, A. Borg, B. Testerink (2022) Approximating Stability for Applied Argument-based Inquiry. *Intelligent Systems with Applications*.

AI for intake – asking the right questions

Extracting observations from complaint form

Paid

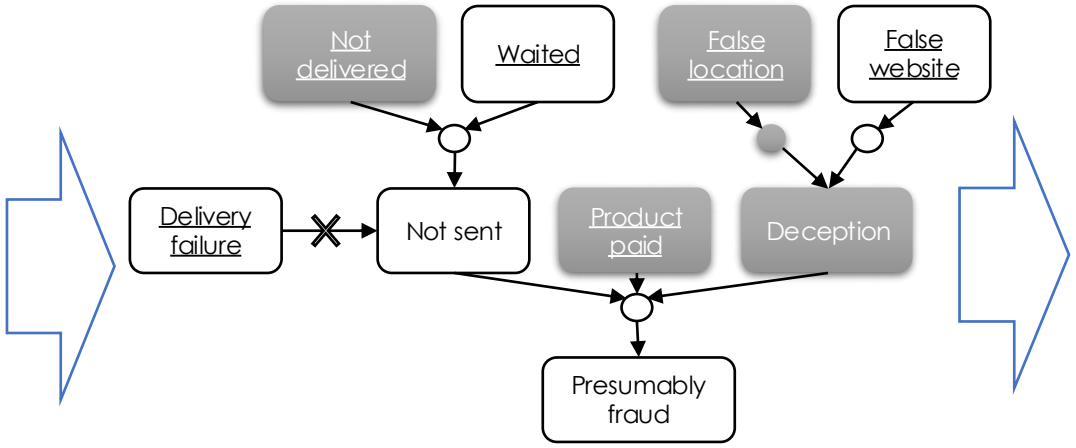
Fictitious example report 1
I would like to report fraud. I recently saw a bicycle for sale on eBay and contacted the advertiser. He said he lived far away, so he would send me the bike. I paid him in good faith, but have still not received anything. I saw on Facebook he lives nearby.

False location

Not delivered

Basic information extraction

Inferring possible fraud (or not)



Computational argumentation
Rules w. exceptions based on DCC & police policy rules

