UNIVERSITY OF TORONTO

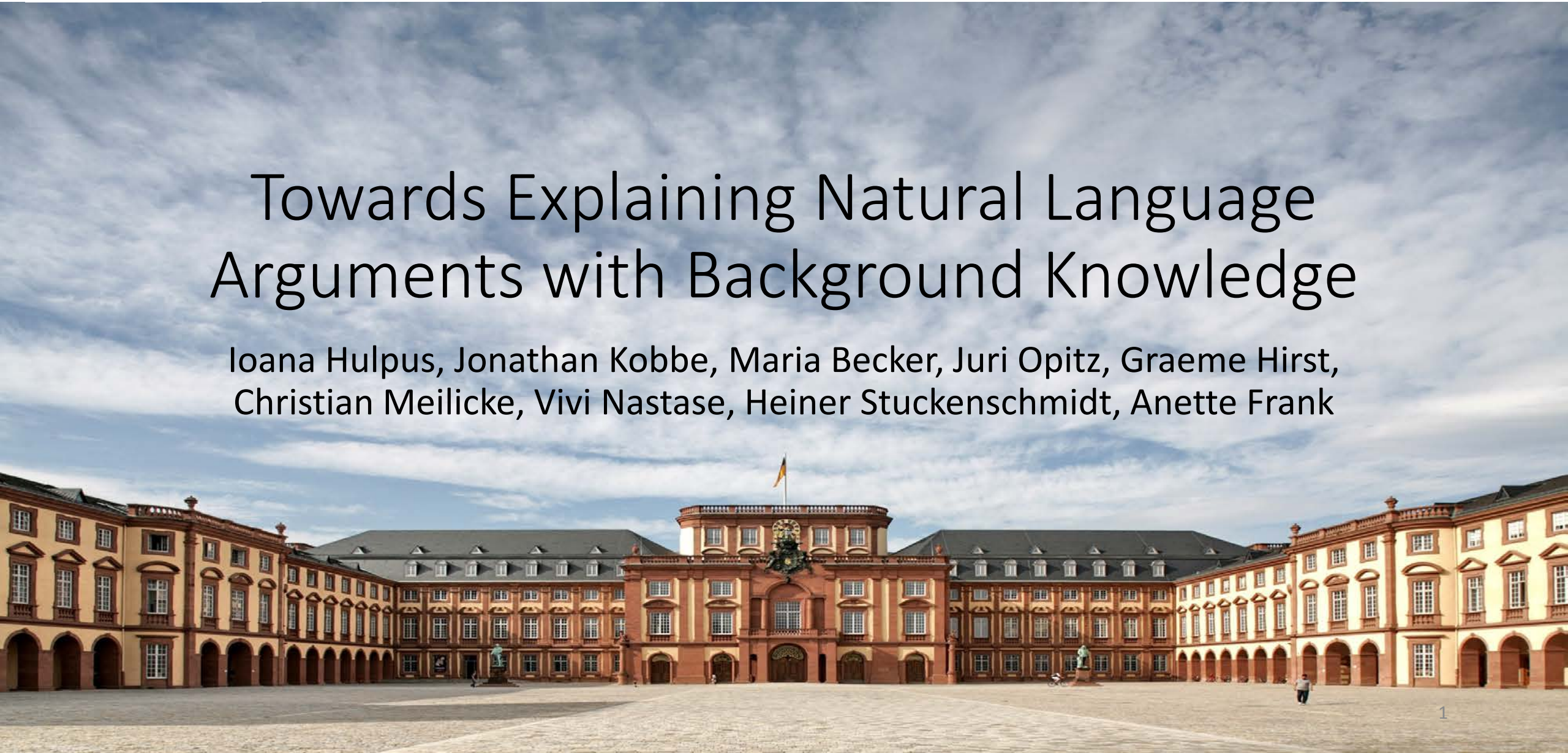
UNIVERSITÄT HEIDELBERG ZUKUNFT SEIT 1386

AI RATIO

UNIVERSITÄT MANNHEIM

# Towards Explaining Natural Language Arguments with Background Knowledge

Ioana Hulpus, Jonathan Kobbe, Maria Becker, Juri Opitz, Graeme Hirst, Christian Meilicke, Vivi Nastase, Heiner Stuckenschmidt, Anette Frank



# Overview

- The Argument Explicitation Task
- Model-based Explicitation
  - Toulmin Model-based Explicitation
  - Walton Scheme-based Explicitation
- Acceptability-based Explicitation
- An Argument Explicitation Framework

# Overview

- The Argument Explicitation Task
- Model-based Explicitation
  - Toulmin Model-based Explicitation
  - Walton Scheme-based Explicitation
- Acceptability-based Explicitation
- An Argument Explicitation Framework

# Motivation

- Support readers to critically judge natural language arguments by:
  - Clear exposure of both implicit and explicit premises that are required for the argument to be **valid** and **sound**

Does the conclusion logically flow from the premises?

