Play and Persuade: An Interactive Exploration of Computational Argumentation

Elfia Bezou-Vrakatseli & Madeleine Waller

October 19th, 2024

27th European Conference on Artificial Intelligence
Santiago de Compostela, Spain

About us

SAFE & TRUSTED AI

UKRI CENTRE FOR DOCTORAL TRAINING

KING'S College LONDON

Elfia

Madeleine

Online Handbook of Argumentation for AI

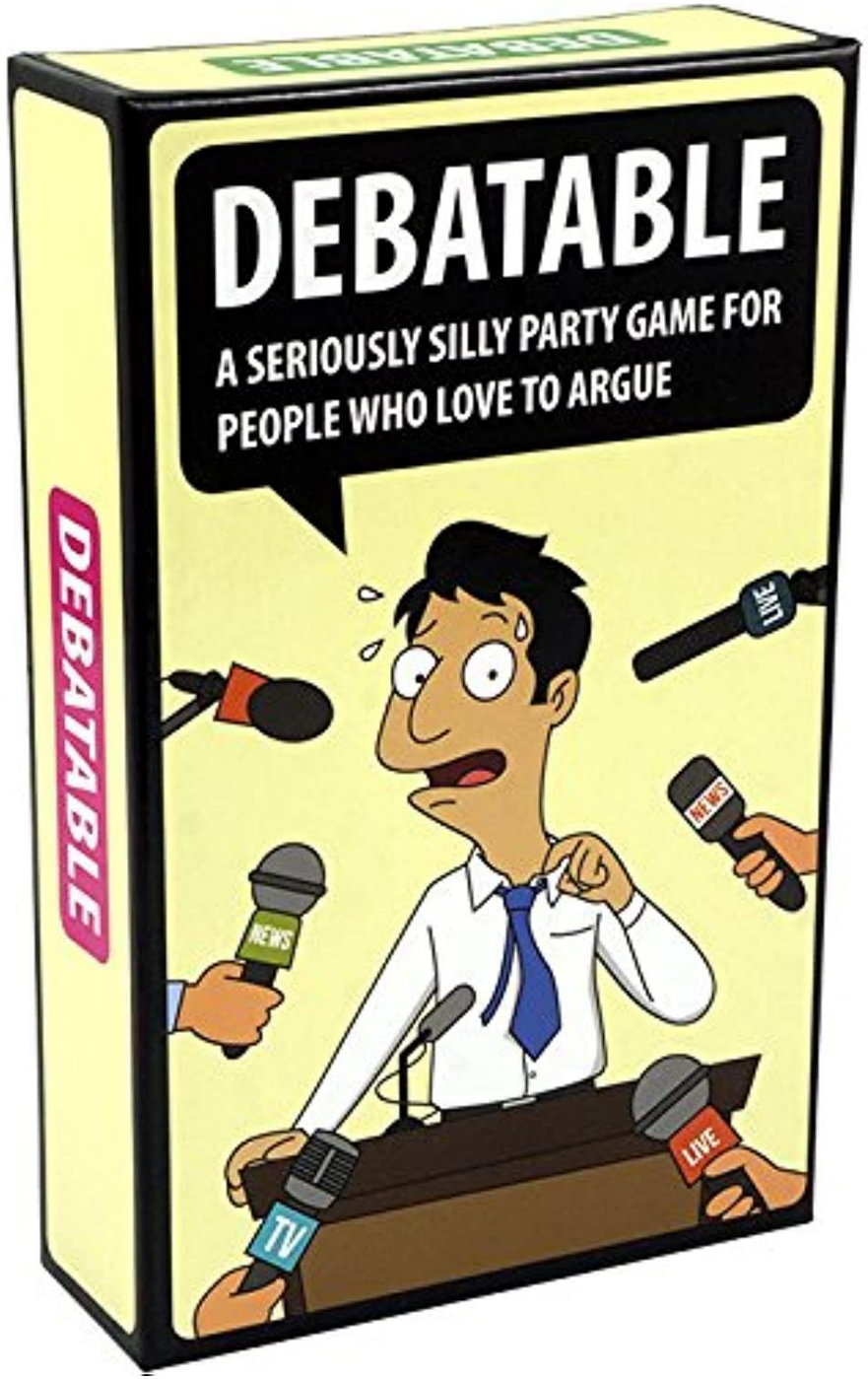
Thanks to Andreas Xydis, Daphne Odekerken, Ameer Saadat-Yazdi for contributions to the slides!

Overview

Introduction to argumentation
Argumentation theory
Abstract argumentation
BREAK (10:30)
Variations of abstract argumentation
Structured argumentation
Argument mining
Applications of argumentation
Discussion and closing

What is argumentation?

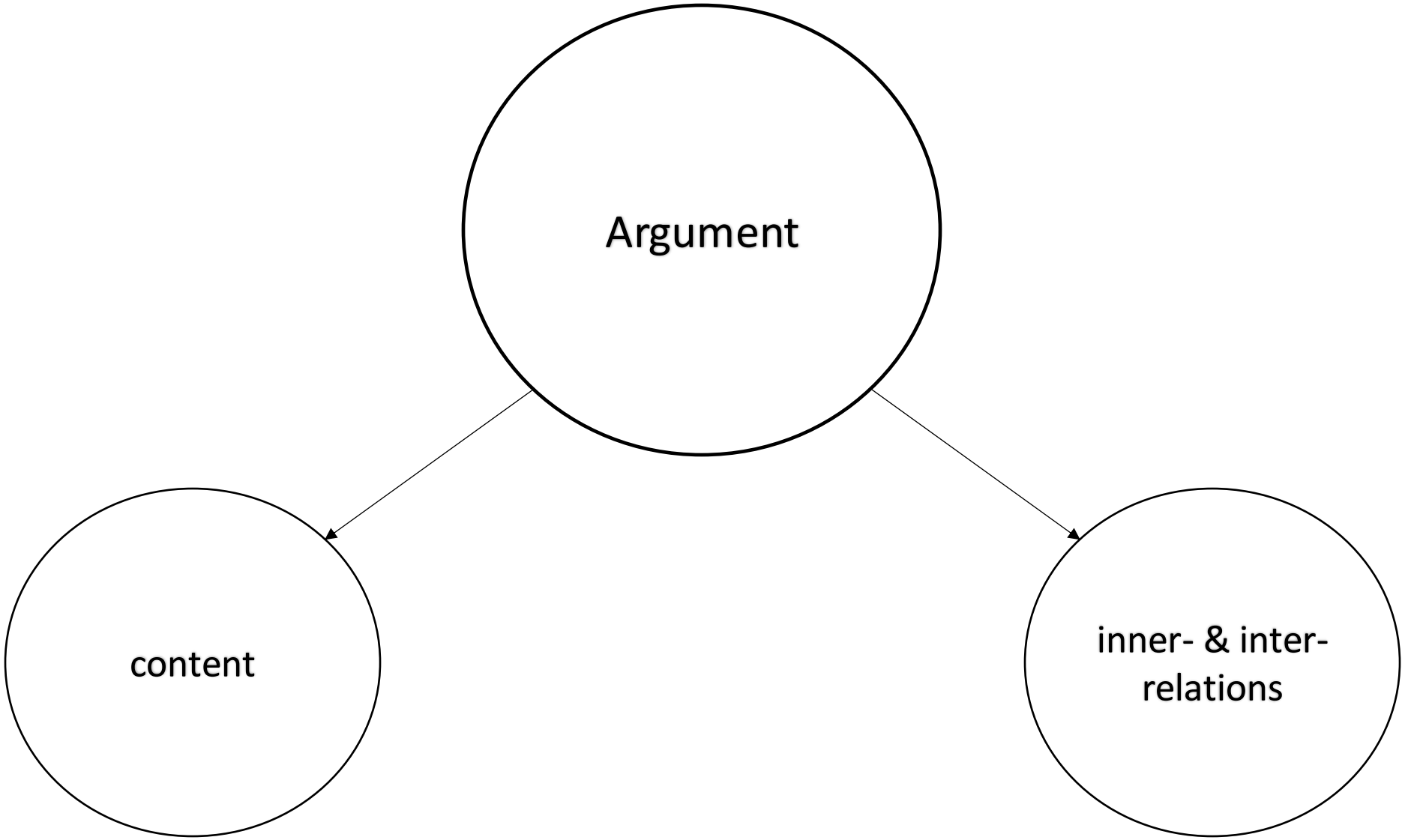
https://www.menti.com/alrmh46gdq27

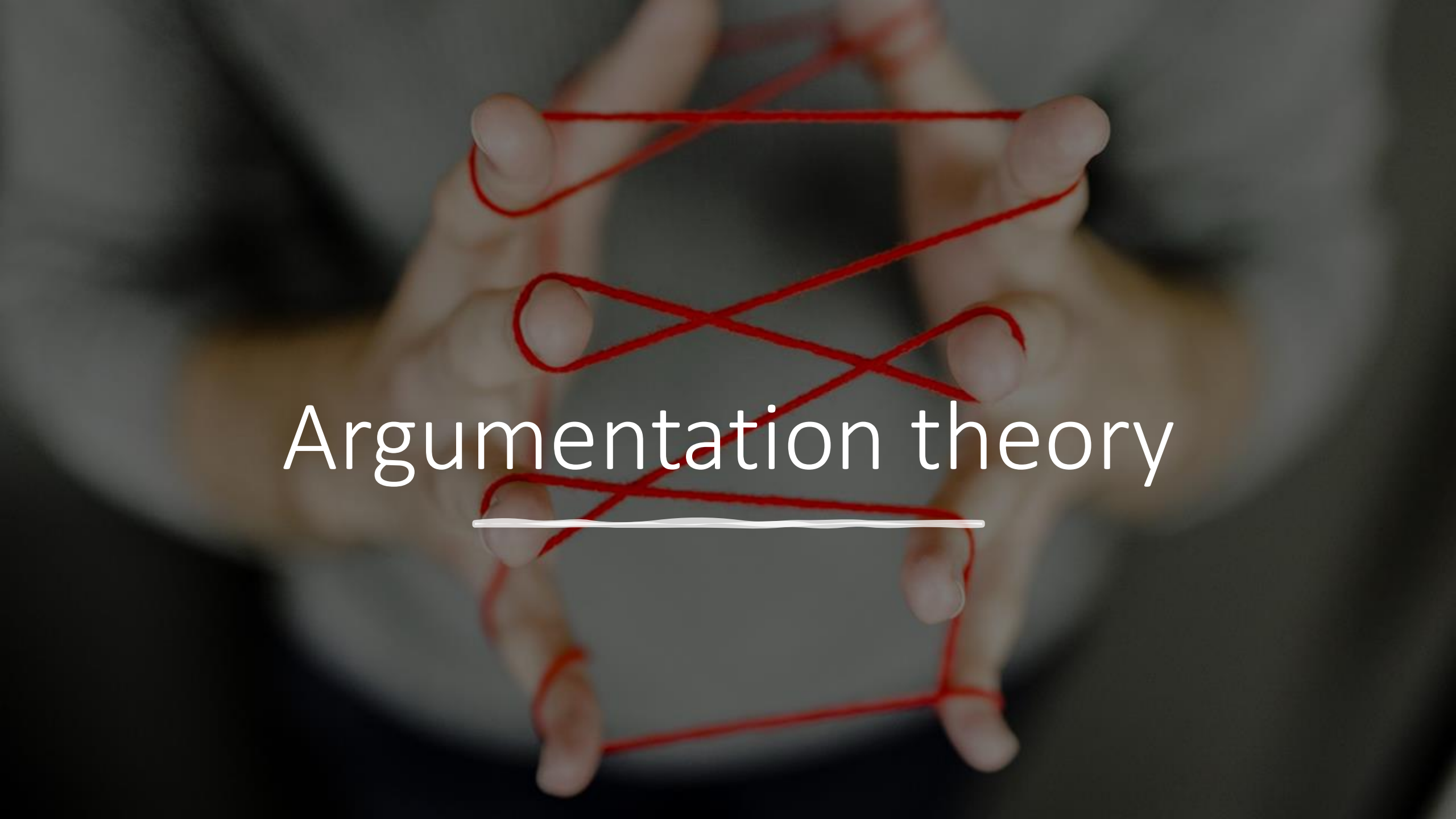


Let's play!

- Get into groups
- Alphabetically, by name:
  1. YES
  2. MODERATOR
  3. NO

1. Moderator:
   - Chairs debate
   - Decides winning side
2. First round:
   - Spin the wheel for a topic
   - 2 min to debate it
3. Second round:
   - Switch roles
   - 2min **per side**

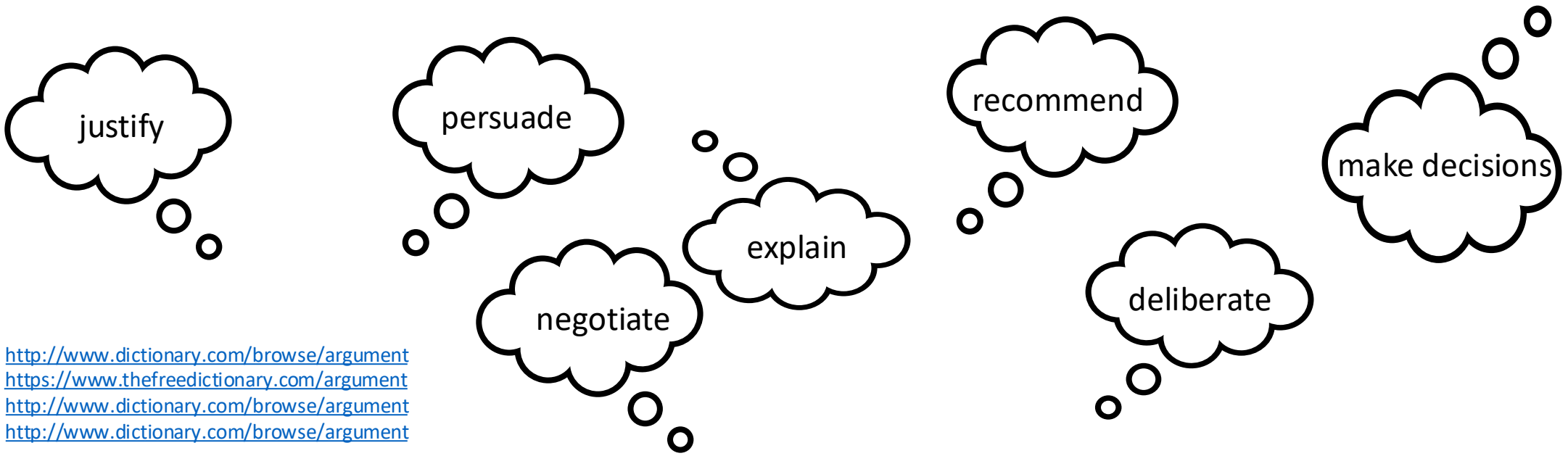

Argument
content
inner- & inter-
relations



Argumentation theory

What is an argument? Why do we argue?

- "A statement, reason, or fact for or against a point" [1]
- "A course of reasoning aimed at demonstrating truth or falsehood" [2]
- "A discussion involving differing points of view" [3]
- "An address or composition intended to convince or persuade" [4]



1. http://www.dictionary.com/browse/argument
2. https://www.thefreedictionary.com/argument
3. http://www.dictionary.com/browse/argument
4. http://www.dictionary.com/browse/argument

Internal Reasoning

- Information processing
- Reasoning about beliefs, goals, intentions

I will go left because it is the fastest route.

p: Left is the fastest route.
c: I will go left.
r: I want to take the fastest route.

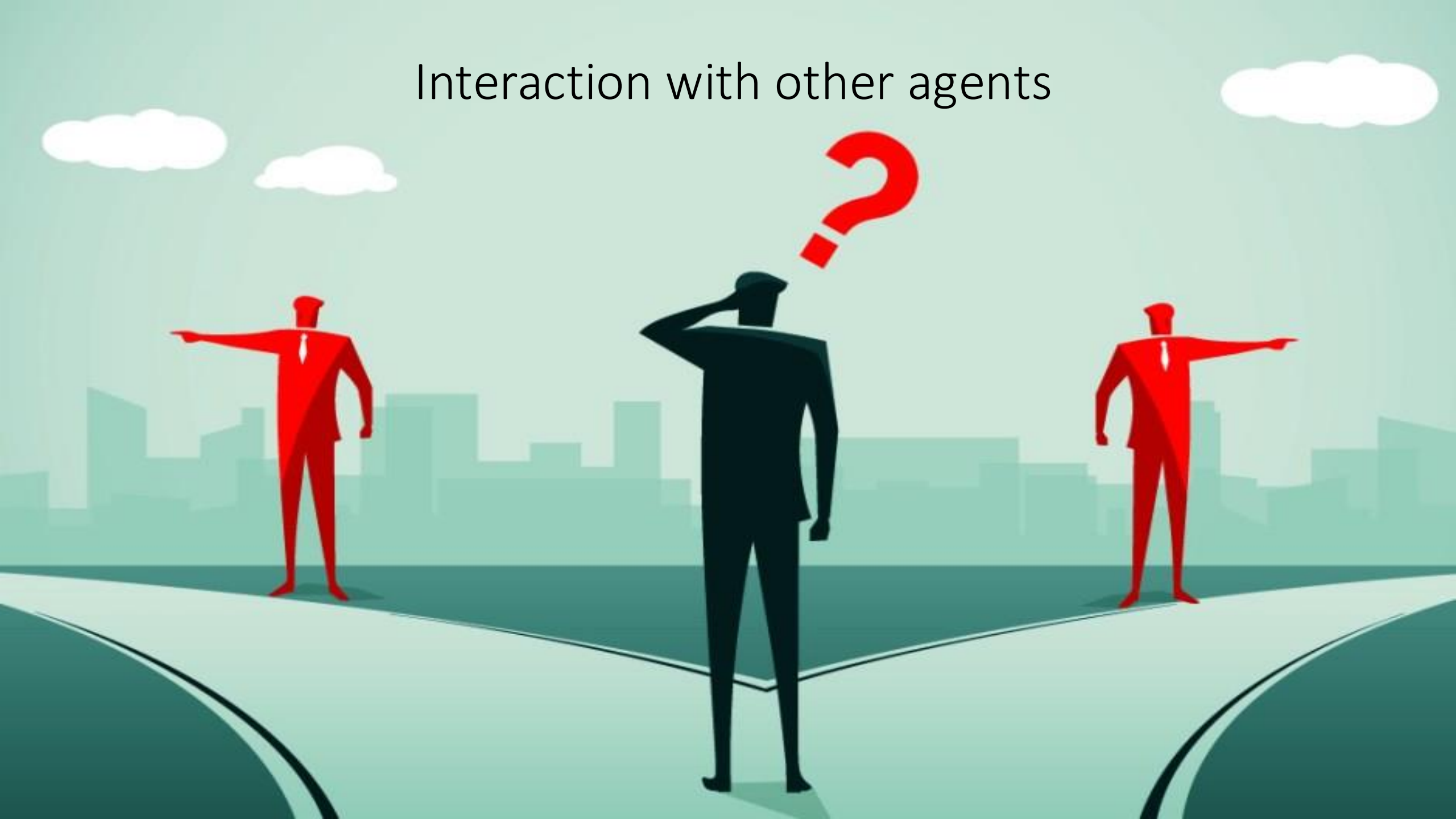
Commonsense reasoning: defeasible

- Inconsistent information
- Knowledge often uncertain or incomplete:
  - conclusions under certain assumptions
  - retract conclusions once learn an assumption is unwarranted

→Non monotonic logic

Actually, today, I will go right because there is an obstacle on the left.

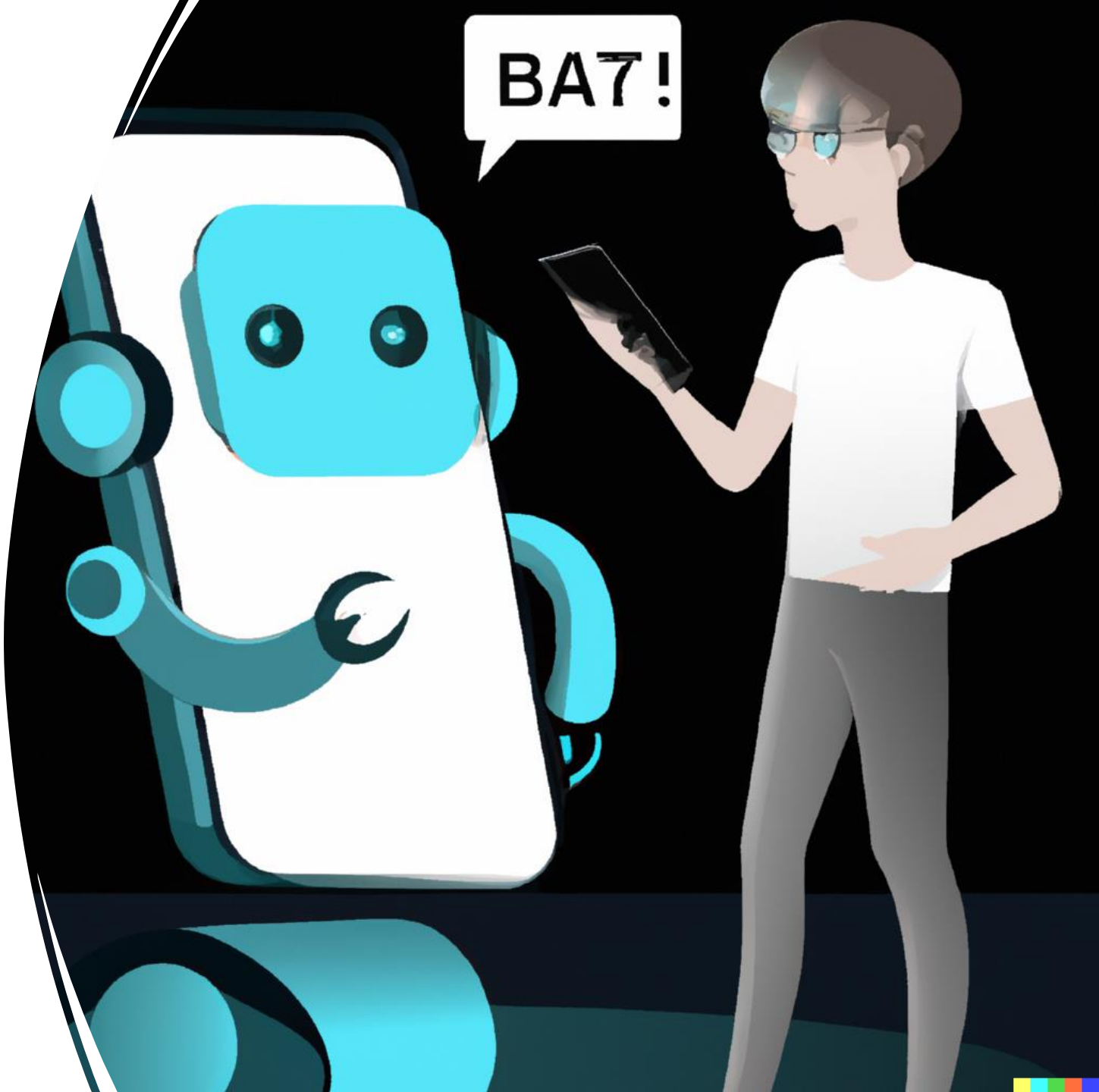
Interaction with other agents



Dialogue

- Tool of interaction & communication
- Enables understanding of both parties involved
  - Information
  - Reasoning exploration

☞ Joint reasoning

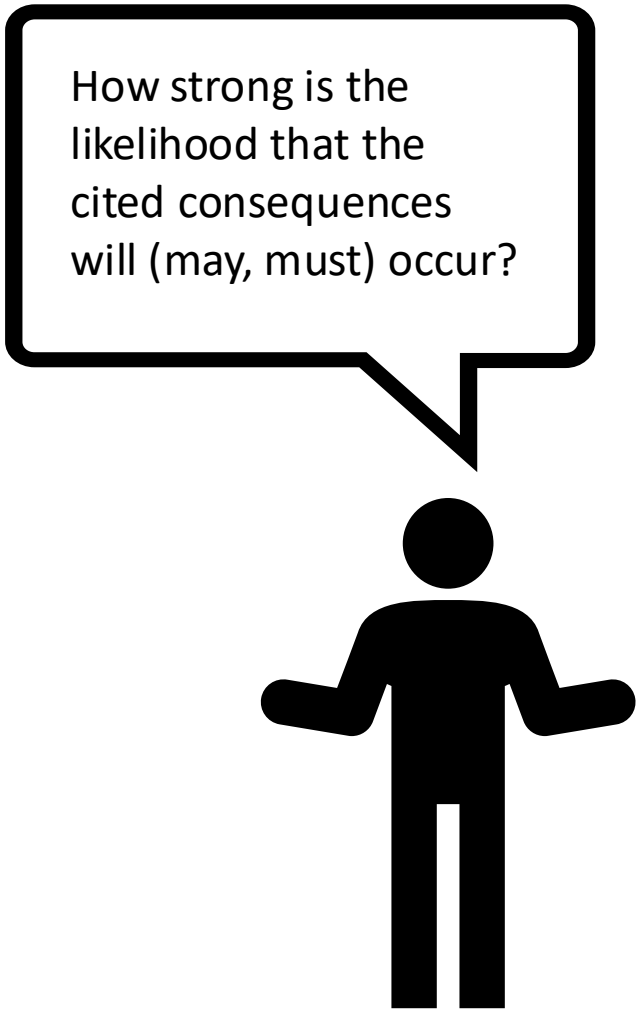


Safe & Trusted AI

- Humans & AI Systems
  - Interaction & Communication
  - Human-AI Dialogue
  - Joint Reasoning
- Argumentation
  - Real-world Reasoning
  - Justification for its claims
  - Explainability & Transparency in Decision Making

Argument Schemes: guiding dialogue

Critical Questions



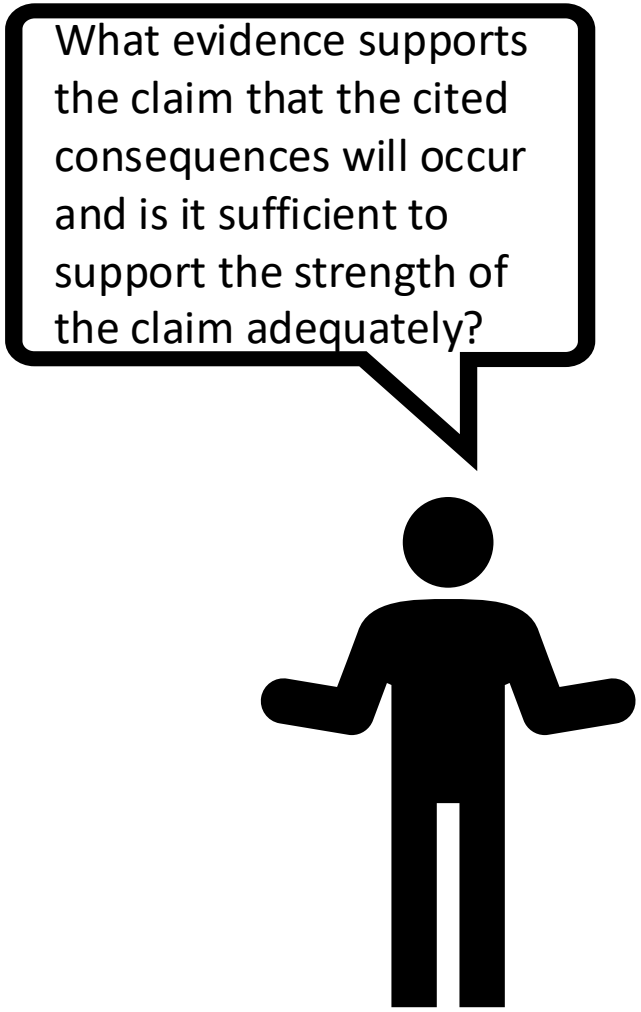
Walton's Argument Schemes

Access to legal abortion improves the health and safety of pregnant people so pregnant people should have the right to choose abortion.

<u>Argument from Positive Consequences</u>

- Premise : If *A* is brought about, good consequences will occur.
- Conclusion : Therefore *A* should be brought about.

Critical Questions



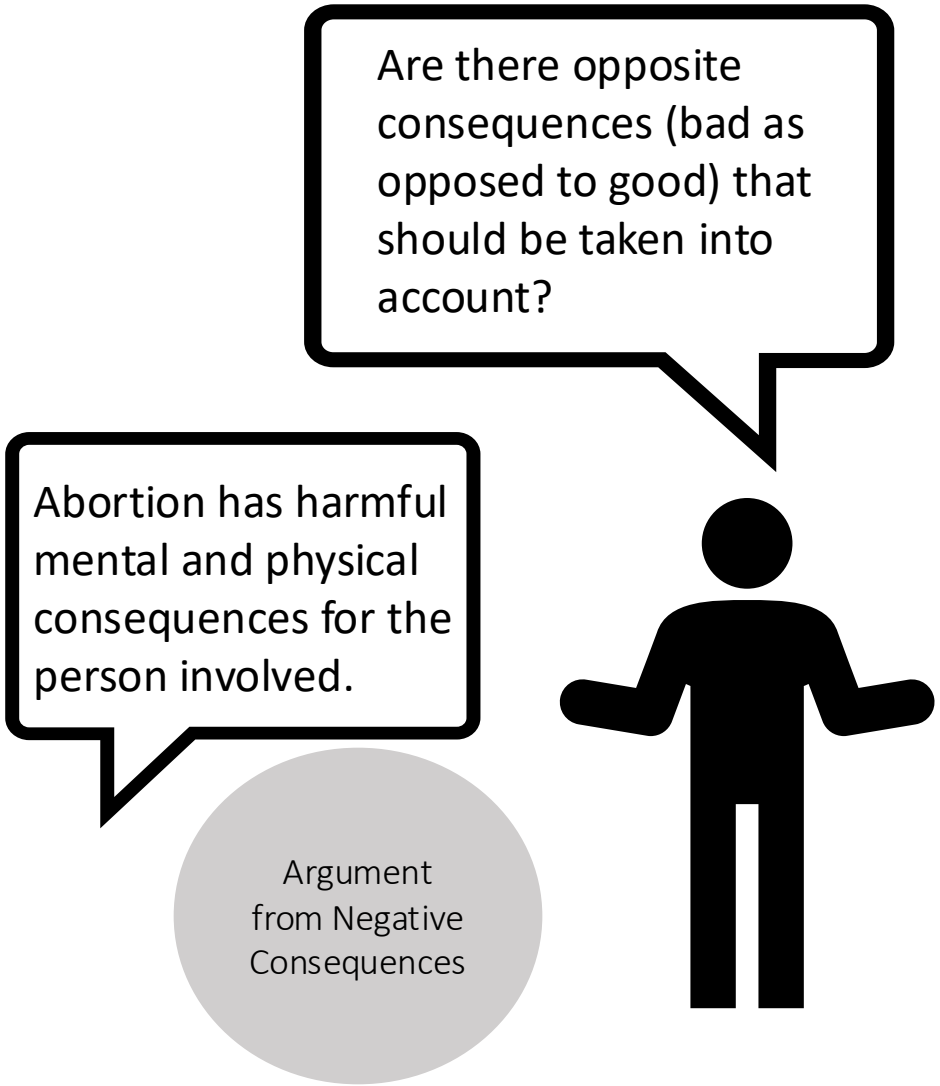
Walton's Argument Schemes

Access to legal abortion improves the health and safety of pregnant people so pregnant people should have the right to choose abortion.

<u>Argument from Positive Consequences</u>

- Premise : If *A* is brought about, good consequences will occur.
- Conclusion : Therefore *A* should be brought about.

Critical Questions

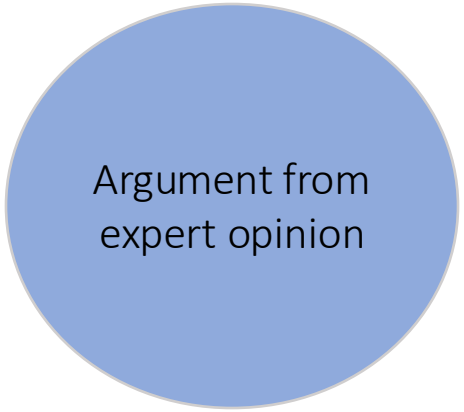


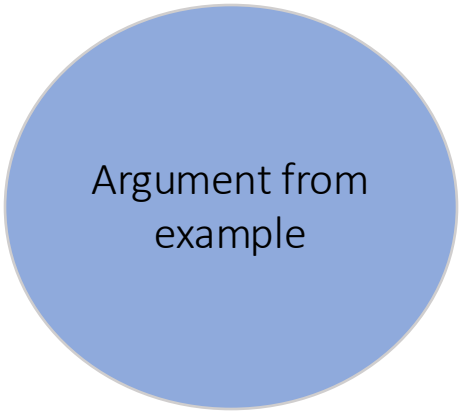
Walton's Argument Schemes

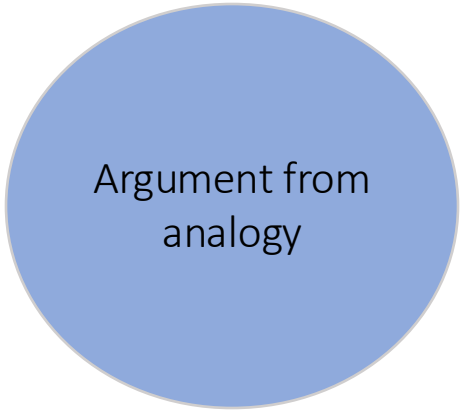
Access to legal abortion improves the health and safety of pregnant people so pregnant people should have the right to choose abortion.

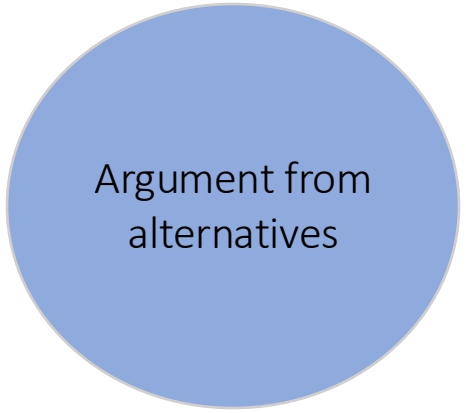
<u>Argument from Positive Consequences</u>

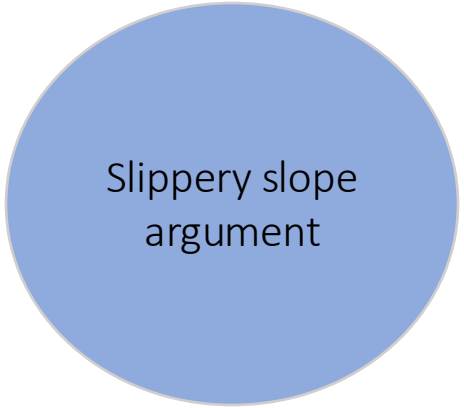
- Premise : If *A* is brought about, good consequences will occur.
- Conclusion : Therefore *A* should be brought about.


