In the Eye of the Beholder: Which Proofs are Best?

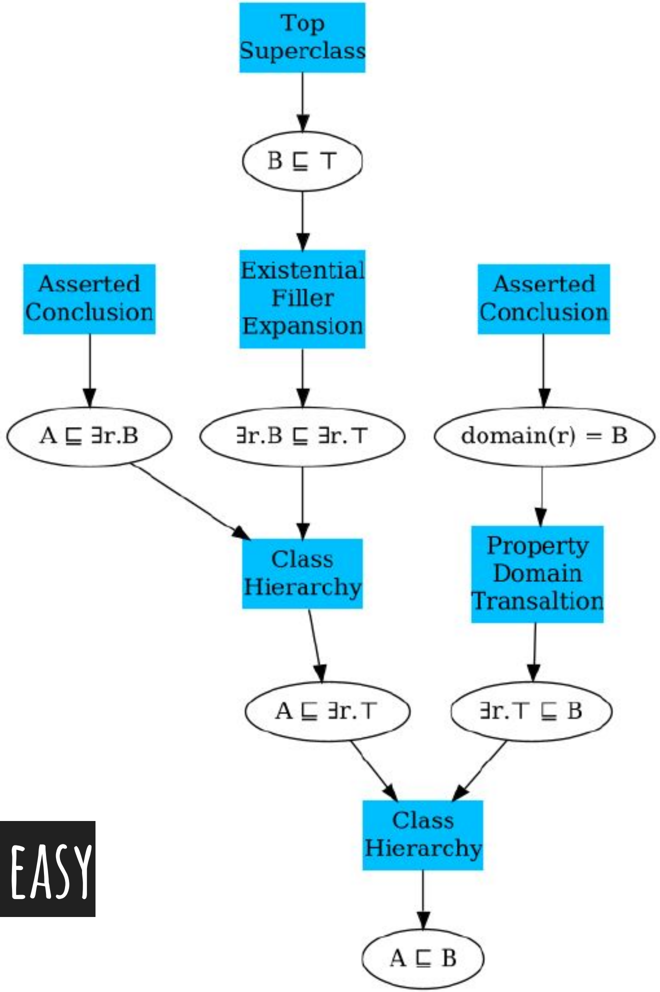
Stefan Borgwardt, Anke Hirsch,
Alisa Kovtunova and Frederik Wiehr

HOW TO EXPLAIN A DECISION MADE BY AN AUTOMATED SYSTEM?

If the knowledge is represented in (some fragment of) first-order logic,

formal proofs!

Challenge



formal proofs* are not easy

*a recap is provided in a few slides

Research Question

How to represent a formal proof in a more understandable way for the user?

Research Question(s)

How to represent a formal proof in a more understandable way for the user?

How to measure understandability?