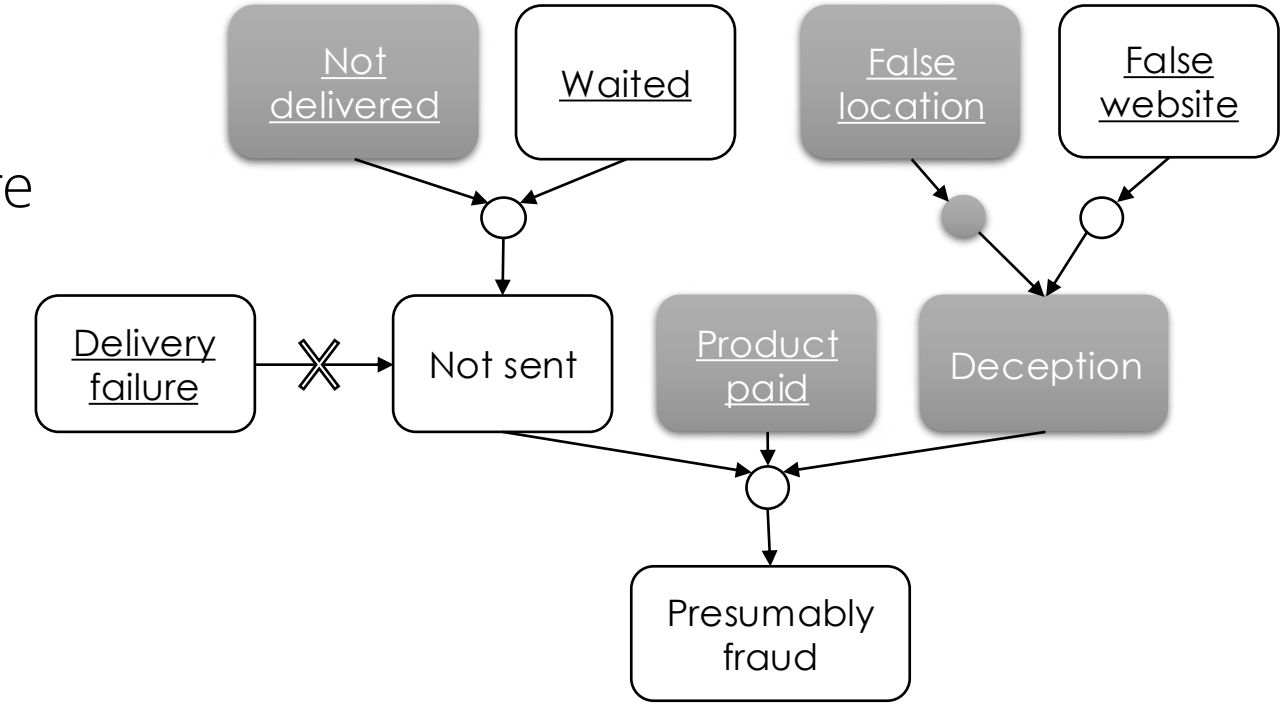
Asking for missing observations

Approximation algorithms
Can new info still change the conclusion (and if so which)?

D. Odekerken, F. Bex, A. Borg, B. Testerink (2022) Approximating Stability for Applied Argument-based Inquiry. *Intelligent Systems with Applications*.

Asking for observations – queryables

- ASPIC+ extended with queryables
  - Elements of $\mathcal{K}$ that are uncertain
- Question: which queryables, if observed, would still change the conclusion?



D. Odekerken, F. Bex, A. Borg, B. Testerink (2022) Approximating Stability for Applied Argument-based Inquiry. *Intelligent Systems with Applications*.

Stability

Future setup of AS: Every setup AS' s.t. $\mathcal{K}(AS) \subseteq \mathcal{K}(AS')$

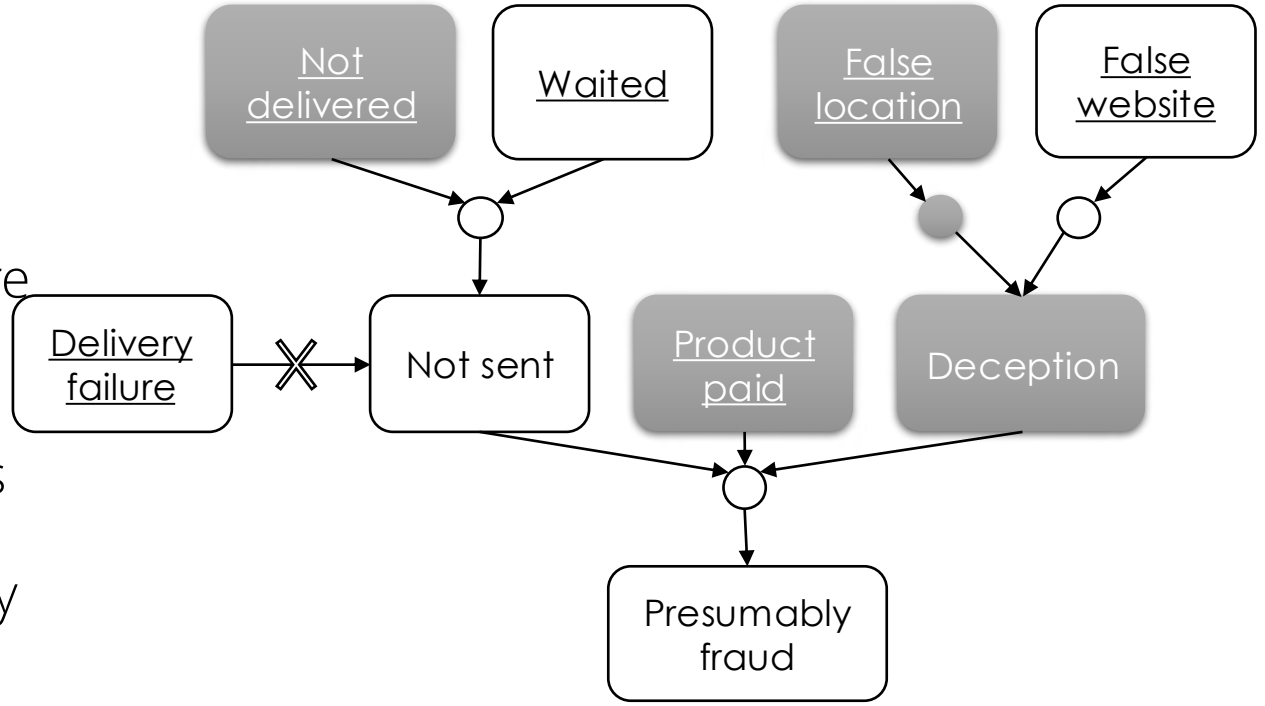
- Stable-Unsatisfiable: No argument for topic τ in any AS'.
- Stable-Defended: In every AS', there is an argument for τ in the grounded extension
- Stable-Out: There is an argument for τ, but in every AS', all arguments for τ are attacked by an argument in the grounded extension.
- Stable-Blocked: There is an argument for τ, but not in the grounded extension and in every AS', it is attacked by an argument that is not in the grounded extension.