Argument from expert opinion

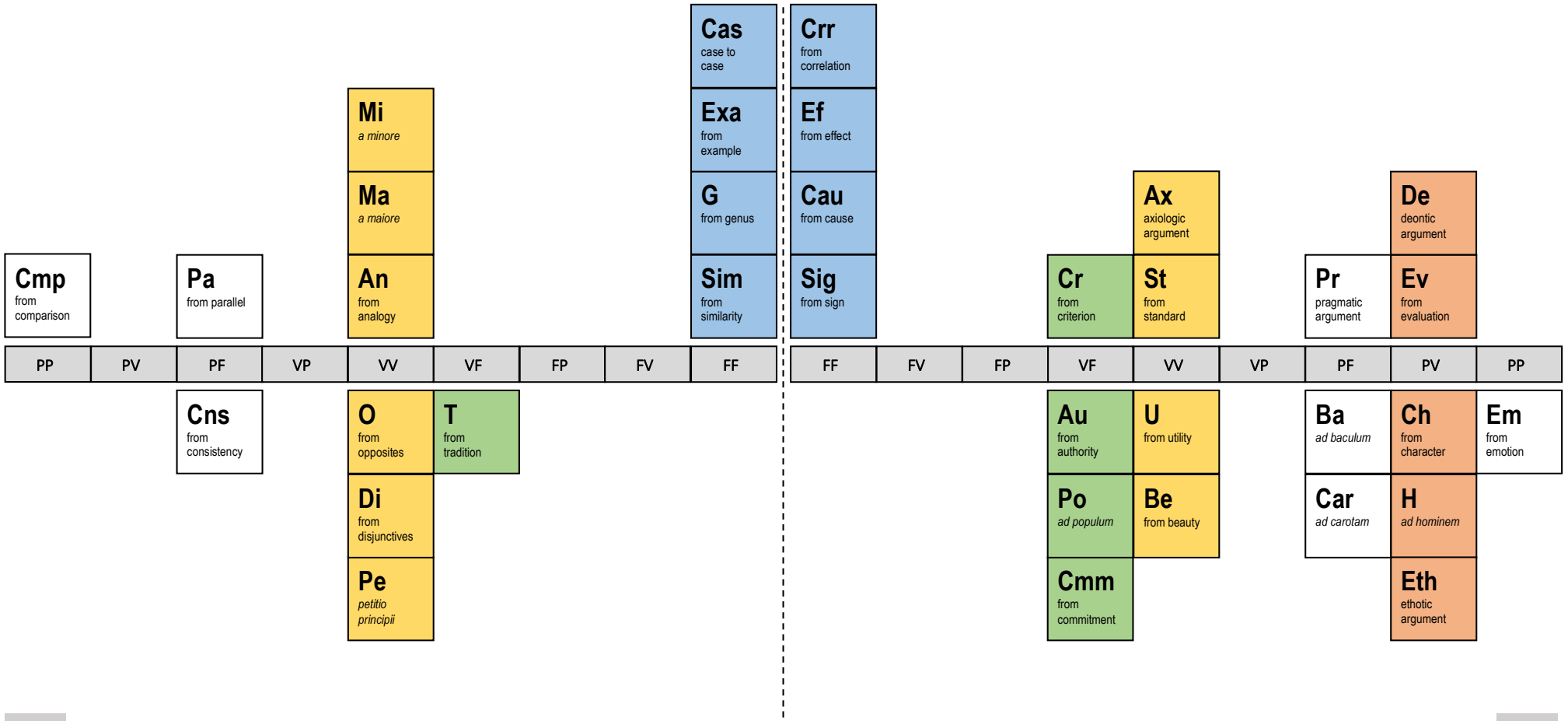

Argument from example


Argument from analogy


Argument from alternatives


Slippery slope argument

PERIODIC TABLE OF ARGUMENTS



VERSION 2.5 © 2020 J.H.M. WAGEMANS

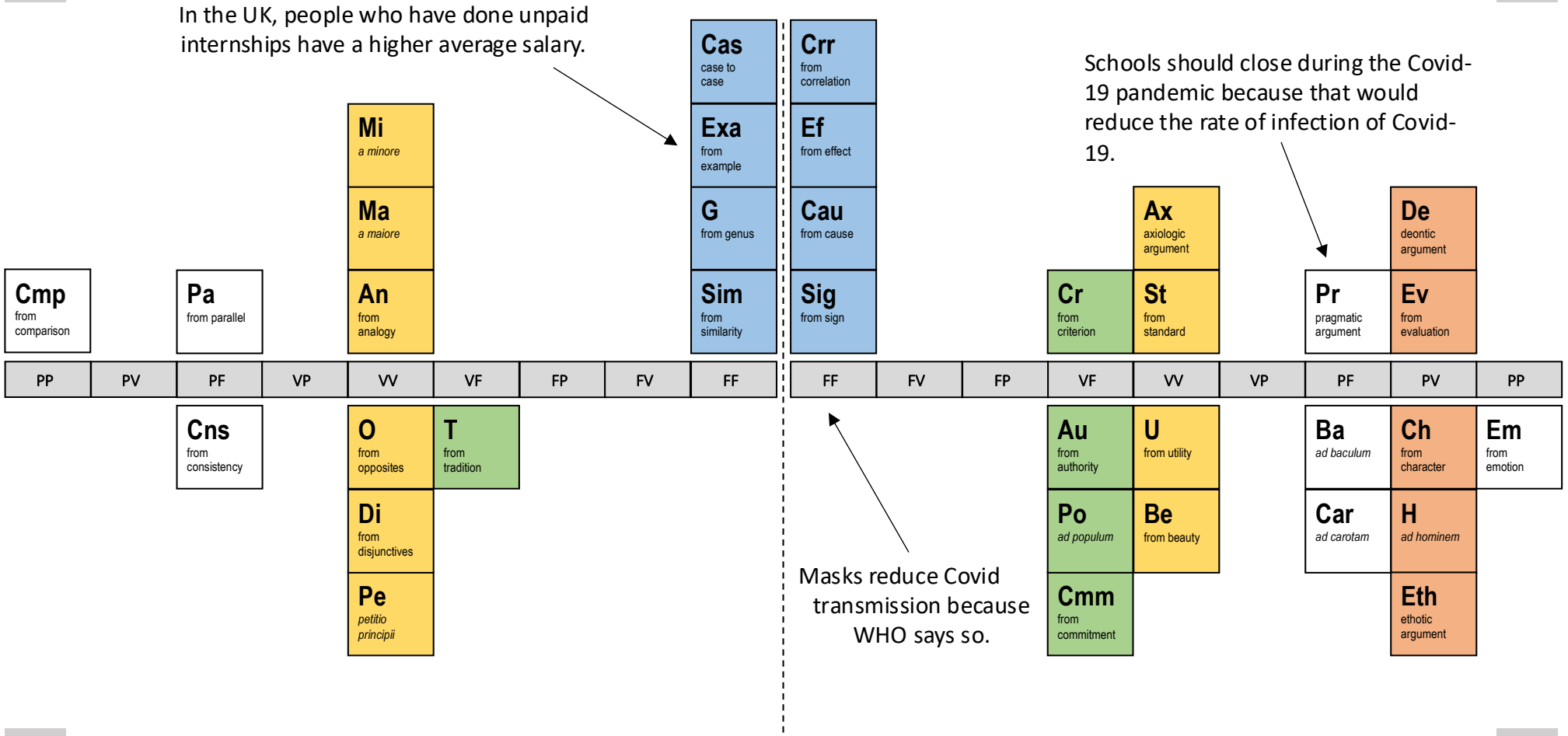
Wagemans' Periodic Table of Arguments

1. Predicate vs Subject: that of which something is said (S) + what is said about the subject (P)
   - Predicate: *Schools should close during the Covid-19 pandemic because that would reduce the rate of infection of Covid-19.*
   - Subject: *Not wearing a mask on the train is forbidden, because not wearing a mask on the bus is forbidden.*
2. First vs Second Order
   - 1st order: *Schools should close during the Covid-19 pandemic because that would reduce the rate of infection of Covid-19.*
   - 2nd order: *Masks reduce Covid transmission because WHO says so.*
3. Substance
   - Policy: *Schools should close during the Covid-19 pandemic.*
   - Value: *Remote learning is a good substitute for in-person education.*
   - Fact: *Closing schools would reduce the rate of infection of Covid-19.*

PERIODIC TABLE OF ARGUMENTS

β subject 1

predicate α 1



2 γ subject

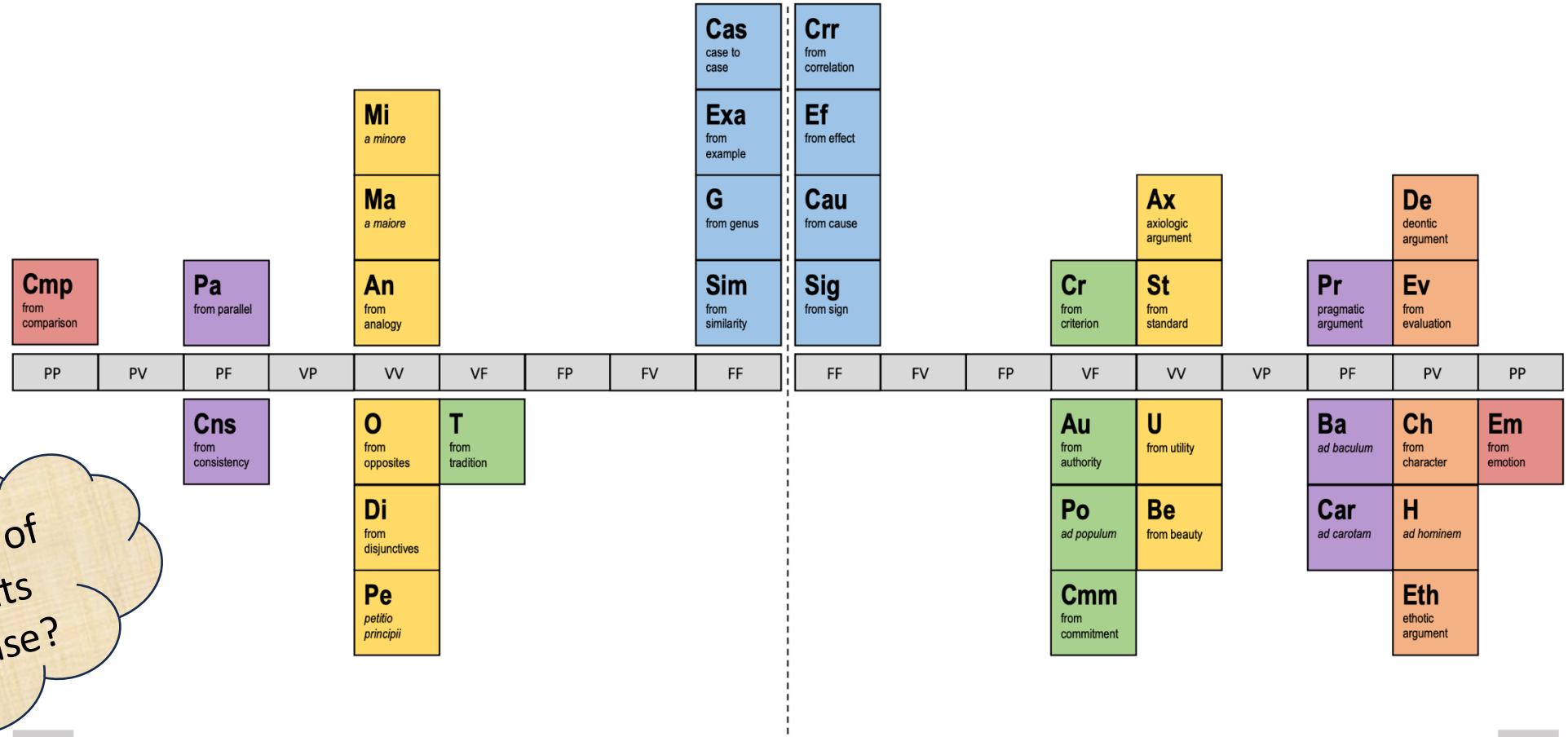
2 predicate δ

VERSION 2.5 © 2020 J.H.M. WAGEMANS

PERIODIC TABLE OF ARGUMENTS

β subject 1

predicate α 1



What type of arguments did you use?

2 γ subject

VERSION 2.5 © 2020 J.H.M. WAGEMANS

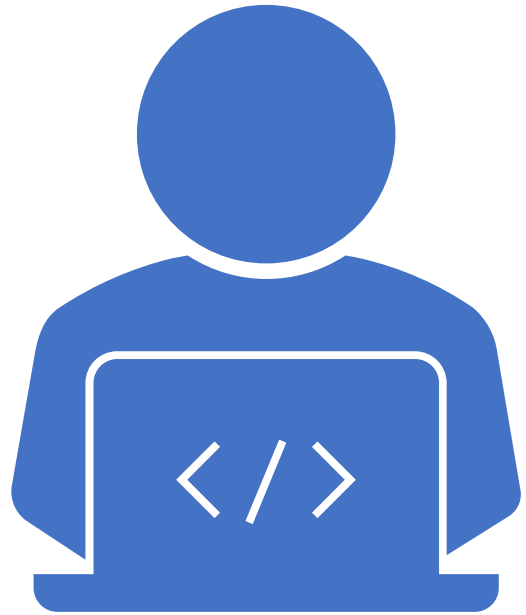
2 predicate δ

For more info please visit www.periodic-table-of-arguments.org

Formalising Argumentation

What is computational argumentation?

- Formalisation of argumentation theory
- Used to support human-computer connections and computer-computer connections
- Applications include:
  - providing reasoning and explaining decision-making
  - natural language processing and generation tasks



Abstract Argumentation

Disregards the internal structure of arguments and focusses on acceptability conditions that allow certain sets of arguments to co-exist in a rational manner based on a **given attack relationship between arguments**.

(P. M. Dung, 1995)

Should I go right or left?

Argument 1 (A1)

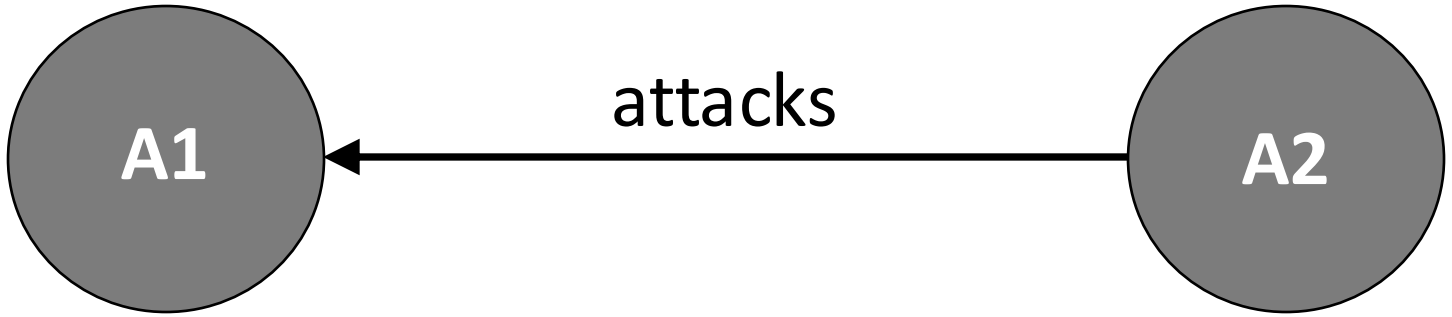
Going left is the fastest route,

therefore I should go left

Argument 2 (A2)

Today there is an obstacle to the left,

therefore I should go right



[1] Phan Minh Dung (1995). "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n–person games". *Artificial Intelligence*. **77** (2): 321–357.

Has social media been good for humanity?

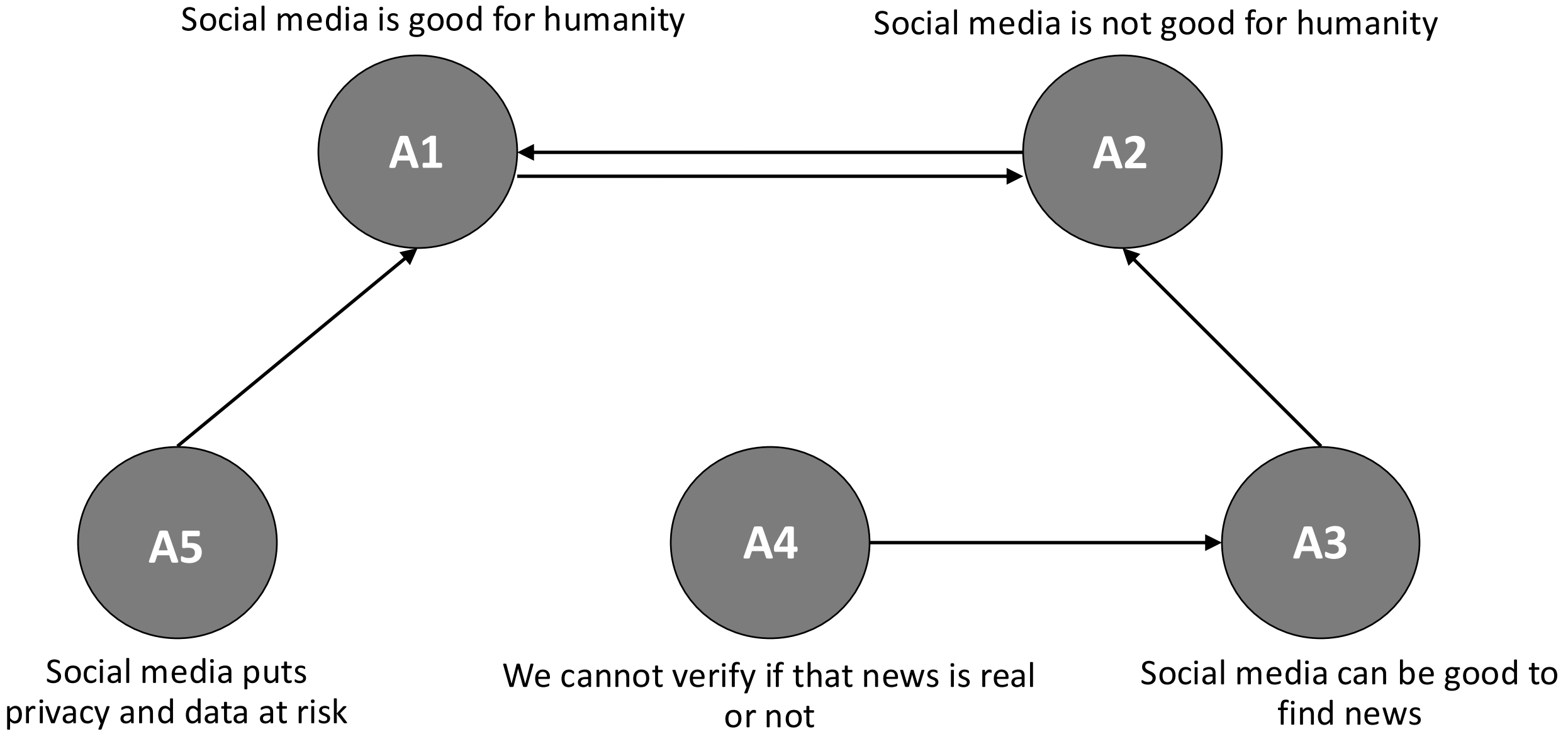
A1: Social media has been good for humanity

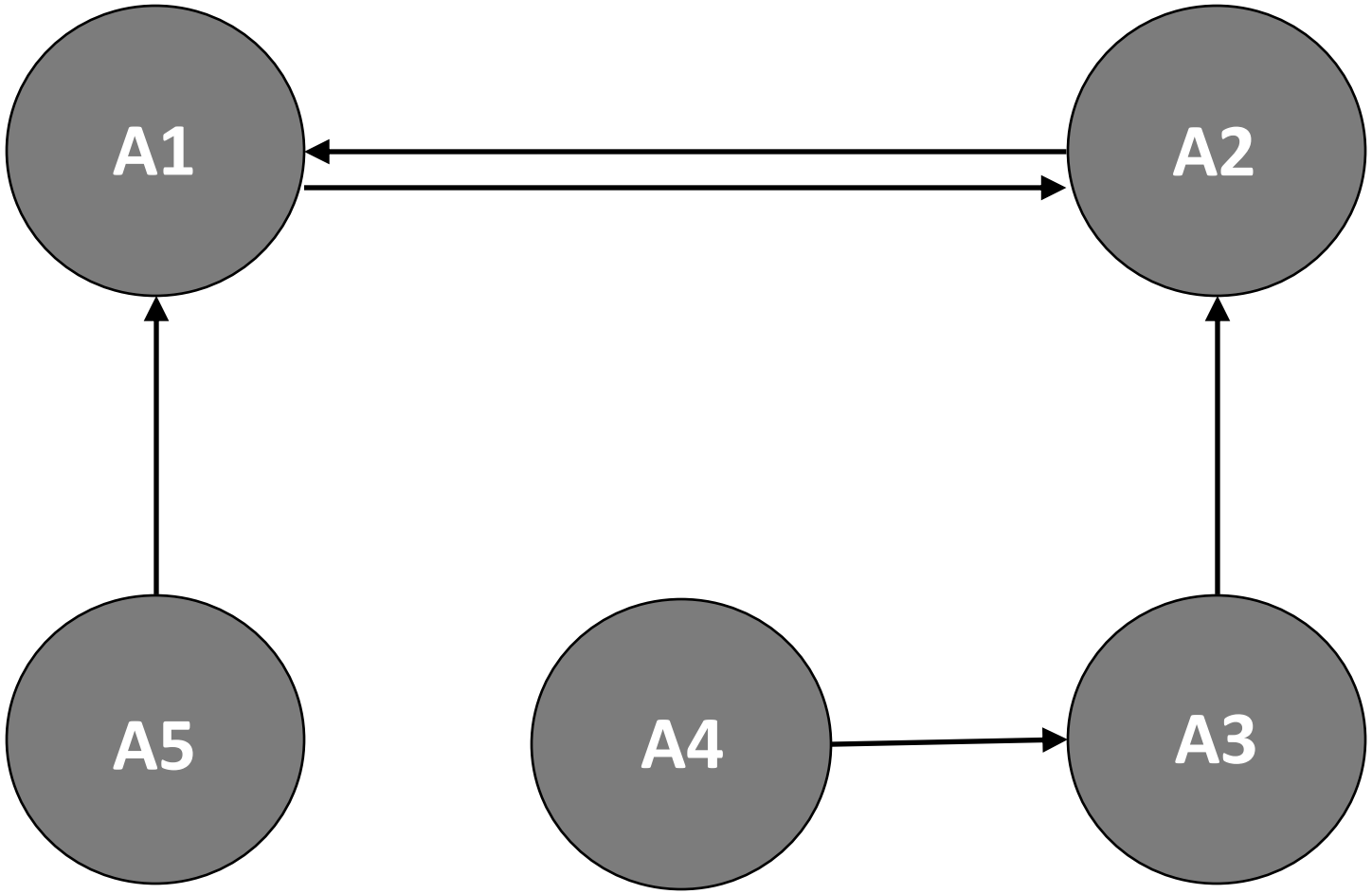
A2: Social media has not been good for humanity

A3: Social media can be good to find news

A4: We cannot verify if that news is real or not

A5: Social media puts privacy and data at risk


Social media is good for humanity
Social media is not good for humanity
A1
A2
A5
A4
A3
Social media puts privacy and data at risk
We cannot verify if that news is real or not
Social media can be good to find news


A1
A2
A5
A4
A3

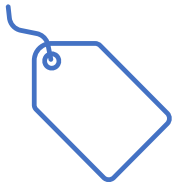
Create your own abstract argumentation graph

- Choose one of the debates you had before
- Extract the key arguments
- Think about how they attack each other
- Draw a graph of arguments and attacks

Abstract Argumentation

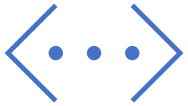
Disregards the internal structure of arguments and **focusses on acceptability conditions** that allow certain sets of arguments to co-exist in a rational manner based on a given attack relationship between arguments.

(P. M. Dung, 1995)



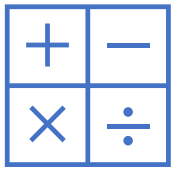
Label-based

Labellings of arguments in the graphs that have specific properties



Extension-based

Subsets of arguments in the graphs that have specific properties



Equations

Solution to a defined set of equations representing the interactions between arguments

Label-based semantics

IN if all its attackers are out (or no attackers)

OUT if it has an attacker that is in

UNDEC if not all its attackers are out
and it does not have an attacker that is in

IN if all its attackers are out

OUT if it has an attacker that is in

UNDEC if not all its attackers are out
and it does not have an attacker that is in

Argument 1 (A1)

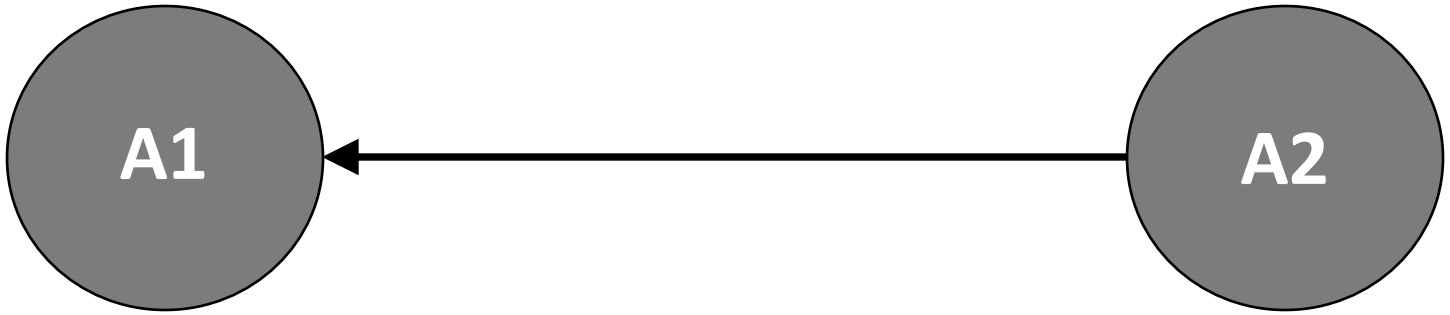
Going left is the fastest route,

therefore I should go left

Argument 2 (A2)

Today there is an obstacle to the left,

therefore I should go right



IN if all its attackers are out

OUT if it has an attacker that is in

UNDEC if not all its attackers are out
and it does not have an attacker that is in

Argument 1 (A1)

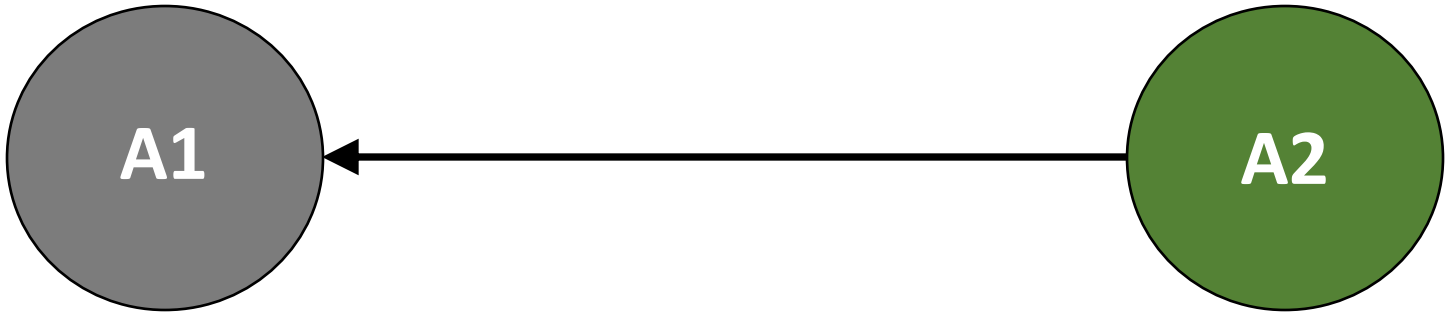
Going left is the fastest route,

therefore I should go left

Argument 2 (A2)

Today there is an obstacle to the left,

therefore I should go right



IN if all its attackers are out

OUT if it has an attacker that is in

UNDEC if not all its attackers are out
and it does not have an attacker that is in

Argument 1 (A1)

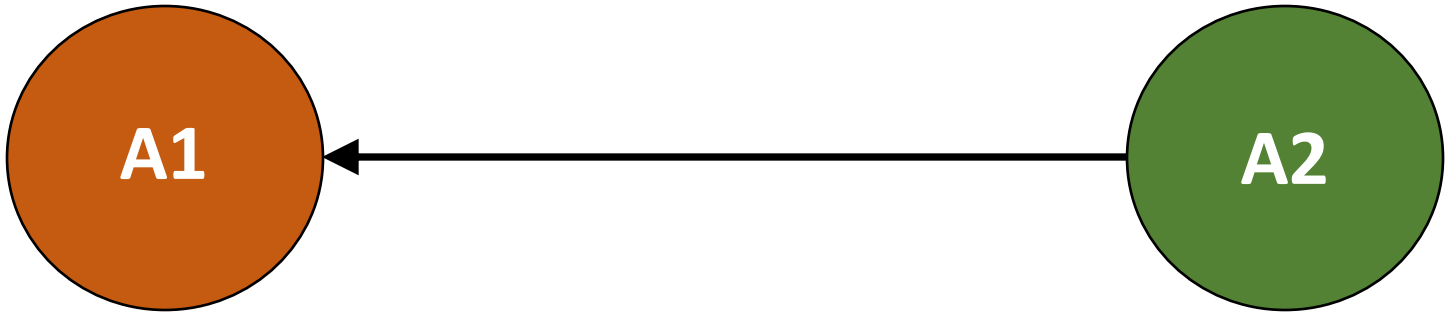
Going left is the fastest route,

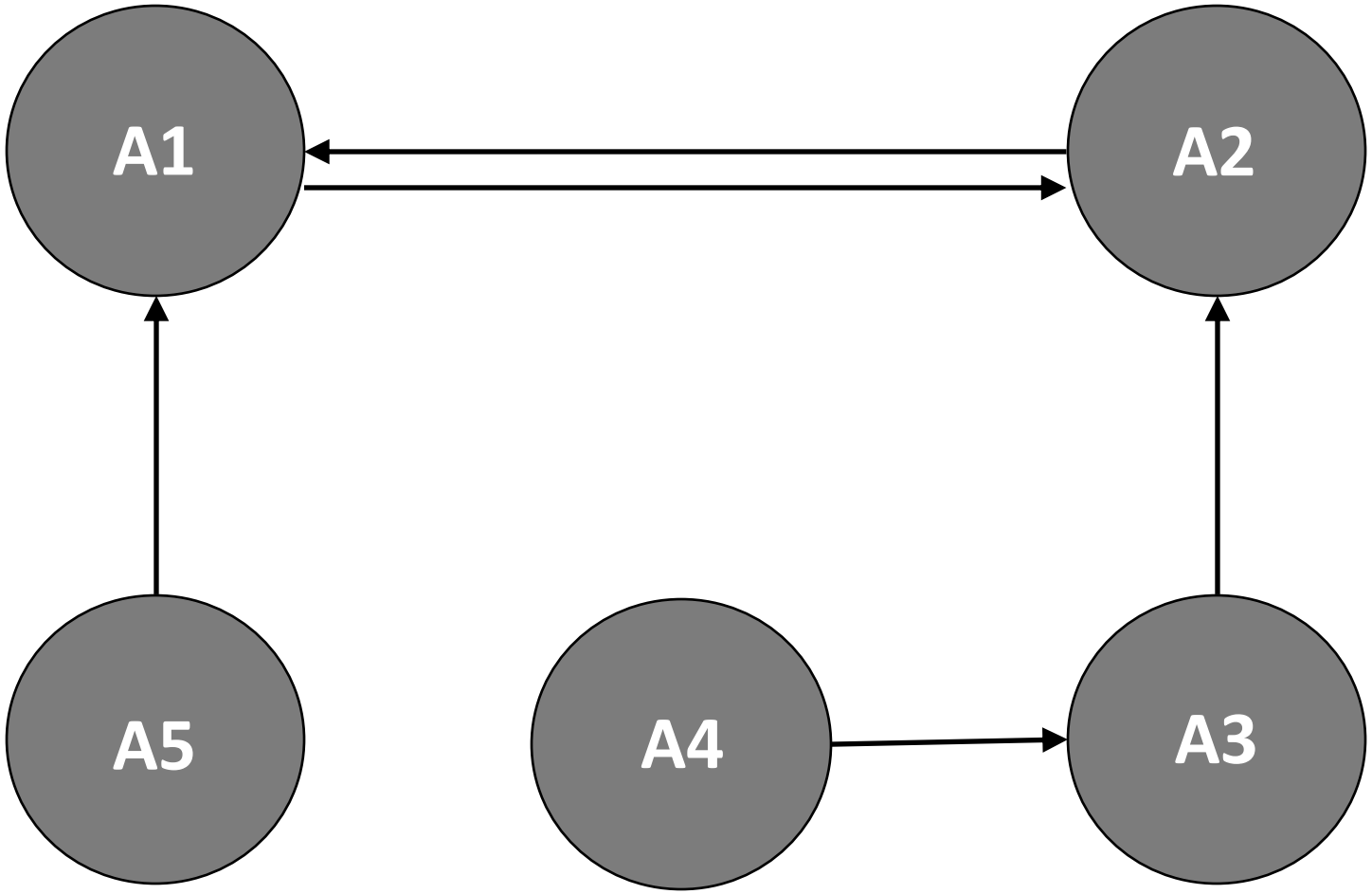
therefore I should go left

Argument 2 (A2)