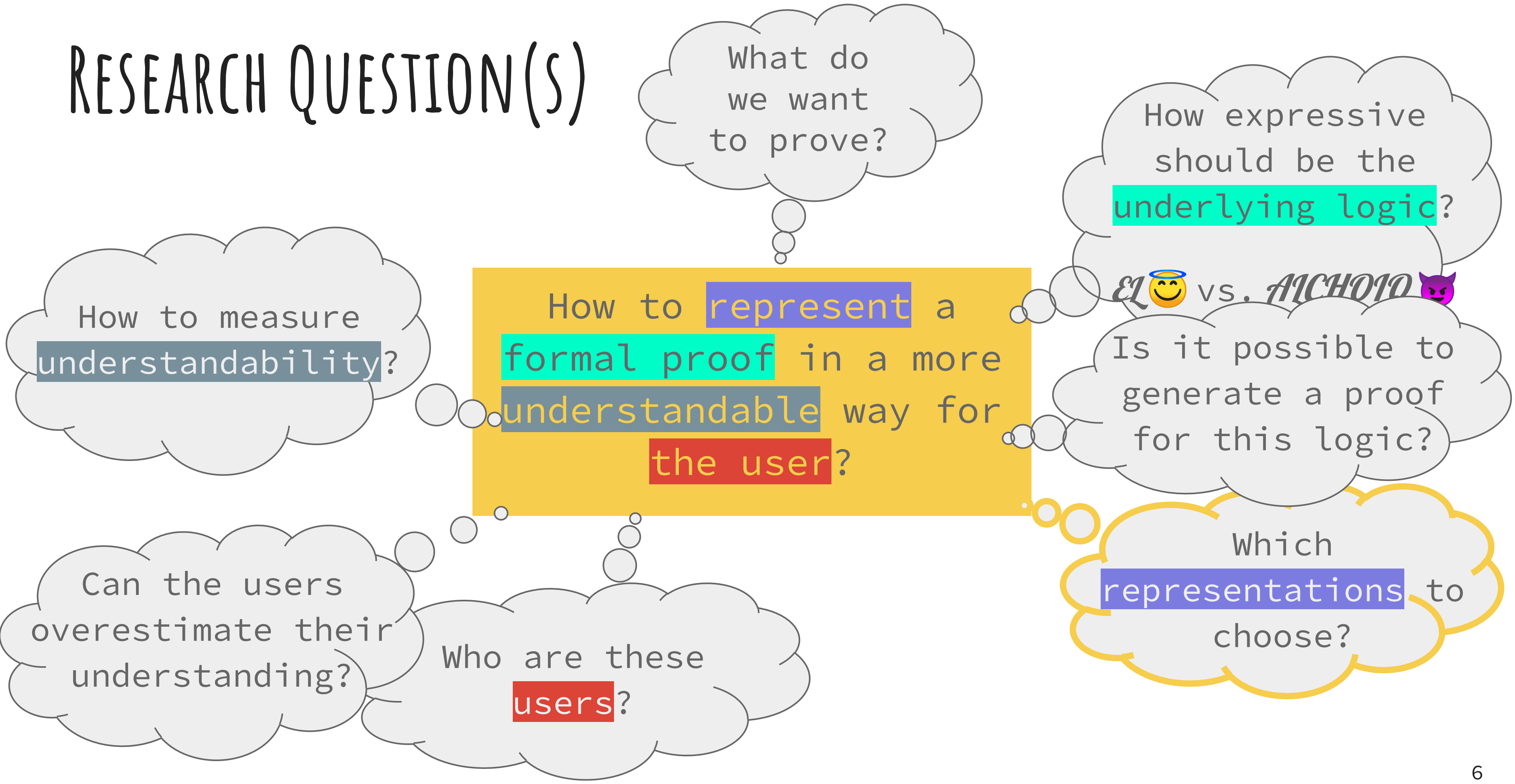
How expressive should be the underlying logic?

$\mathcal{EL}$ 😇 vs. $\mathcal{ALCHOIQ}$ 😈

Which representations to choose?

Who are these users?

Research Question(s)



DL Theory*:

- Bird, Cage, Egg — **Concept Names**
- sitsIn, lays — **Roles**
- $\exists$sitsIn.Cage — **Complex Concepts**
- $\forall$lays.Egg
- Bird $\sqcap$ $\exists$sitsIn.Cage

Tweety $\sqsubseteq$ CageBird — **Axioms**

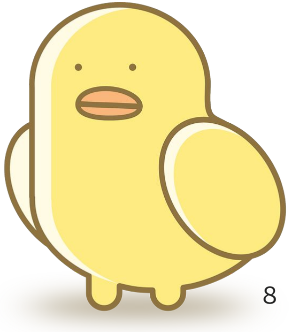
CageBird $\equiv$ Bird $\sqcap$ $\exists$sitsIn.Cage

*also formally defined in the previous talk "Finding Proofs for Description Logic Entailments in Practice"

An example of an Ontology

Ontology axioms

$$\begin{aligned}
\mathsf{Penguin} &\equiv \exists \mathsf{hasPart}.\mathsf{Wings} \sqcap \mathsf{NotFlying}\\
\mathsf{Bird} &\sqsubseteq \exists \mathsf{hasPart}.\mathsf{Wings} \sqcap \forall \mathsf{lays}.\mathsf{Egg}\\
\mathsf{CageBird} &\equiv \mathsf{Bird} \sqcap \exists \mathsf{sitsIn}.\mathsf{Cage}\\
\mathsf{Tweety} &\sqsubseteq \mathsf{CageBird}\\
\mathsf{Tweety} &\sqsubseteq \mathsf{NotFlying}
\end{aligned}$$