D. Odekerken, F. Bex, A. Borg, B. Testerink (2022) Approximating Stability for Applied Argument-based Inquiry. *Intelligent Systems with Applications*.

Asking for observations – queryables

- Stable-Defended
  - Deception, $\mathcal{K}$
- Rest is not stable

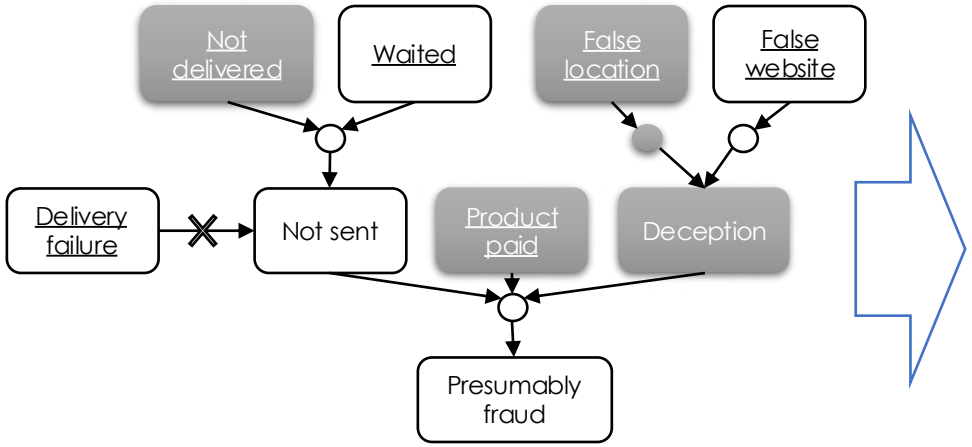
- Calculating all possible future extensions is expensive, so approximate
  - If Delivery Failure, topic is Stable-Out
  - If Waited and not Delivery failure, topic is Stable-In



A. Borg & F. Bex (2021) Explaining Arguments at the Dutch National Police. *Explainable AI for Law (XAILA)*.

AI for intake – recommendation & explanation

Inferring possible fraud (or not)



Response
Thank you for your complaint. In your case, the system has concluded that it is not a case of fraud, since you did not wait for at least 5 days. We recommend you do not file an official report at this point.