Today there is an obstacle to the left,

therefore I should go right

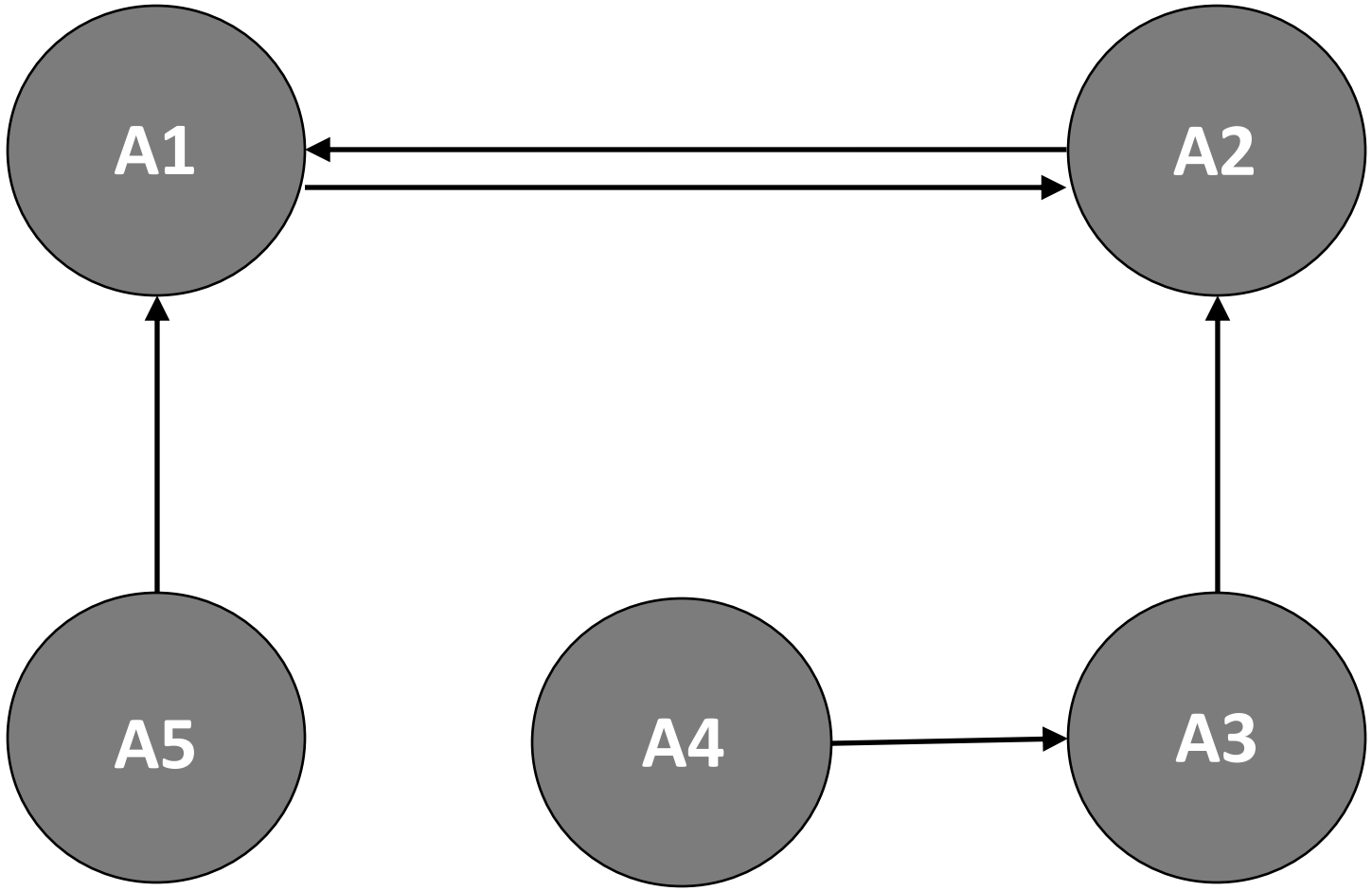



A1
A2
A5
A4
A3

IN if all its attackers are out

OUT if it has an attacker that is in

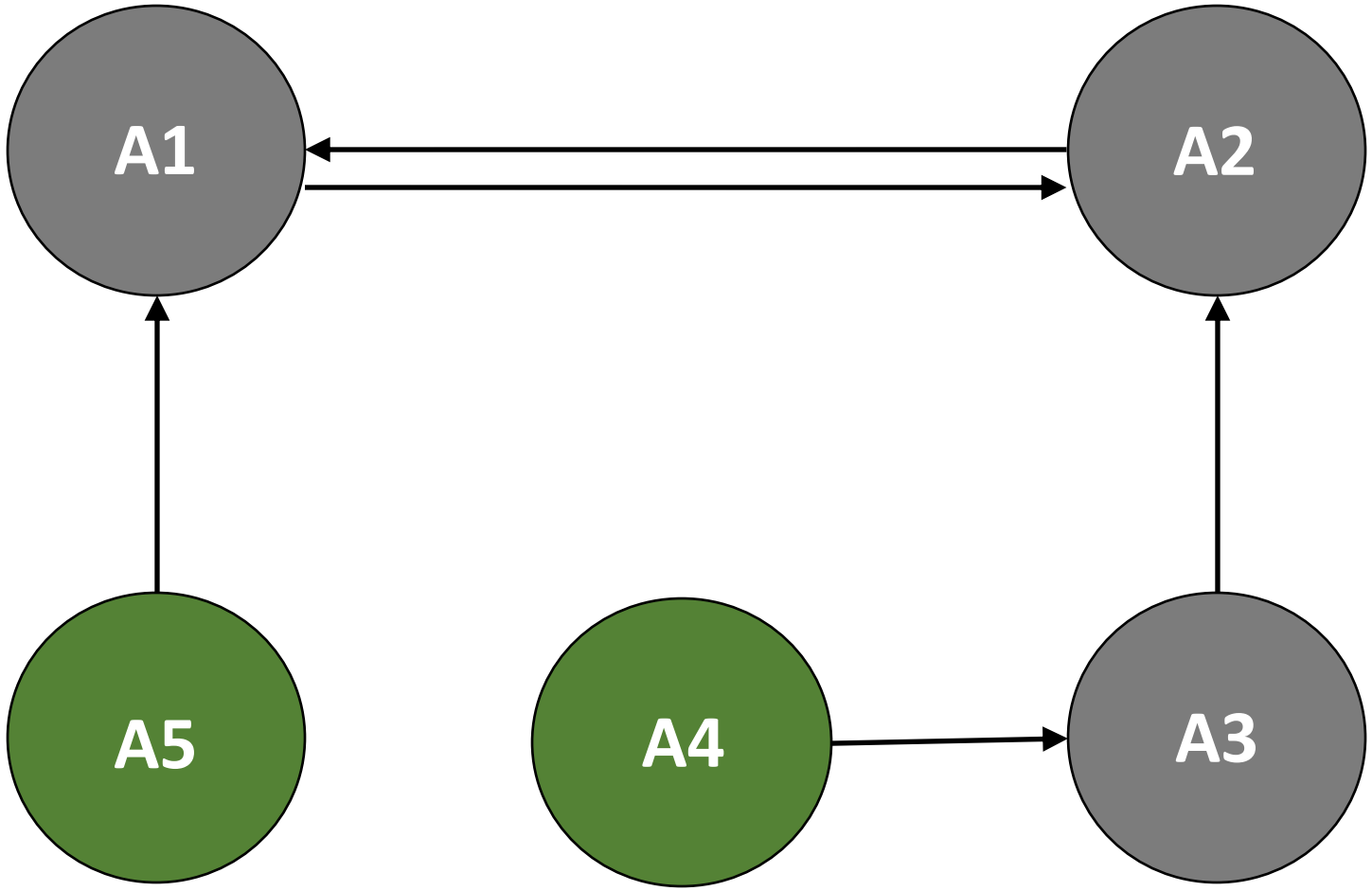
UNDEC if not all its attackers are out
and it does not have an attacker that is in



IN if all its attackers are out

OUT if it has an attacker that is in

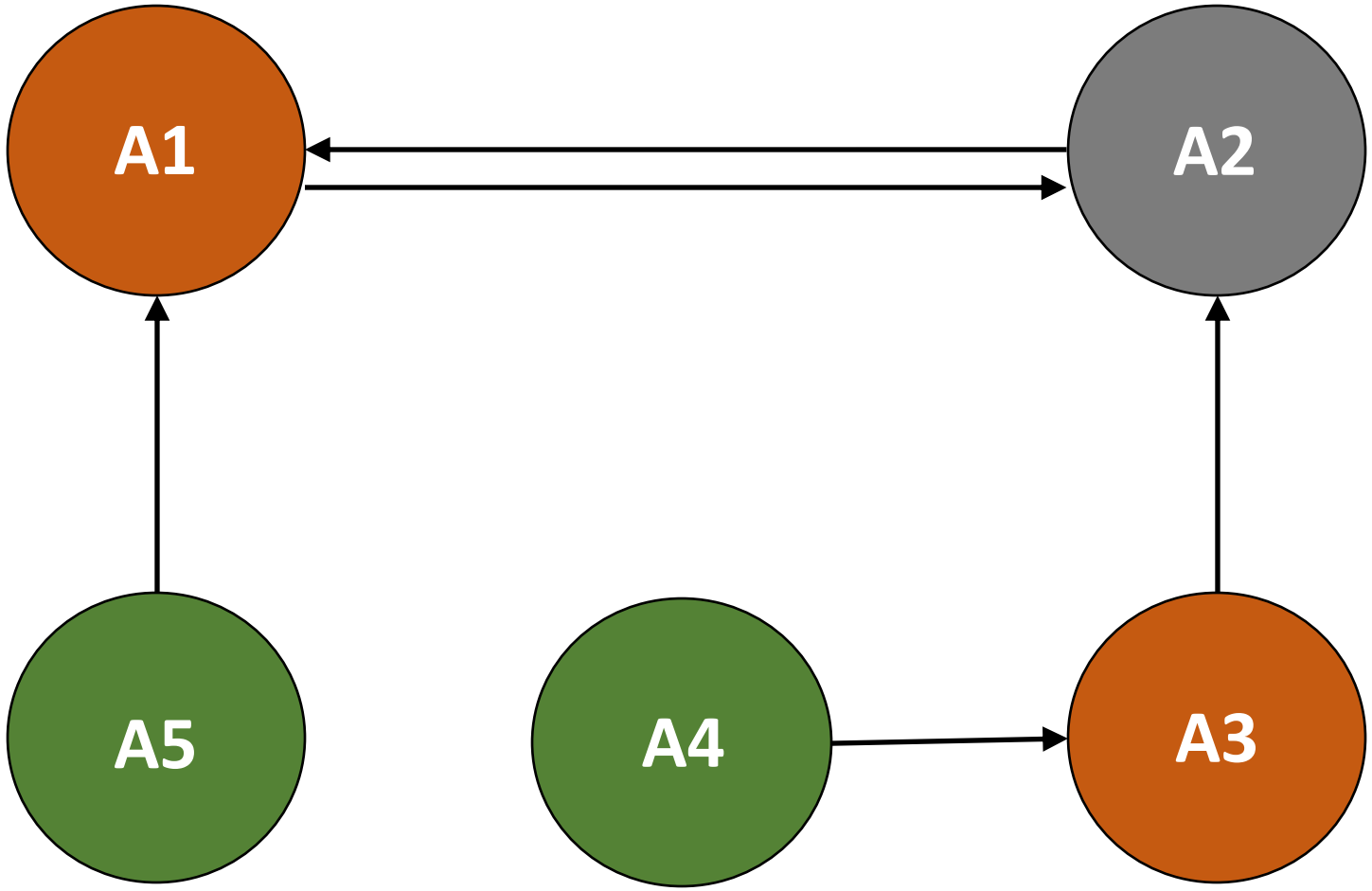
UNDEC if not all its attackers are out
and it does not have an attacker that is in



IN if all its attackers are out

OUT if it has an attacker that is in

UNDEC if not all its attackers are out
and it does not have an attacker that is in

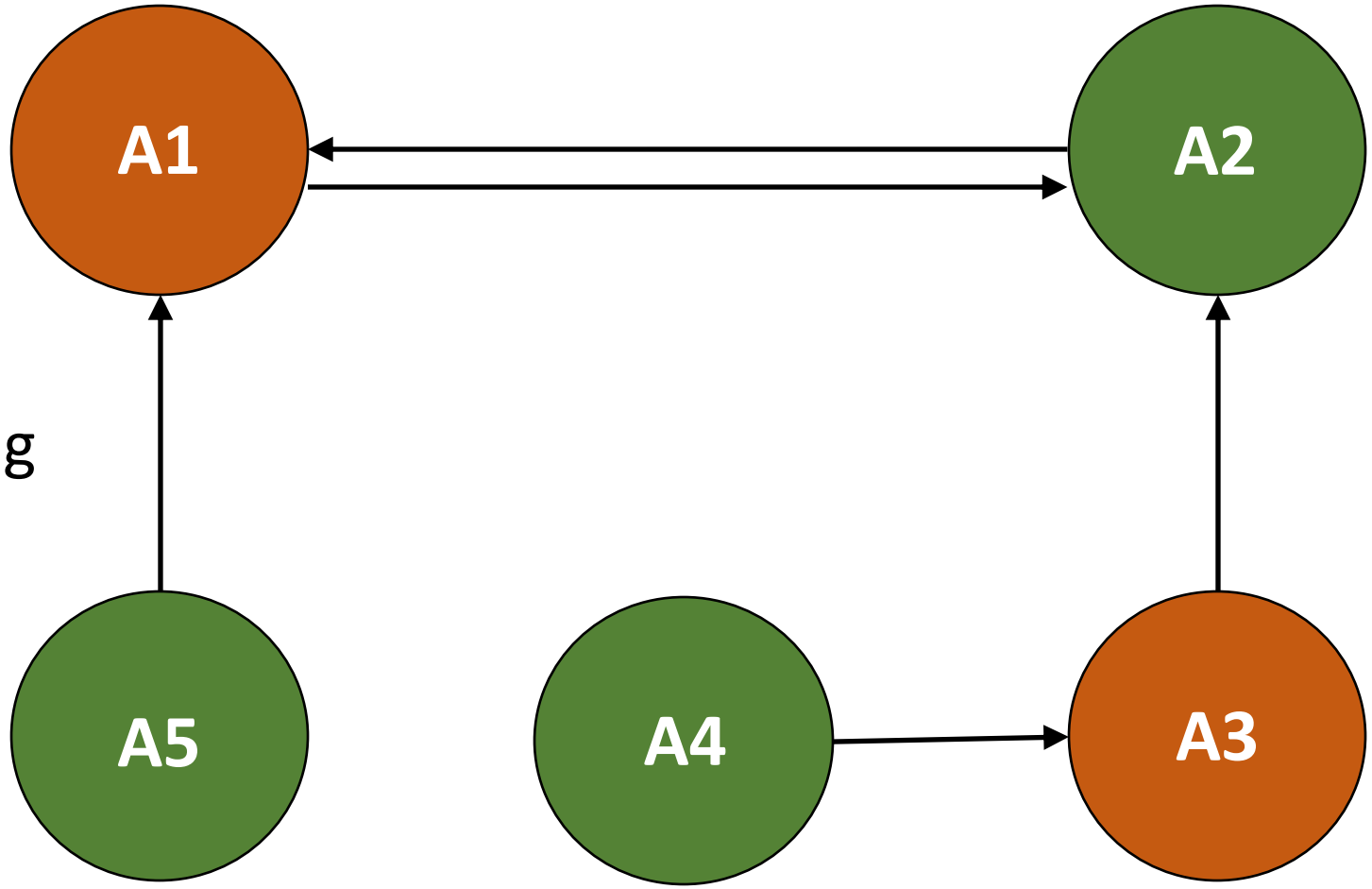


IN if all its attackers are out

OUT if it has an attacker that is in

UNDEC if not all its attackers are out
and it does not have an attacker that is in

We call this a complete labelling



Has social media been good for humanity?

A1: Social media has been good for humanity

A2: Social media has not been good for humanity

A3: Social media can be good to find news

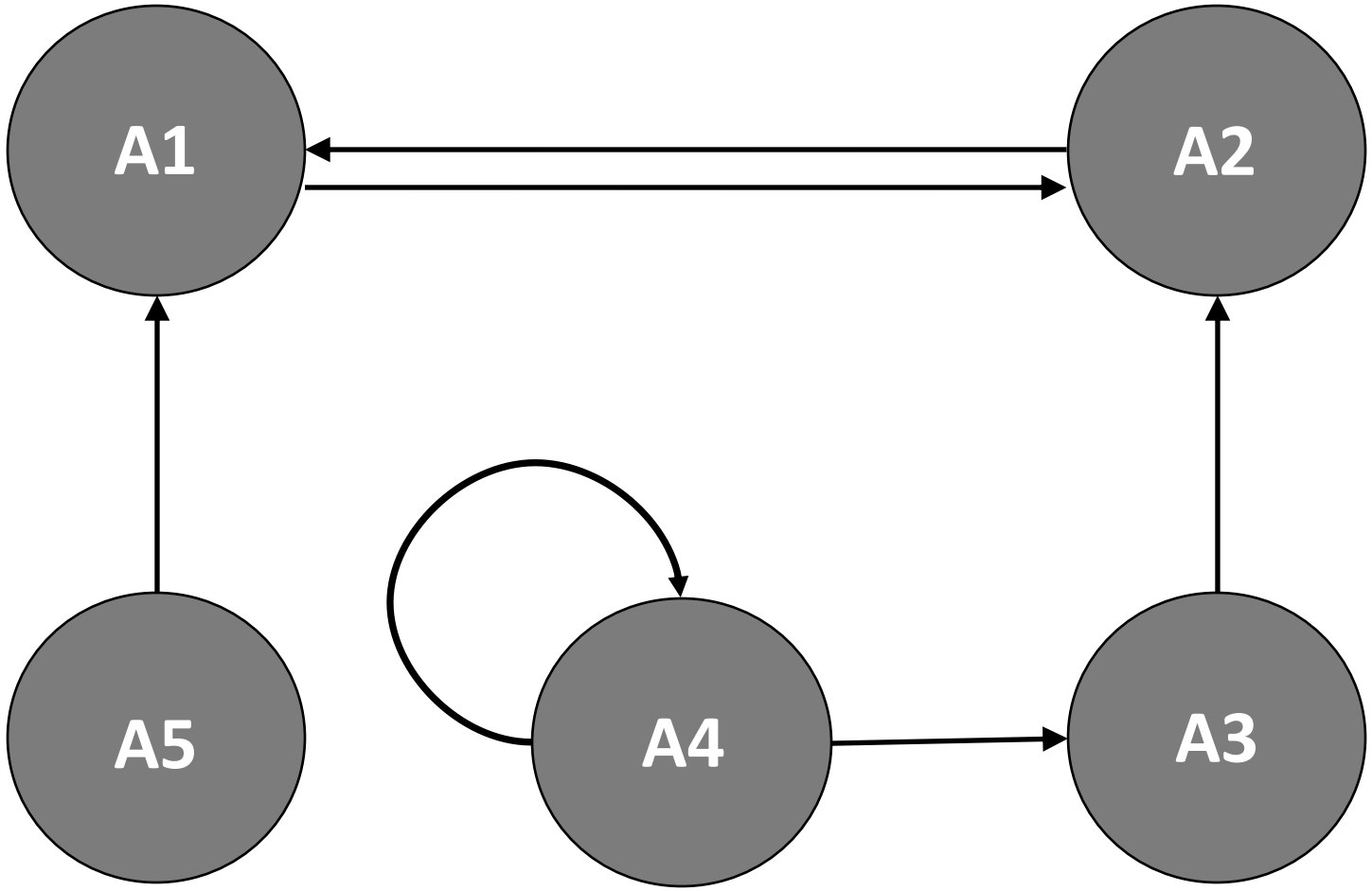
A4: I read on social media that we cannot verify whether news on social media is real or fake

A5: Social media puts privacy and data at risk

IN if all its attackers are out

OUT if it has an attacker that is in

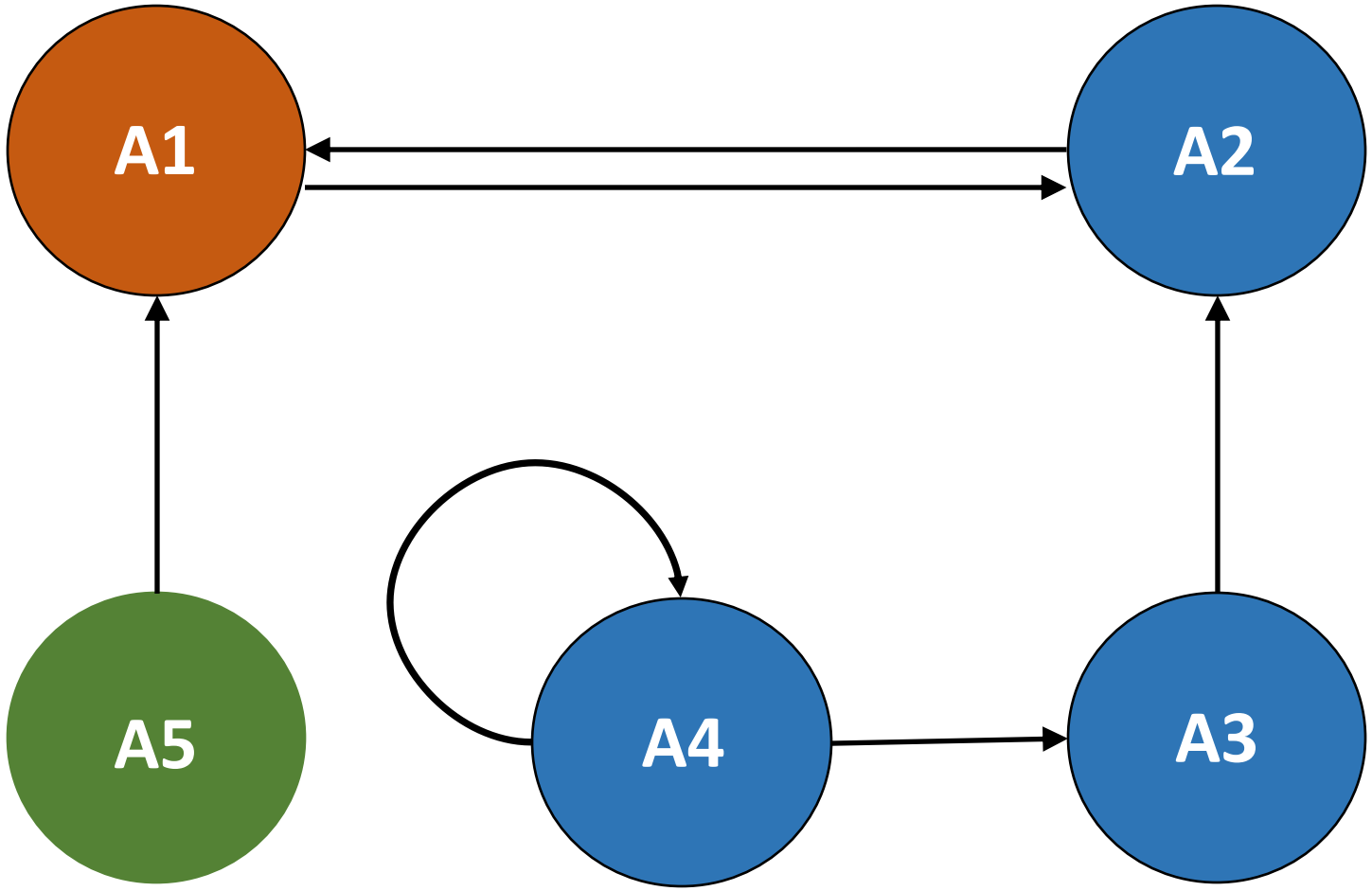
UNDEC if not all its attackers are out
and it does not have an attacker that is in



IN if all its attackers are out

OUT if it has an attacker that is in

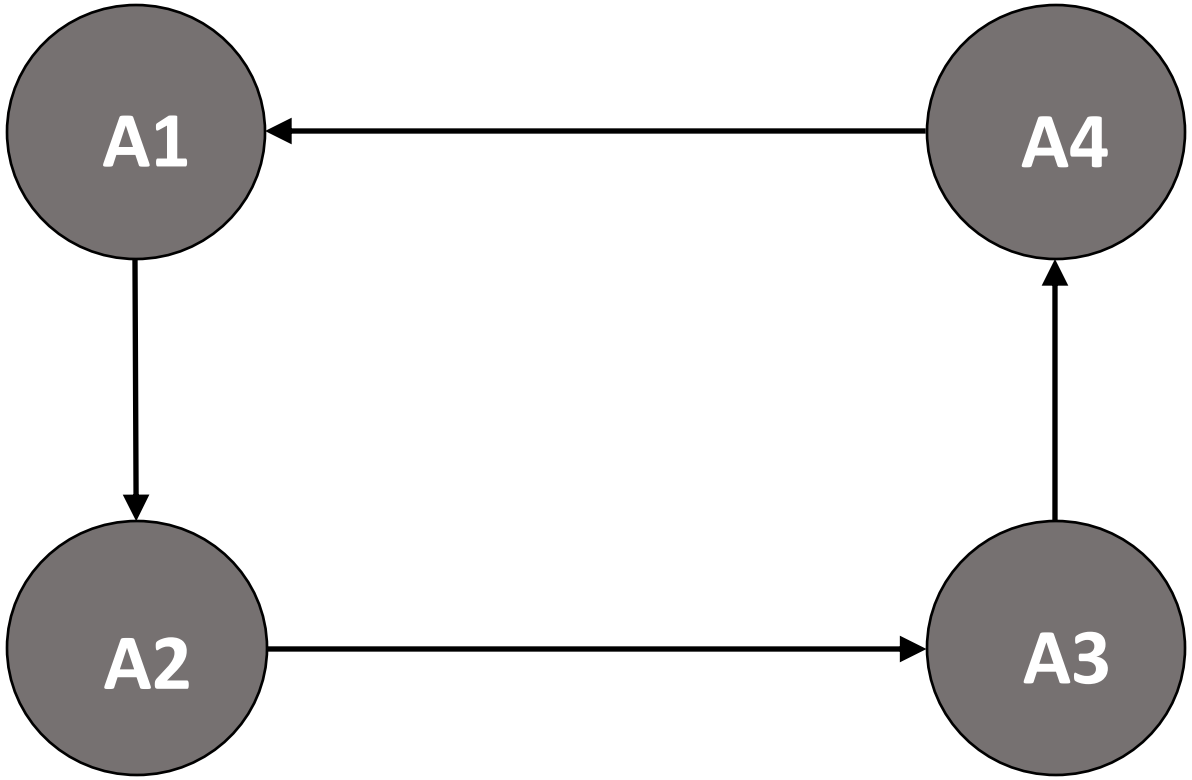
UNDEC if not all its attackers are out
and it does not have an attacker that is in

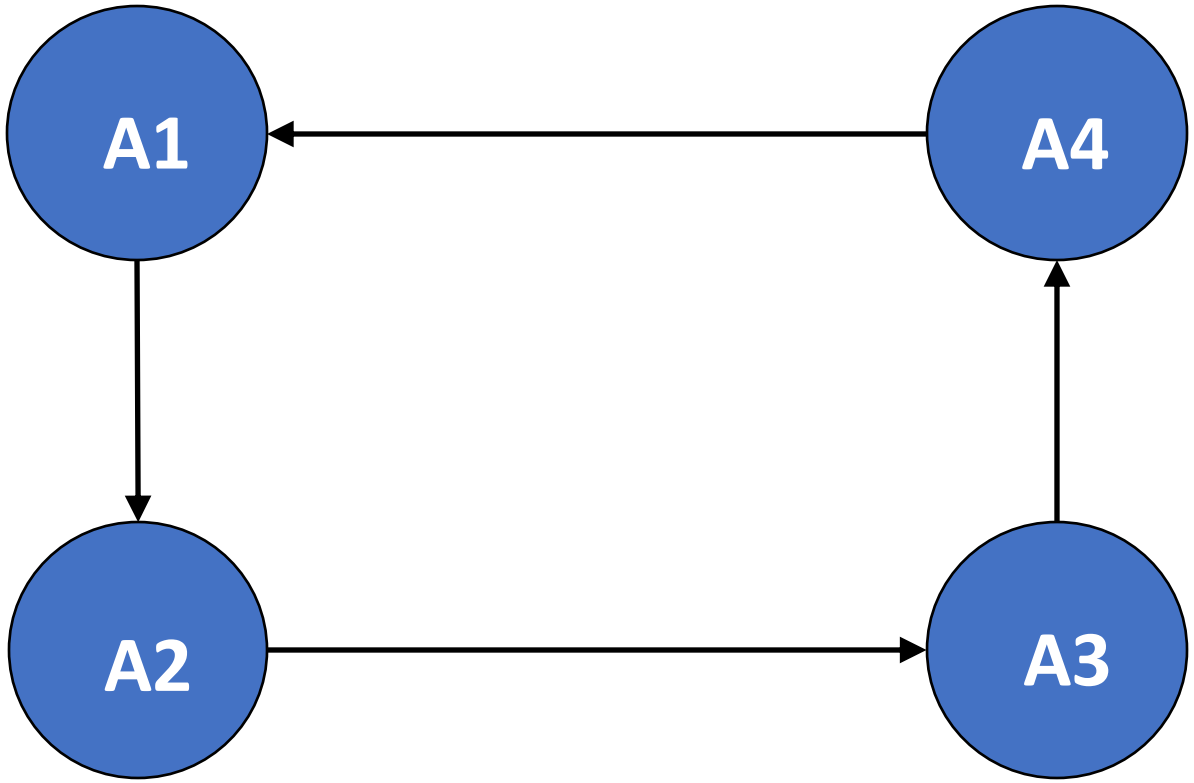


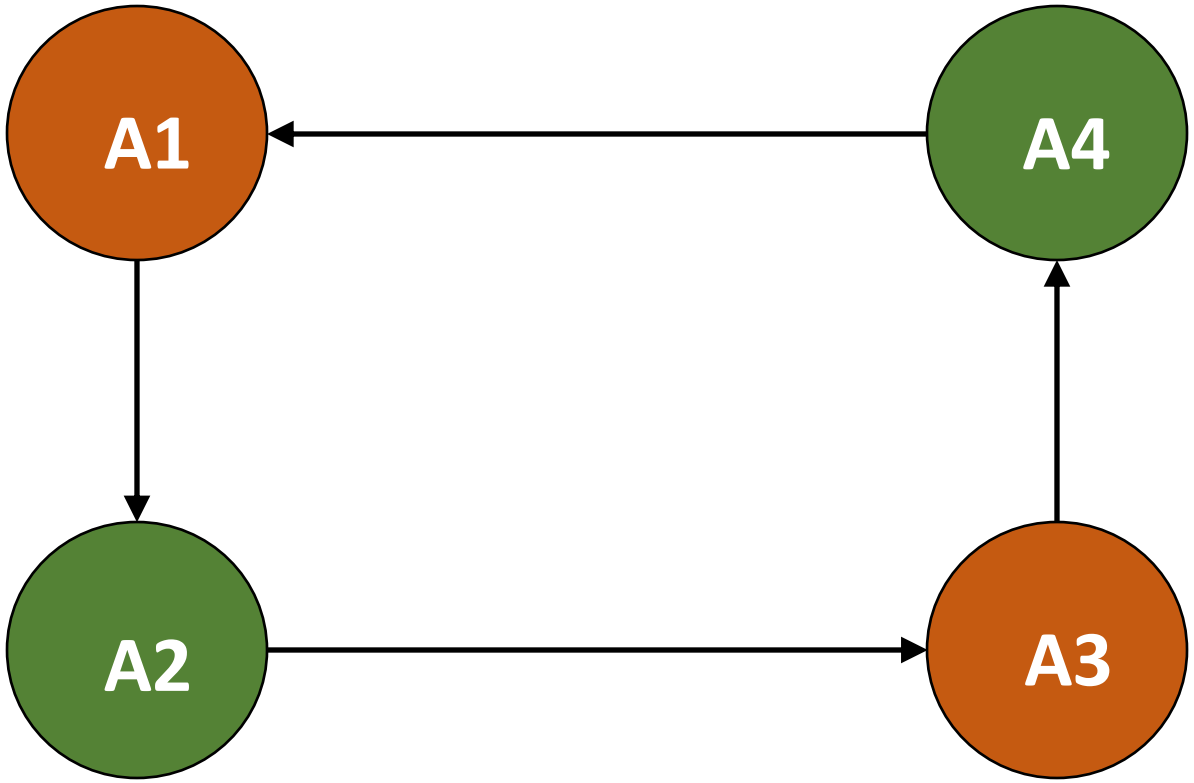
IN if all its attackers are out

OUT if it has an attacker that is in

UNDEC if not all its attackers are out
and it does not have an attacker that is in