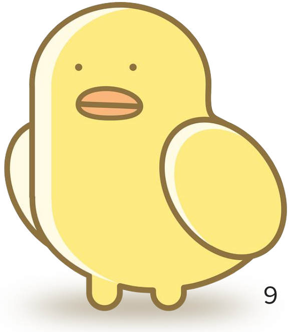


An example of an Ontology

Ontology axioms

$$\begin{aligned} \mathsf{Penguin} &\equiv \exists \mathsf{hasPart.Wings} \sqcap \mathsf{NotFlying} \\ \mathsf{Bird} &\sqsubseteq \exists \mathsf{hasPart.Wings} \sqcap \forall \mathsf{lays.Egg} \\ \mathsf{CageBird} &\equiv \mathsf{Bird} \sqcap \exists \mathsf{sitsIn.Cage} \\ \mathsf{Tweety} &\sqsubseteq \mathsf{CageBird} \\ \mathsf{Tweety} &\sqsubseteq \mathsf{NotFlying} \end{aligned}$$

$\boxed{\mathsf{Tweety} \sqsubseteq \mathsf{Penguin}}$



An Example of Consequence-Based reasoning rules

$$\frac{C \equiv D}{C \sqsubseteq D}$$

$$\frac{C \sqsubseteq D \sqcap E \sqcap F}{C \sqsubseteq D,\ C \sqsubseteq E,\ C \sqsubseteq F}$$

$$\frac{}{C \sqsubseteq C}$$

$$\frac{C \sqsubseteq D,\ D \sqsubseteq F}{C \sqsubseteq F}$$

An Example of Consequence-Based reasoning rules

$$\frac{C \equiv D}{C \sqsubseteq D}$$

$$\frac{C \sqsubseteq D \sqcap E \sqcap F}{C \sqsubseteq D, \quad C \sqsubseteq E, \quad C \sqsubseteq F}$$

$$\frac{}{C \sqsubseteq C}$$

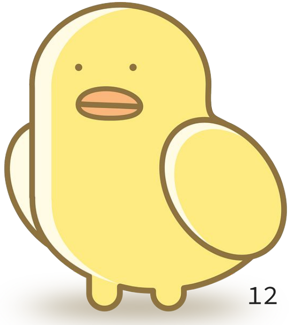
Transitive rule

$$\frac{C \sqsubseteq D, \quad D \sqsubseteq F}{C \sqsubseteq F}$$

An example of a formal proof

$$
\dfrac{
  \dfrac{
    \dfrac{
      \dfrac{
        \mathsf{Tweety} \sqsubseteq \mathsf{CageBird}
        \qquad
        \dfrac{\mathsf{CageBird} \equiv \mathsf{Bird} \sqcap \exists \mathsf{sitsIn}.\mathsf{Cage}}{\mathsf{CageBird} \sqsubseteq \mathsf{Bird} \sqcap \exists \mathsf{sitsIn}.\mathsf{Cage}}
      }{\mathsf{Tweety} \sqsubseteq \mathsf{Bird} \sqcap \exists \mathsf{sitsIn}.\mathsf{Cage}}
    }{\mathsf{Tweety} \sqsubseteq \mathsf{Bird}}
    \qquad
    \dfrac{\mathsf{Bird} \sqsubseteq \exists \mathsf{hasPart}.\mathsf{Wings} \sqcap \forall \mathsf{lays}.\mathsf{Egg}}{\mathsf{Bird} \sqsubseteq \exists \mathsf{hasPart}.\mathsf{Wings}}
  }{\mathsf{Tweety} \sqsubseteq \exists \mathsf{hasPart}.\mathsf{Wings}}
  \qquad
  \mathsf{Tweety} \sqsubseteq \mathsf{NotFlying}
}{\mathsf{Tweety} \sqsubseteq \exists \mathsf{hasPart}.\mathsf{Wings} \sqcap \mathsf{NotFlying}}
\qquad
\dfrac{\mathsf{Penguin} \equiv \exists \mathsf{hasPart}.\mathsf{Wings} \sqcap \mathsf{NotFlying}}{\exists \mathsf{hasPart}.\mathsf{Wings} \sqcap \mathsf{NotFlying} \sqsubseteq \mathsf{Penguin}}
$$

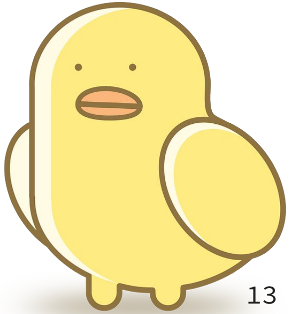
$$
\boxed{\mathsf{Tweety} \sqsubseteq \mathsf{Penguin}}
$$