Computational argumentation
Rules w. exceptions based on DCC & police policy rules

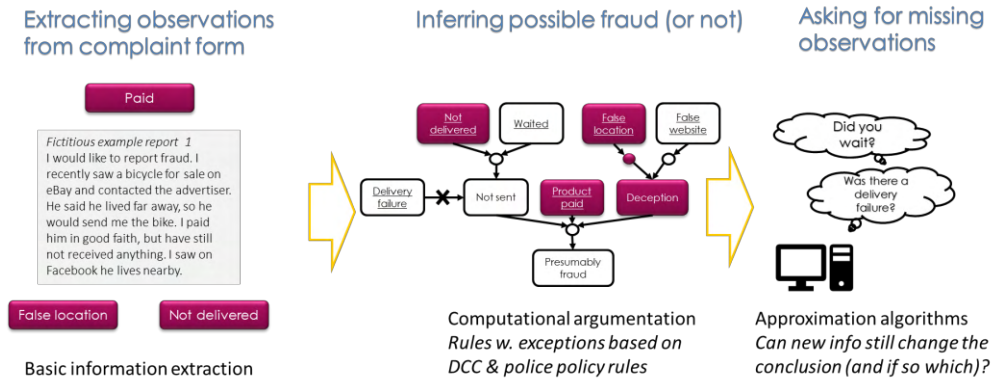
Explanations
Explaining (non-)acceptance in terms of arguments and counterarguments

E. Nieuwenhuizen, A. Meijer, F. Bex, S. Grimmelikhuijsen Explanations increase citizen trust in police algorithmic recommender systems: Findings from two experimental tests. *Under Review*

AI for intake – evaluation with citizens

- Do citizens trust the system with and without an explanation?
  - Controlled experiment 1700+ participants
  - Recommendation: do not file report
  - Explanation: the webshop is whitelisted
- Do users follow the recommendation (trusting behaviour)?
  - Without explanation (*computer says no*), 40-60% followed recommendation, trusted the system
  - With explanation, 65-80% followed recommendation, trusted the system

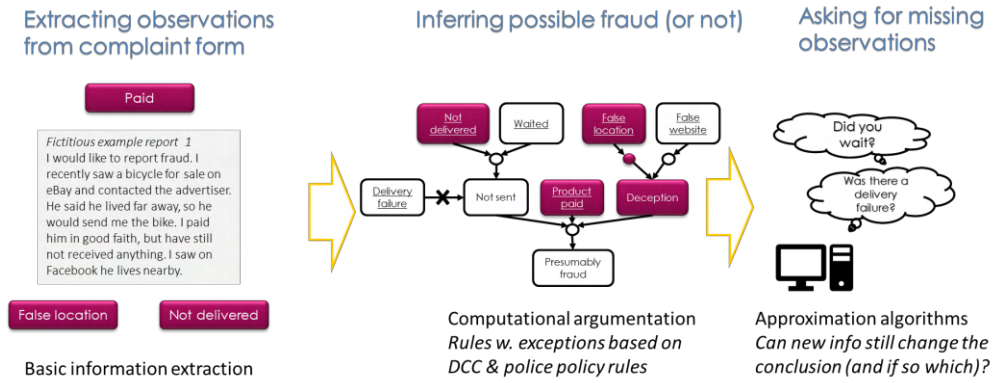
AI for intake – evaluation with case workers



- Observe case workers at the Dutch Police
  - Before the system: manually go through the process
  - With the system:
    - Still assess each submitted case
    - Observation extraction and question asking is done by system
      - Case worker gets form + list of observations
    - Conclusion/recommendation given to case worker without explanation

Carlos Soares, Stephan Grimmelikhuijsen, and Albert Meijer. 2023. Screen-level bureaucrats in the age of algorithms. An ethnographic study of algorithmically supported public service workers in the Netherlands Police. *Information Polity*

AI for intake – evaluation with case workers



- Case workers were helped by automated extraction & question asking
  - "The process of assessing an online fraud report is automated on the front side of the process."
  - "I think that I spend around five percent of the time assessing an online fraud report compared to [before the system]."
- Conclusion fraud or not ignored
  - No explanation – professional opinion & discretion
    - "If it says [fraud], it tells me nothing, I cannot say, 'sure [this is fraud]'. I still want to read it all."

