



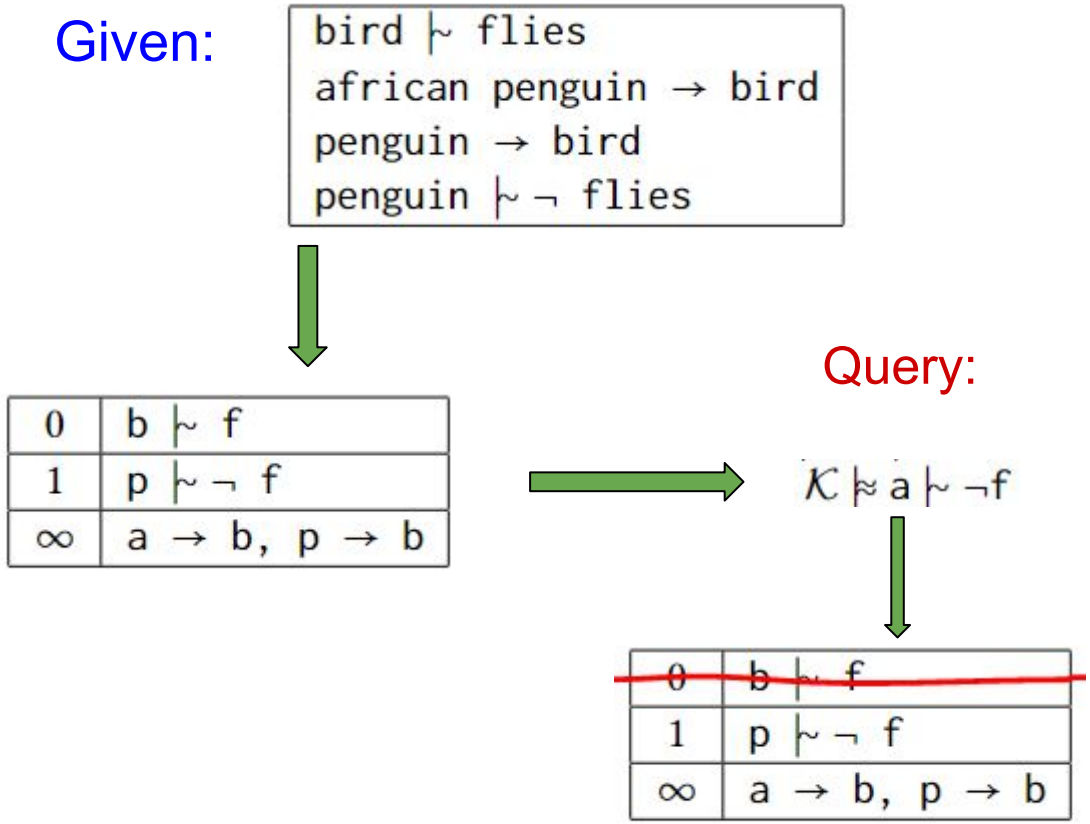
# Generalised weak justifications for rational defeasible entailment and the *potential* connection to strong explanations