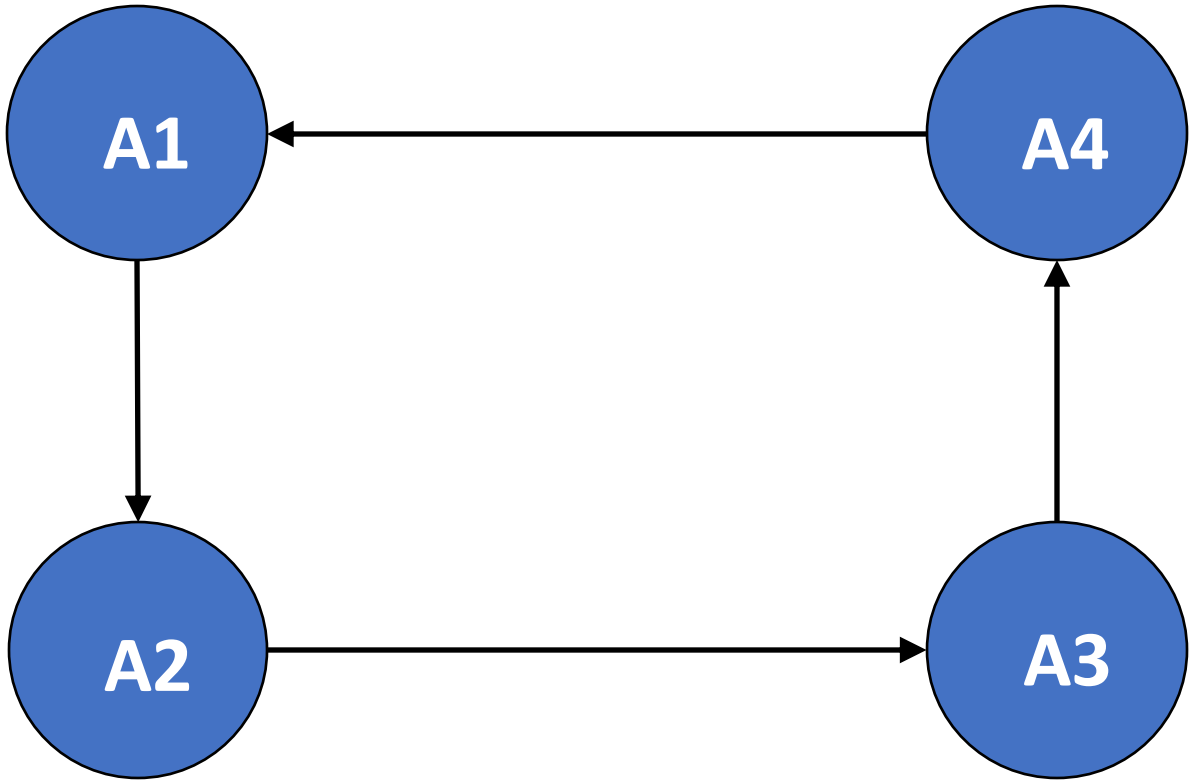

A1
A4
A2
A3


A1
A4
A2
A3

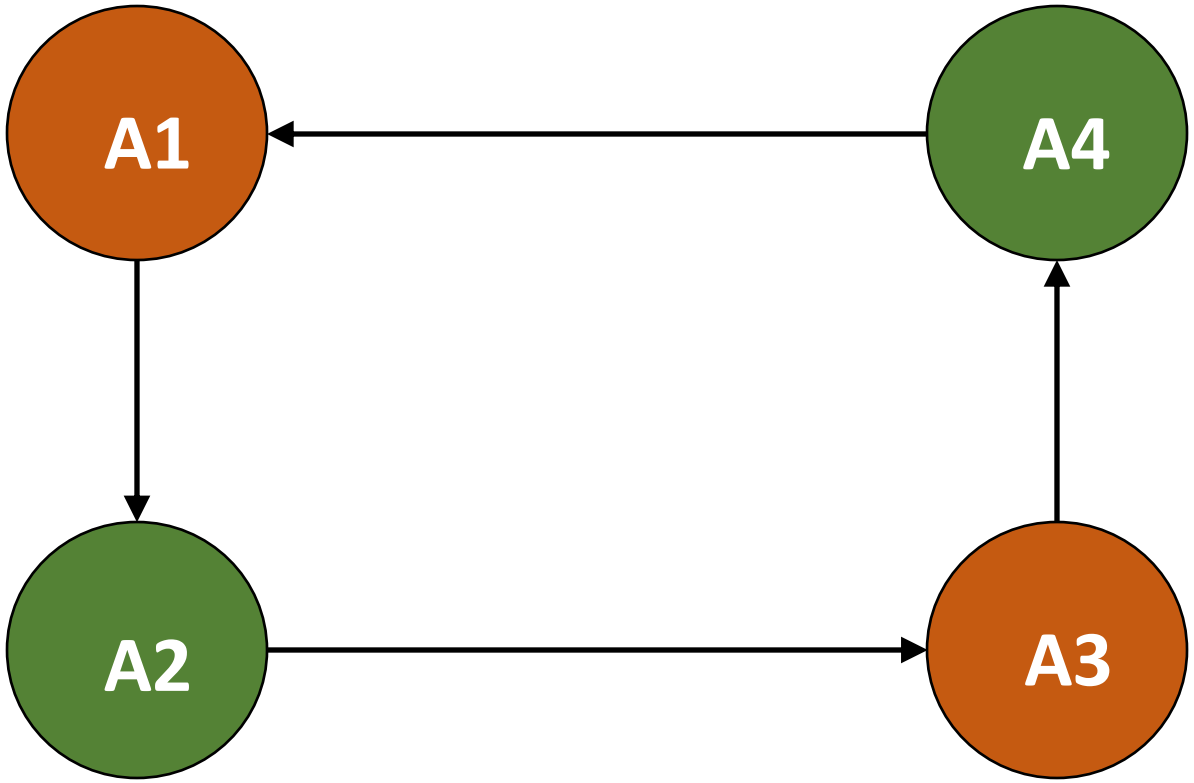
Labellings

- Grounded labelling – minimise the arguments that are IN
- Preferred labelling – maximise the arguments that are IN
- Stable labelling – no UNDEC arguments
- Semi-stable labelling – minimise the arguments that are UNDEC

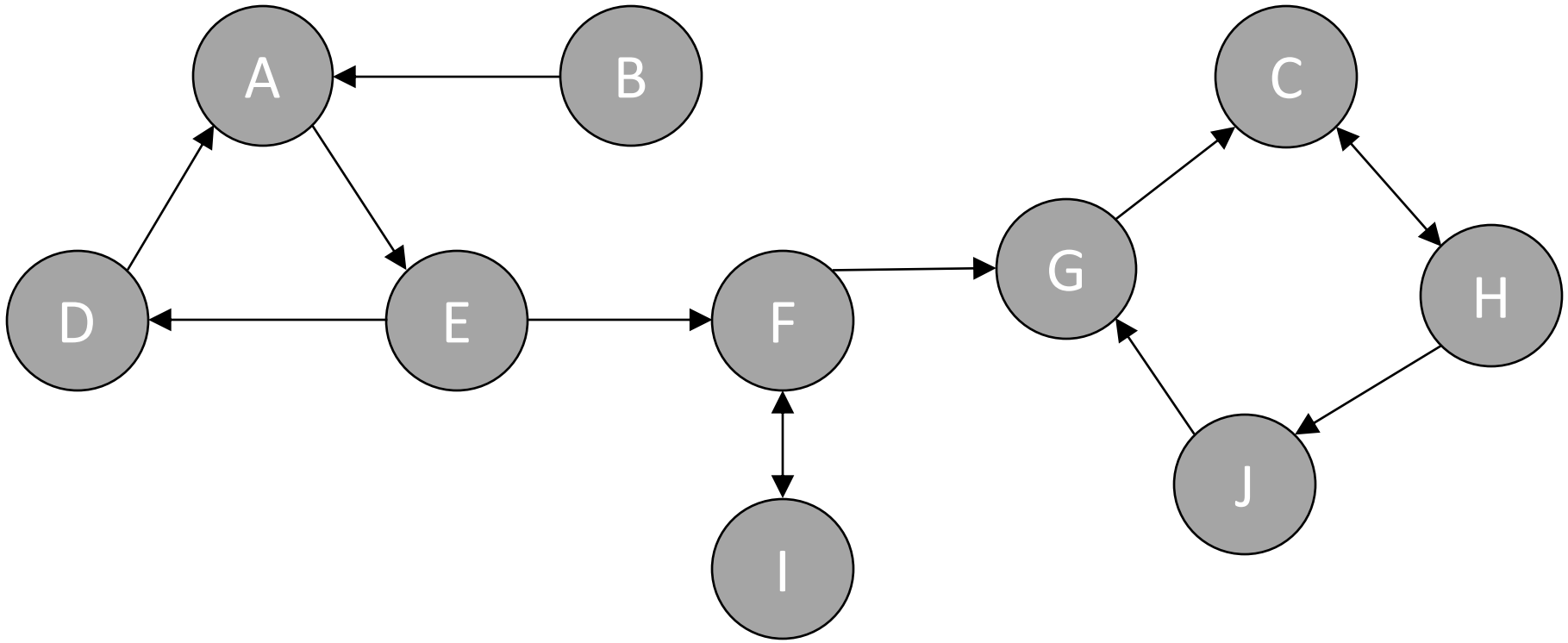
Grounded labelling



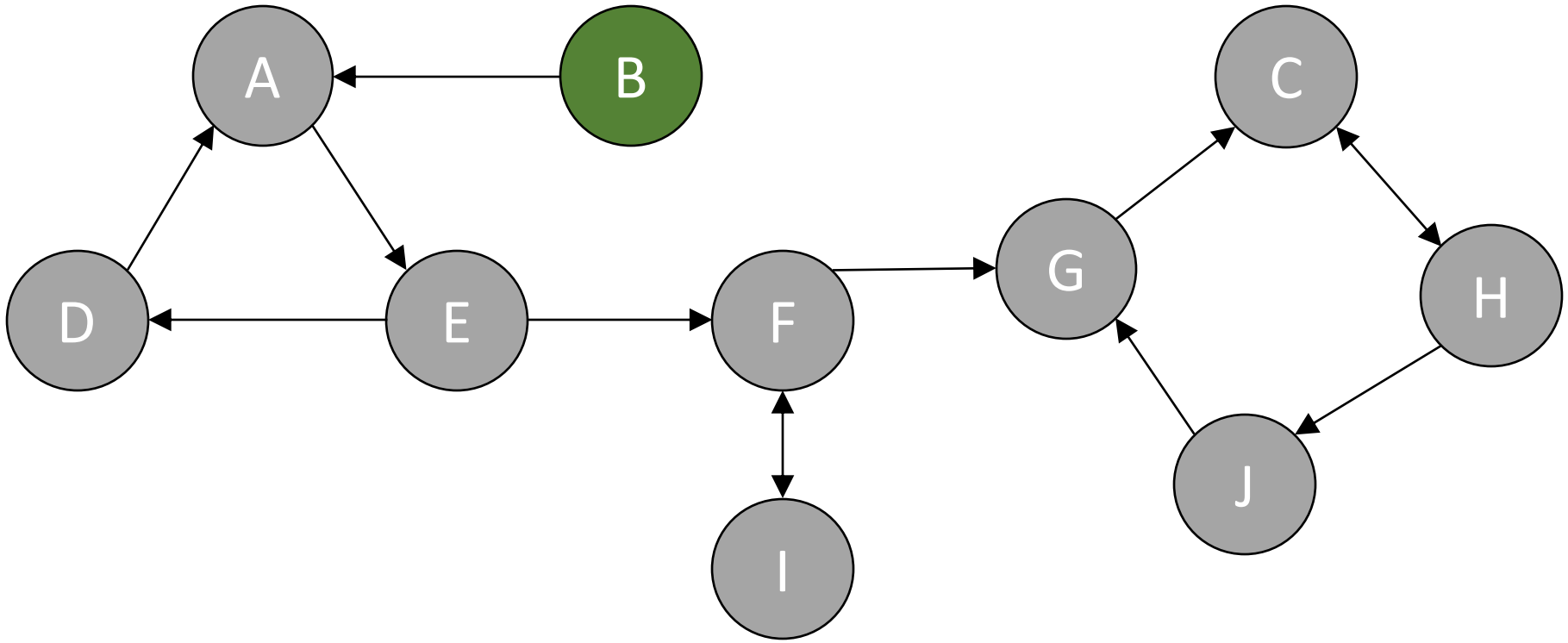
Preferred labelling
Stable labelling
Semi-stable labelling



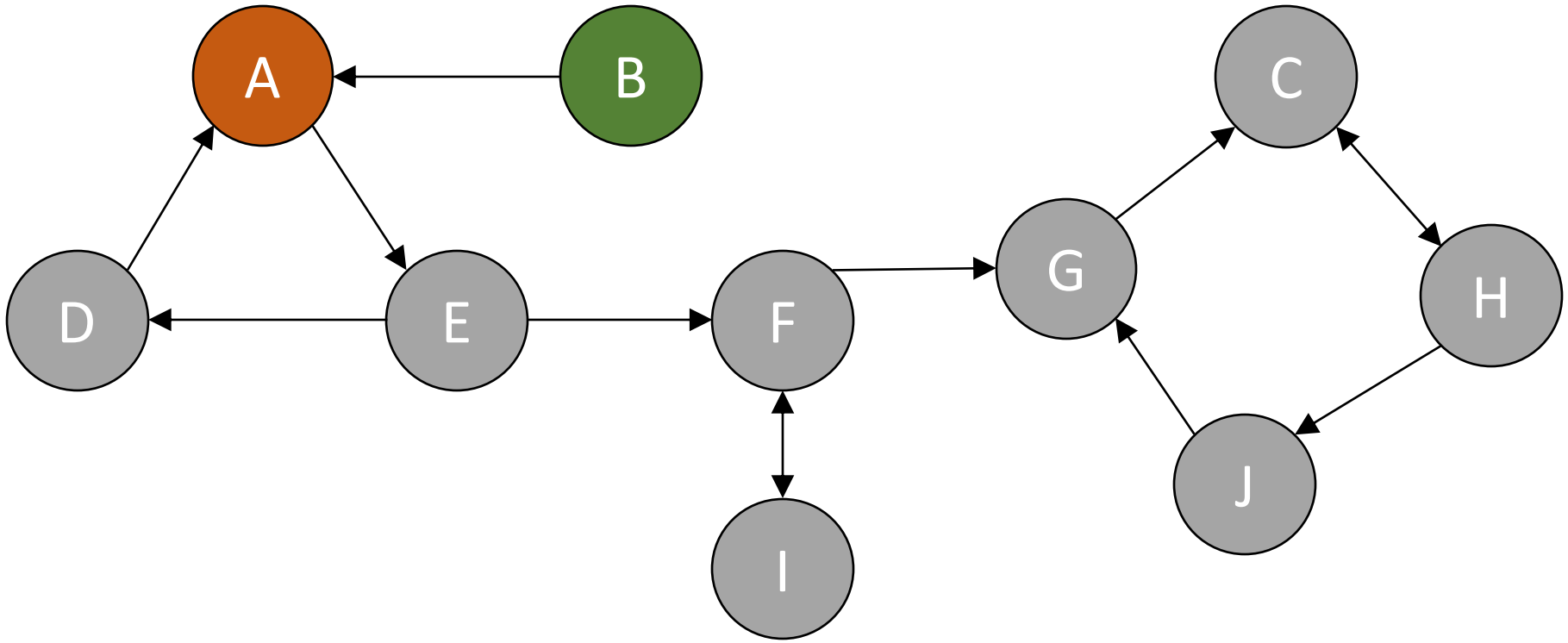
Which are the grounded labellings?



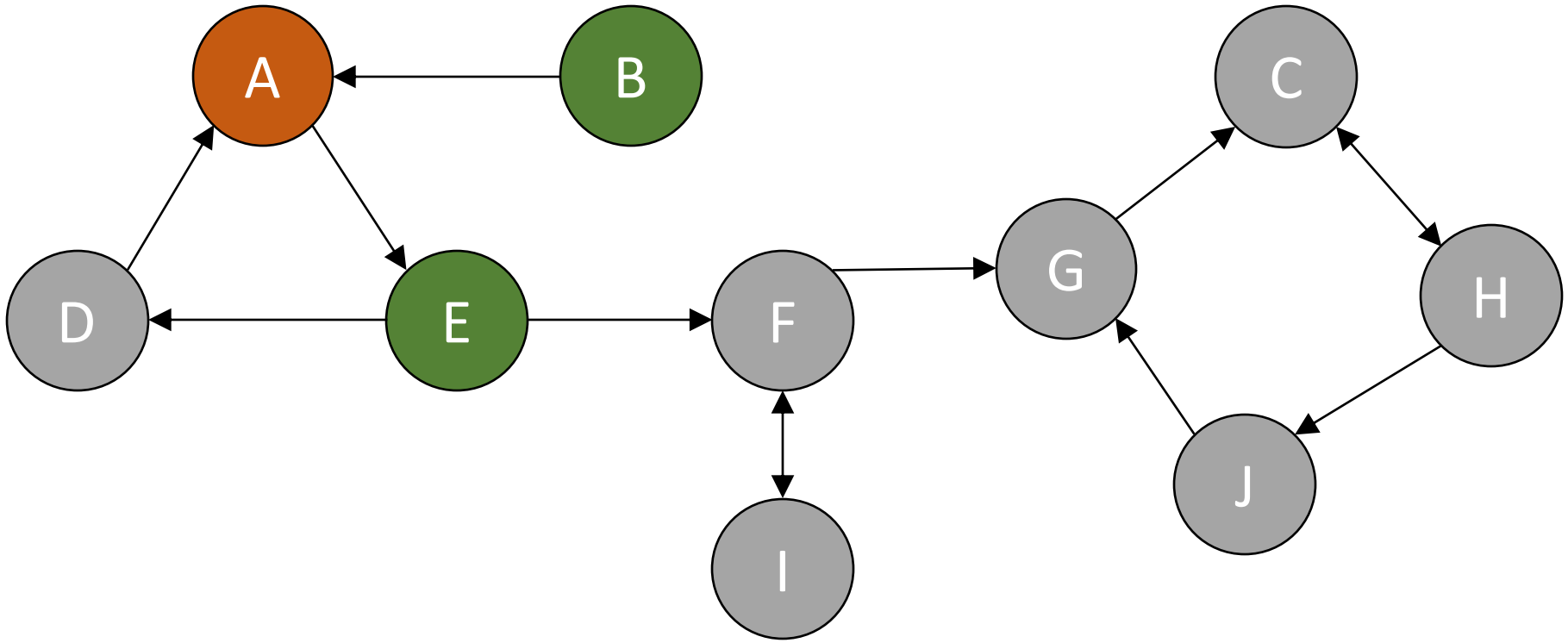
Which are the grounded labellings?



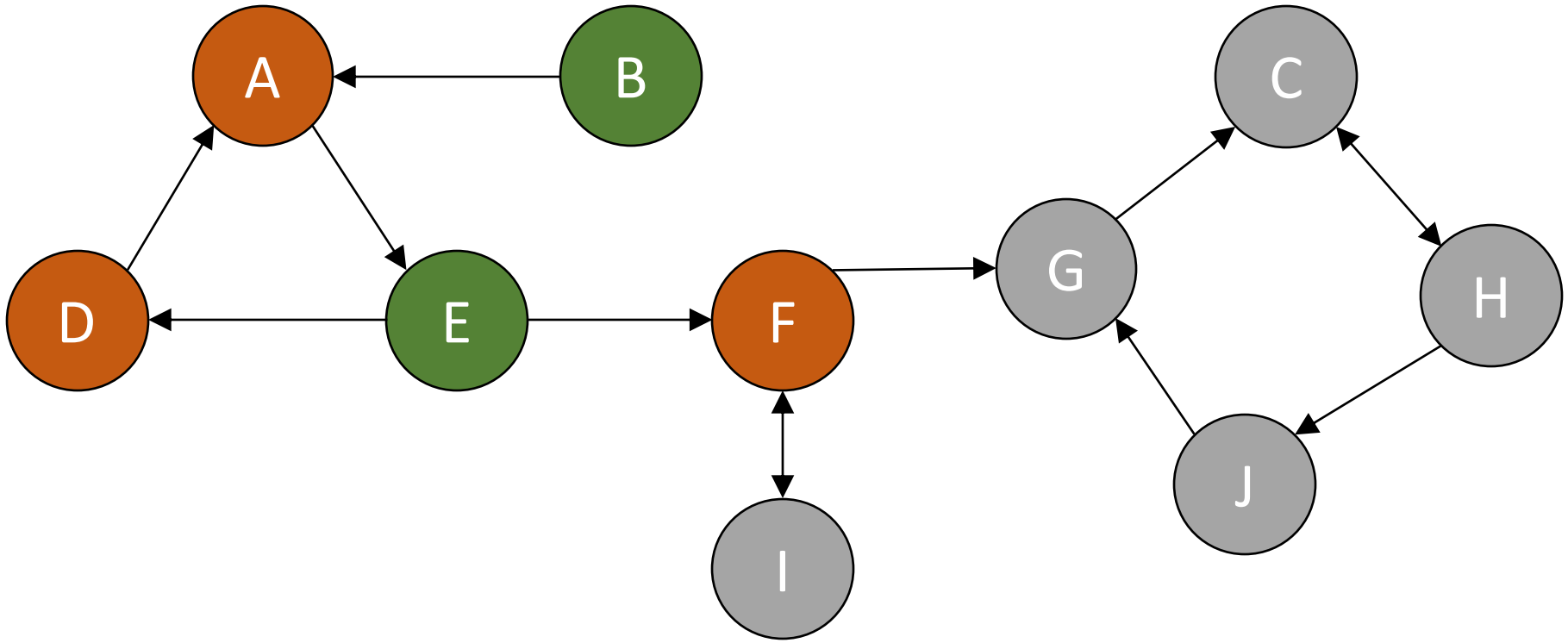
Which are the grounded labellings?



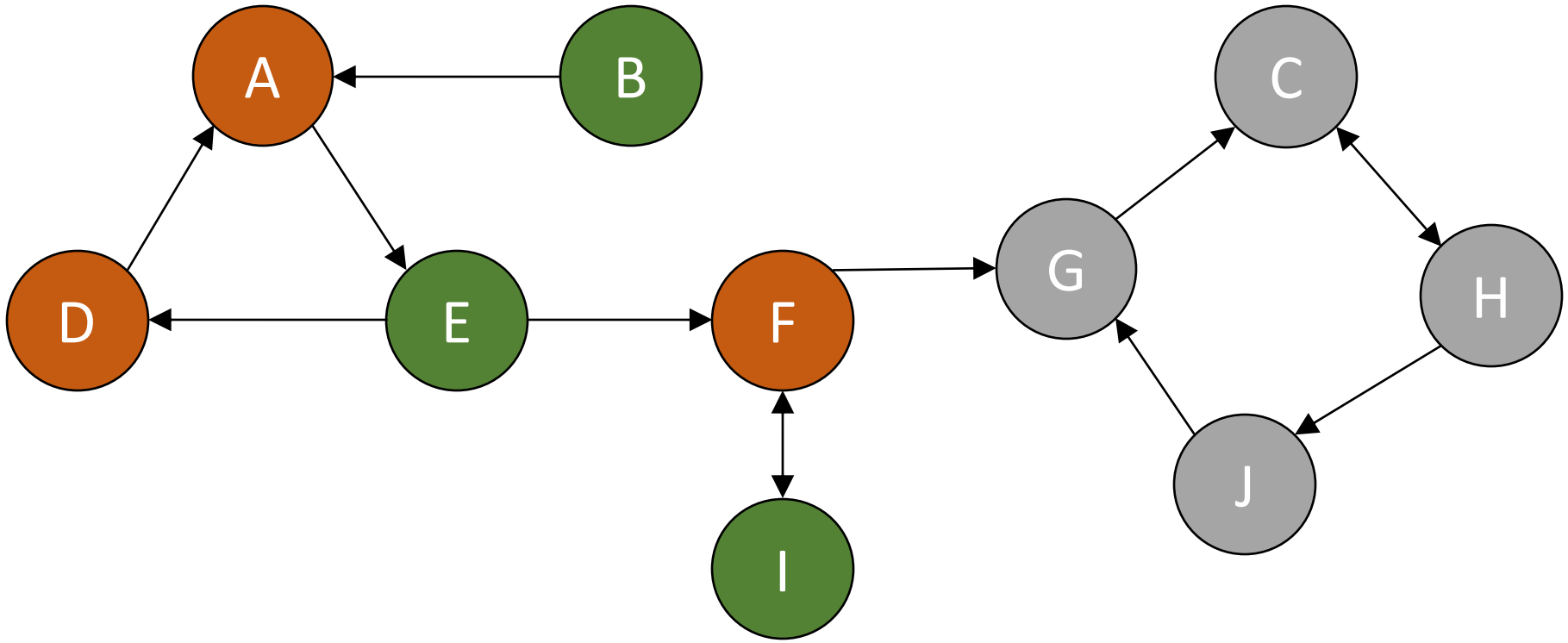
Which are the grounded labellings?



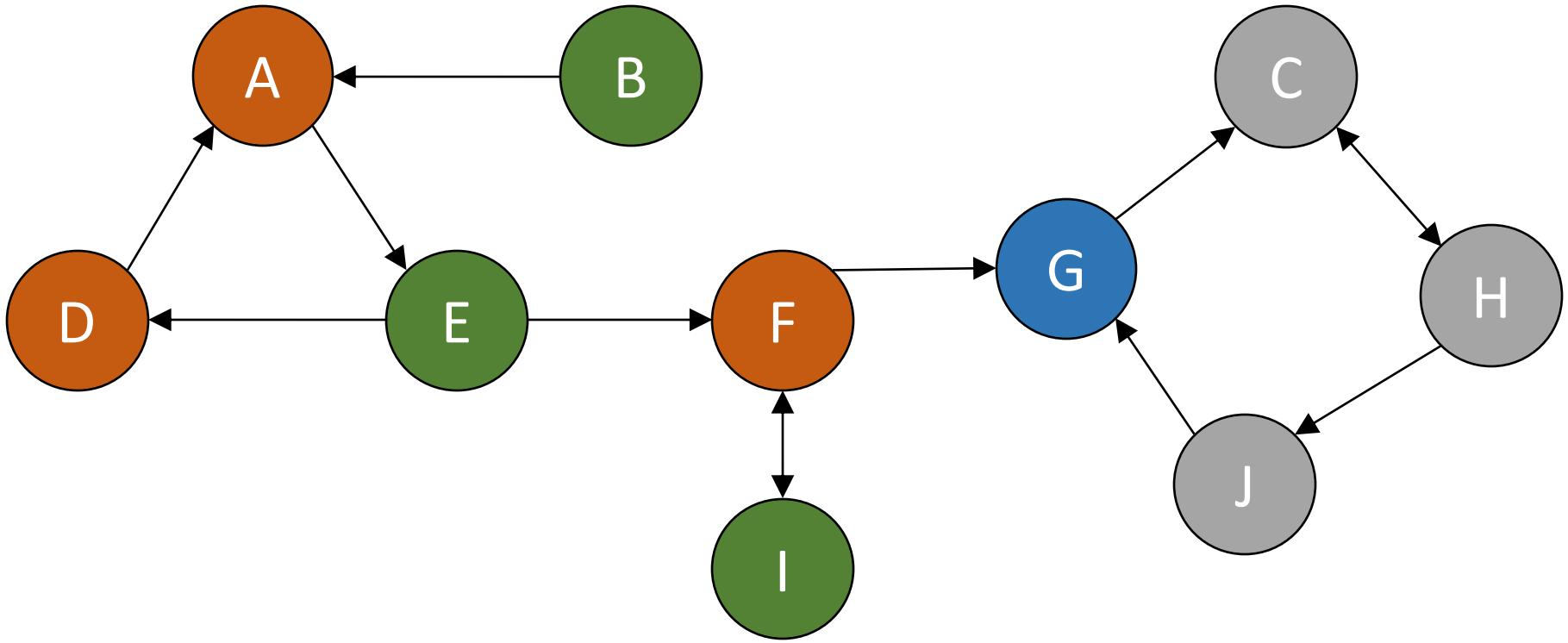
Which are the grounded labellings?



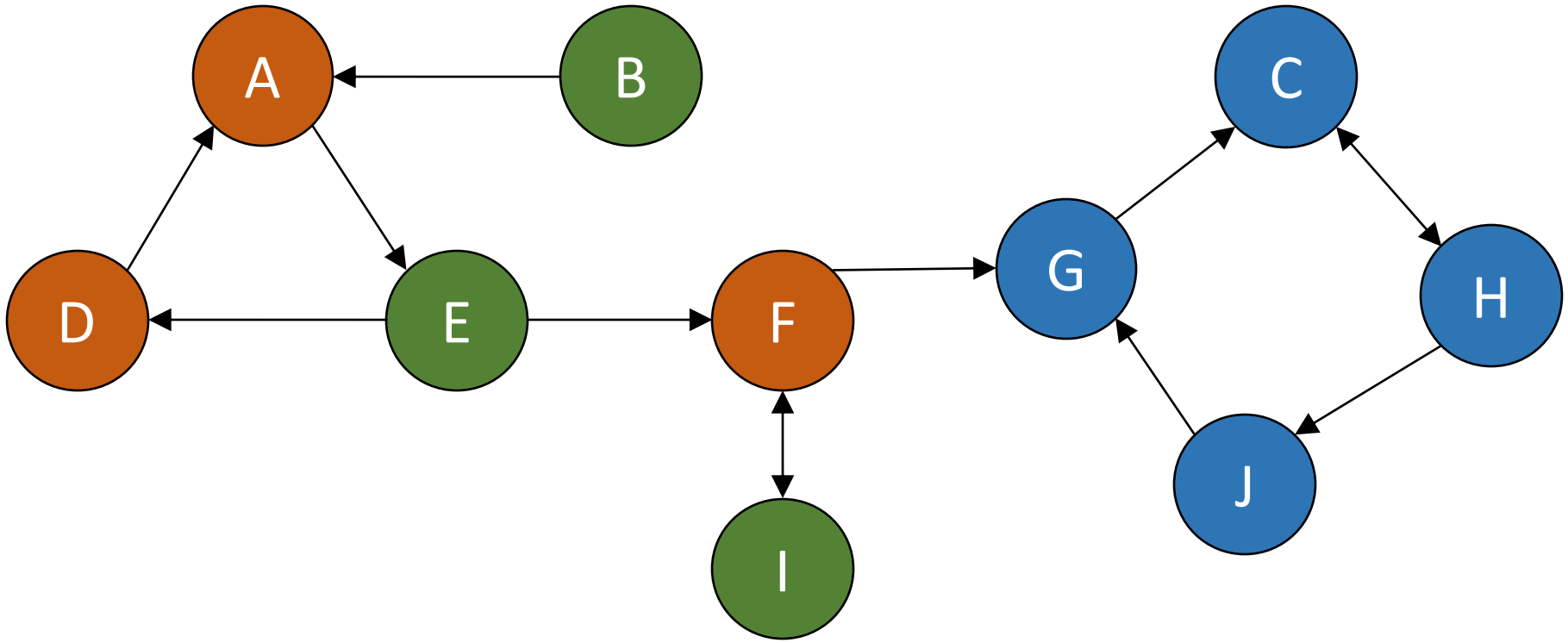
Which are the grounded labellings?



Which are the grounded labellings?



Which are the grounded labellings?



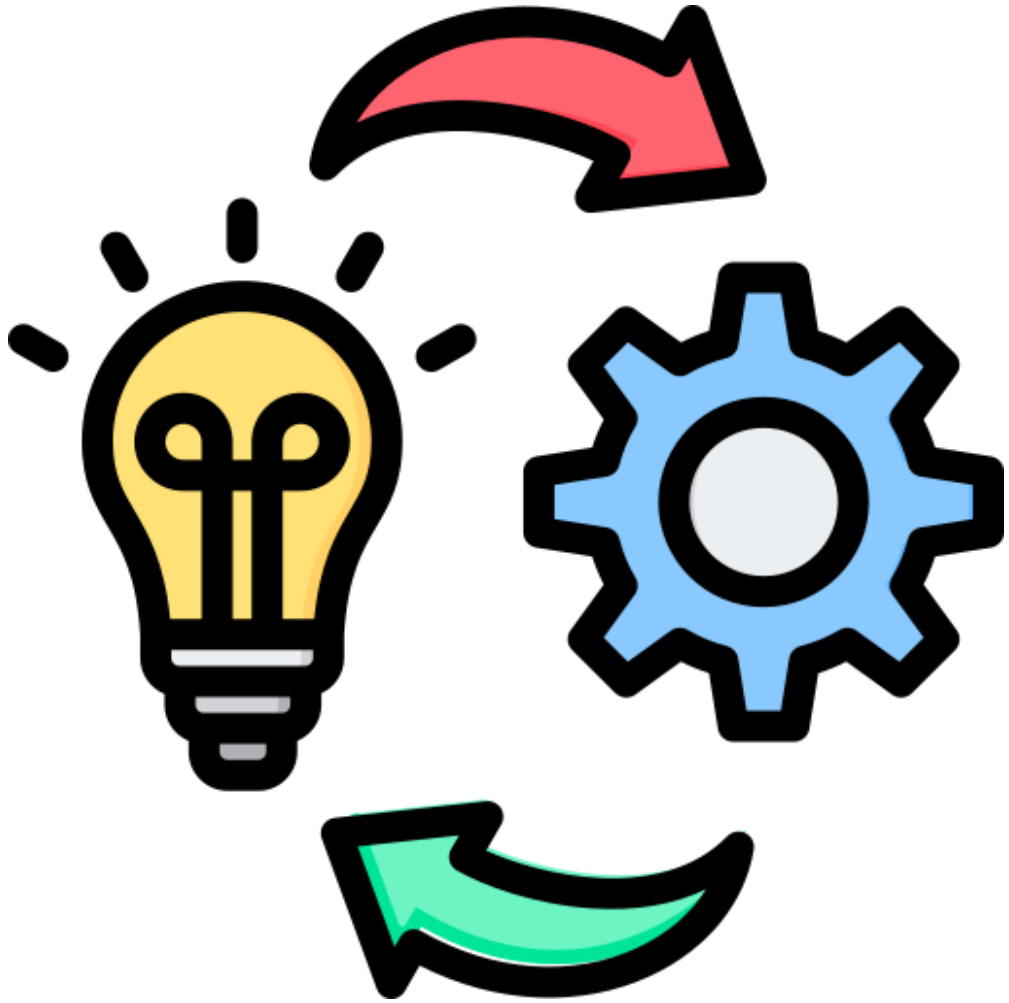
What are the labellings of the graphs you created for your debate?

- Grounded labelling – minimise the arguments that are IN
- Preferred labelling – maximise the arguments that are IN
- Stable labelling – no UNDEC arguments
- Semi-stable labelling – minimise the arguments that are UNDEC

Break

Implementation time!

http://argteach.herokuapp.com



Variations of abstract argumentation frameworks

Bipolar argumentation frameworks (BAFs)

- Adds support relations to abstract argumentation frameworks
- Semantics defined differently to account for this:
  - An argument is accepted only if it is directly defended or supported by arguments that are themselves already accepted in a grounded manner.

Has social media been good for humanity?