Are the premises true?

  - Uncovering of potential weak points of the argument

# Argument Explicitation

- The Task of Argument Explicitation:
  - Given a natural language argument, provide a structured and semantically founded analysis of the argument, and enrich it with the information that is crucial for **understanding the reasoning** behind the argument.

# Example

*Any drastic decline in animal life is cause for alarm, and the current decline in frogs and toads is drastic. (*from [1]*)*

- Although it is a quite simple deductive argument for humans, its analysis poses challenges:
  - The conclusion is left implicit (‘*The current decline in frogs and toads is cause for alarm*.’ )
  - The argument’s **validity** relies on the **unstated premise** that *frogs and toads* are *animal life*
  - The argument’s **soundness** relies on the truthfulness of:
    - the required but unstated premise that *frogs and toads* are *animal life*
    - the explicit premise that there is a *drastic decline* in frogs and toads.

[1] Walton, D., Reed, C.A.: Argumentation schemes and enthymemes. Synthese 145(3), 339-370 (2005)

# Use-case for Argument Explicitation

*Any drastic decline in animal life is cause for alarm, and the current decline in frogs and toads is drastic.*

- Show the user the claims that must be true for the argument to hold

1. **Explicit:** *The current decline in frogs and toads is drastic*

2. **Implicit:** *Frogs and toads are animal life*

# Use-case for Argument Explicitation with Fact Checking

*Any drastic decline in animal life is cause for alarm, and the current decline in frogs and toads is drastic.*

- Show the user the claims that must be true for the argument to hold, highlighting the results from fact-checking

1. **Explicit:** *The current decline in frogs and toads is drastic*
   Results of fact checking in knowledge bases:
   - *"The decline in amphibian populations is recognized as one of the most severe examples of the Holocene extinction….", provenance: dbr:Decline_in_amphibian_populations*

2. **Implicit:** *Frogs and toads are animal life*
   Results of fact checking in knowledge bases:
   - *dbr:Toad rdf:type dbo:Animal, provenance: DBpedia*
   - *dbr:Frog rdf:type dbo:Animal, provenance: DBpedia*

# Two Sub-tasks

- Argument analysis
  - Concerned with analyzing the text of the argument in order to identify argument components (e.g. premises and conclusion) and the structure of the argument;

- Argument reconstruction
  - Concerned with making explicit any unstated but implicit premises / conclusion;

# Two Sub-tasks

- Argument analysis
  - Concerned with analyzing the text of the argument in order to identify argument components (e.g. premises and conclusion) and the structure of the argument;

Most related work (particularly in NLP) deals with this task

- Argument reconstruction
  - Concerned with making explicit any unstated but implicit premises / conclusion;

We stress the importance of this task (which is quite ambitious and challenging), and the opportunity brought by the availability of external background knowledge

# Overview

- The Argument Explicitation Task
- Model-based Explicitation
  - Toulmin Model-based Explicitation
  - Walton Scheme-based Explicitation
- Acceptability-based Explicitation
- An Argument Explicitation Framework

# Guiding Argument Explicitation with Argument Models

- **Argument models** have been proposed particularly in Philosophy to represent patterns of **defeasible** arguments. Two of the most popular:
  - The Toulmin Model
  - The Walton Schemes
- Our position: Mapping arguments and their components to such models reveals the role of each component, and can guide the reconstruction of the implicit components

# Explicitation based on the Toulmin Model

- Toulmin Model [2]
  - Six ingredients (types of argument components):