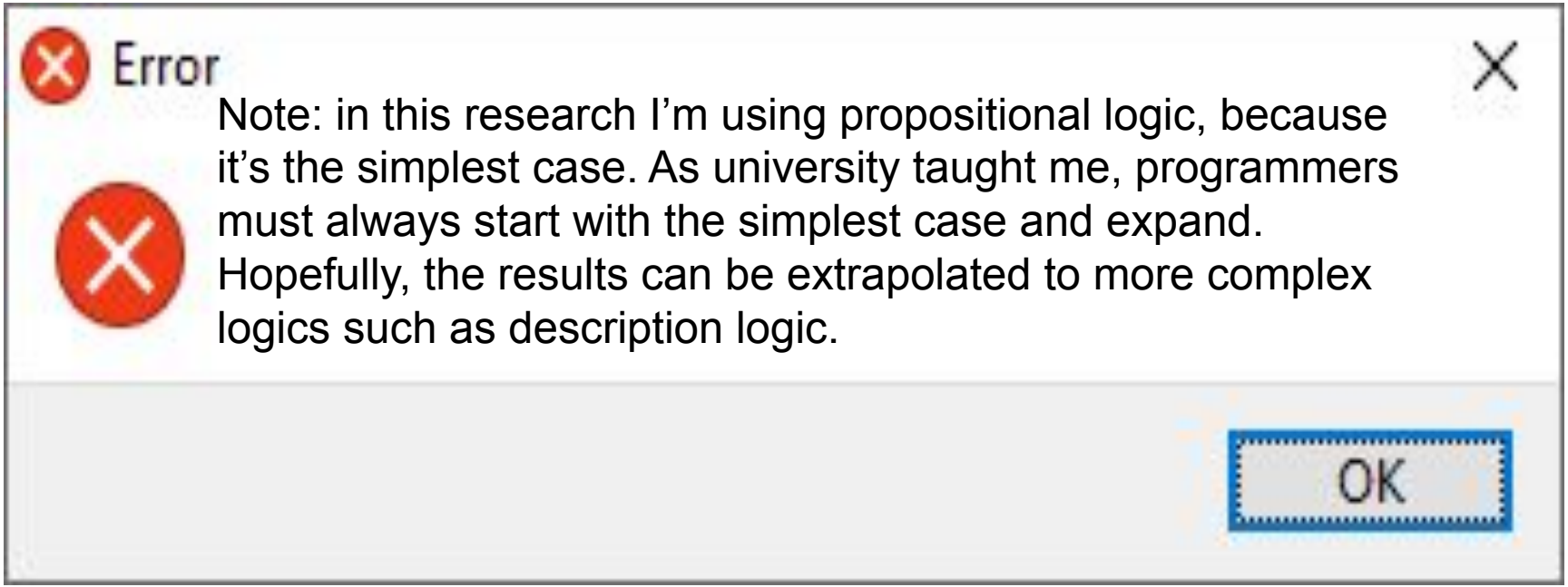
MSc by Dissertation
Jane Imrie
01-2024

# Table of Contents

# Defeasible Logic

- Rational closure is one form of nonmonotonic entailment
  - Rank statements
- Query checking: first check the whole ranking, then subsets if necessary
  - After that, use classical entailment checking
- Rational defeasible entailment - outlines characteristics that entailment relations should have




Error
Note: in this research I’m using propositional logic, because it’s the simplest case. As university taught me, programmers must always start with the simplest case and expand. Hopefully, the results can be extrapolated to more complex logics such as description logic.
OK

# A problem of trust

- Black-box nature of AI systems
- Potential for prejudicial choices made by algorithms within the AI system
- Explanatory facilities = crucial component of KR systems
    - Arguably all AI systems

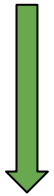
# Classical Justifications

- Justifications are just one form of explanations
- Justifications: smallest subset from our knowledge base which allows an entailment to hold

Given:

$$\text{bird} \rightarrow \text{flies}$$
$$\text{bird} \rightarrow \text{wings}$$
$$\text{robin} \rightarrow \text{bird}$$
$$\text{penguin} \rightarrow \text{bird}$$
$$\text{penguin} \rightarrow \neg\ \text{flies}$$

Query: $\mathcal{K} \models \mathrm{r} \rightarrow \mathrm{w}?$



$$\mathcal{J} = \{\mathrm{r} \rightarrow \mathrm{b},\ \mathrm{b} \rightarrow \mathrm{w}\}$$

# Defeasible Justifications

- Nonmonotonicity adds complexity
- Not as well studied as classical justifications
- Weak justifications - form of defeasible explanation that uses rational closure
- Strong explanations 💪 - logic agnostic formalism, generalisable to monotonic and nonmonotonic logics
  - Adds a constraint for justifications - info from within the knowledge base but outside of a justification - cannot render said justification false
    - If it does, then the justification is invalid

Given:

$$\begin{aligned} &\text{bird} \mathrel{|\!\sim} \text{flies} \\ &\text{penguin} \mathrel{|\!\sim} \text{bird} \\ &\text{penguin} \mathrel{|\!\sim} \neg\ \text{flies} \\ &\text{special penguin} \mathrel{|\!\sim} \text{bird} \\ &\text{special penguin} \mathrel{|\!\sim} \text{flies} \end{aligned}$$

Query: $s \mathrel{|\!\sim} f$

(1) $\mathcal{J}_1 = \{s \mathrel{|\!\sim} f\}$
(2) $\mathcal{J}_2 = \{s \mathrel{|\!\sim} p,\ p \mathrel{|\!\sim} b,\ b \mathrel{|\!\sim} f\}$

However, $\mathcal{J}_2$ does not qualify as a strong justification since adding $\{p \mathrel{|\!\sim} \neg f\}$ causes the entailment to fail.

# Research Motivations and Objectives

- Currently, defeasible justifications have only been formalised for rational closure i.e. weak justifications
- We want to investigate the possibility of **generalisation of weak justifications to rational defeasible entailment**
  - We want to test if these generalised properties can satisfy the properties of strong explanations
- The above leads to two scenarios:
  - One where they do fully satisfy the properties - which warrants further investigation of the connection between strong and weak explanations
  - One where they do not - again, this requires research into which do and do not and why

# Important Papers


(1) Gerhard Brewka and Markus Ulbricht. 2019. *Strong explanations for nonmonotonic reasoning*. Description Logic, Theory Combination, and All That: Essays Dedicated to Franz Baader on the Occasion of His 60th Birthday (2019), 135–146

(2) Victoria Chama. 2020. *Explanation for defeasible entailment*. Master's thesis. Faculty of Science

(3) Lloyd Everett, Emily Morris, and Thomas Meyer. 2021. *Explanation for KLM-Style Defeasible Reasoning*. In Southern African Conference for Artificial Intelligence Research. Springer, 192–207.

(4) Steve Wang, Thomas Meyer, and Deshendran Moodley. 2022. *Defeasible Justification Using the KLM Framework*. In Southern African Conference for Artificial Intelligence Research. Springer, 187–201.

# Conclusion

- Future plans: PhD
  - Not really sure on topic but have some ideas
    - The integration between graph theory, strong explanations and weak justifications
    - Optimal ways to inject explanatory facilities into mixed AI systems (i.e. ones that use a combination of ML and Symbolic AI)
      - Based on Calegari, R., Omicini, A. and Sartor, G., 2020. *Argumentation and logic programming for explainable and ethical AI*. In CEUR WORKSHOP PROCEEDINGS (Vol. 2742, pp. 55-68). Sun SITE Central Europe, RWTH Aachen University.
    - Something else depending on the results of my research