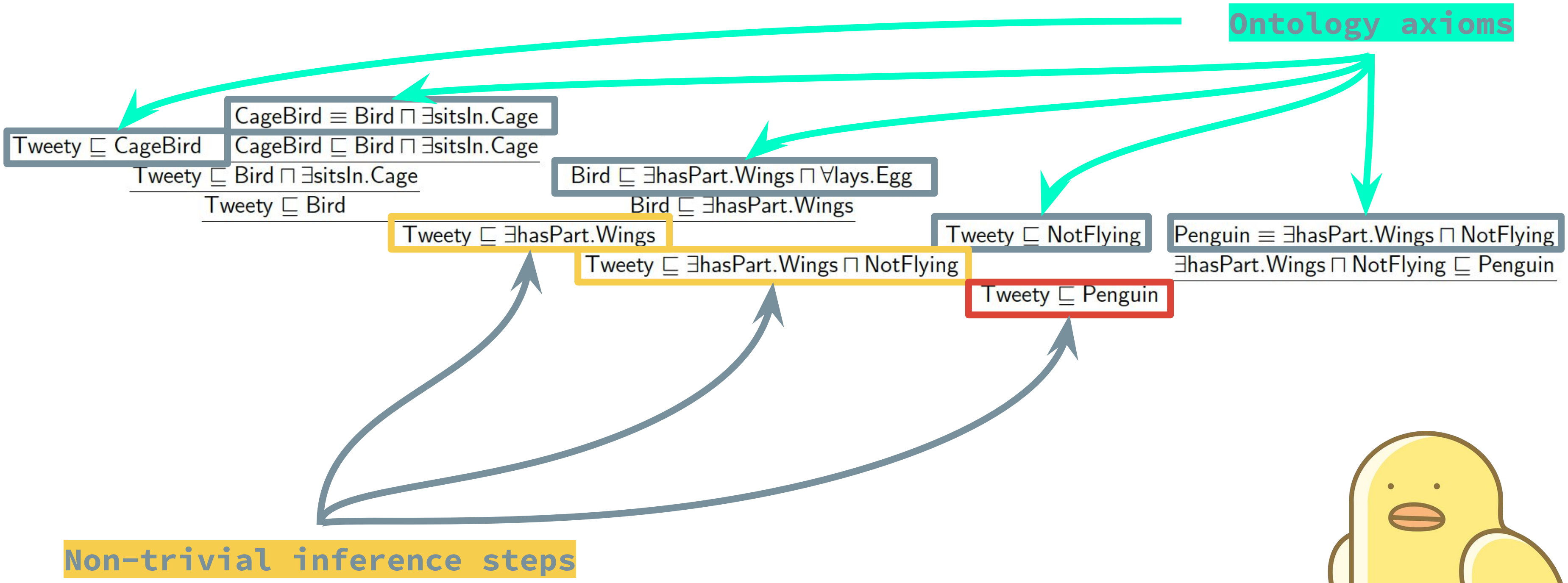
An example of a formal proof

Ontology axioms

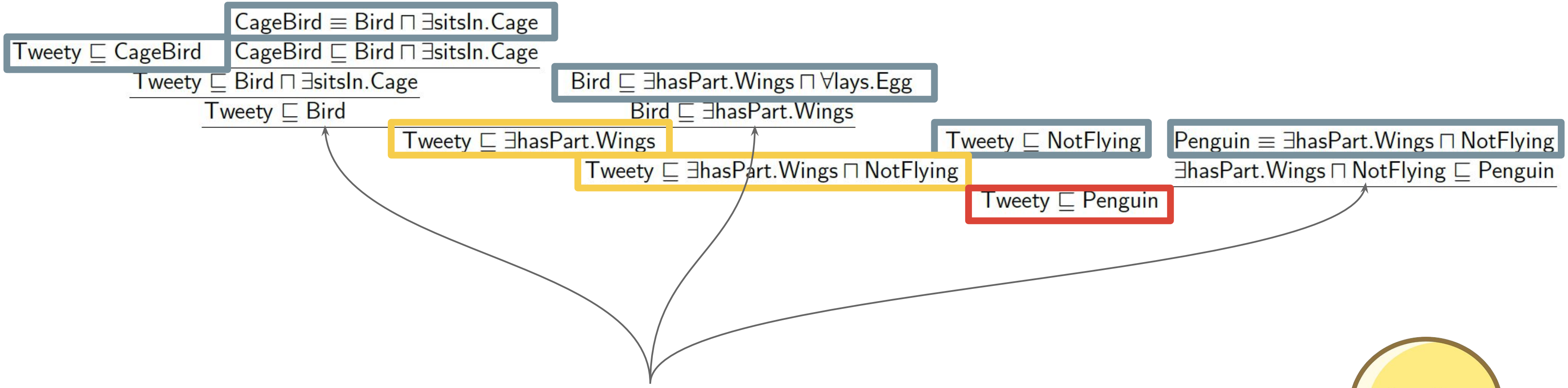
$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\boxed{\text{Tweety} \sqsubseteq \text{CageBird}} \quad \dfrac{\boxed{\text{CageBird} \equiv \text{Bird} \sqcap \exists \text{sitsIn}.\text{Cage}}}{\text{CageBird} \sqsubseteq \text{Bird} \sqcap \exists \text{sitsIn}.\text{Cage}}}{\text{Tweety} \sqsubseteq \text{Bird} \sqcap \exists \text{sitsIn}.\text{Cage}}}{\text{Tweety} \sqsubseteq \text{Bird}} \quad \dfrac{\boxed{\text{Bird} \sqsubseteq \exists \text{hasPart}.\text{Wings} \sqcap \forall \text{lays}.\text{Egg}}}{\text{Bird} \sqsubseteq \exists \text{hasPart}.\text{Wings}}}{\text{Tweety} \sqsubseteq \exists \text{hasPart}.\text{Wings}} \quad \boxed{\text{Tweety} \sqsubseteq \text{NotFlying}}}{\text{Tweety} \sqsubseteq \exists \text{hasPart}.\text{Wings} \sqcap \text{NotFlying}} \quad \dfrac{\boxed{\text{Penguin} \equiv \exists \text{hasPart}.\text{Wings} \sqcap \text{NotFlying}}}{\exists \text{hasPart}.\text{Wings} \sqcap \text{NotFlying} \sqsubseteq \text{Penguin}}}{\boxed{\text{Tweety} \sqsubseteq \text{Penguin}}}$$