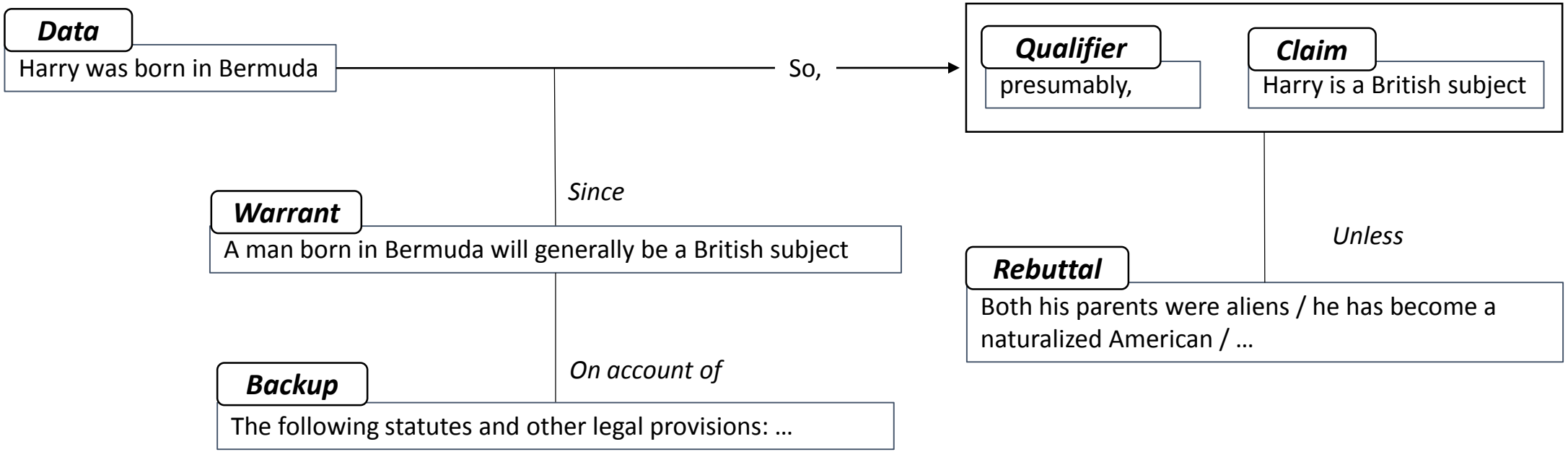
| | |
|---|---|
| ***Claim*** | - the conclusion |
| ***Datum*** | - fact or evidence that supports the claim |
| ***Warrant*** | - why the datum supports / implies the claim |
| ***Backup*** | - why the warrant holds |
| ***Qualifier*** | - the extent to which the warrant holds |
| ***Rebuttal*** | - exceptional case that would remove the authority of the warrant |

[2] Toulmin, S.E.: The uses of argument. Cambridge university press (2003)

# Explicitation based on the Toulmin Model

*Harry was born in Bermuda. So, presumably, Harry is a British subject.*

**Human Explicitation of the Argument**



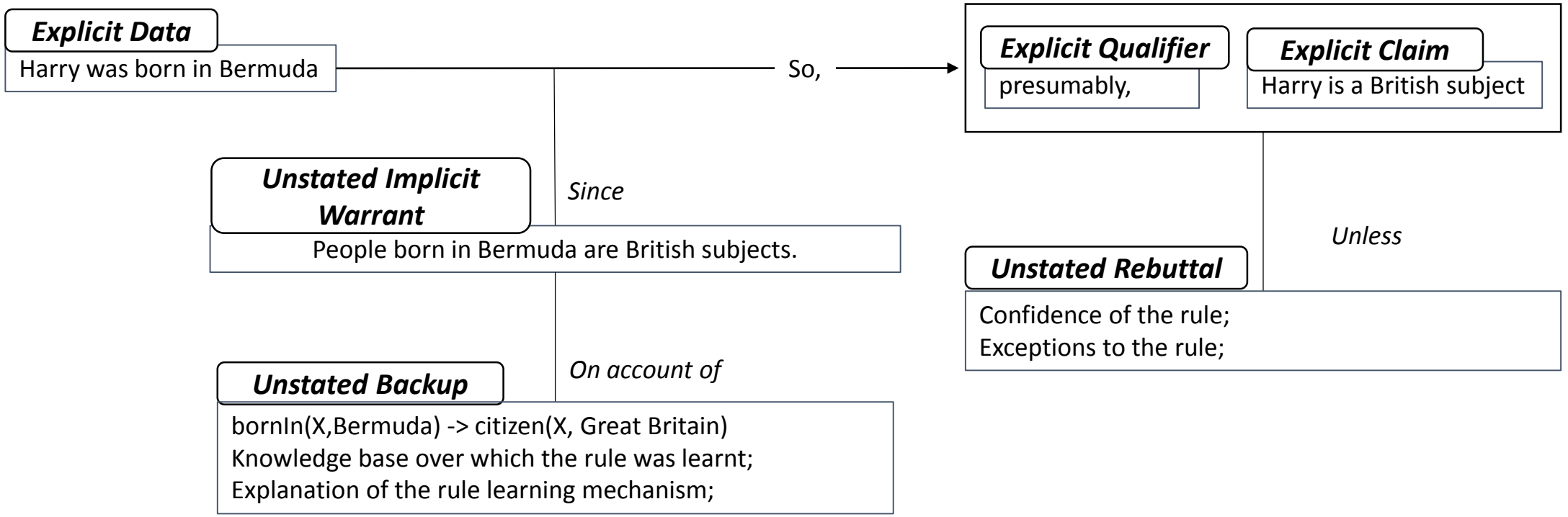
# Explicitation based on the Toulmin Model