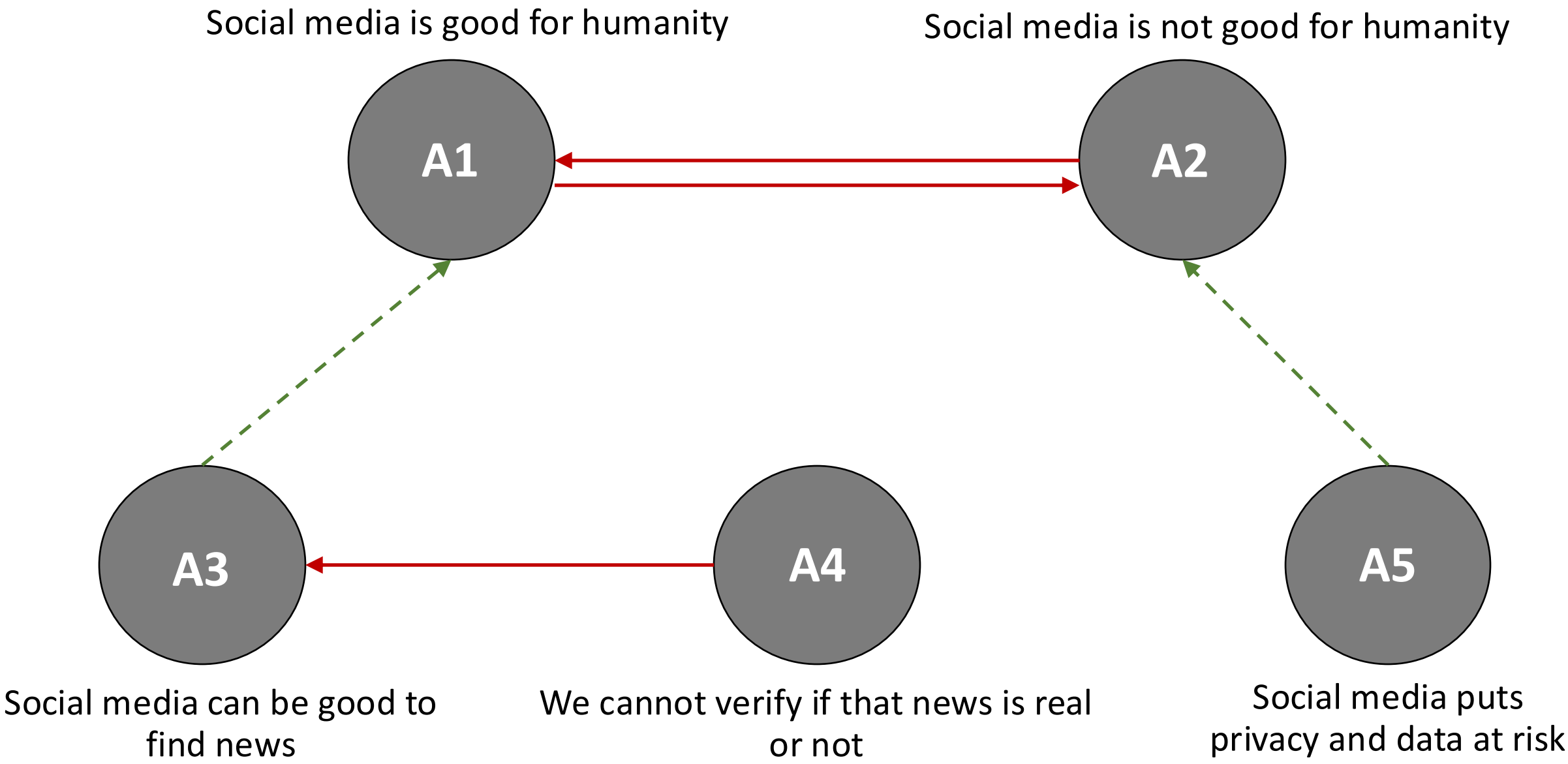
A1: Social media has been good for humanity

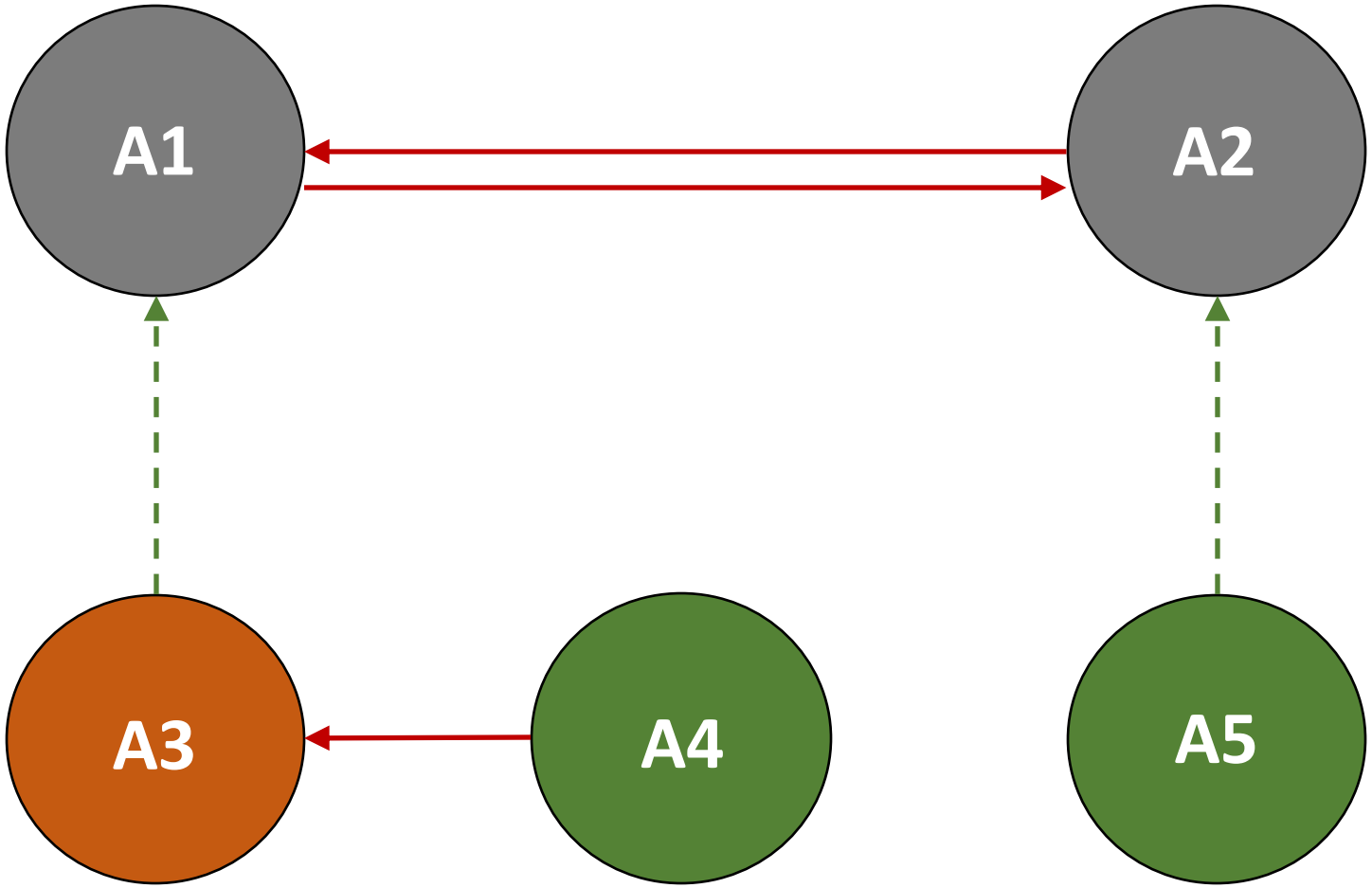
A2: Social media has not been good for humanity

A3: Social media can be good to find news

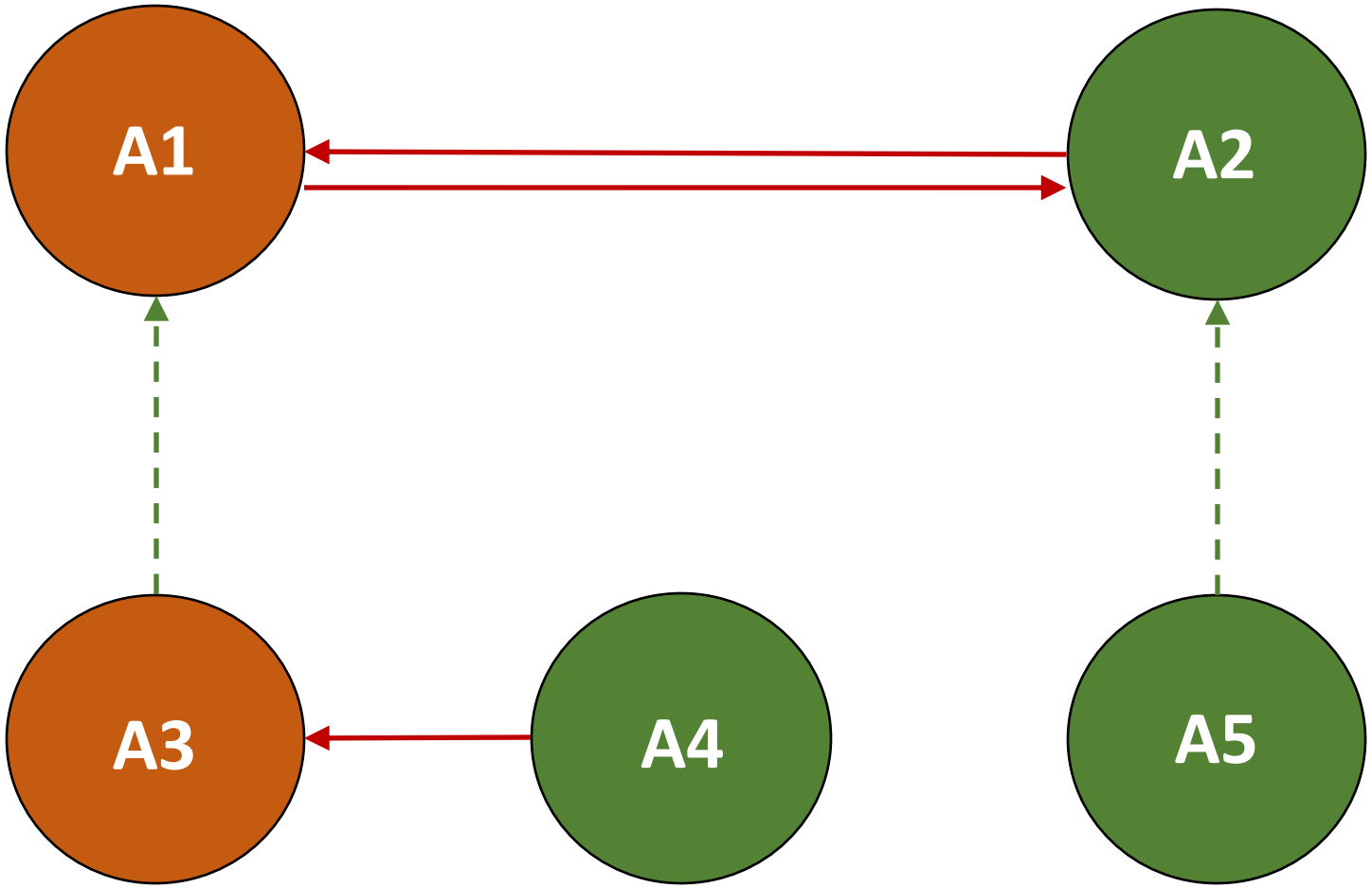
A4: We cannot verify if that news is real or not

A5: Social media puts privacy and data at risk

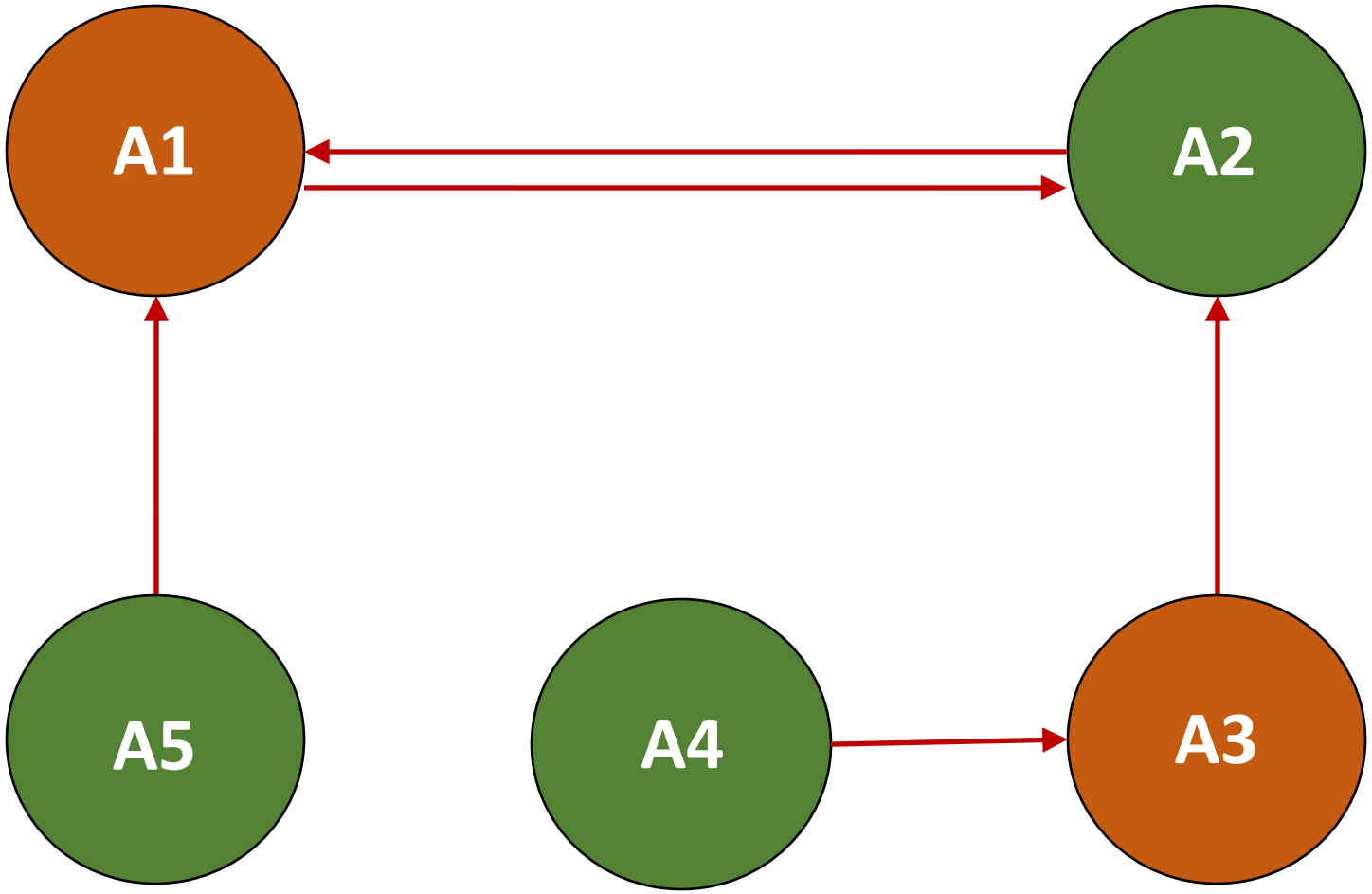

Social media is good for humanity
Social media is not good for humanity
A1
A2
A3
A4
A5
Social media can be good to find news
We cannot verify if that news is real or not
Social media puts privacy and data at risk


A1
A2
A3
A4
A5

Graph with attack and support relations



Graph with only attack relations



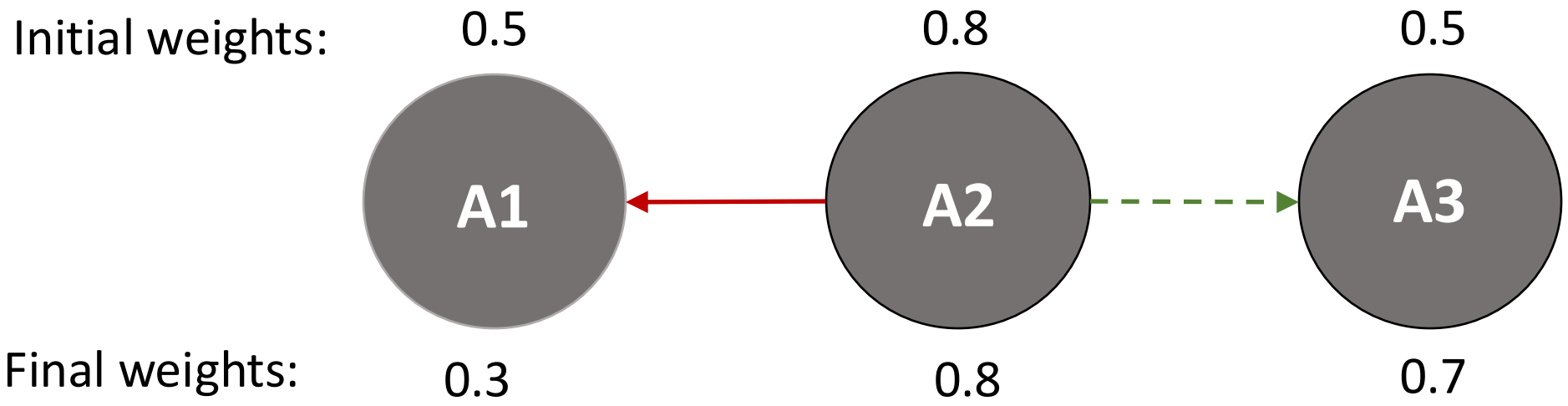
Weighted argumentation frameworks (WAFs)

- Adds numerical values to the abstract argumentation graph
- Intrinsic weights assigned to arguments/attacks/supports representing their initial strength
- Higher weights indicate stronger arguments/attacks/supports and therefore have more influence on the final acceptability calculated
- Semantics used to calculate final weights of arguments based on the weights of incoming arguments/attacks/supports

WAF semantics

1. Strength values are attracted by their initial weight [0,1]
2. Attackers force the strength value towards 0 proportionally to their strength
3. Supporters force the strength value towards 1 proportionally to their strength

Using the Quadratic Energy Model Semantics (Potyka 2019)



Applications

- Legal reasoning
- Decision-making systems
  - Bias detection
- Multi-agent systems

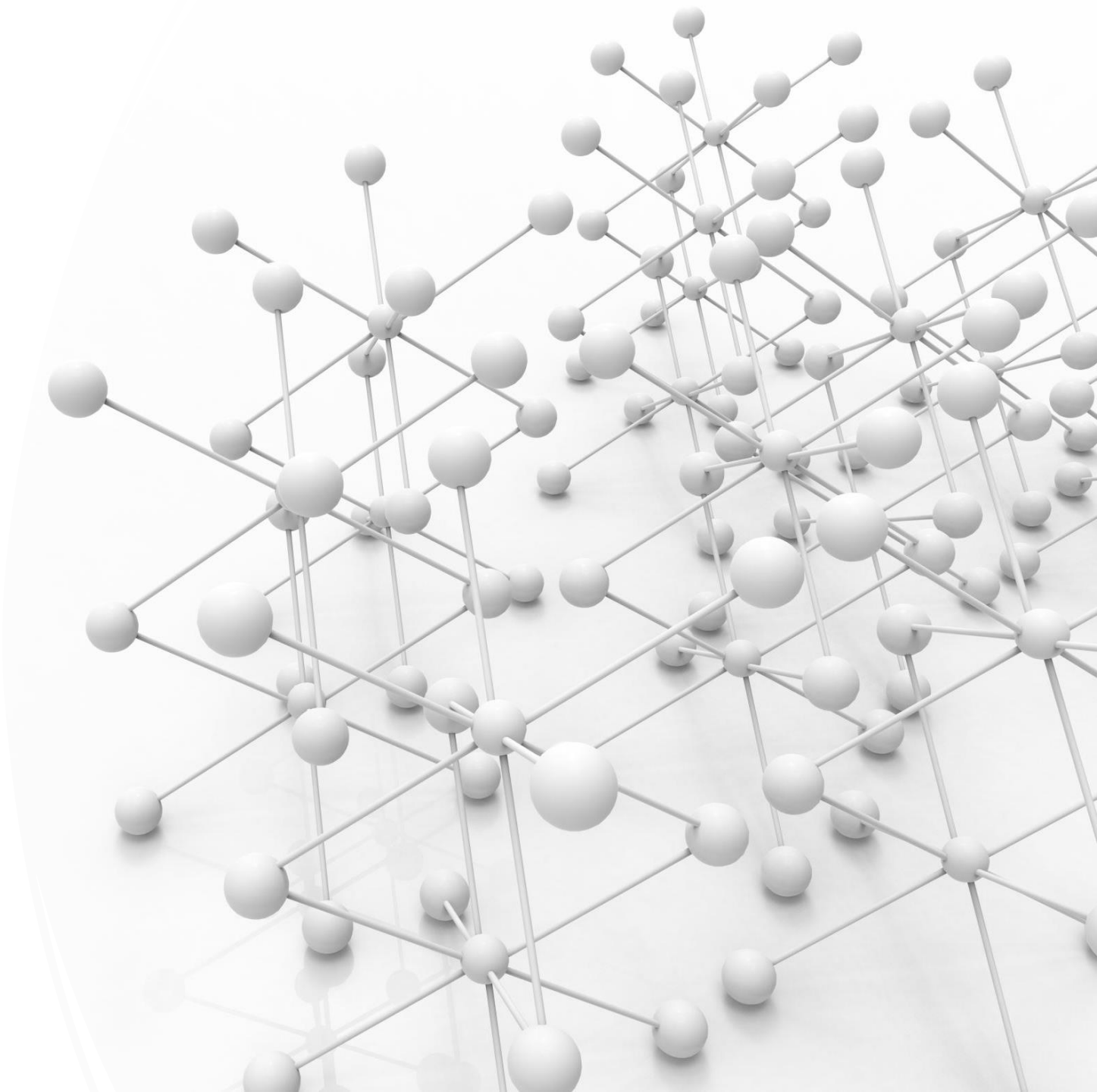
Look back at your debate graphs

- Would it be easier to create with support relations?
  - Think about how you would change it
- How could you add weights to your graphs? On the arguments or relationships.
  - What would these weights represent?
- Would it be helpful to have any other relationship represented in your graph?

Structured argumentation

Building on abstract argumentation, structured argumentation additionally details:

- the *internal* structure of arguments
- the explicit relationships among them



Structured argumentation models premises, conclusions, and inferential rules

p: Left is the fastest route.

c: I will go left.

r: I want to take the fastest route.

Strict and defeasible rules

Strict Rules

If A and B then **always** C

All tuba players are mortal.
I am a tuba player.
 I am mortal.

Defeasible rules

If A and B then **usually** C

After complaints, I don’t play the tuba.
Neighbours complained.
 I don’t play the tuba…

Unless I have a concert.

Statements, axioms, ordinary premises

Statements

Facts, beliefs, derivations, …

Axioms (necessary premises)

Facts

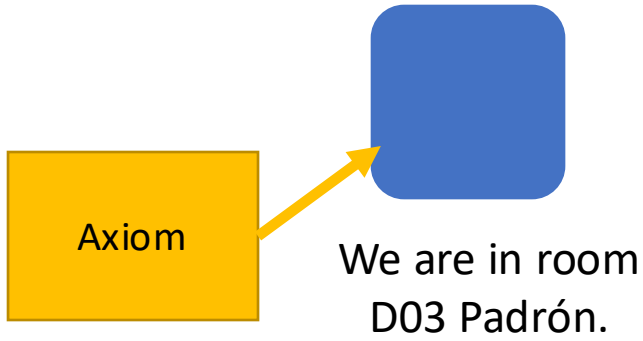
We are in room D03 Padrón.

Ordinary premises

Things you believe

Room D03 Padrón is beautiful.

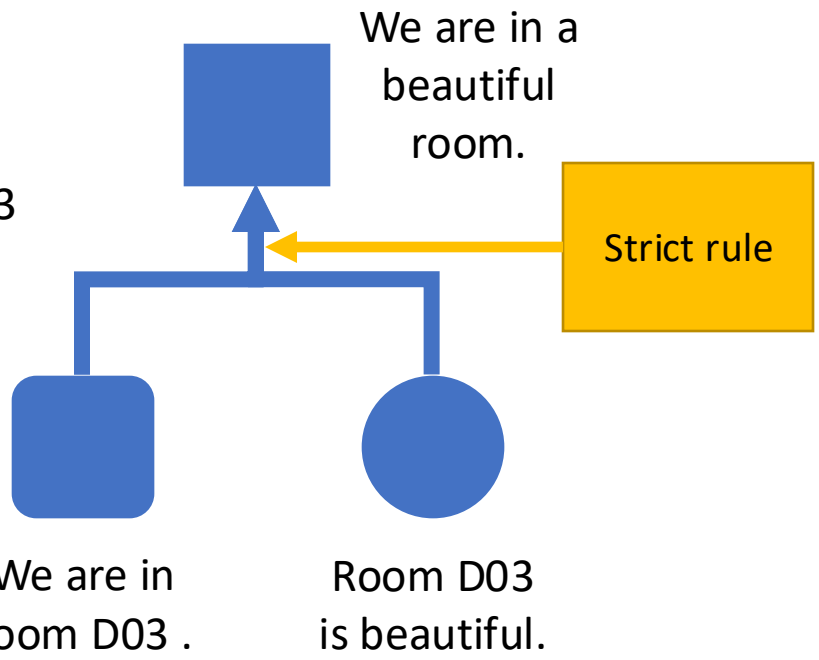
Constructing arguments



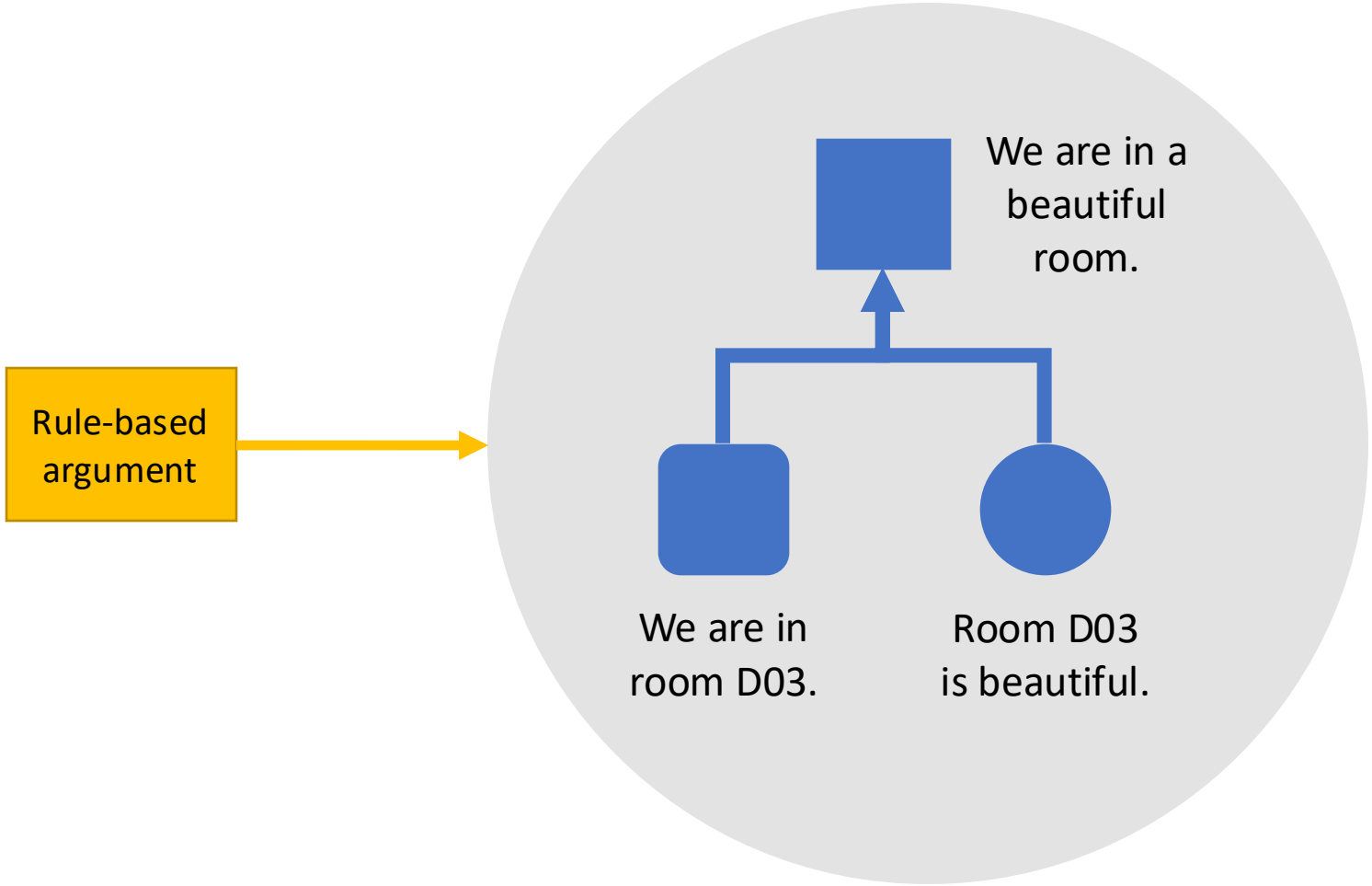
Constructing arguments

Constructing arguments

If we are in room D03 and room D03 is beautiful,
then we are in a beautiful room.



Constructing arguments



Structured argumentation (ASPIC$^+$): An argument consists of a conclusion deductively and/or defeasibly inferred from some premises.

e.g.

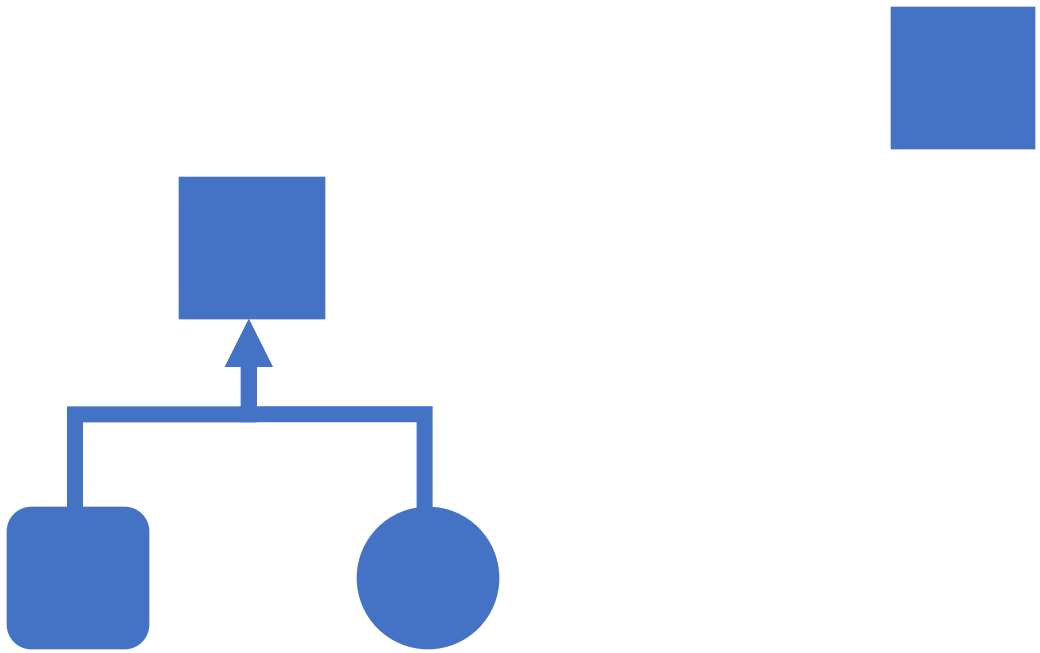
(conclusion)	b
(inference rule)	$a \Rightarrow b$
(premise)	a

Take an umbrella with you.

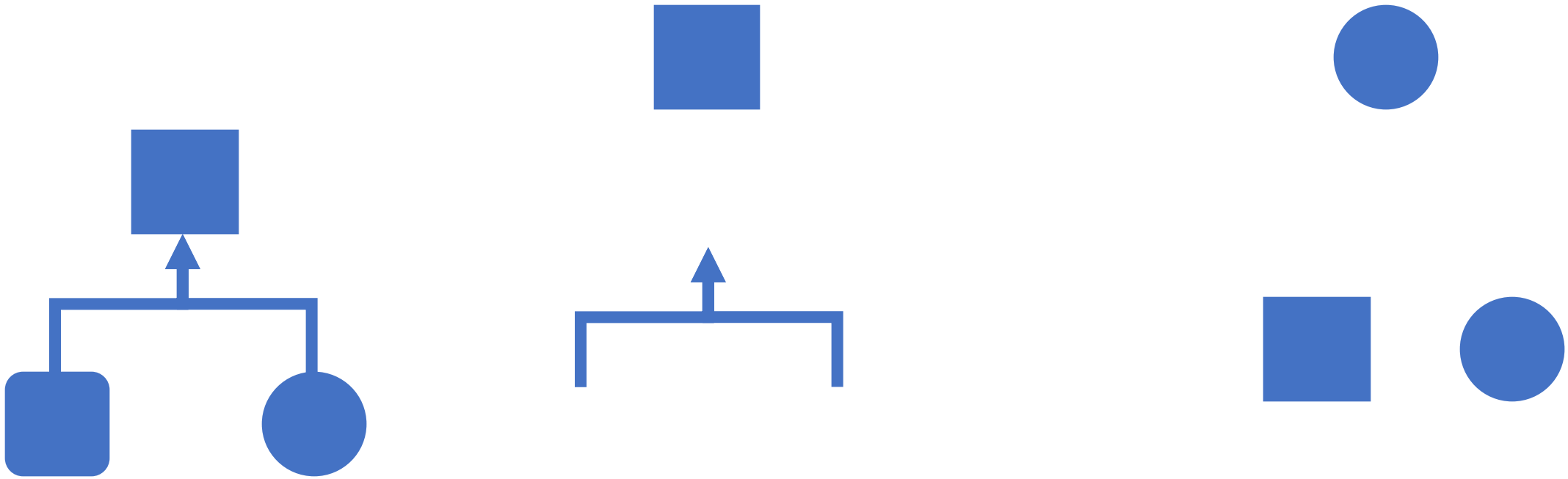
If it's raining outside, then take an umbrella with you.

It's raining outside.

Enthymemes



Enthymemes



Arguments with **incomplete logical structure**.

Modelling enthymemes

Argument from previous slide:

b
|
a ⇒ b
|
a

1.

b
|
a ⇒ b

If it's raining outside, then take an umbrella with you.
Take an umbrella with you.

2.

a ⇒ b
|
a

It’s raining outside.
If it's raining outside, then take an umbrella with you.

3.

b

a

It’s raining outside.
Take an umbrella with you.

4.

a

It’s raining outside.

5.

b

Take an umbrella with you.

6.

a ⇒ b

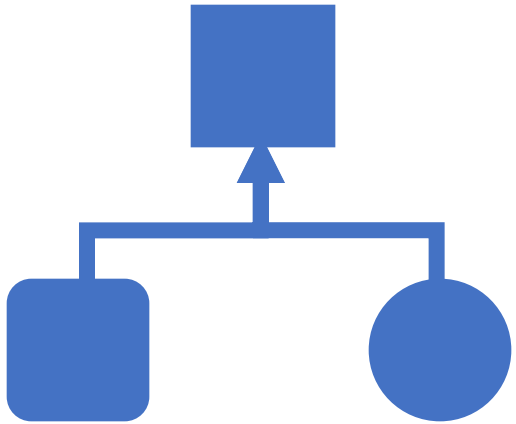
If it's raining outside, then take an umbrella with you.