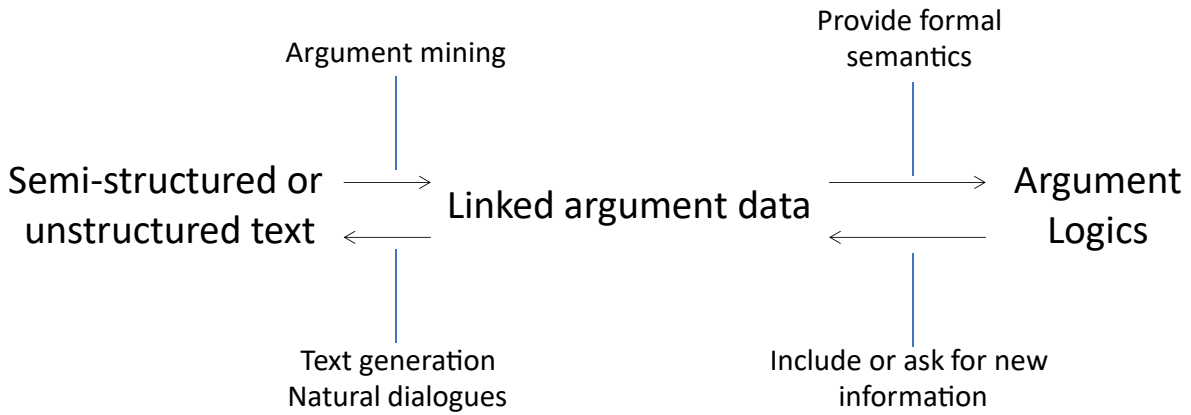
Carlos Soares, Stephan Grimmelikhuijsen, and Albert Meijer. 2023. Screen-level bureaucrats in the age of algorithms. An ethnographic study of algorithmically supported public service workers in the Netherlands Police. *Information Polity*

Computational argumentation, automated reasoning & applications

- Argumentation technologies involving ontologies, logic, are very much “GOFAI”
  - Work for bounded, well defined domains
    - Knowledge acquisition & modelling
  - Do what you expect (or prove) them to do, are relatively easy to explain
  - Can be computationally expensive
  - Need NLP to interact with real users, natural language text

Applications of computational argumentation – the future

- Integration with neural (language) models
  - Allows reasoning with natural language
  - Can help with computational complexity
- Formal models can be used to check for correctness & consistency, to explain



D. Craandijk & F. Bex (2020) Deep Learning for Abstract Argumentation Semantics. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI 2020)*

Y. Guo, T. Yu, L. Bai, J. Tang, Y. Ruan and Y. Zhou, "Argumentative Explanation for Deep Learning: A Survey," *2023 IEEE International Conference on Unmanned Systems (ICUS)*

An application to help the court

- The Dutch courts are being overwhelmed by appeals to traffic fines, often written by ChatGPT. They want you to design a system for them that makes the process more efficient by (semi-)automatically handling cases

Citizens submit an appeal to the court

Violation of art. 54 Traffic law – exceeding the speed limit by 20 km/h.
"I agree that I drove too fast on the motorway. However, I have a good reason, because my wife was in labour and we had to get to the hospital as soon as possible. We arrived just in time."
Decision by prosecution 27-7-2023; Appeal submitted on 2-8-2023; fine paid.

Violation of art. 31 Traffic law – not stopping for a red light.
"I did not see the red light, because it was hidden behind the leaves of a tree. I've attached a photograph showing that the traffic light is obscured."
Decision by prosecution 13-4-2023; Appeal submitted on 18-4-2023; fine paid.

Violation of art. 19 Traffic law – parking outside a designated parking zone.
"I have not paid the fine because I am in debt and have no money."
Decision by prosecution 17-8-2023; Appeal submitted on 20-8-2023; fine not paid.