- State of the Art:
  - Most approaches in literature focus on classifying *claims* and *premises*, without making the distinction between data, warrant and backup;
  - The recent work of Lugini and Litman (2018)[3] tackles classification of *claims*, *data* and *warrant*;
  - There is no work that aims reconstruction of implicit argument components;

[3] Lugini, L., Litman, D.: Argument component classification for classroom discussions. In: Workshop on Argument Mining. pp. 57-67 (2018)

# Explicitation based on the Toulmin Model

*Harry was born in Bermuda. So, presumably Harry is a British subject.*



# Explicitation based on Walton Schemes

- Walton proposed about 50 argument schemes organized in a hierarchy [4];
- Examples of Walton schemes of argumentation

**Argument from Analogy**
**Premise 1.** Generally, case C1 is similar to case C2.
**Premise 2.** A is true (or false) in case C1.
**Conclusion.** A is true (or false) in case C2.

**Argument from Verbal Classification**
**Premise 1.** *a* has property *P.*
**Premise 2.** for all *x,* if *x* has property *P, x* can be classified as having property *G.*
**Conclusion.** *a* has Property G.

[4] Walton D, Reed C, Macagno F. Argumentation schemes. Cambridge University Press; 2008 Aug 4.

# Explicitation based on Walton Schemes

**Argument from Verbal Classification**
**Premise 1.** *a* has property *P.*
**Premise 2.** for all *x,* if *x* has property *P, x* can be classified as having property *G.*
**Conclusion.** *a* has Property G.

*Any drastic decline in animal life is cause for alarm, and the current decline in frogs and toads is drastic.*

**Premise 1.** *current decline in frogs and toads* has property *is drastic decline in animal life.*
**Premise 2.** for all *x,* if *x* has property *is drastic decline in animal life, x* can be classified as having property *is cause for alarm.*
**Unstated Implicit Conclusion.** *current decline in frogs and toads* has Property *is cause for alarm*.

Even in this apparently simple argument:

- **Premise 1** is only partly explicit in the text: the fact that *frogs and toads* are *animal life* is background knowledge left implicit.
- **Conclusion** is left implicit altogether.

# Explicitation based on Walton Schemes

- State of the Art:
  - Several works have already attempted automated classification of arguments in Walton schemes, mostly using hand-made lexicons of sentence connectors, specific to each scheme.
  - Among them, Lawrence and Reed (2016) [5] also classify the specific premises for each Walton scheme; For example in the previous example, it could classify the first part of the argument as Premise 2 and the second part as Premise 1.
  - No work:
    - considers the semantics of the content of the argument;
    - aims reconstruction of missing premises;
    - aims classification of the elements (e.g., *a*, *P*, G in the previous example).
- Our position: the classification of the scheme elements can guide the reconstruction of the missing premises;

[5] Lawrence, J., Reed, C.: Argument mining using argumentation scheme structures.In: COMMA. pp. 379-390 (2016)

# Explicitation based on Walton Schemes

*The PowerShot SX510 is a fantastic camera as all Canon cameras have great image stabilisation.*

**Walton Scheme:** *Argument from Verbal Classification*

1. Classify the scheme elements

**Premise 1: -**

**Premise 2:** *for all x, if x has property is Canon camera, x can be classified as having property has great image stabilisation*

