Computation & Cognition Lab

ngoodman@stanford.edu

How we come to know

Noah D. Goodman
Stanford University

College de France
Feb. 11, 2019

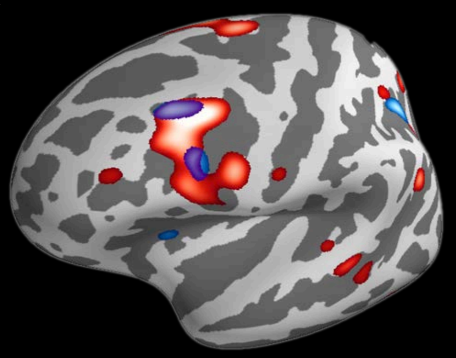
The problem of induction

20 years

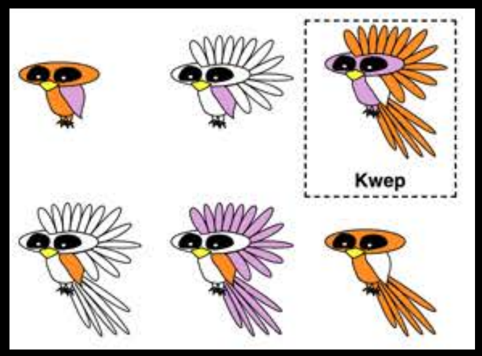
- In just a few years, babies go from knowing very little to building rocket ships and twitter accounts.
- How do people learn so much from so little?

$$f_1(x) = 1 \wedge f_3(x) = 0$$

An innate "language of thought" in which complex concepts can be built from simple pieces