Enthymemes

Backward extension

Enthymemes

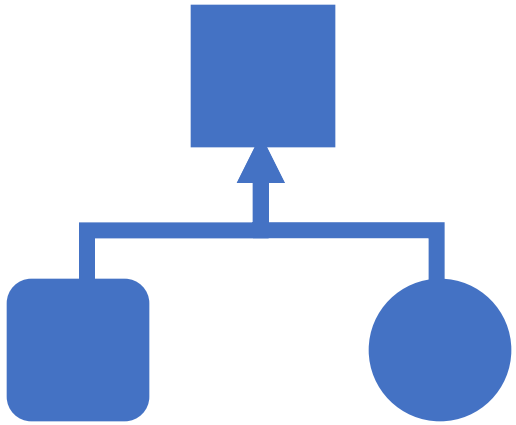
Backward extension

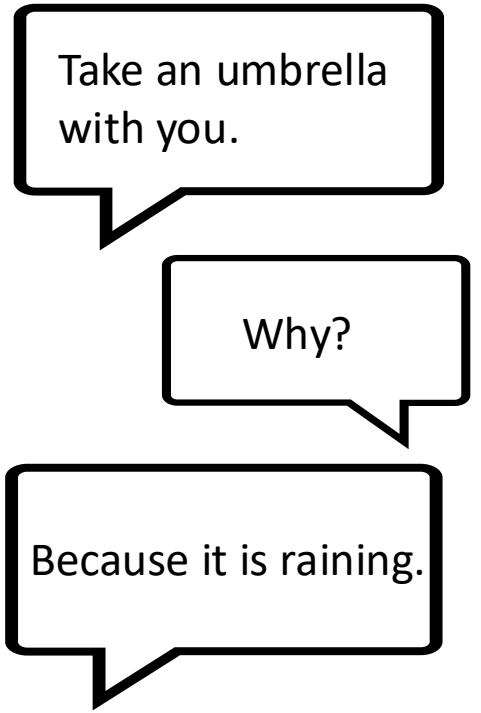


Enthymemes

Backward extension

Locutions: *why*, *because*





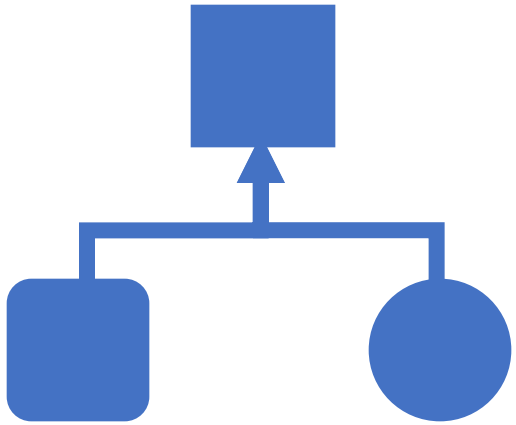
Enthymemes

Enthymemes

Forward extension

Enthymemes

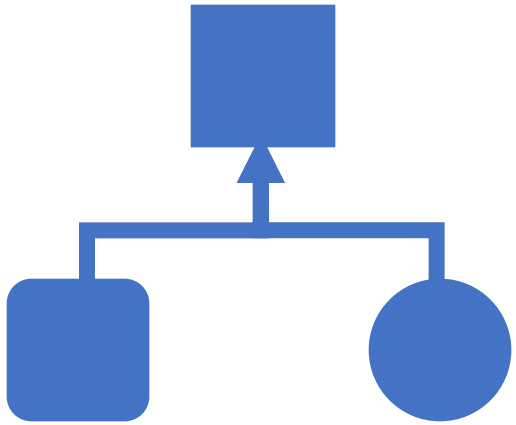
Forward extension

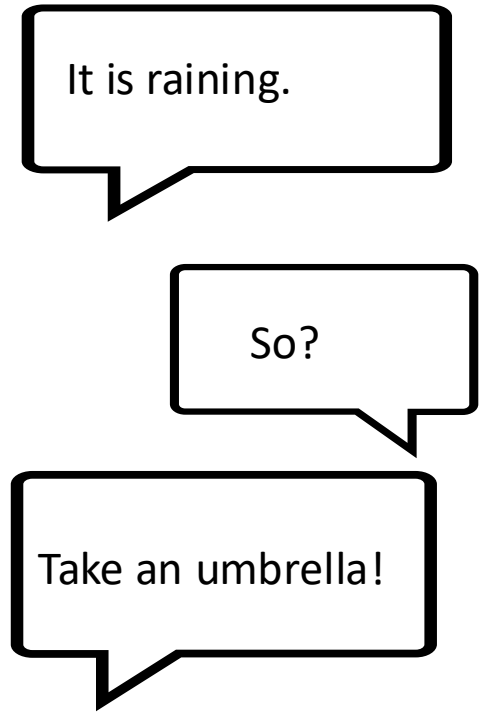


Enthymemes

Forward extension

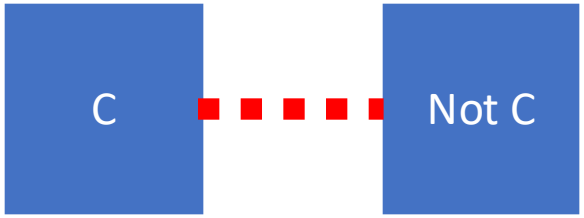
Locutions: *and-so*, *hence*



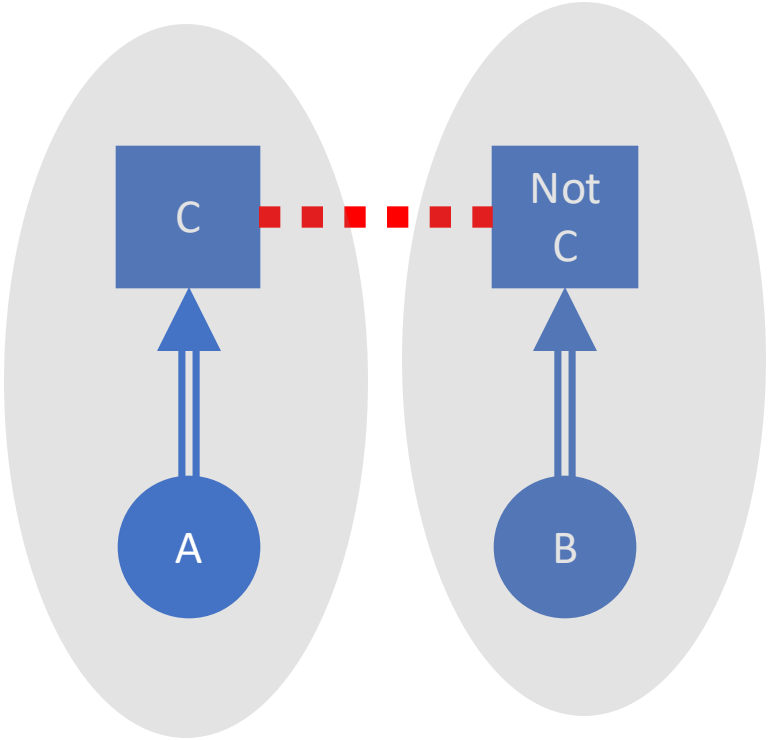


Attacks

C: The pencil is green.



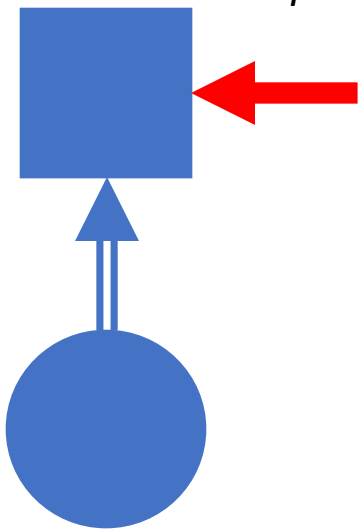
Attacking Arguments

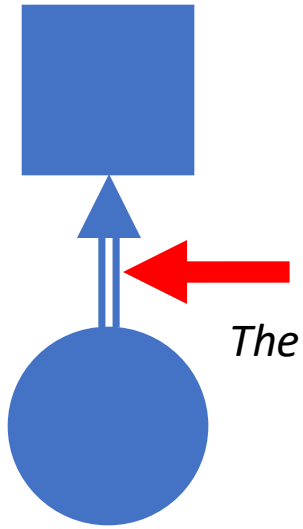


A: Elfia sees that the pencil is green.
B: Maddie sees that the pencil is not green.
r: If I see the pencil color X, then it is color X.

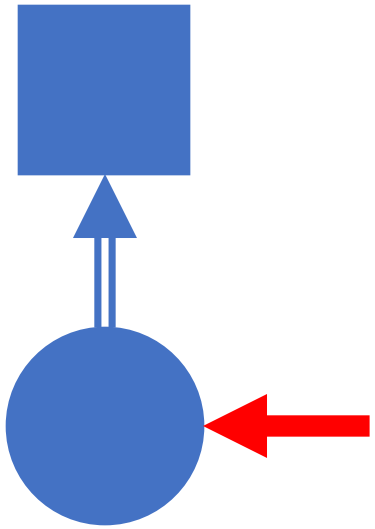
Points of attack

The pencil is not green.



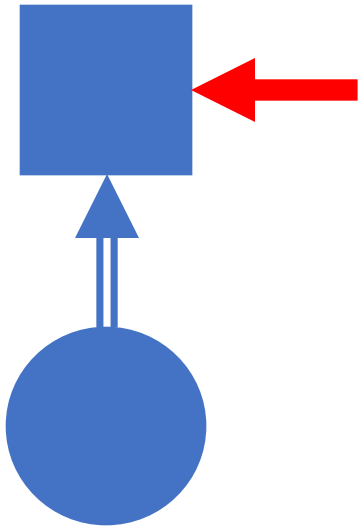


The pencil is lit with a green light.

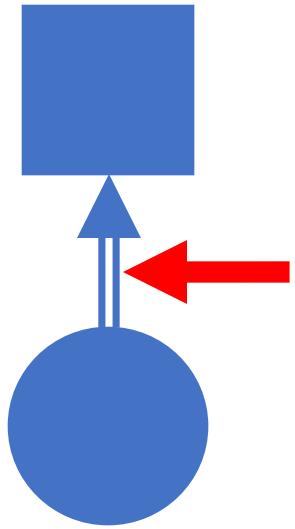


You don’t see the pencil green (eg, you are lying).

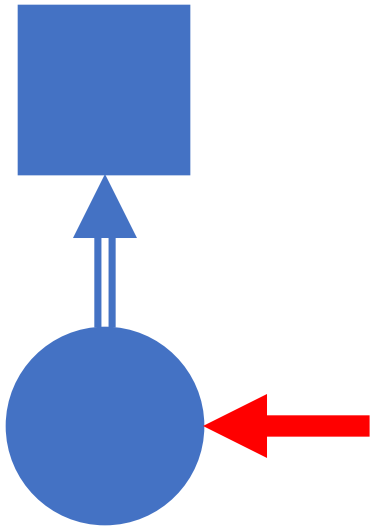
Points of attack



Rebuttal

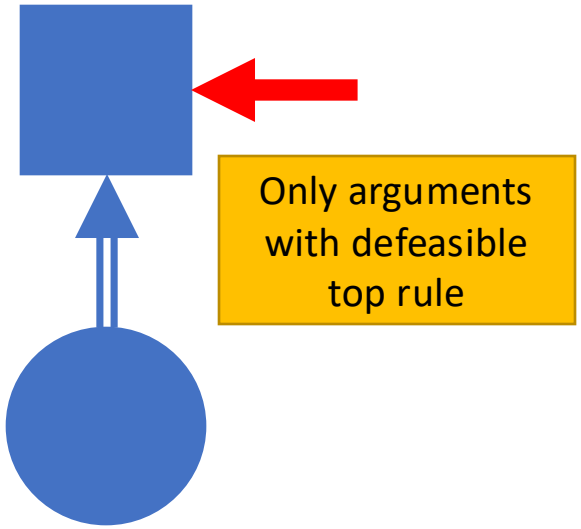


Undercutting

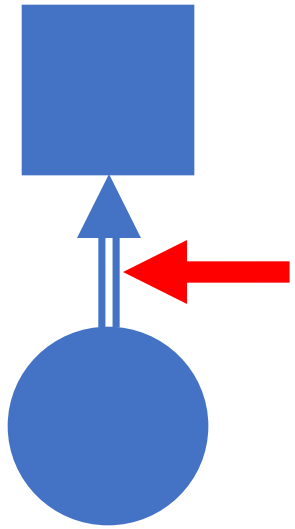


Undermining

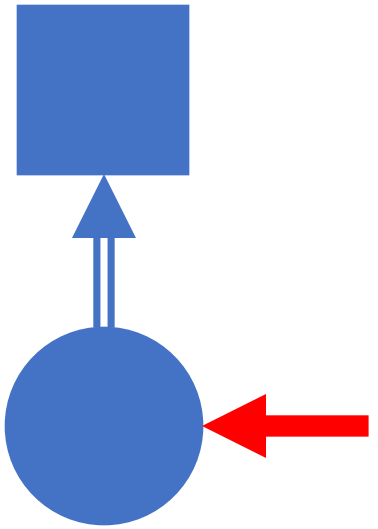
Points of attack



Rebuttal

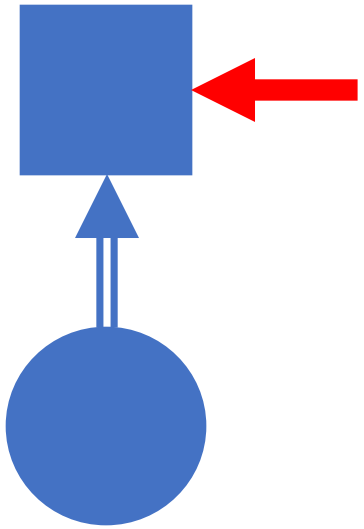


Undercutting

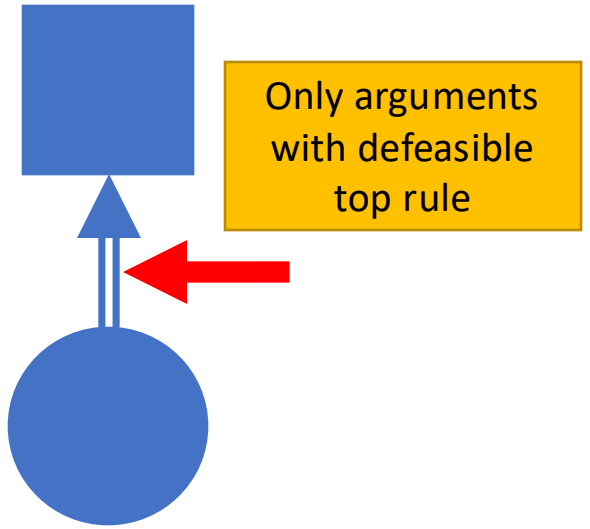


Undermining

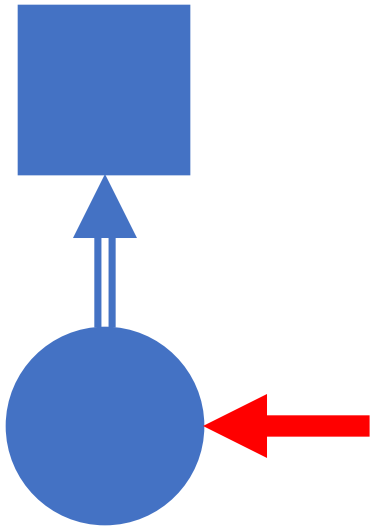
Points of attack



Rebuttal

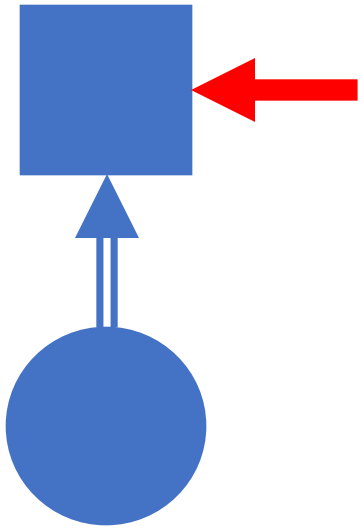


Undercutting

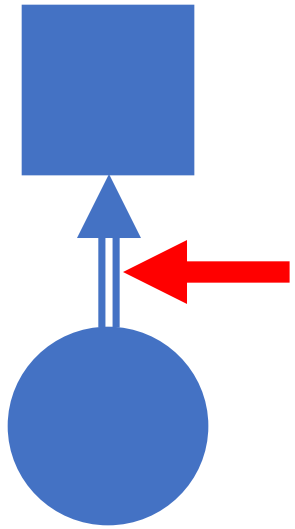


Undermining

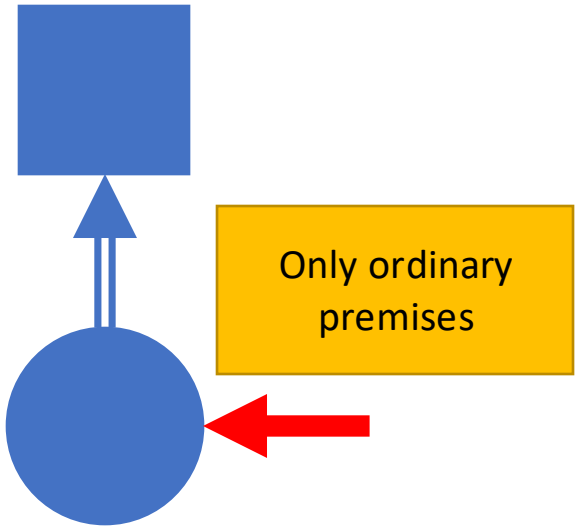
Points of attack



Rebuttal

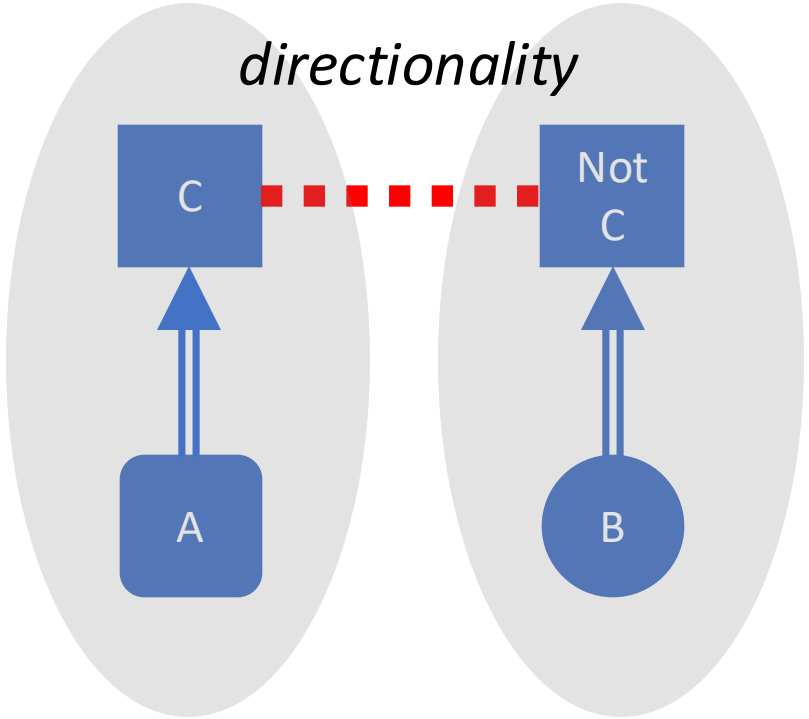


Undercutting



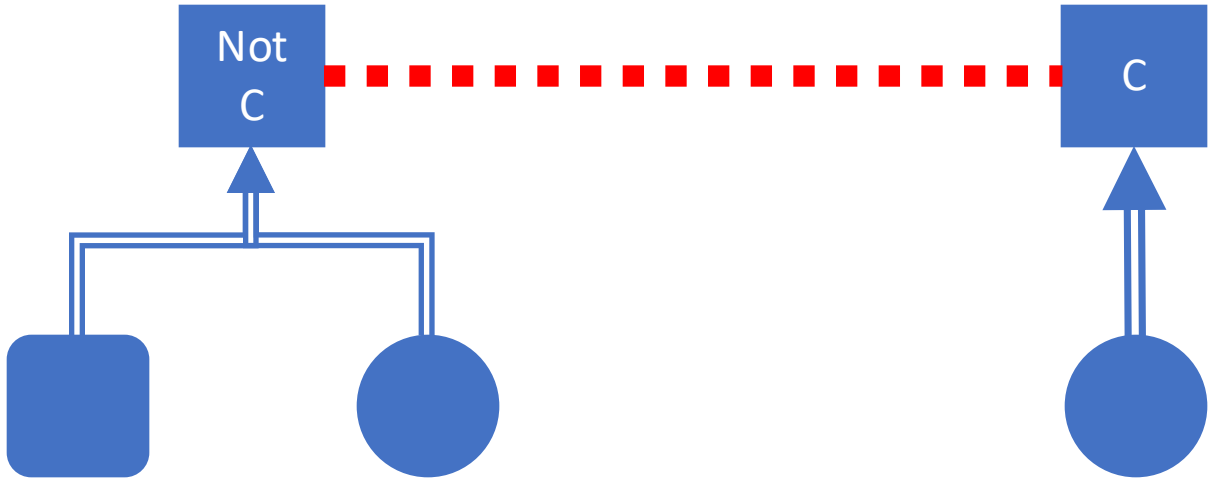
Undermining

Attacking Arguments



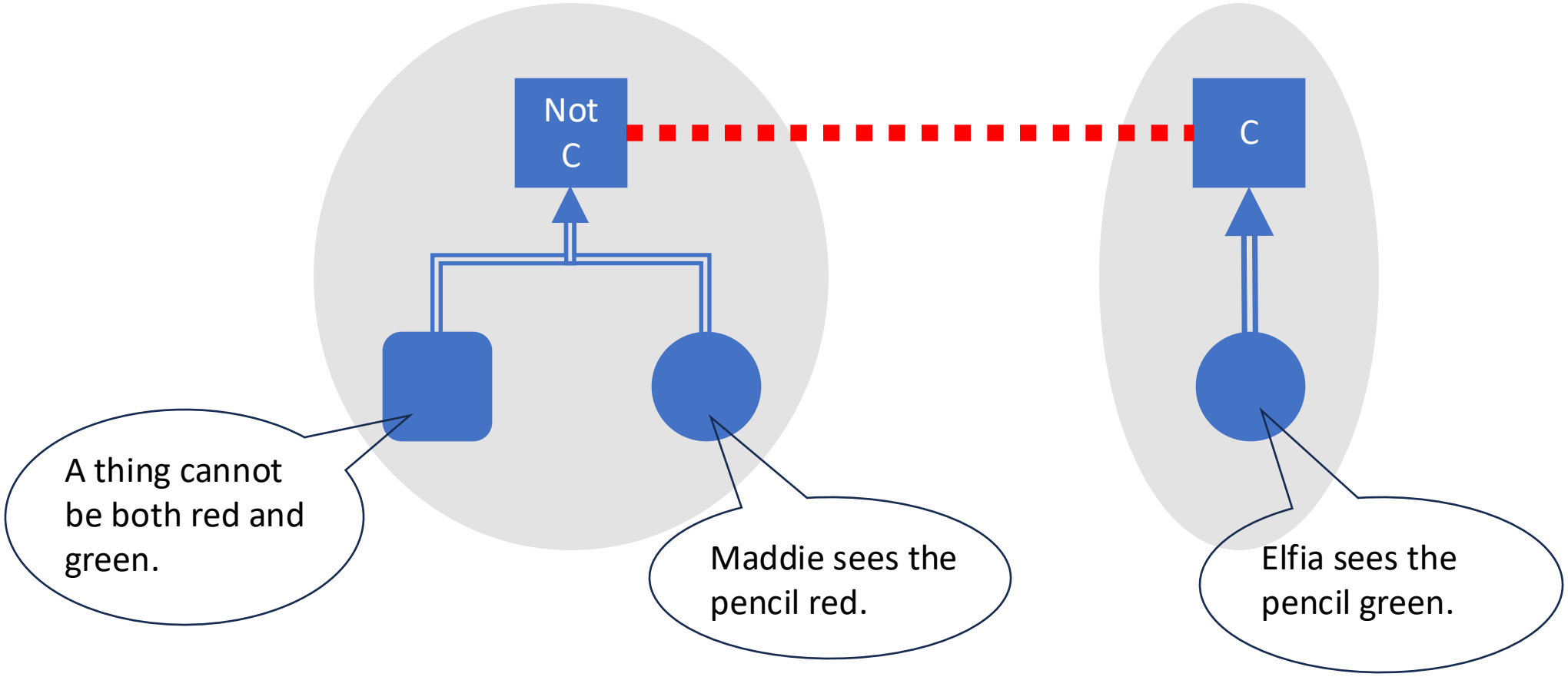
How do we resolve?

Rebuttal attack (symmetric)

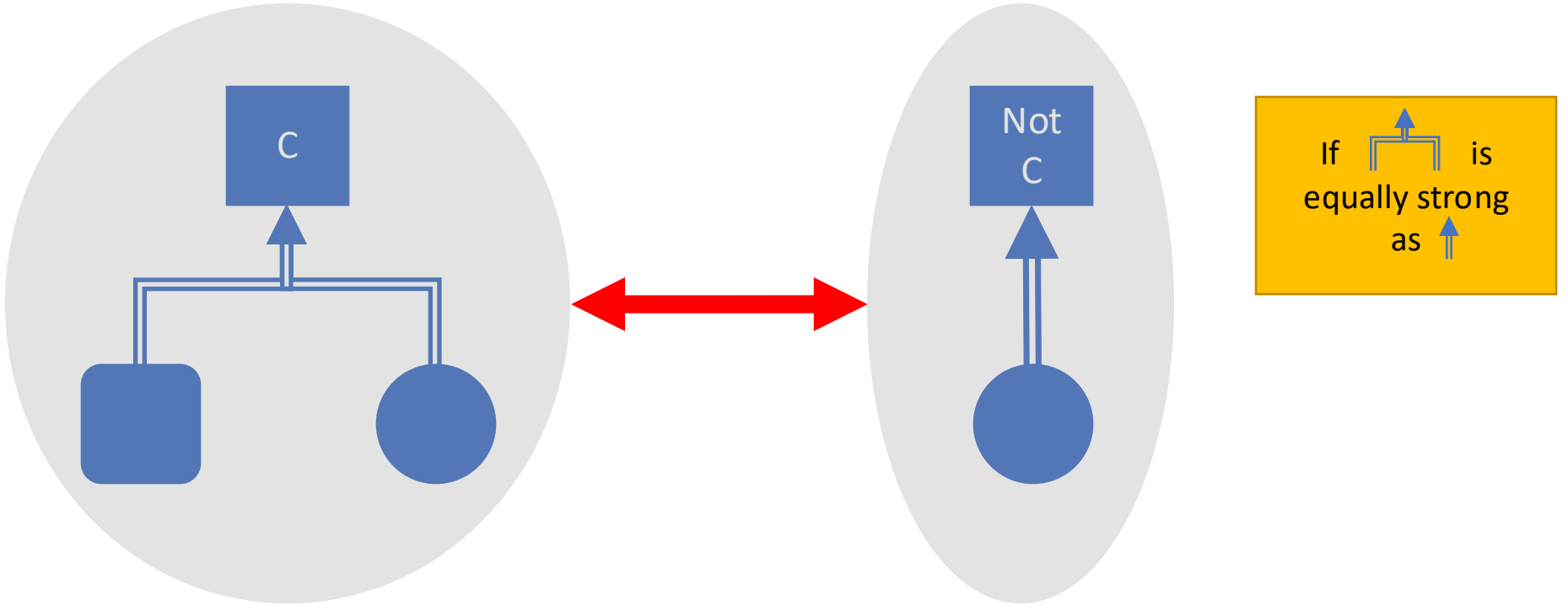


Rebuttal attack (symmetric)

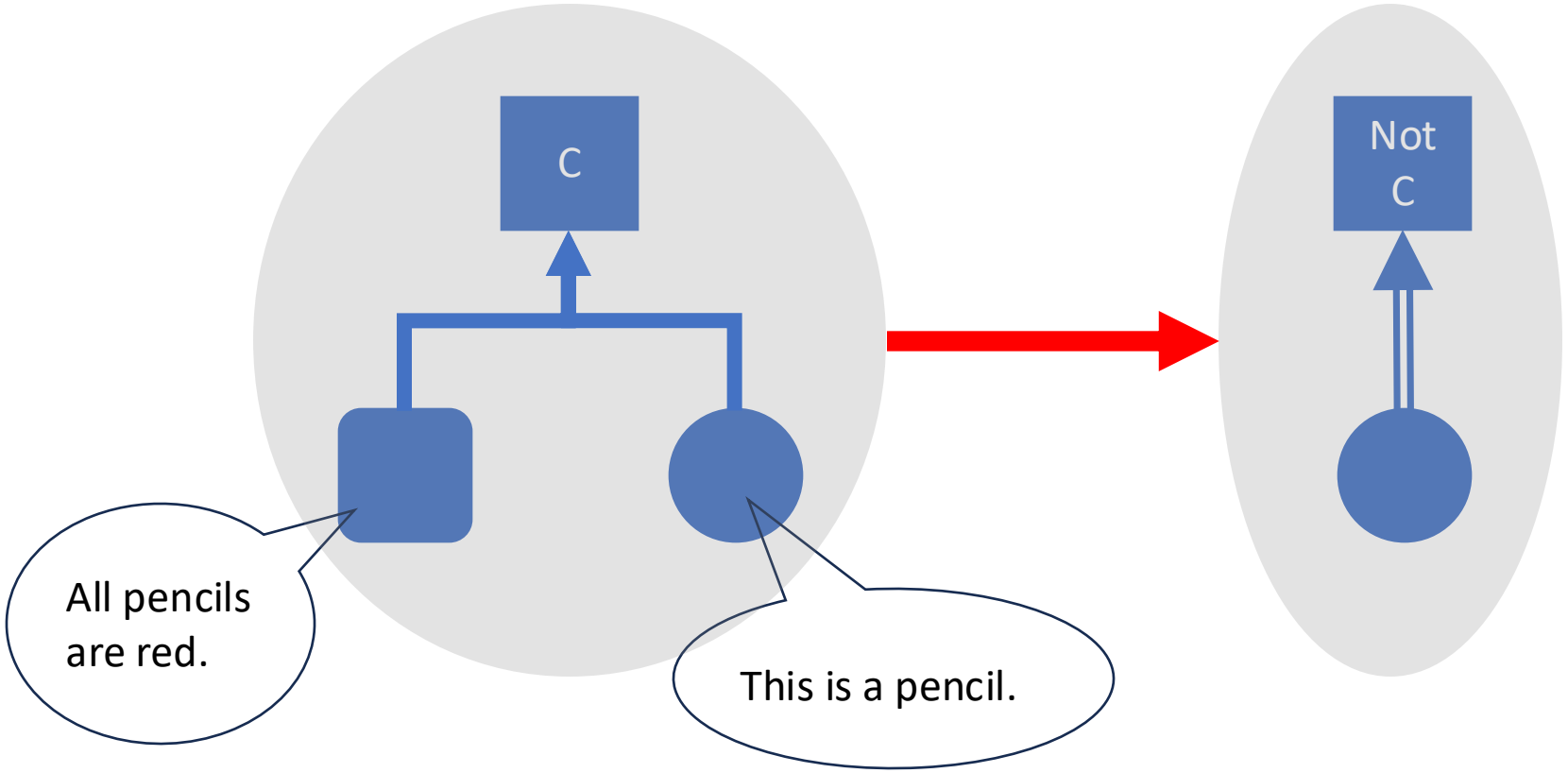
C: The pencil is green.