**Conclusion:** *The PowerShot SX510 has property is a fantastic camera.*

# Explicitation based on Walton Schemes

2. Use the classified elements to fill in the placeholders in the unstated premise

**Unstated Premise 1:** *a* has property *P* ⇔ *The PowerShot SX510 has property is Canon camera.*

**Premise 2:** *for all x, if x has property is Canon camera, x can be classified as having property has great image stabilisation*

**Conclusion:** *The PowerShot SX510 has property is a fantastic camera.*

# Explicitation based on Walton Schemes

3. Resolve conflicts

**Unstated Premise 1:** *The PowerShot SX510 has property is Canon camera.*

**Premise 2:** *for all x, if x has property is Canon camera, x can be classified as having property has great image stabilisation*

**Conclusion:** *The PowerShot SX510 has property is a fantastic camera.*

# Explicitation based on Walton Schemes

3. Resolve conflicts

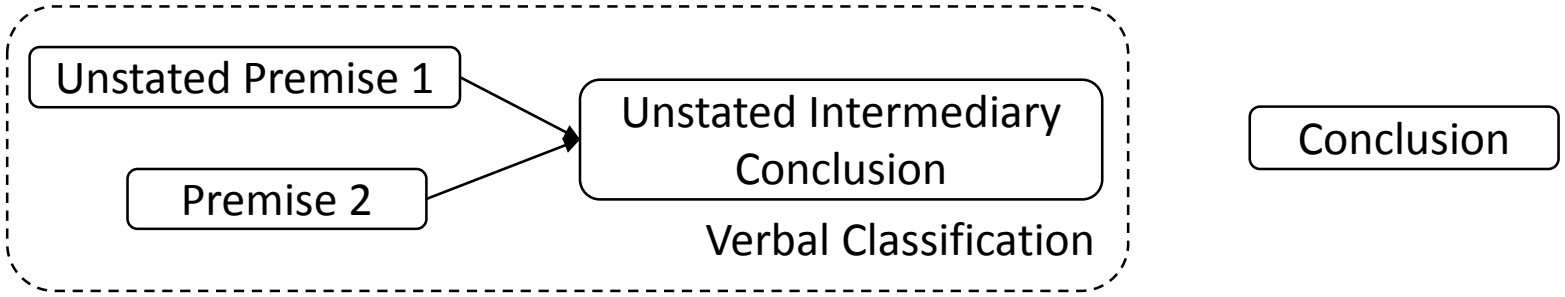
3.1 Draw unstated intermediary conclusions;

**Unstated Premise 1:** *The PowerShot SX510 has property is Canon camera.*

**Premise 2:** *for all x, if x has property is Canon camera, x can be classified as having property has great image stabilisation*

**Unstated Intermediary Conclusion:** *The PowerShot SX510 has property has great image stabilisation.*

**Conclusion:** *The PowerShot SX510 has property is a fantastic camera.*



# Explicitation based on Walton Schemes

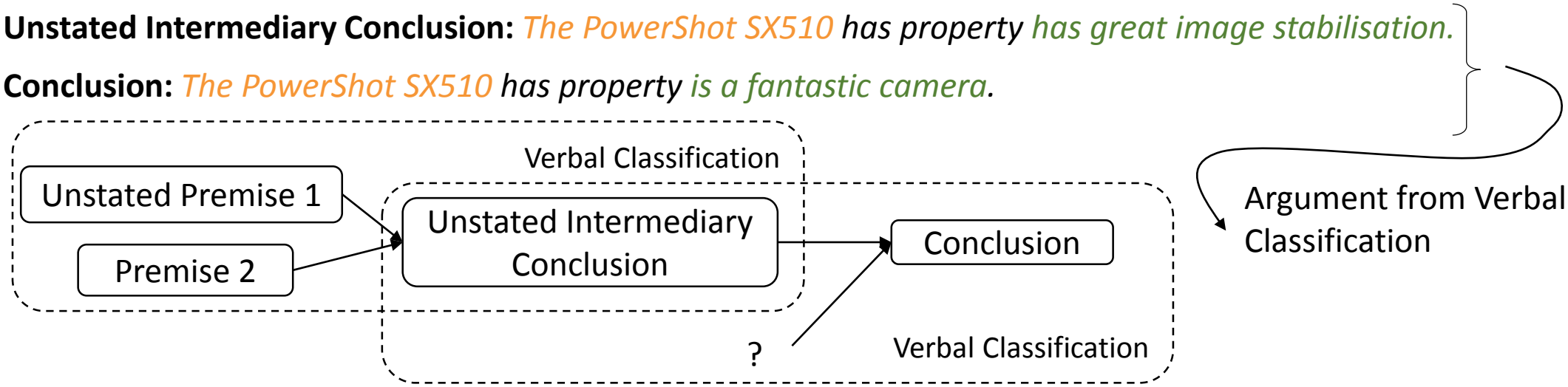
3.2 Reconstruct a valid Walton scheme argument that has as premise the intermediary conclusion and as conclusion the original conclusion