An example of formal proof shortening

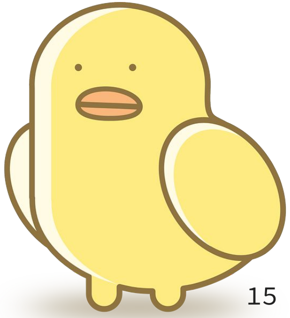


An example of formal proof shortening



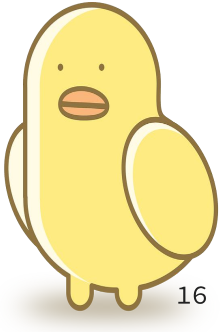
Instances of 3 types of inference steps:

- '≡' implies '⊑'
- intermediate elements in a chain of transitive rules,
- '⊓' on the right side.



An example of formal proof shortening

$$\dfrac{\dfrac{\dfrac{\mathsf{Tweety} \sqsubseteq \mathsf{CageBird} \quad \mathsf{CageBird} \equiv \mathsf{Bird} \sqcap \exists \mathsf{sitsIn.Cage} \quad \mathsf{Bird} \sqsubseteq \exists \mathsf{hasPart.Wings} \sqcap \forall \mathsf{lays.Egg}}{\mathsf{Tweety} \sqsubseteq \exists \mathsf{hasPart.Wings}} \quad \mathsf{Tweety} \sqsubseteq \mathsf{NotFlying}}{\mathsf{Tweety} \sqsubseteq \exists \mathsf{hasPart.Wings} \sqcap \mathsf{NotFlying}} \quad \mathsf{Penguin} \equiv \exists \mathsf{hasPart.Wings} \sqcap \mathsf{NotFlying}}{\mathsf{Tweety} \sqsubseteq \mathsf{Penguin}}$$



Axiom Verbalisation

$$\frac{\mathsf{Tweety} \sqsubseteq \exists \mathsf{hasPart.Wings} \sqcap \mathsf{NotFlying} \qquad \mathsf{Penguin} \equiv \exists \mathsf{hasPart.Wings} \sqcap \mathsf{NotFlying}}{\mathsf{Tweety} \sqsubseteq \mathsf{Penguin}}$$