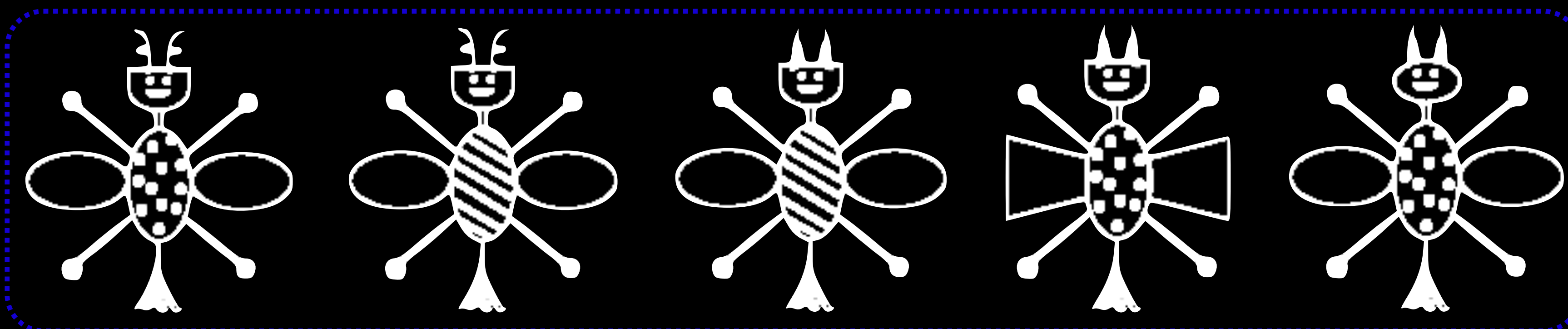


A powerful learning mechanism to go from *examples* to *concepts*

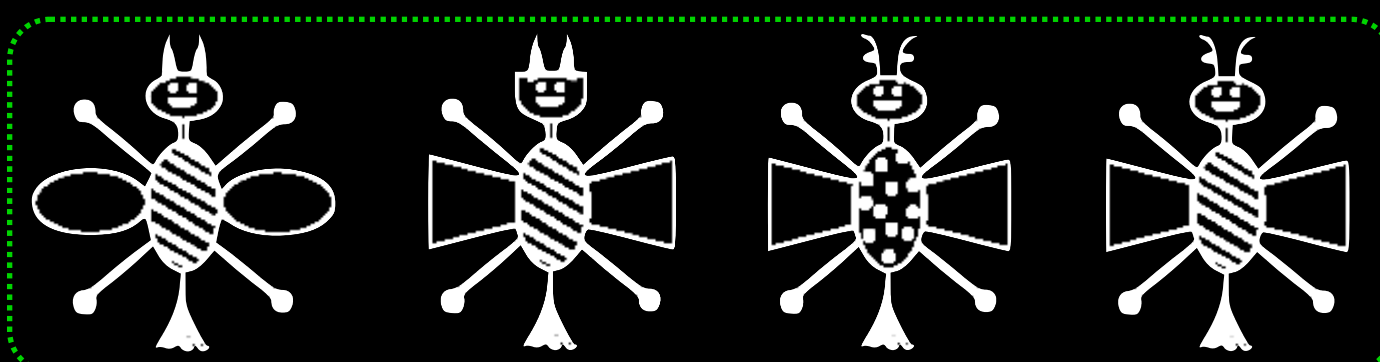


Concept learning

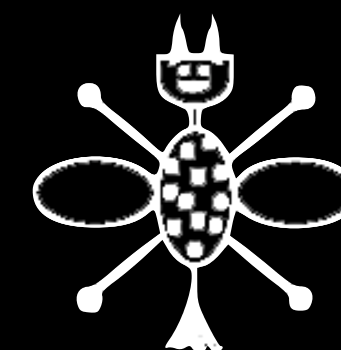
Medin & Schaffer (1978):



“These are **Kweps**”



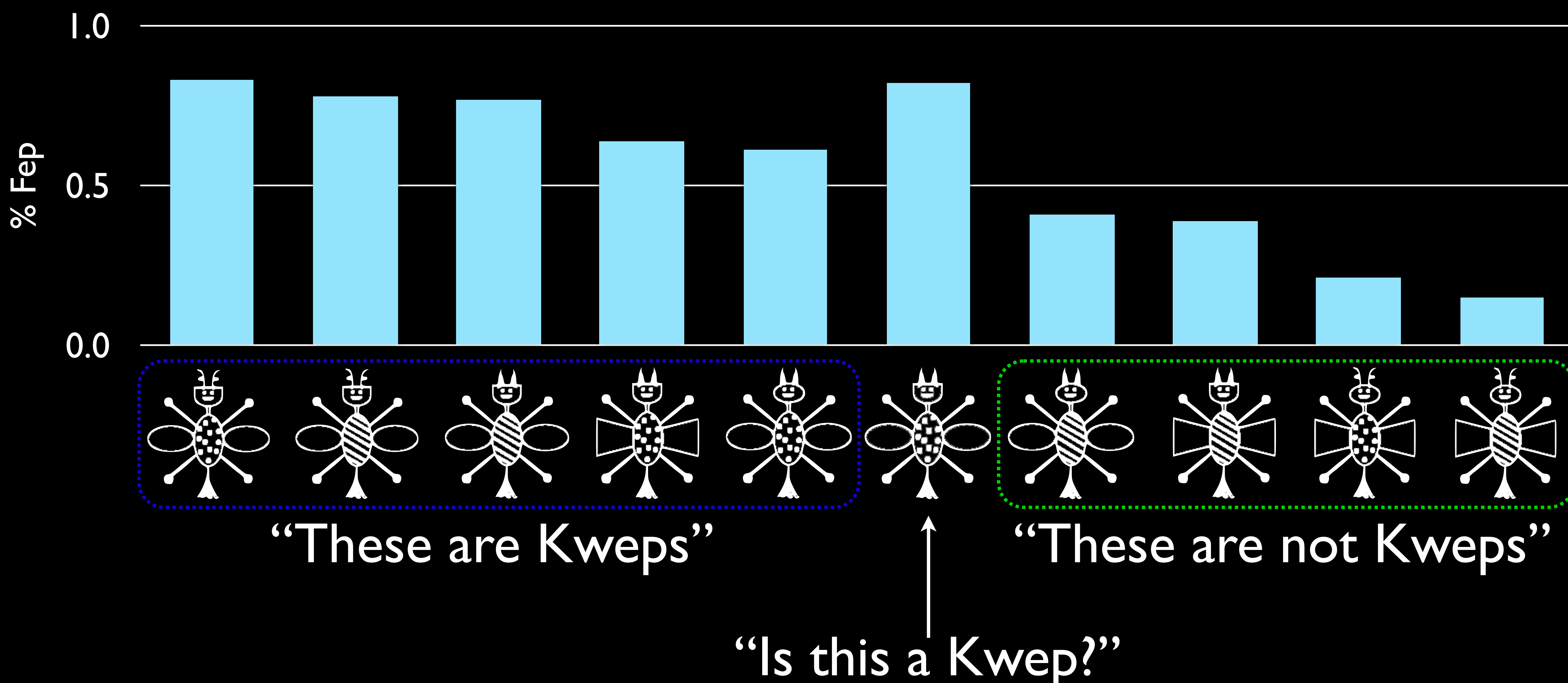
“These are **not Kweps**”



“Is this a Kwep?”