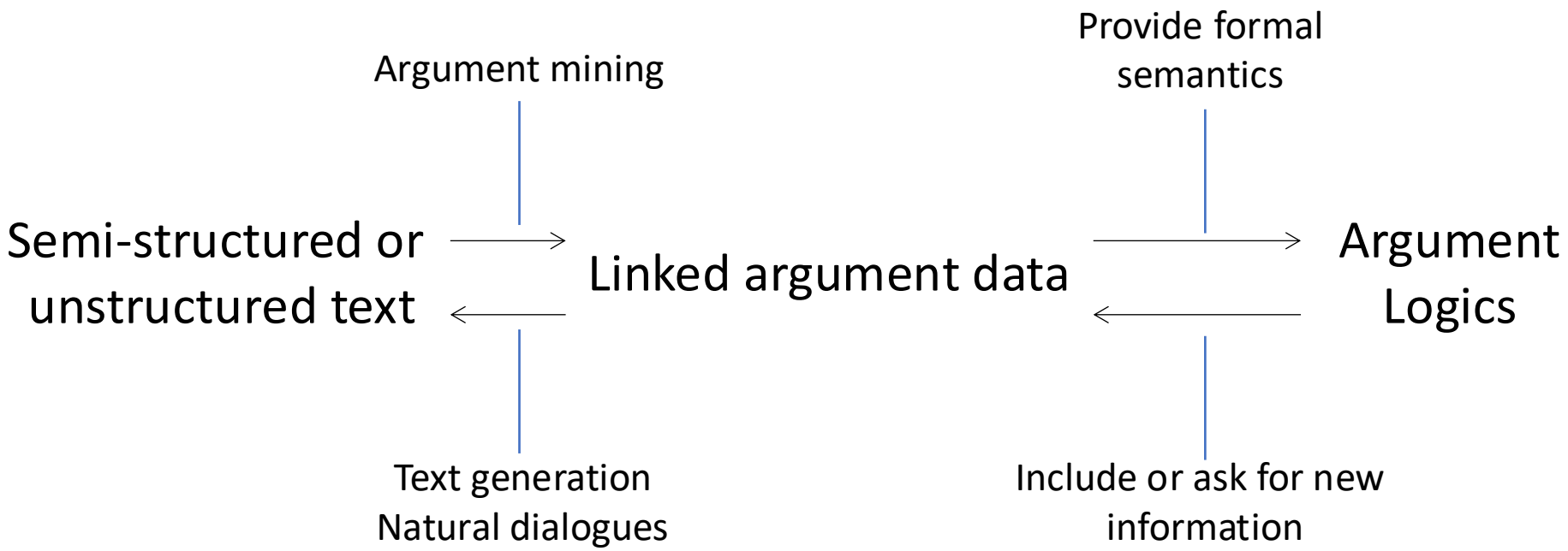
Violation of art. 20 Traffic law – endangering traffic by parking on the road.
"The prosecution argues I parked in the road. However, I was parked on the side of the road with two wheels on pavement and the traffic could easily pass my car, so I was in violation of art 19 Traffic law, for which the fine is significantly lower"
Decision by prosecution 8-2-2023; Appeal submitted on 12-2-2023; fine paid.

Paralegals decide on the appeals

- Once the appeal comes in, the paralegal reads it and makes a decision: approve – reject – change – inadmissible
  - Has the appeal been submitted within 4 weeks? Has the fine been paid?
    - If not, the appeal is inadmissible.
  - If the appeal is admissible, it depends on the motivation.
    - Different article of traffic law applies: change
    - Situation was such that the appellant could not know they broke the law: approve
    - Otherwise: reject
  - Note that paralegals always have discretionary authority to decide differently in a case, for example if it very unreasonable for the person to have to pay the fine.

Design an application to help the court

- An argumentation system that receives as input the appeals.
- Describe in 1 slide what the system does, and what you need to build it
  - Rules? Arguments? Language processing & generation?


Argument mining
Provide formal semantics
Semi-structured or unstructured text
Linked argument data
Argument Logics
Text generation
Natural dialogues
Include or ask for new information

Conclusion

- Designing AI for practice is really difficult!

- Combine knowledge & data
  - Use new techniques without forgetting the old ones
- Engage with stakeholders from practice
  - Is AI really the solution (and does it matter whether it is AI?)
- Combine different disciplines
  - Building and evaluating AI for law from different perspectives