Since / From the facts that

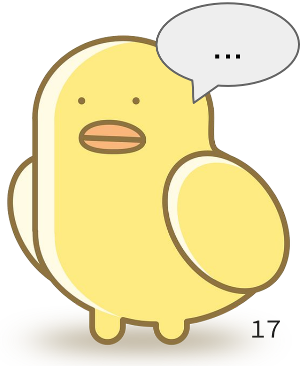
Tweety has wings and does not fly

AND

a penguin is a synonym for something that has wings and does not fly

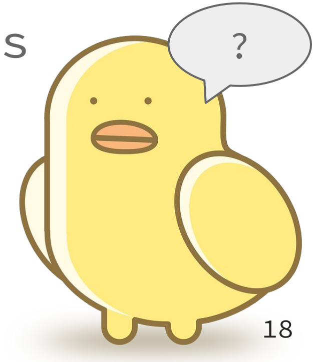
,

Tweety is a penguin.



Which proofs are best?

$$\begin{tikzcd} \boxed{\text{long formal proofs}} \arrow[r] \arrow[d] & \text{long textual proofs} \\ \text{short formal proofs} \arrow[r] & \text{short textual proofs} \end{tikzcd}$$



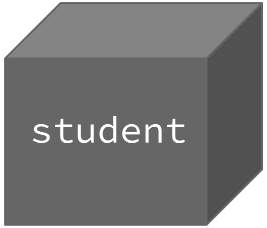
Hypothesis

- It is easier to understand a short, concise explanation than a longer version (in the same representation format).
- Users with less experience in logic can understand the longer text better than a short formal proof.
- Users with more experience in logic can understand a long formal proof better than a long text.

Solution found: DL meets Psychology!

Let's ask ~~people~~ students!

{"Cell Culture",
"DNA",
"T-Cells",
"Amputation"}
×
{long, short}
×
{text, formal}



experience in logic

- Training Phase
- Study Phase:
  Explaining
  +
  [Proof Rating,
  Ranking]

Bonus Slide 1/3: "Amputation" example

$$\mathsf{AmputationOfFinger} \equiv \mathsf{Amputation} \sqcap \exists \mathsf{locatedIn.FingerStructure}$$

$$\mathsf{AmputationOfHand} \equiv \mathsf{Amputation} \sqcap \exists \mathsf{locatedIn.HandStructure}$$

$$\mathsf{PartOfHand} \sqsubseteq \mathsf{HandStructure} \sqcap \exists \mathsf{partOf.EntireHand}$$

$$\mathsf{FingerStructure} \sqsubseteq \mathsf{PartOfHand}$$

1. Amputation of a finger is a synonym for an amputation that is located in a finger structure.
2. Amputation of a hand is a synonym for an amputation that is located in a hand structure.
3. Every part of a hand is a hand structure that is part of an entire hand.
4. Every finger structure is a part of a hand.

Bonus Slide 2/3: "Amputation" example

The ontology above implies that every amputation of a finger is an amputation of a hand:

$$\dfrac{\dfrac{\mathsf{AmputationOfFinger} \equiv \mathsf{Amputation} \sqcap \exists \mathsf{locatedIn}.\mathsf{FingerStructure} \qquad \dfrac{\mathsf{FingerStructure} \sqsubseteq \mathsf{PartOfHand} \qquad \mathsf{PartOfHand} \sqsubseteq \mathsf{HandStructure} \sqcap \exists \mathsf{partOf}.\mathsf{EntireHand}}{\mathsf{FingerStructure} \sqsubseteq \mathsf{HandStructure}}}{\mathsf{AmputationOfFinger} \sqsubseteq \mathsf{Amputation} \sqcap \exists \mathsf{locatedIn}.\mathsf{HandStructure}} \qquad \mathsf{AmputationOfHand} \equiv \mathsf{Amputation} \sqcap \exists \mathsf{locatedIn}.\mathsf{HandStructure}}{\mathsf{AmputationOfFinger} \sqsubseteq \mathsf{AmputationOfHand}}$$

short formal proof

Bonus Slide 3/3: "Amputation" example

Since every finger structure is a part of a hand and every part of a hand is a hand structure that is part of an entire hand, every finger structure is a hand structure.

Since amputation of a finger is a synonym for an amputation that is located in a finger structure and every finger structure is a hand structure, every amputation of a finger is an amputation that is located in a hand structure.