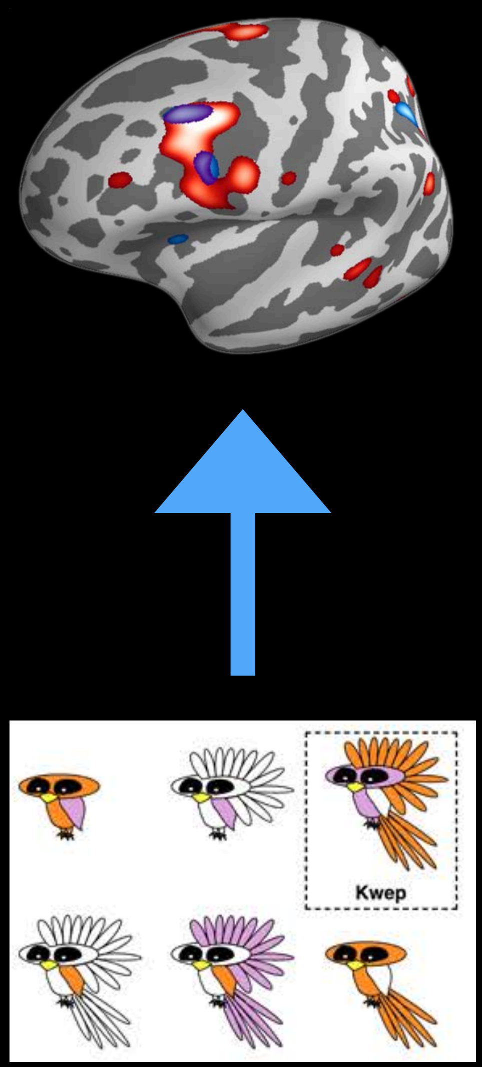
Concept learning

Medin & Schaffer, 1978 (data from Nosofsky, et al.,1994):



- Typicality
- Prototype enhancement

- Graded judgements



A powerful learning mechanism to go from *examples* to *concepts*

Bayes' rule

$$P(h|d) \propto P(d|h) \cdot P(h)$$

The probability of a hypothesis, h, given observed data, d.

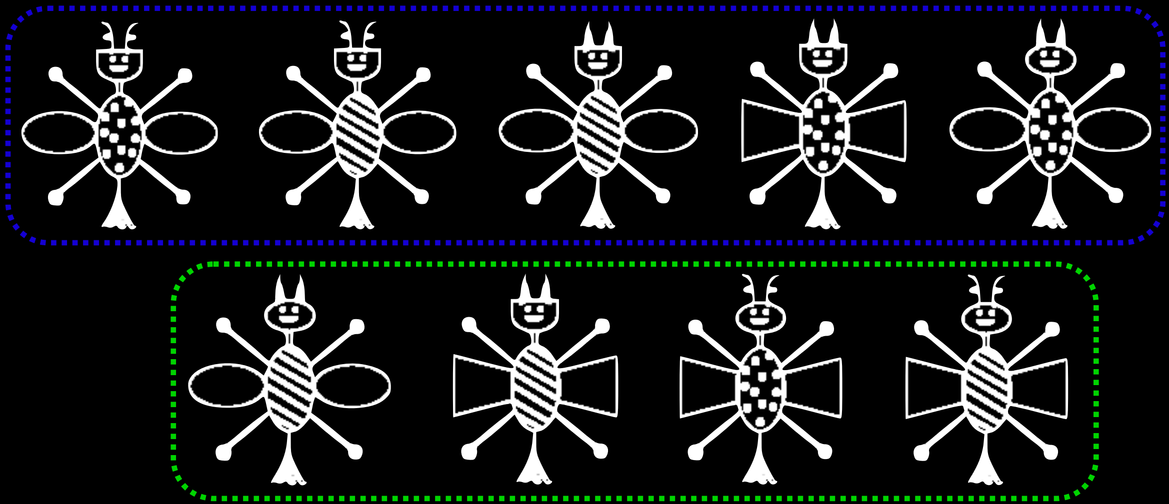
The likelihood of that data, if the hypothesis is true.

How much we believe the hypothesis *a priori*.

- Bayes' rule tells us *what* to learn from observations, given *prior* and *likelihood*.

Rule hypotheses

"It's a Kwep if it has flat head and round wings"



"Kwep if *rule*."

$$\textit{rule} \longrightarrow \text{feature}$$

$$\textit{rule} \longrightarrow \textit{rule} \text{ and } \textit{rule}$$

$$\textit{rule} \longrightarrow \textit{rule} \text{ or } \textit{rule}$$

$$f_1(x) = 1 \wedge f_3(x) = 0$$

- We can derive an infinite set of possible rules from finite features and simple combinations (a *grammar*).

Rule prior probability

"Kwep if *rule*."