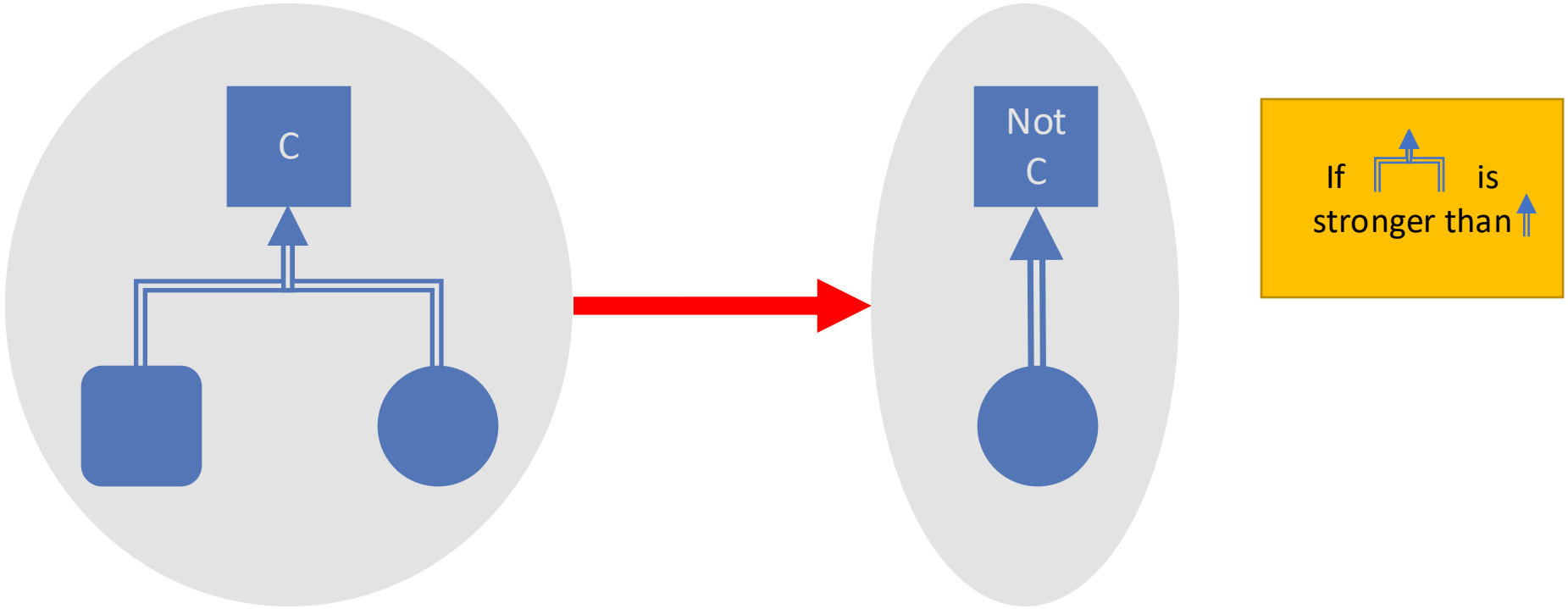
Rebuttal attack (symmetric)



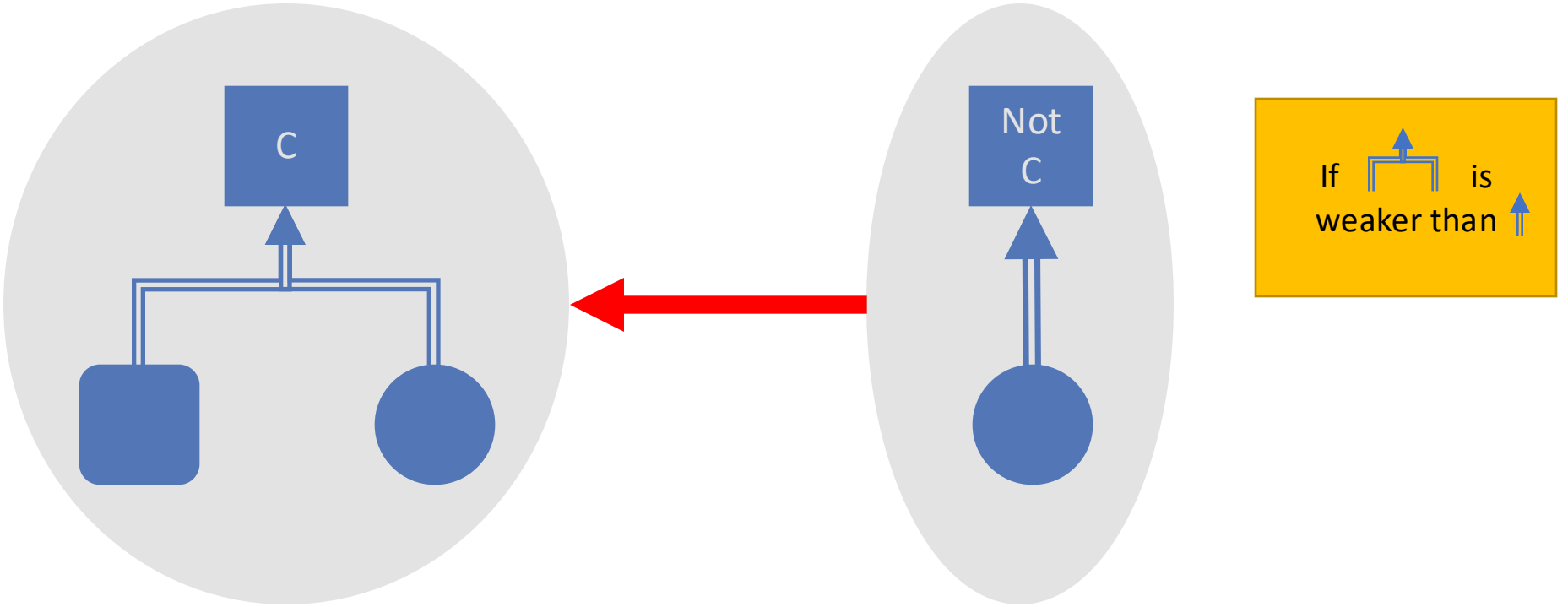
Rebuttal attack (asymmetric)



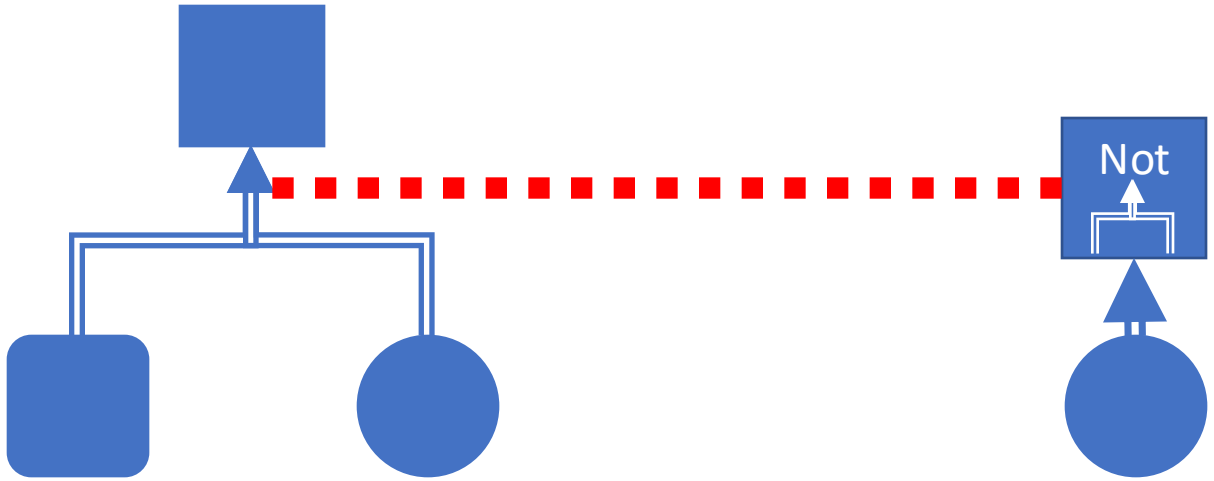
Rebuttal attack (asymmetric)



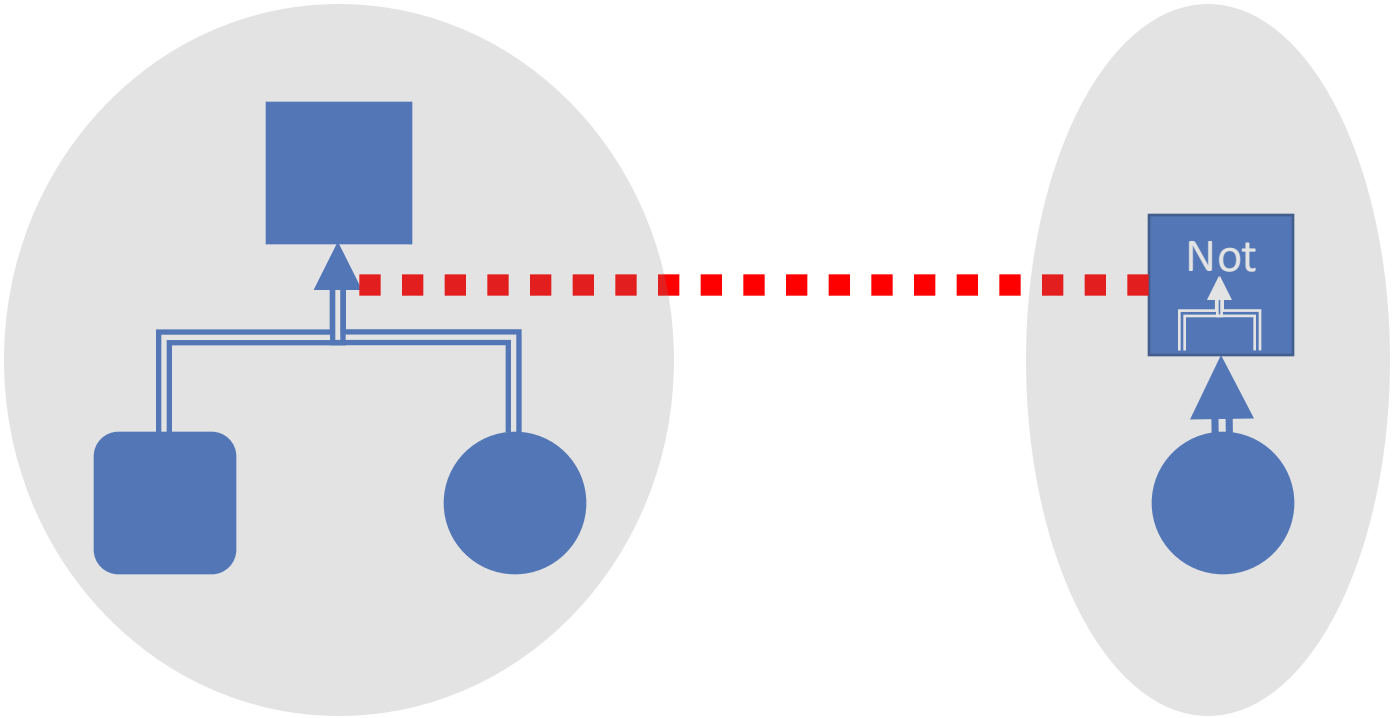
Rebuttal attack (asymmetric)



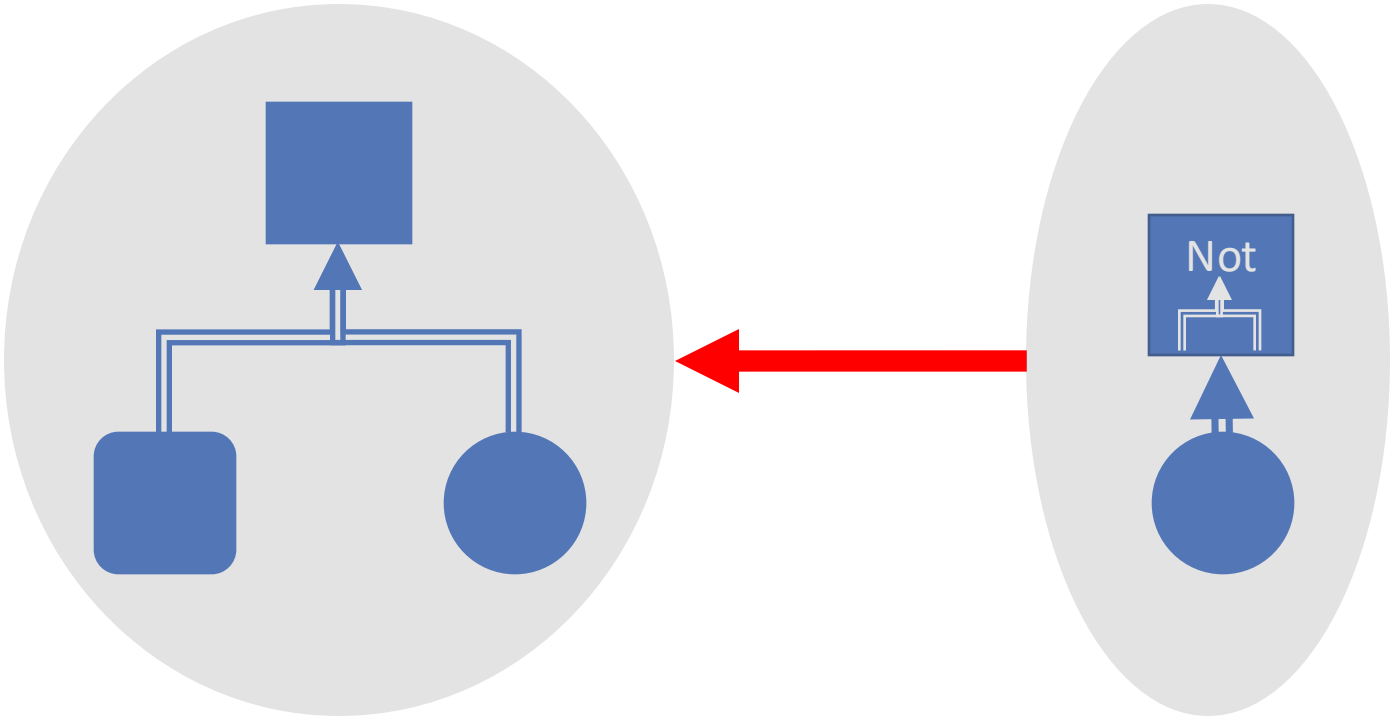
Undercutting attack (always asymmetric)



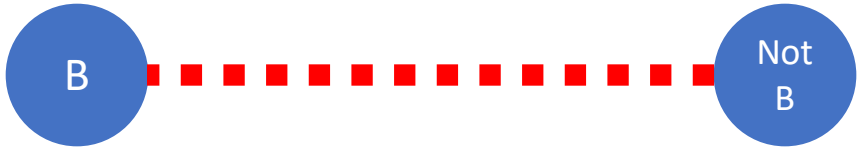
Undercutting attack (always asymmetric)



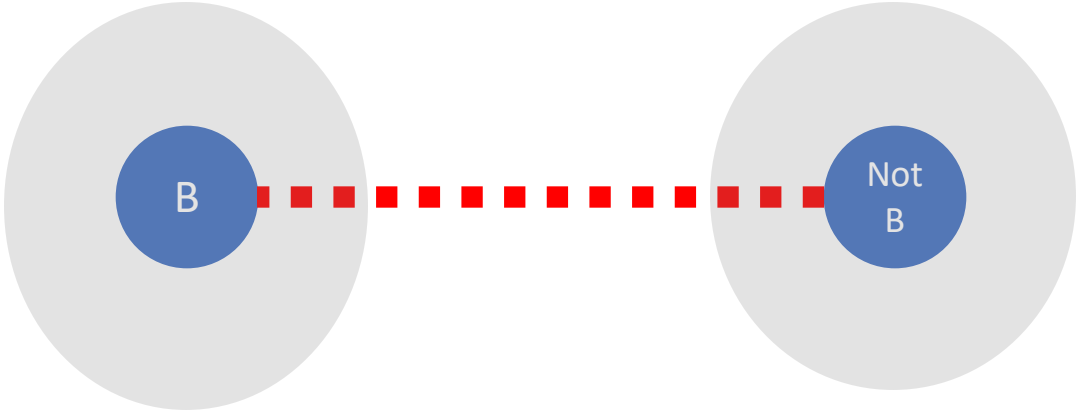
Undercutting attack (always asymmetric)



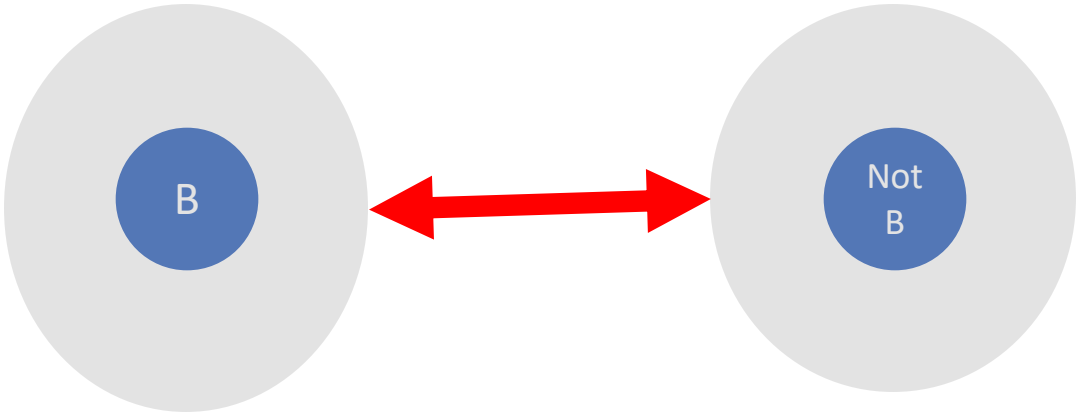
Undermining attack (symmetric)



Undermining attack (symmetric)

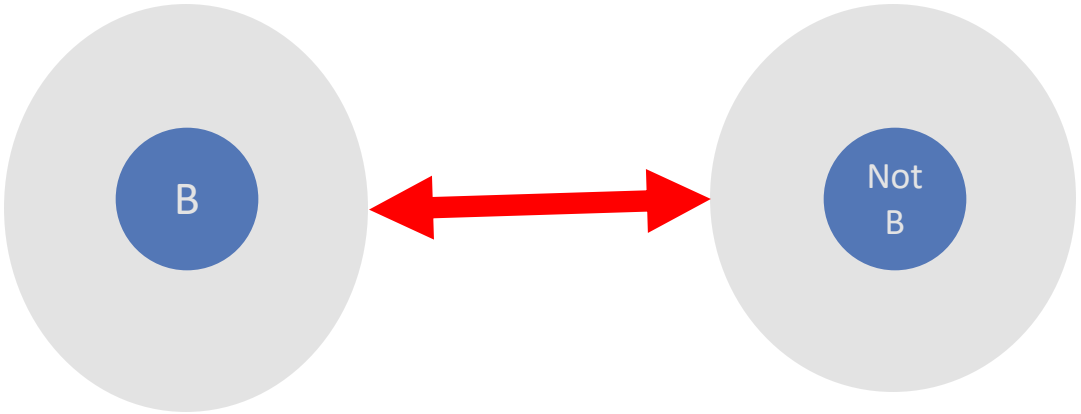


Undermining attack (symmetric)



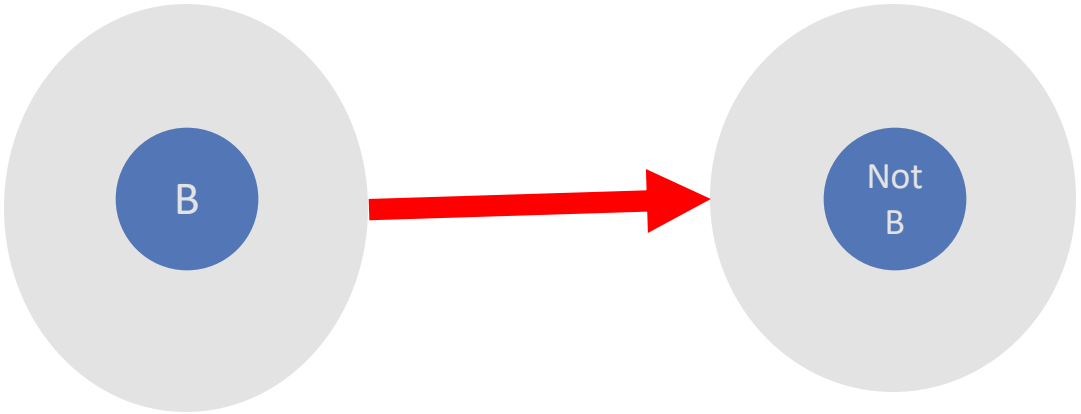
Undermining attack (symmetric)

If “B” is equally strong as “Not B”



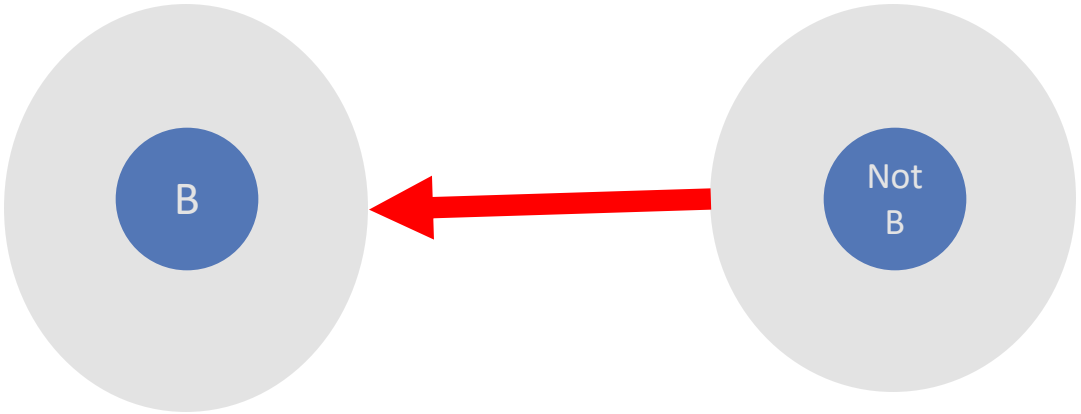
Undermining attack (asymmetric)

If “B” stronger than “Not B”

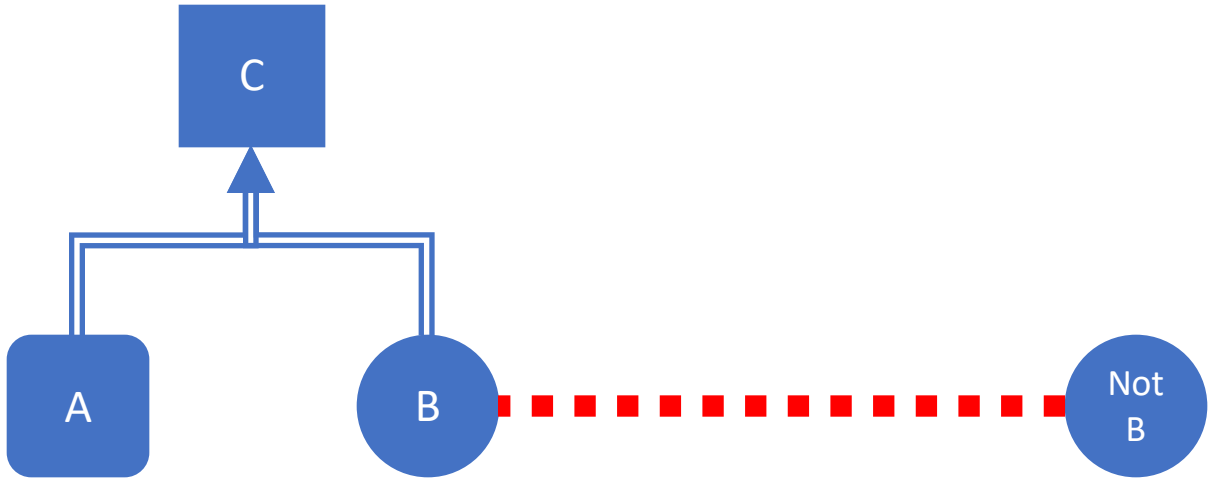


Undermining attack (asymmetric)

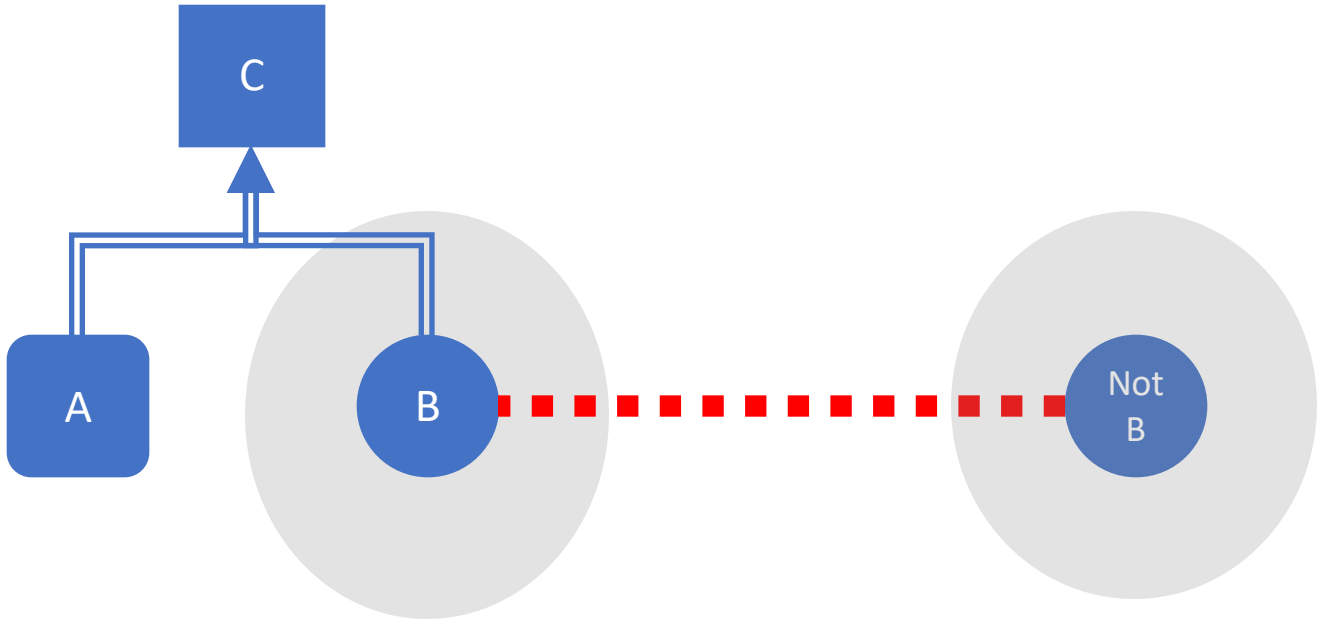
If “B” weaker than “Not B”



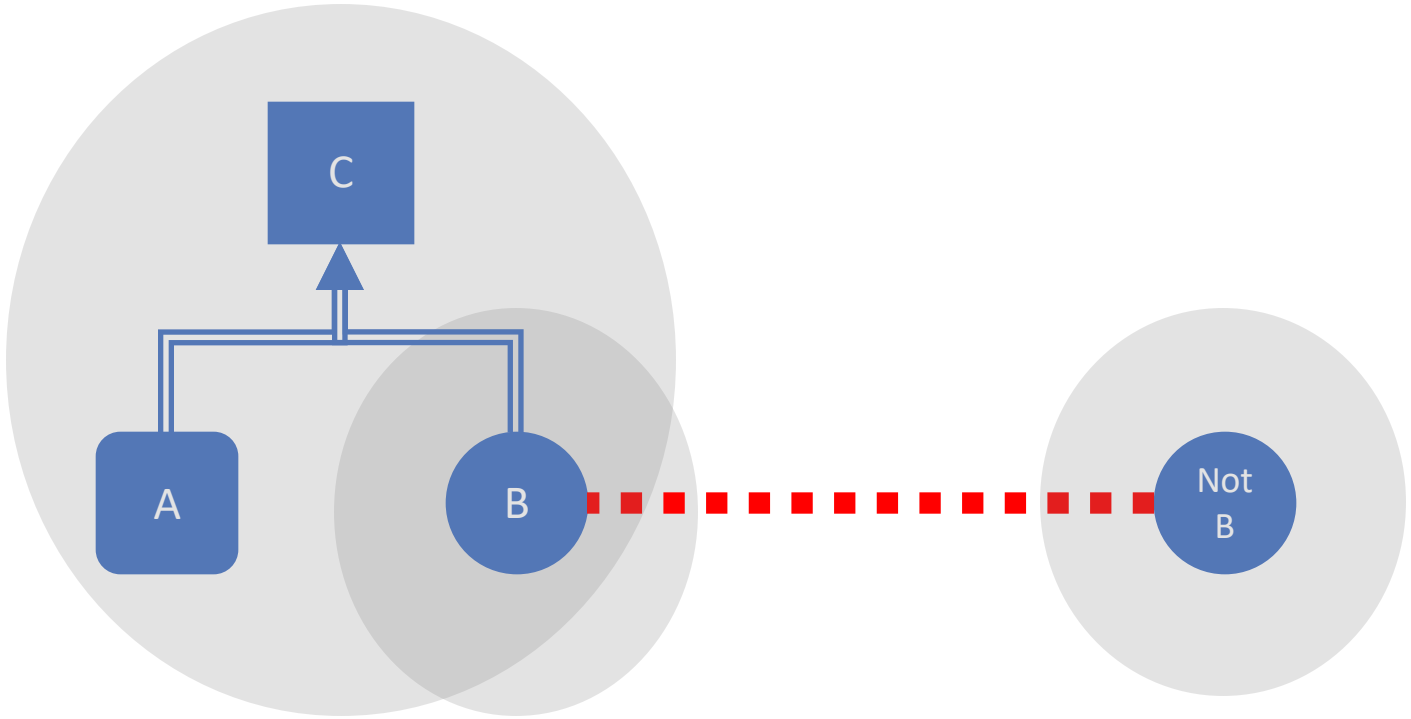
Indirect attacks



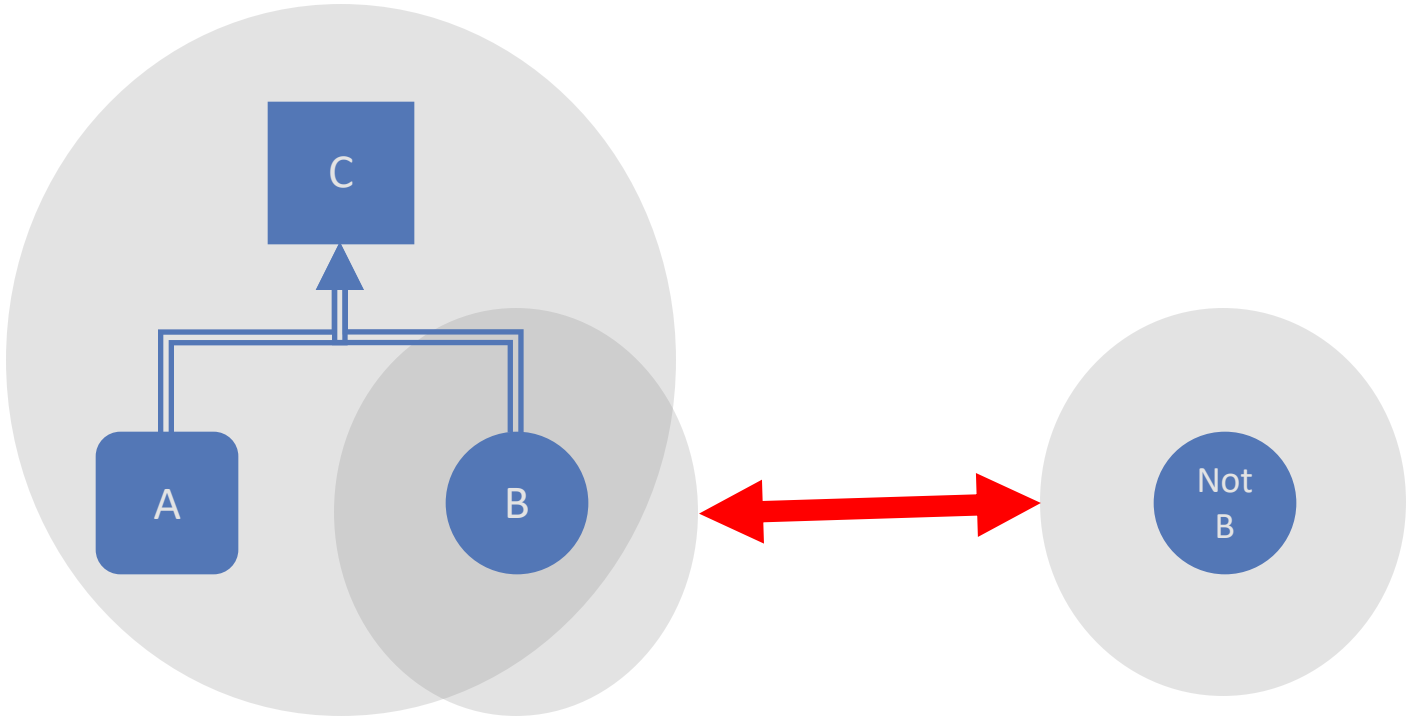
Indirect attacks



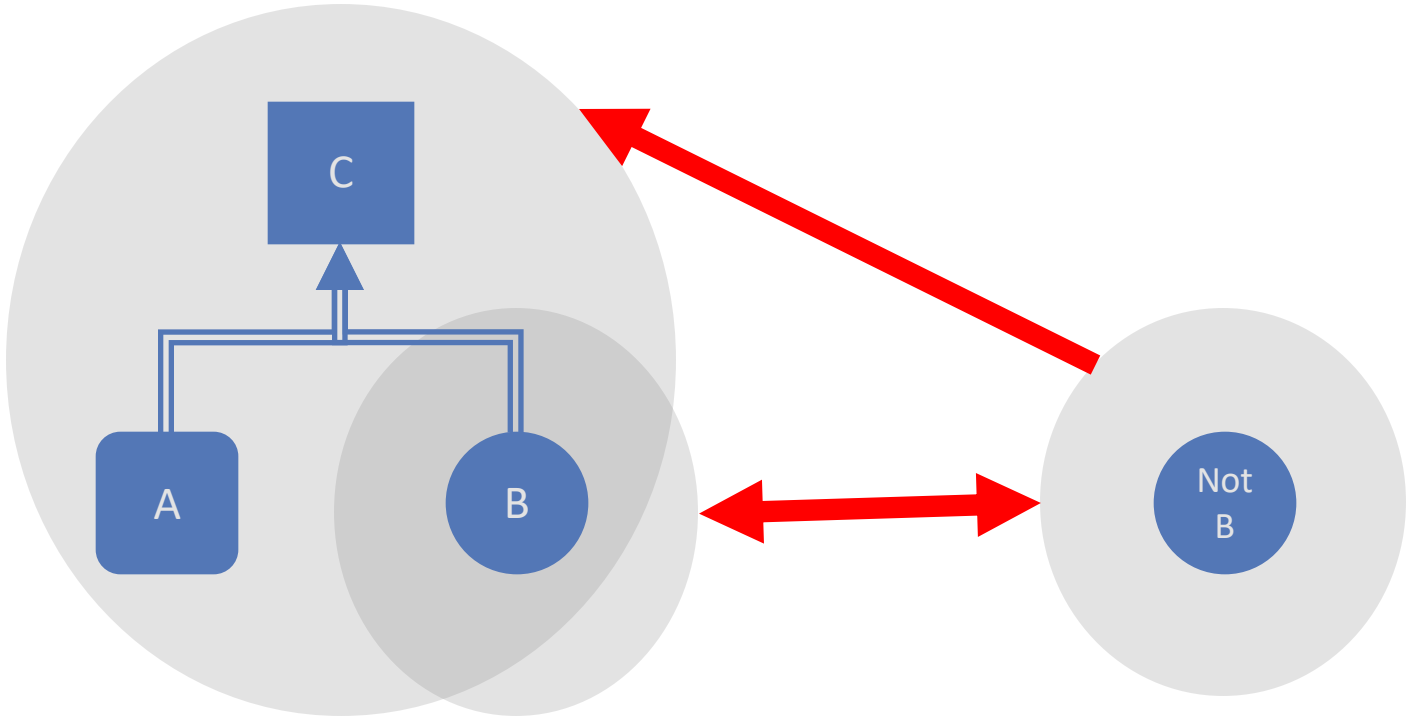
Indirect attacks



Indirect attacks



Indirect attacks



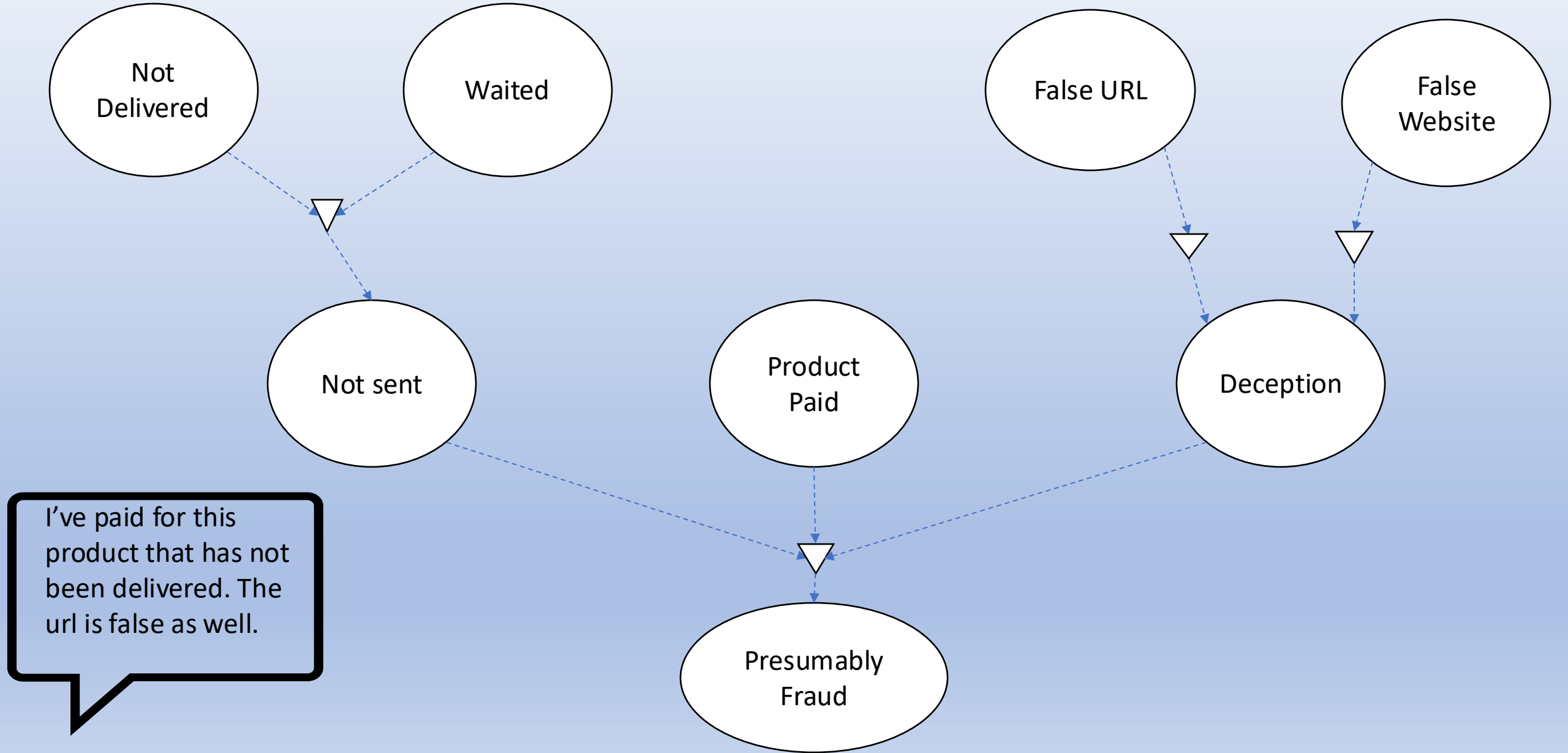
Quest Time!