Since every amputation of a finger is an amputation that is located in a hand structure and amputation of a hand is a synonym for an amputation that is located in a hand structure, every amputation of a finger is an amputation of a hand.

short textual proof

PROCEDURE

Where: Online with Zoom and Go2Meeting

Who: 16 students with some experience in formal logic,
Mean Age = 23, SD = 1.71

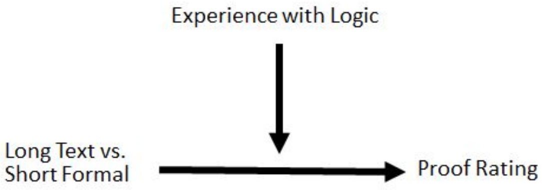
Statistics: Multiple linear regression with contrast coding,
Friedman's ANOVA for the ranking

	Mean	*SD*	*Range*
Propositional Logic	3.25	1.00	1-5
First-Order Logic	2.94	0.68	2-4
Description Logic	2.31	0.79	1-4

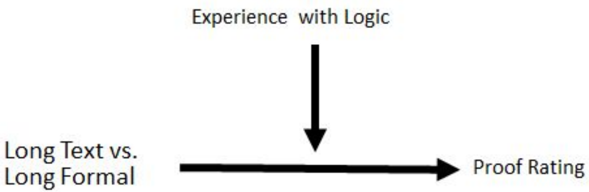
Quantitative results

- Hypothesis 1: shorter proofs rated as easier than the longer ones (independent of the presentation format)
  - 14.2% explained variance in the rating after each proof, $R^2 = .14$, $F(3,60) = 3.30$, $p < .05$, $\beta = -.29$, $t(60) = -2.42$, $p < .05$

- Hypothesis 2:

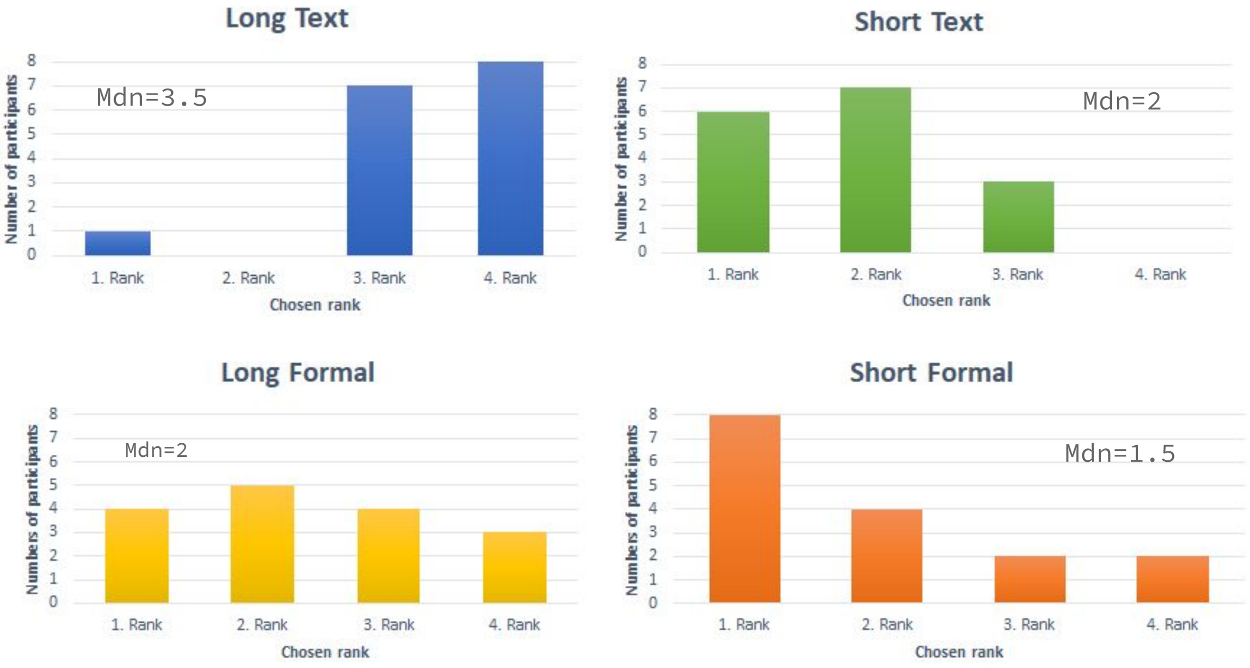


- Hypothesis 3:



Quantitative results - Ranking

- Significant model, χ2(3) = 15.29, p<.01; moderate effect size (Kendall's W=.32)
- Short Text > Long Text, Z = 1.53, p<.008, Moderate effect size, r=0.38
- Short Formal > Long Text, Z = 1.50, p<.008, Moderate effect size, r=0.38