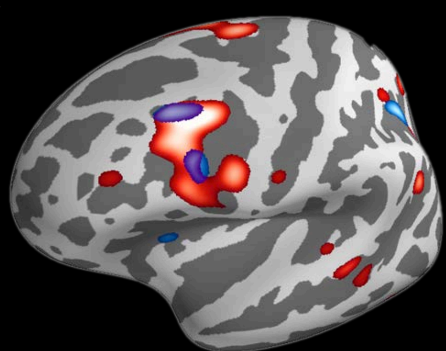
$\text{rule} \xrightarrow{50\%} \text{feature}$

$\text{rule} \xrightarrow{30\%} \text{rule and rule}$

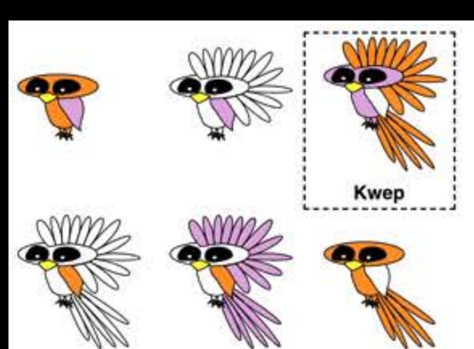
$\text{rule} \xrightarrow{20\%} \text{rule or rule}$

- Assign a probability to each rule-building step (a *probabilistic* grammar).
- The overall probability of a rule is the probability of all choices to make it.
  - Longer rules are less likely *a priori*.

$$f_1(x) = 1 \wedge f_3(x) = 0$$



$$P(h|d) \propto P(d|h) \cdot P(h)$$



Grammar gives "language of thought" for rules together with prior probabilities.

"Kwep if *rule*."