Go back to your arguments and discuss within your groups:

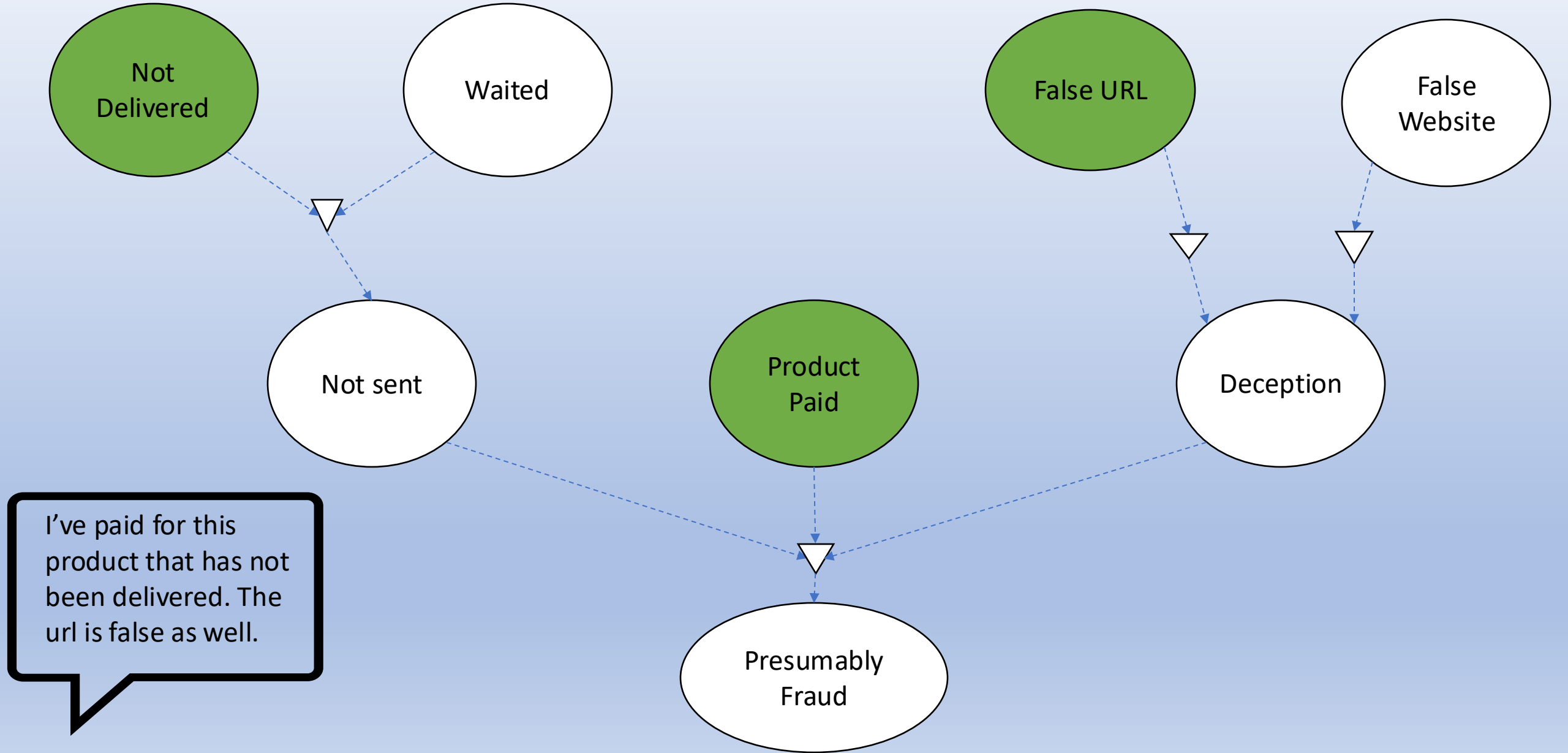
- Can you represent them as premise, rule, conclusion?
- Did you use enthymemes?
  - How did you find out the missing elements?
- How (where) do they attack each other (which attack types)?
  - Rebuttal (attacking conclusion)
  - Undercutting (attacking rule)
  - Undermining (attacking premises)

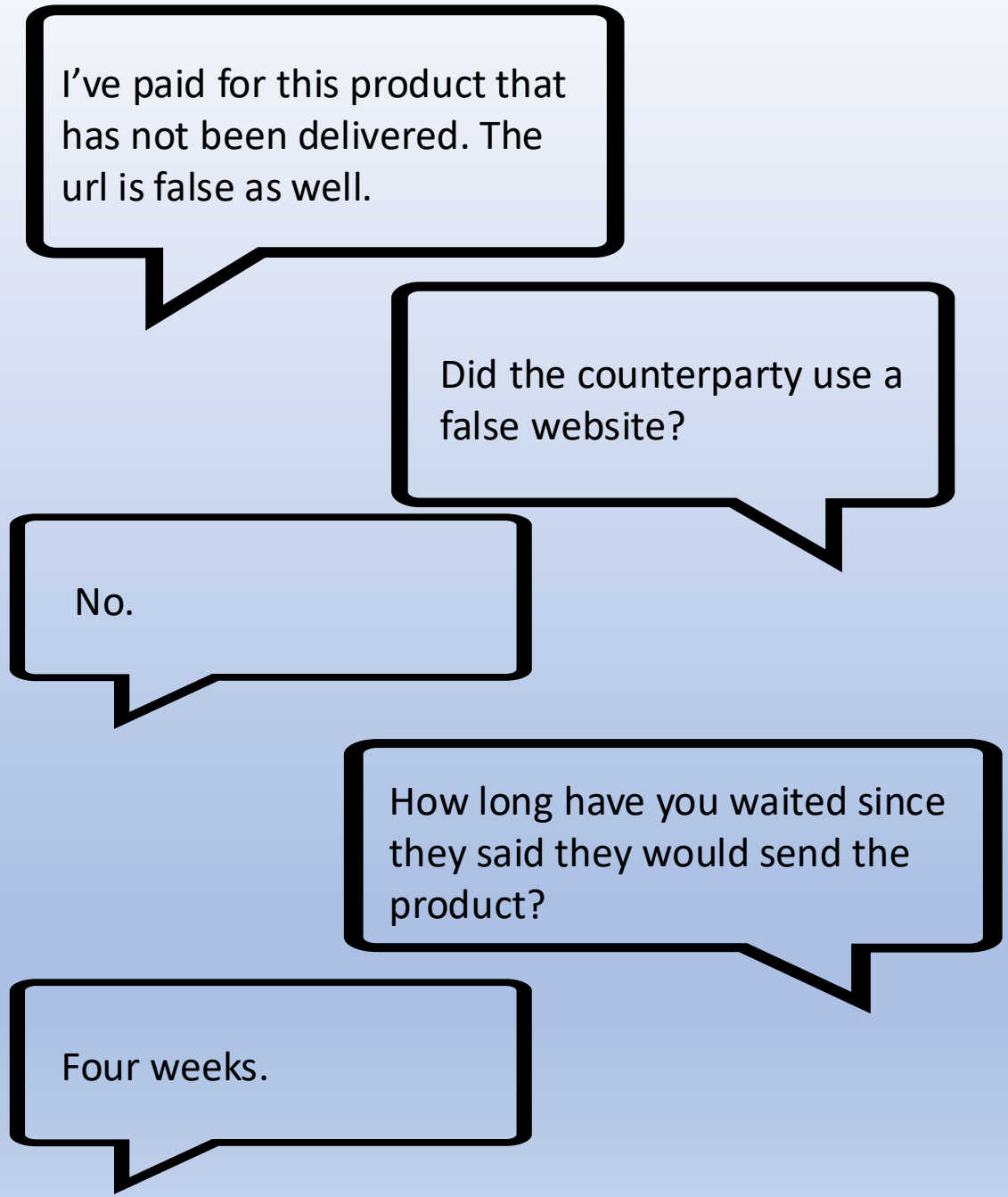

FROM THEORY TO PRACTICE:
ARGUMENTATION IN ACTION
POLICE
POLICE
POLICE

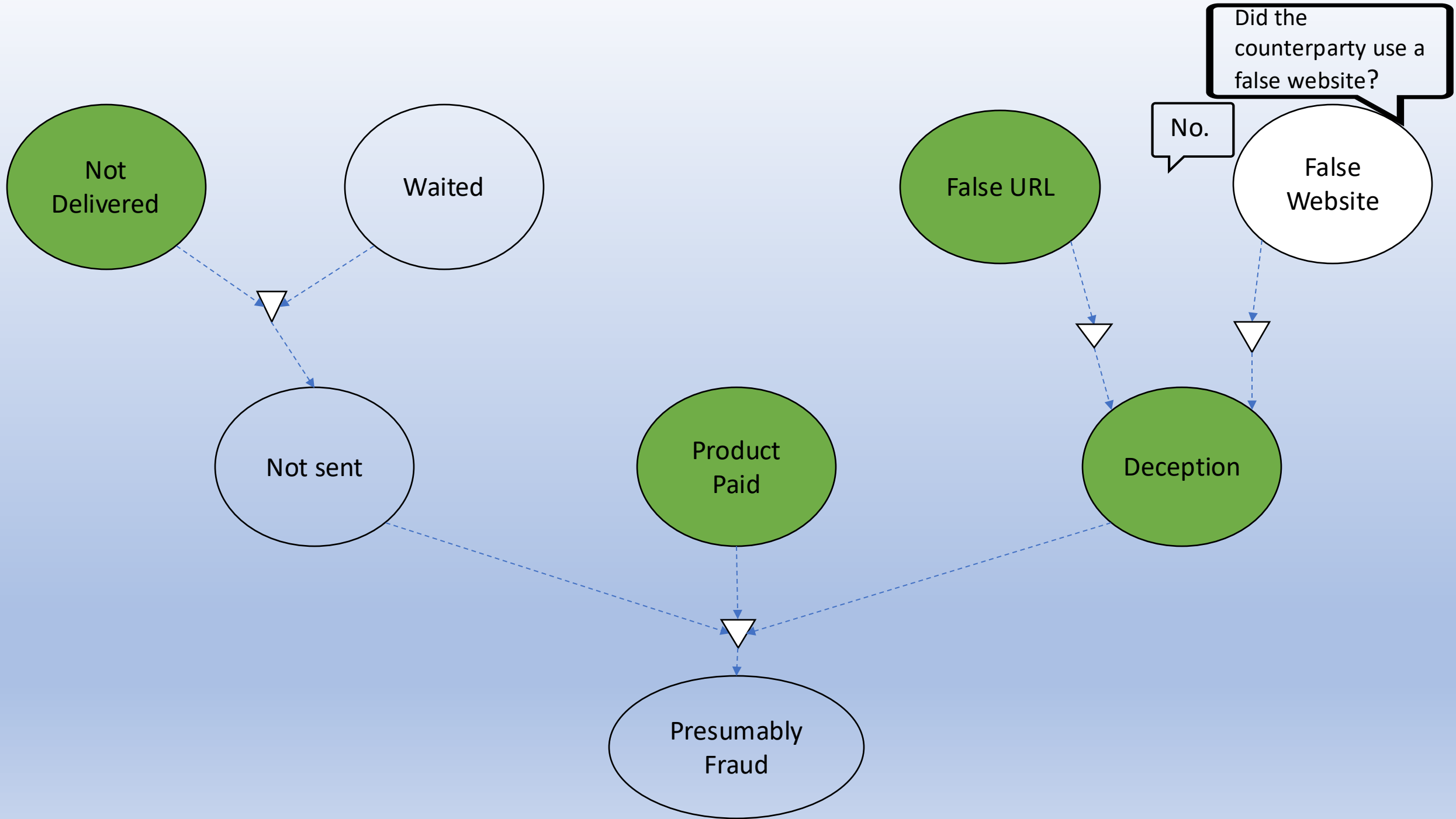
Application: Dutch Police

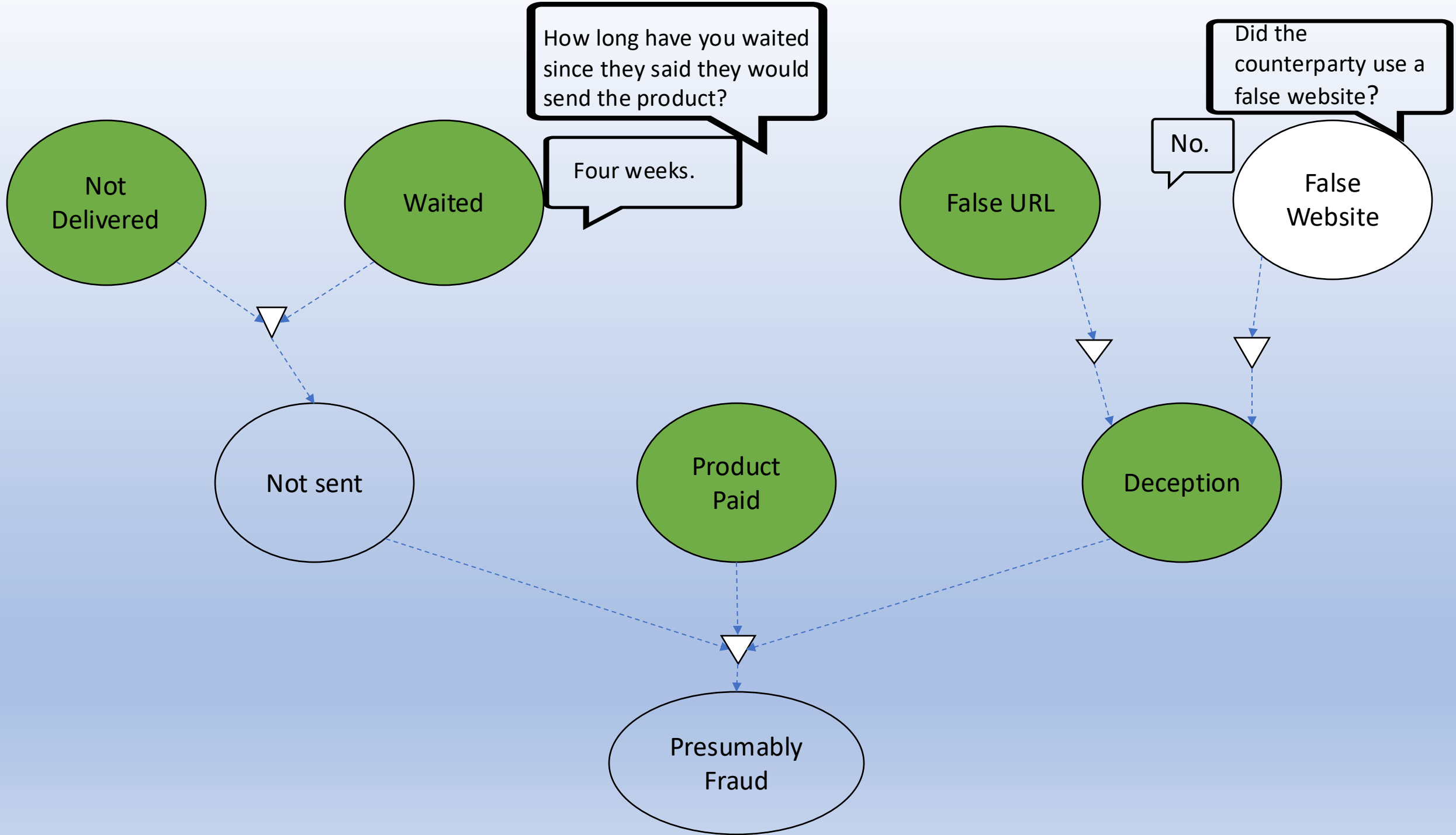


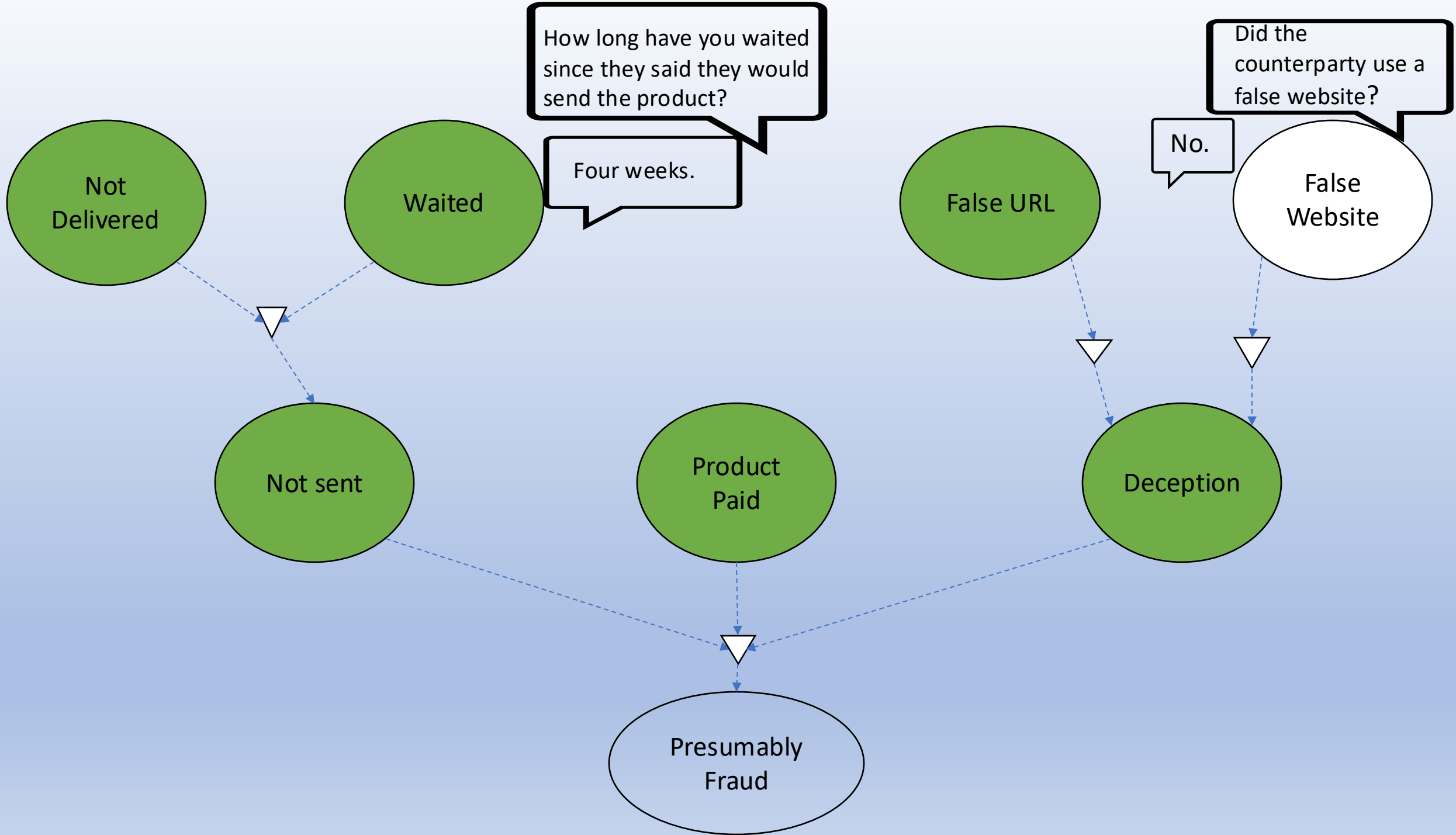
Application: Dutch Police

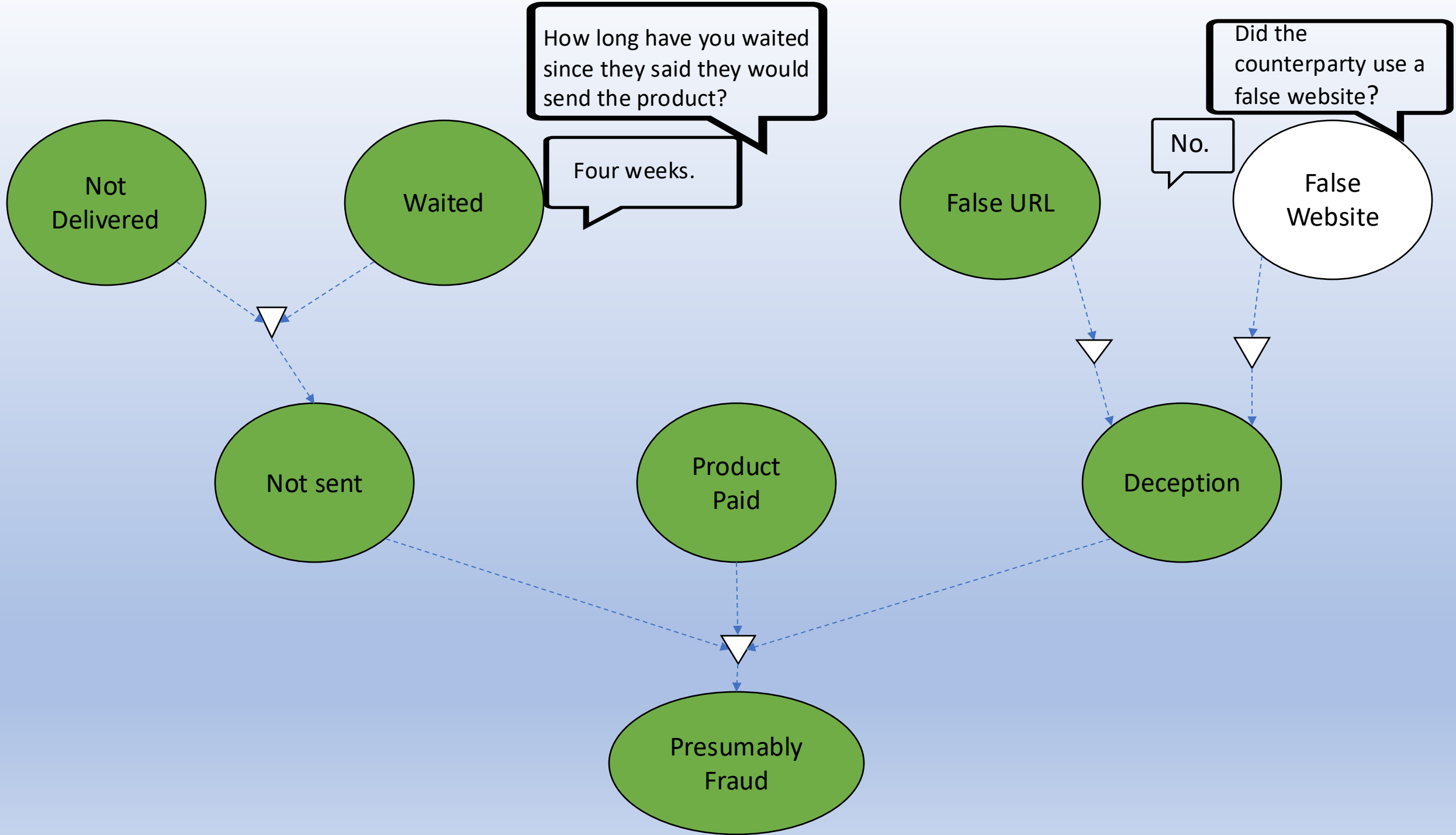


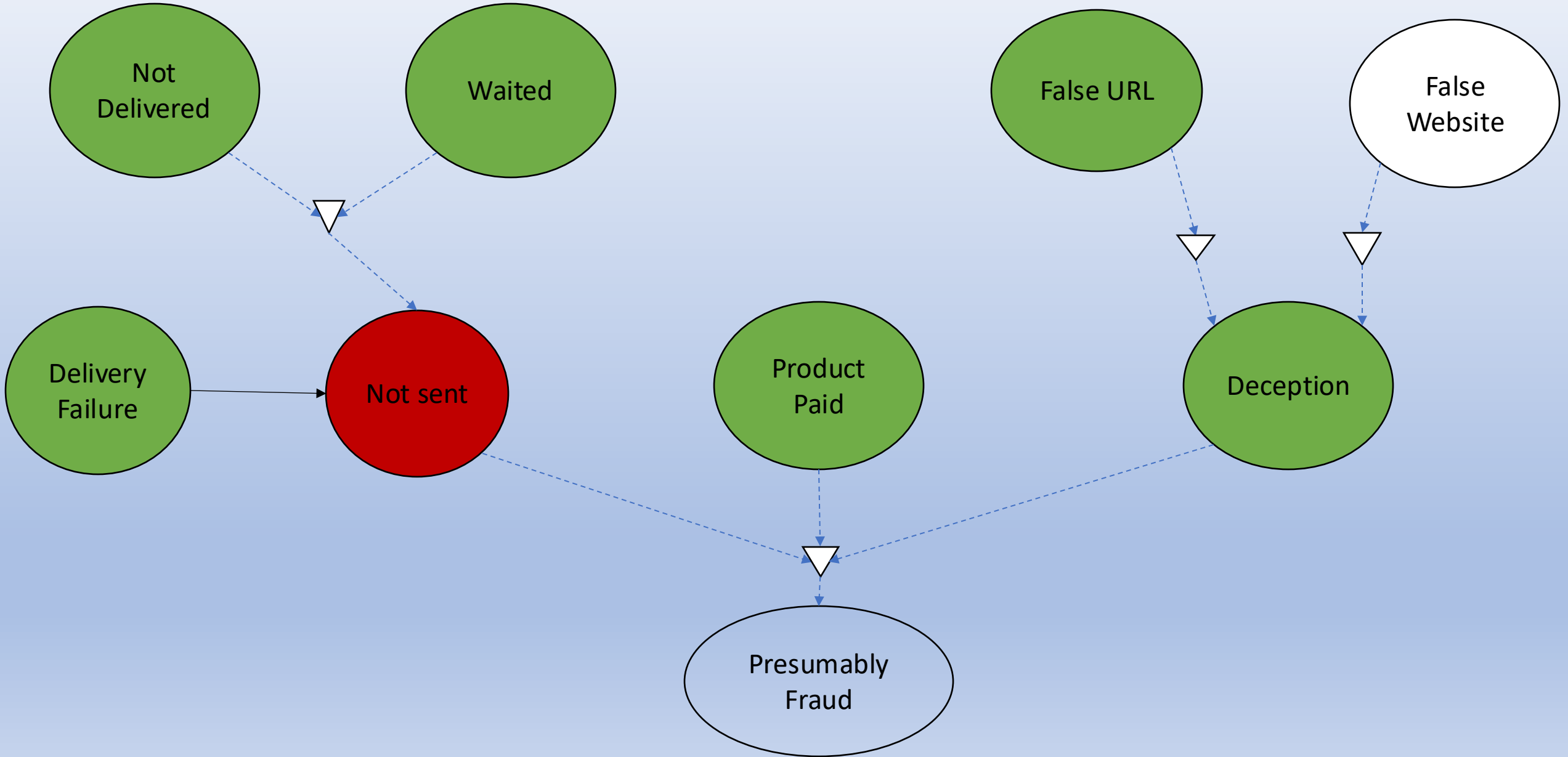

I’ve paid for this product that has not been delivered. The url is false as well.
Did the counterparty use a false website?
No.
How long have you waited since they said they would send the product?
Four weeks.

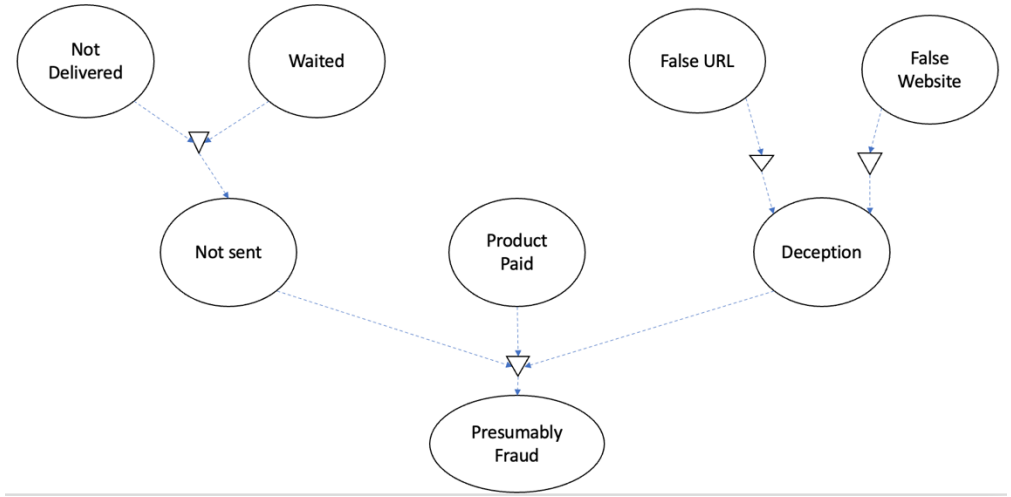

Not Delivered
Waited
Not sent
Product Paid
False URL
False Website
Did the counterparty use a false website?
No.
Deception
Presumably Fraud


How long have you waited since they said they would send the product?
Four weeks.
Did the counterparty use a false website?
No.
Not Delivered
Waited
False URL
False Website
Not sent
Product Paid
Deception
Presumably Fraud


How long have you waited since they said they would send the product?
Four weeks.
Did the counterparty use a false website?
No.
Not Delivered
Waited
False URL
False Website
Not sent
Product Paid
Deception
Presumably Fraud


Not Delivered
Waited
How long have you waited since they said they would send the product?
Four weeks.
False URL
Did the counterparty use a false website?
No.
False Website
Not sent
Product Paid
Deception
Presumably Fraud


Not Delivered
Waited
False URL
False Website
Delivery Failure
Not sent
Product Paid
Deception
Presumably Fraud



https://pyarg.npai.science.uu.nl/50-chat

Argument mining

Argument Mining

“The automatic identification and extraction of the structure of inference and reasoning expressed as arguments presented in natural language.”

(J. Lawrence and C. Reed, 2020)

Argument Mining

Focus on:

"Developing methods to process textual data and reconstruct argumentative content, specifically, extracting arguments along with their relations from natural language texts to the end of providing machine-processable structured data that can be used by computational models."

(E. Cabrio and S. Vilata, 2018)

Argumentation in Natural Language

Challenges:

- Logic of natural language is intractable
- Natural language often carries a lot of implicit information
- Not all natural language arguments are well formed or sound
- Many competing notions of argument strength, quality and soundness

Argument Extraction

- Component classification: Identifying argument components (e.g., claim and premise)
- Component identification: Textual boundaries related to arguments
- Structure identification: Relations between the identified arguments (e.g., attack and support)

(C. Stab and I. Gurevych , 2017)

Text Segmentation

- Text segmentation involves the extraction of the fragments of text from the original piece that will form the constituent parts of the resulting argument structure

- Elementary Discourse Units (EDUs): non-overlapping spans of text corresponding to self-contained piece of argumentative content (e.g., premise, conclusion)

(J. Lawrence and C. Reed, 2020)

Argument / Non-Argument Classification

"Determining which of the segments previously identified are part of the argument being presented and which are not."

(J. Lawrence and C. Reed, 2020)

Example:

Michael Buerk: John Lamiday, thank you very much indeed for joining us this evening. Our third witness is Nick Dearden, who is director of the Jubilee Debt Campaign. *Mr Dearden, you'd like people not to have to pay their debts.* Where's the morality in that?

Nick Dearden: *I wouldn't like people not to have to pay their debts across the board.* But I think what we say is that *this isn't simply a matter of individual morality. Debt is used time and again as a set of economic decisions, and political decisions, to achieve certain things in society.* And *very often what high levels of debt can mean, and especially when the debt is on very unjust terms, is a massive redistribution of wealth in society, from the poorest to the richest.*

Within AI

The extensive use of artificial intelligence (AI) drives the need of developing artificial cognitive systems that can reason in a manner like those of humans. *(L. Michael and A.C. Kakas, 2016)*

- Due to its logical foundations and rule-governed mechanisms, argumentation provides the appropriate support for computational reasoning engines.
- The dialectical nature of argumentation and its similarity with common-sense reasoning makes it easier for users to understand its concepts and interrogate AI systems.
- Joint reasoning through dialogues between individuals and AI agents can be used so that the decision-making process of AI agents and its ethical implications are well-adjusted to adhere to human values.

Explainable AI

Argumentation can translate the decision of an AI system in an argumentation procedure, which shows step by step how it concludes to a result

Given a set of possible decisions, the decisions can be mapped to a graphical representation, with predefined attack properties that subsequently will lead to the winning decision and will show the steps that were followed to reach it

Using argumentation for providing explanations makes an AI system friendlier and more trustworthy to the user as its reasoning mechanism for explaining its decision(s) is closer to the human way of thinking.

OHAAI

Online Handbook of Argumentation for AI

Trends in argumentation research

Theory	65,12%
Application	41,86%
Abstract Argumentation	55,81%
Structured Argumentation	37,21%
Argument Mining; NLP	16,28%
Dialogues	34,88%
Explainable/Responsible AI	25,58%
Logic	18,60%
Neural Networks	9,30%
Complexity	9,30%
Multi-Agent Systems	6,98%
Enthymemes	9,30%
Other	30,23%

Thank you!