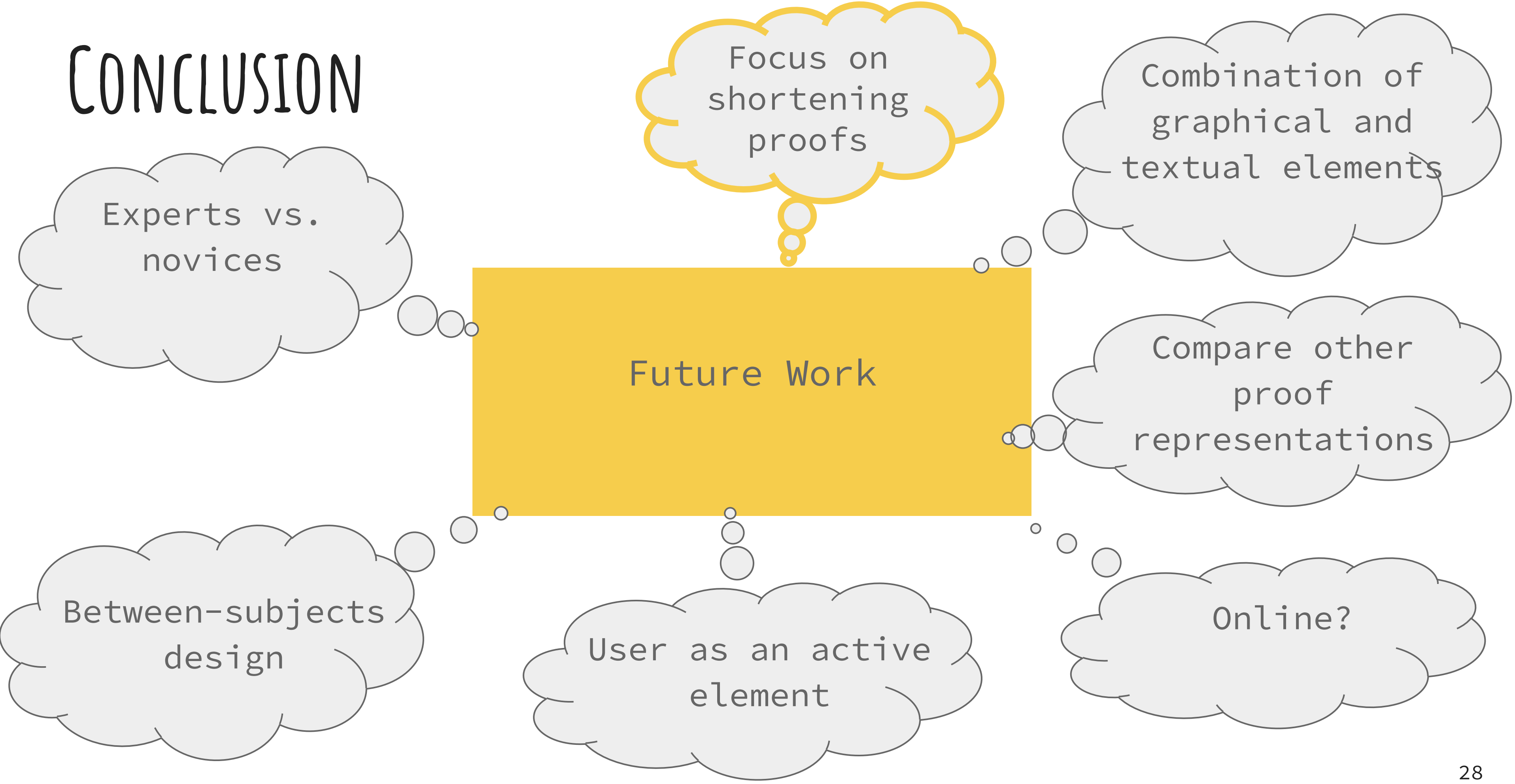


Qualitative Results

- Formal proofs preferred over textual proofs

Formal Proof	Textual Proof
“easier to understand” “clearer” “easier to find certain parts” “orientation is better” within the proof “easy to follow the proof”	Inconvenient “less understandable” “hard to understand” “annoying”

Conclusion


Future Work
Focus on shortening proofs
Experts vs. novices
Combination of graphical and textual elements
Compare other proof representations
Between-subjects design
User as an active element
Online?

Thank you
for your attention