rule $\longrightarrow$ feature

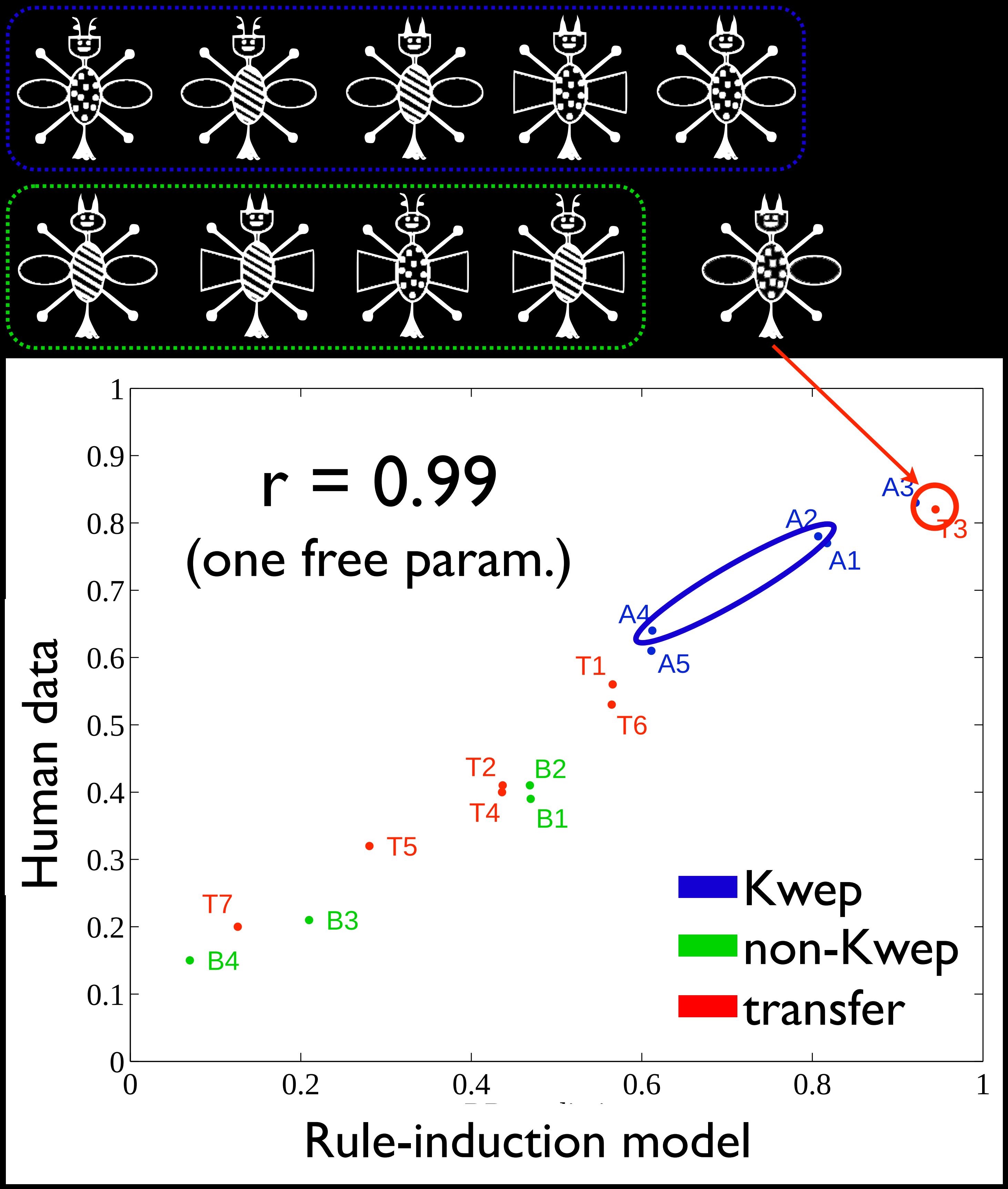
rule $\longrightarrow$ rule and rule

rule $\longrightarrow$ rule or rule

Simple noise likelihood: the rule is right with a high probability.