**Unstated Premise 1:** *The PowerShot SX510 has property is Canon camera.*

**Premise 2:** *for all x, if x has property is Canon camera, x can be classified as having property has great image stabilisation*

**Unstated Intermediary Conclusion:** *The PowerShot SX510 has property has great image stabilisation.*

**Conclusion:** *The PowerShot SX510 has property is a fantastic camera.*



# Explicitation based on Walton Schemes

3.2 Reconstruct a valid Walton scheme argument that has as premise the intermediary conclusion, and as conclusion the original conclusion

**Unstated Premise 1a:** *The PowerShot SX510 has property is Canon camera.*

**Premise 2a:** *for all x, if x has property is Canon camera, x can be classified as having property has great image stabilisation*

**Unstated Intermediary Conclusion a:** *The PowerShot SX510 has property has great image stabilisation.*

**Unstated Premise 1b:** *The PowerShot SX510 has property has great image stabilisation.*

**Unstated Premise 2b:** *for all x, if x has property has great image stabilisation, x can be classified as having property is a fantastic camera.*

**Conclusion b:** *The PowerShot SX510 has property is a fantastic camera.*

# Explicitation based on Walton Schemes

*The PowerShot SX510 is a fantastic camera as all Canon cameras have great image stabilisation.*

4. Classify premises as *fact* or *subjective* explaining the result;

**Unstated Premise 1a:** *The PowerShot SX510 has property is Canon camera. fact (URL to knowledge base)*

**Premise 2a:** *for all x, if x has property is Canon camera, x can be classified as having property has great image stabilisation. subjective (great)*

**Unstated Intermediary Conclusion a:** *The PowerShot SX510 has property has great image stabilisation.*

**Unstated Premise 1b:** *The PowerShot SX510 has property has great image stabilisation.*

**Unstated Premise 2b:** *for all x, if x has property has great image stabilisation, x can be classified as having property is a fantastic camera. subjective (great, fantastic)*

**Conclusion b:** *The PowerShot SX510 has property is a fantastic camera.*

# Overview

- The Argument Explicitation Task
- Model-based Explicitation
  - Toulmin Model-based Explicitation
  - Walton Scheme-based Explicitation
- Acceptability-based Explicitation
- An Argument Explicitation Framework

# Acceptability-based Explicitation

- As opposed to the previous Explicitation strategies, the *acceptability-based Explicitation* focuses on the relation between arguments, rather than the internal structure of one atomic argument;

- The most influential framework on relations between arguments is Dung's framework that defines the *attack* relation;
  - In Dung's framework, an argument is acceptable only to the extent to which it is able to defend against attacking arguments;
  - More recent work extends the framework with *support* relations;
  - Pollock (1987) identifies two types of attacks:
    - **Rebuttal –** direct attack to a premise / conclusion;
    - **Undercut –** attack to the logical entailment of the conclusion given the premise (the *warrant* in Toulmin's terminology ).

# Acceptability-based Explicitation

- State of the art:
  - Probably one of the most prolific areas in argument mining is to detect attack/support relations between argument components;
  - There is though much less work on detecting such relations between entire arguments, although this is the aim of Dung's framework;
- Our position: an attack relation between two argument components of the same monological argumentation text, indicates a reference to a (anticipated) counter-argument