$$P(\text{Kwep}|\text{rule}(x)) = \epsilon$$

Example: concept learning

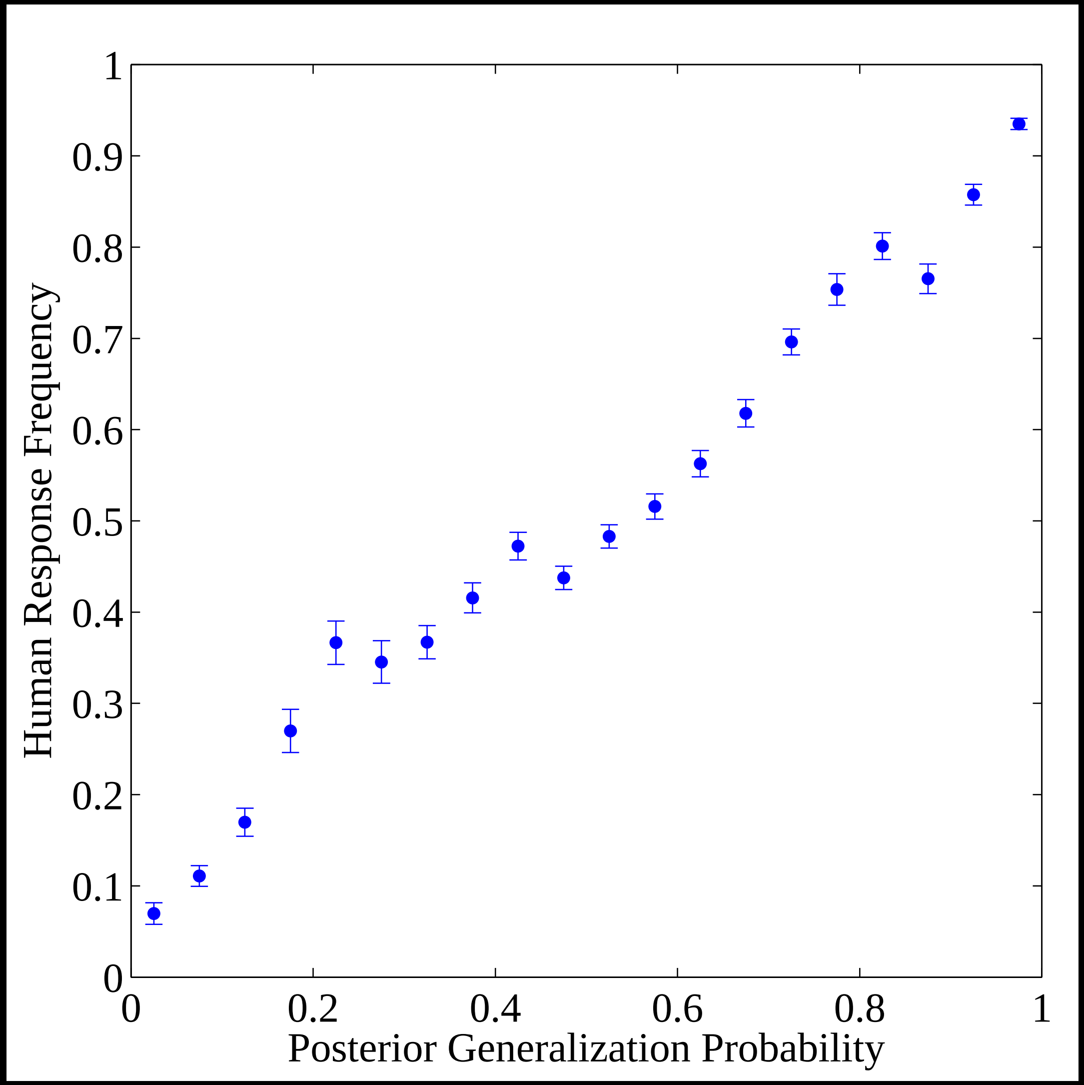


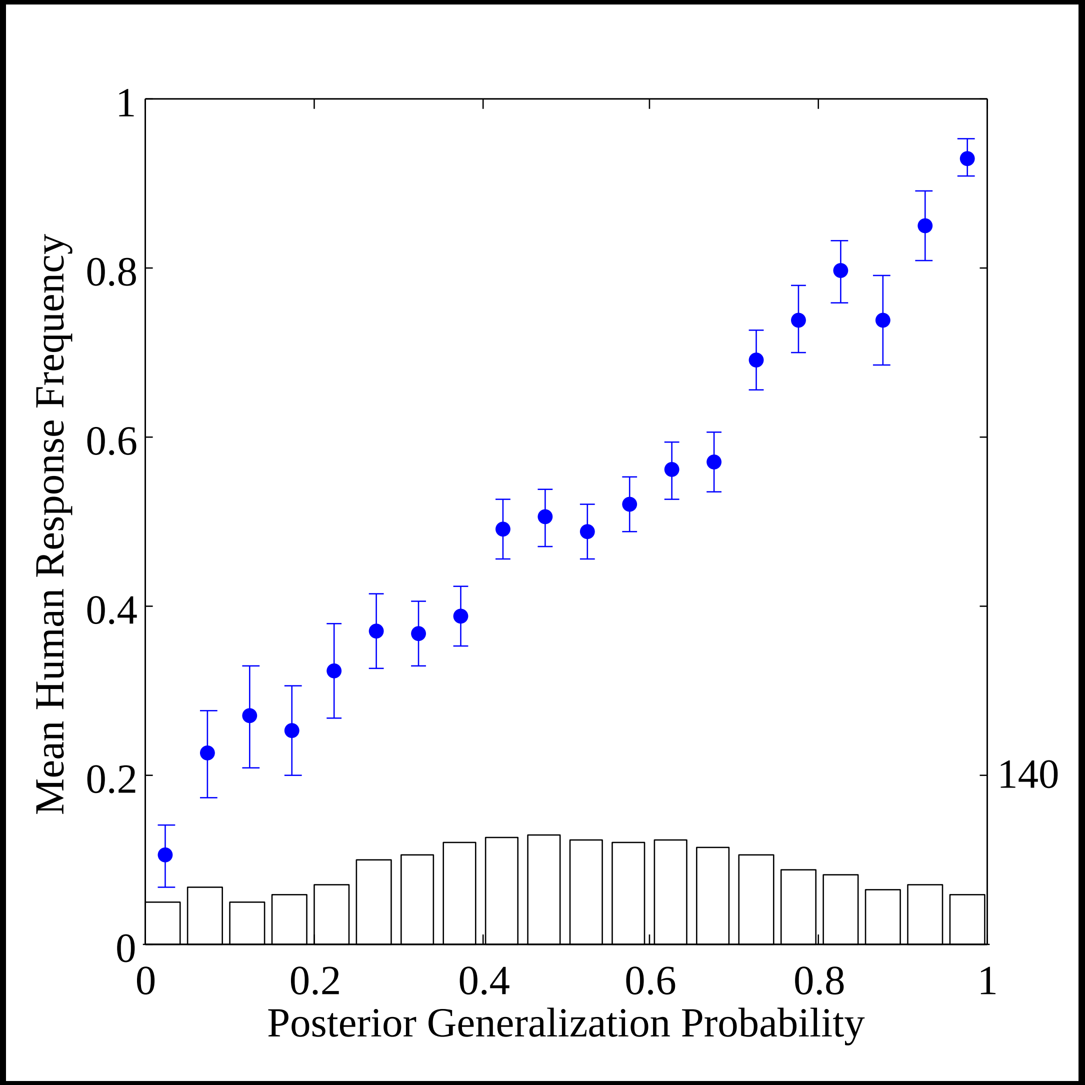
- Graded judgments
- Typicality
- Prototype enhancement

Goodman, et al.
(2007, 2008a, 2008b)