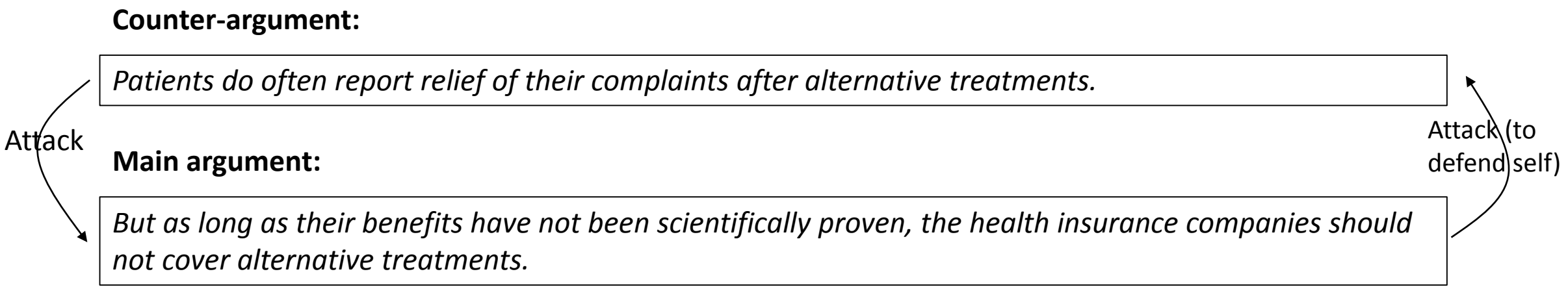
# Acceptability-based Explicitation

- The purpose of the Acceptability-based type of Explicitation is to identify the atomic arguments in an argumentative text:
  - The main argument – the atomic argument whose conclusion is the main conclusion of the text;
  - Supporting arguments – atomic arguments whose conclusions are intermediary and act as premises to the main argument (for example, the first argument in the Walton scheme Explicitation is a supporting argument).
  - Counter-arguments – atomic arguments whose conclusions attack a premise or the conclusion of the main argument, or of a supporting argument;

# Acceptability-based Explicitation

*Patients do often report relief of their complaints after alternative treatments. But as long as their benefits have not been scientifically proven, the health insurance companies should not cover alternative treatments.*

1. Identify attacking pairs of statements and isolate the counter-argument(s) from the main argument;



# Acceptability-based Explicitation

*Patients do often report relief of their complaints after alternative treatments. But as long as their benefits have not been scientifically proven, the health insurance companies should not cover alternative treatments.*

2. Reconstruct the counter-argument by generating its conclusion which contradicts the main conclusion.

**Counter-argument:**

**Premise:** Patients do often report relief of their complaints after alternative treatments.
**Implicit Conclusion:** The health insurance companies **should cover** alternative treatments.

Attack

Attack counter argument to defend self

**Main argument:**

**Premise:** As long as their benefits have not been scientifically proven.
**Conclusion:** The health insurance companies **should not cover** alternative treatments.

# Acceptability-based Explicitation

*Patients do often report relief of their complaints after alternative treatments. But as long as their benefits have not been scientifically proven, the health insurance companies should not cover alternative treatments.*

3. Explicitate the arguments, potentially using models such as Toulmin / Walton, and classify the attacks as rebuttal or undercut (example below uses Toulmin-based Explicitation for each atomic argument).

**Counter-argument:**

**Premise (Data):** Patients do often report relief of their complaints after alternative treatments.
**Implicit Premise (Warrant):** Health insurance companies should cover treatments after which patients report relief.
**Implicit Conclusion (Claim):** The health insurance companies should cover alternative treatments.

Undercut

**Main argument:**

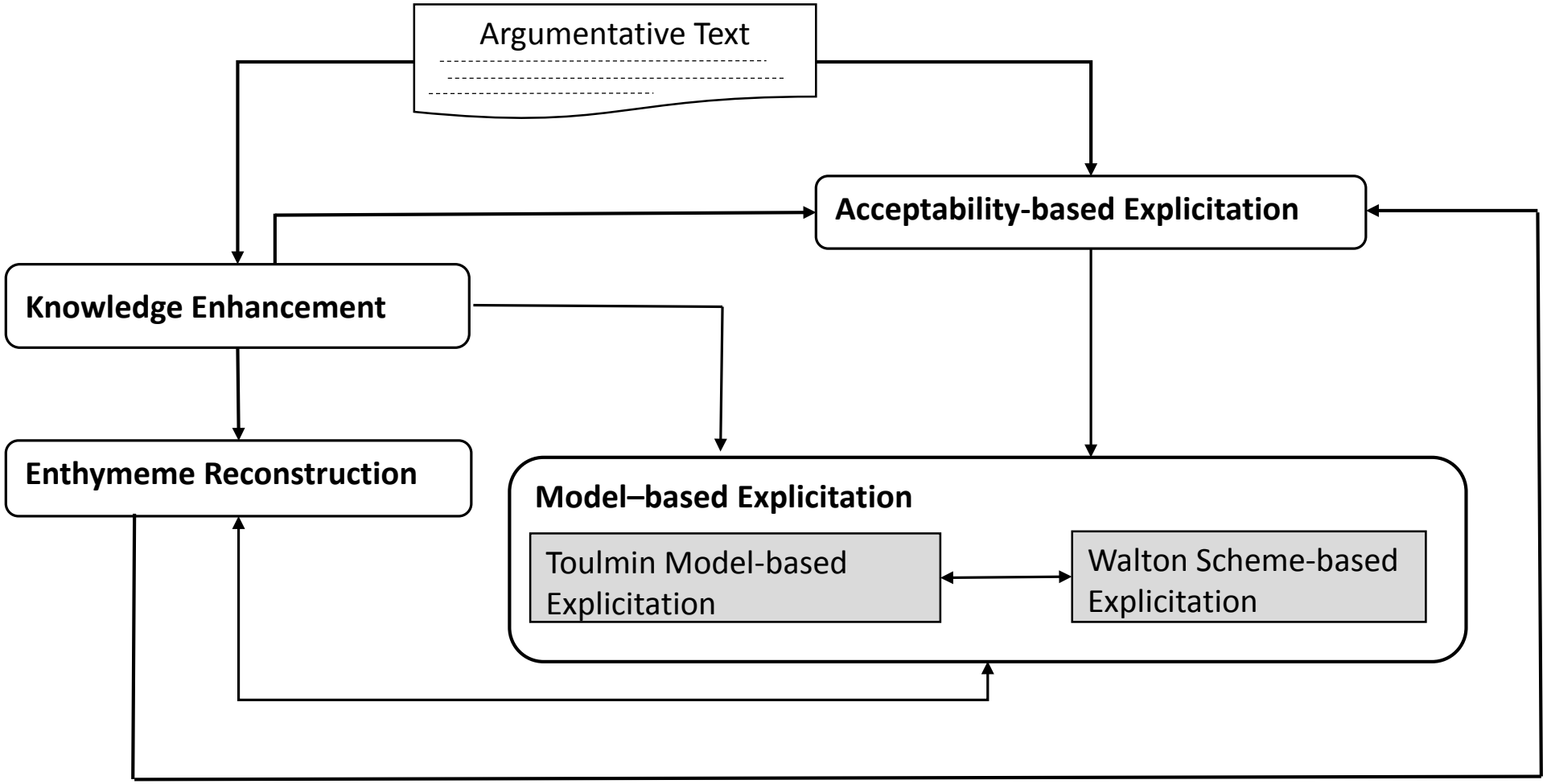
Rebuttal

**Premise (Data):** Benefits of alternative treatments have not been scientifically proven.
**Implicit Premise (Warrant):** Health insurance companies should cover treatments that are scientifically proven.
**Conclusion (Claim):** The health insurance companies should not cover alternative treatments.

# Overview

- The Argument Explicitation Task
- Model-based Explicitation
  - Toulmin Model-based Explicitation
  - Walton Scheme-based Explicitation
- Acceptability-based Explicitation
- An Argument Explicitation Framework

# Argument Explicitation Framework



# Conclusion

- We introduce the problem of argument explicitation that puts a particular focus on reconstructing the chain of reasoning behind arguments;

- The argument explicitation task
  - is challenging – beyond state of the art in multiple AI domains;
  - is required in order to truly make sense of arguments;
  - places reasoning (back) to the center of semantics:
    - not formal reasoning (what ontologies are good at), but informal reasoning (what people are good at), guided by informal reasoning models rather than formal logics;

# Acknowledgements


- This work has been funded by the Deutsche Forschungsgemeinschaft (DFG) within the project ExpLAIN, Grant Number STU 266/14-1 and FR 1707/-4-1, as part of the Priority Program \Robust Argumentation Machines (RATIO)“ (SPP-1999) as well as by the Leibniz Science Campus Empirical Linguistics & Computational Language Modeling, supported by Leibniz Association grant no. SAS2015-IDS-LWC and by the Ministry of Science, Research, and Art of Baden-Wurttemberg.