Broader test

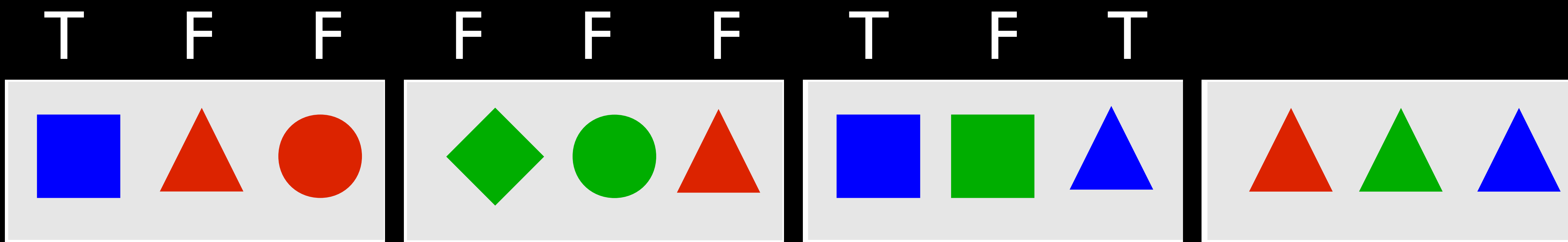
- 7 Boolean features.
- 43 randomly generated concepts (3-6 pos. + 2 neg. exs)
- 128 judgements (~122 transfer questions)





Goodman, et al (2008)

Complex concepts

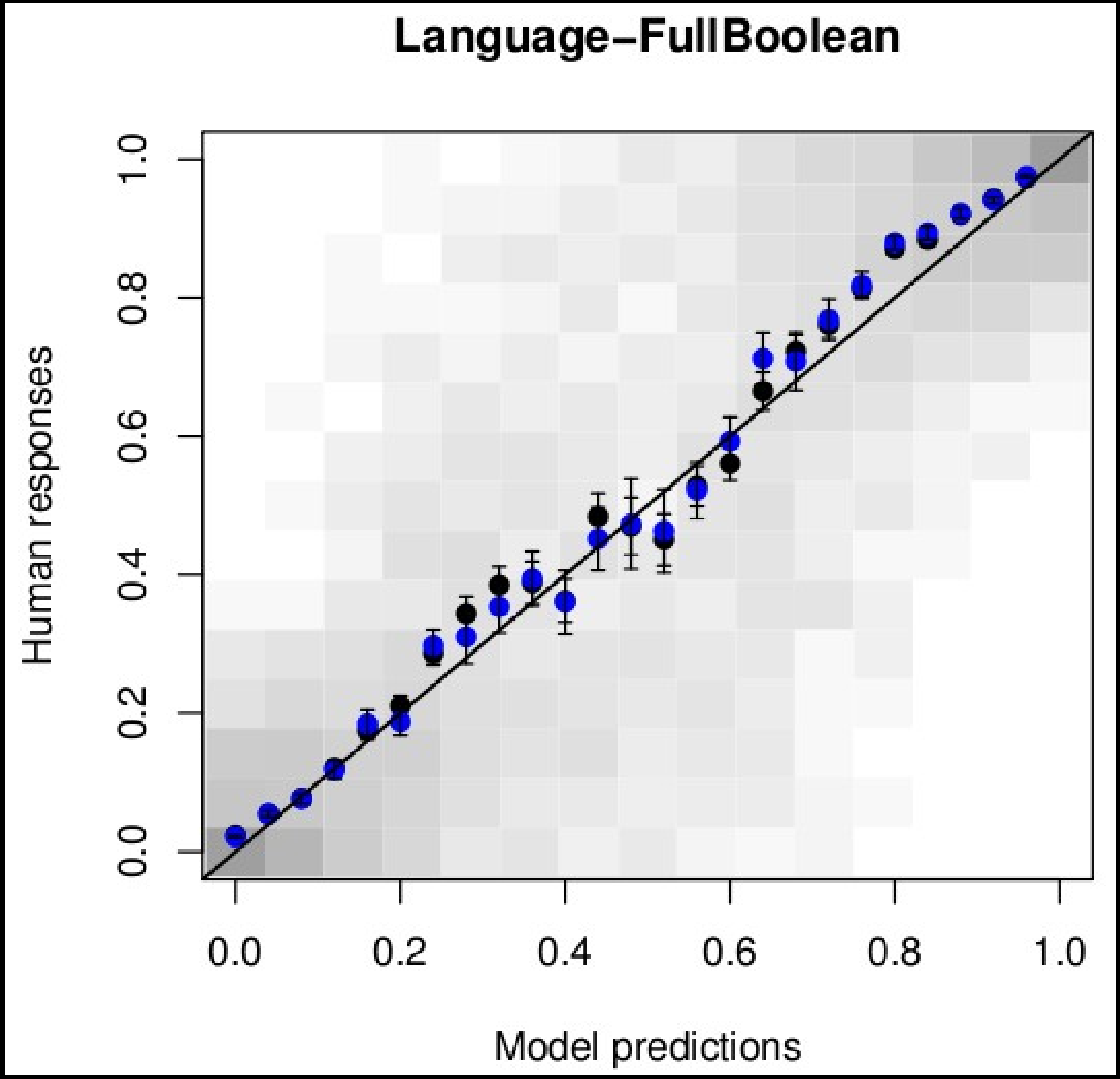


- Big online experiment.
  - 108 concepts,
    - Boolean (*circle or red*)
    - Context-dependent (“Determiners”)
      (*unique largest , exists another with same shape*)
  - 2 orders per concept,
  - 1596 participants.

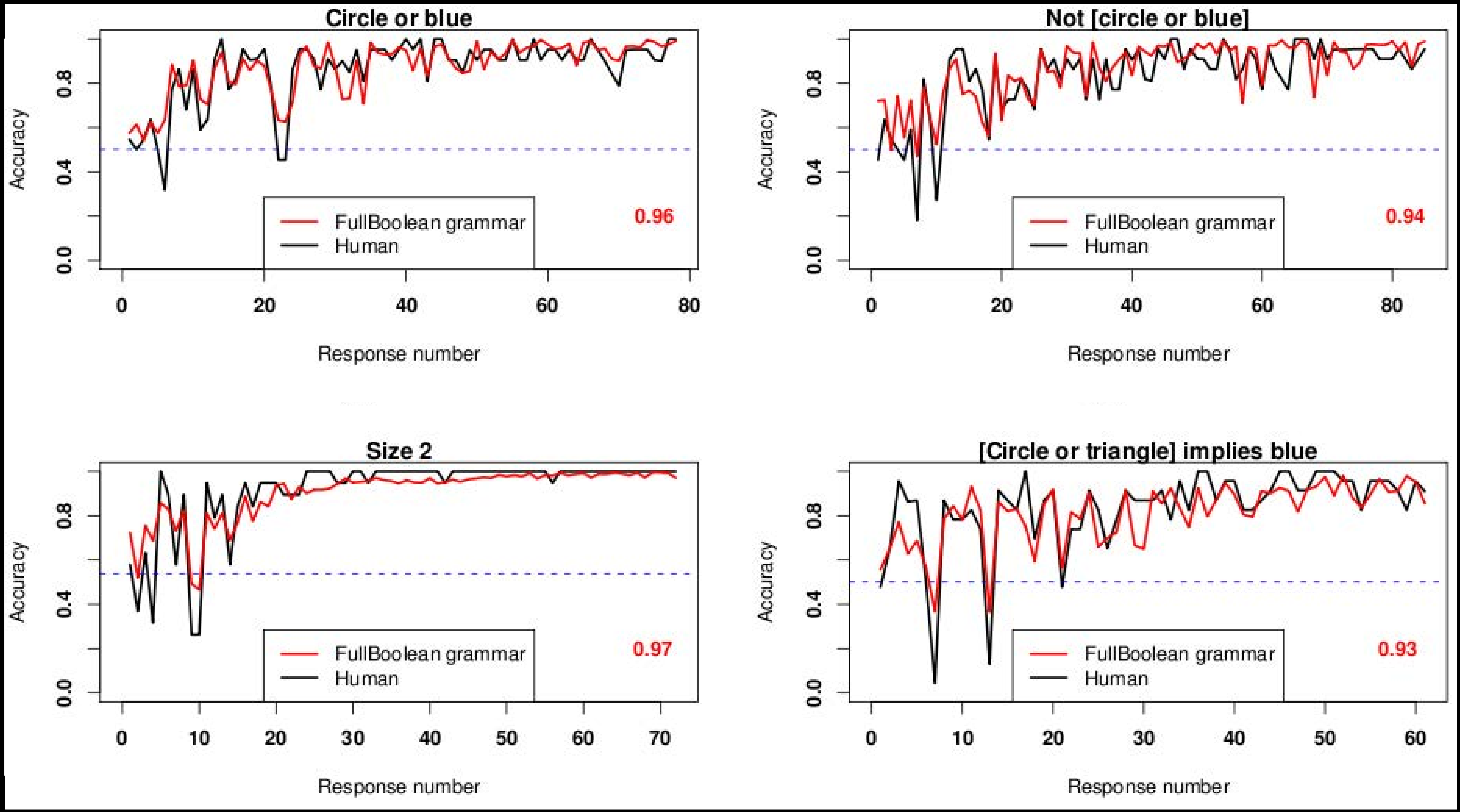
Piantadosi, Goodman, Tenenbaum (2016)

Boolean concepts

Learning Boolean concepts,
model performance on Boolean concepts:

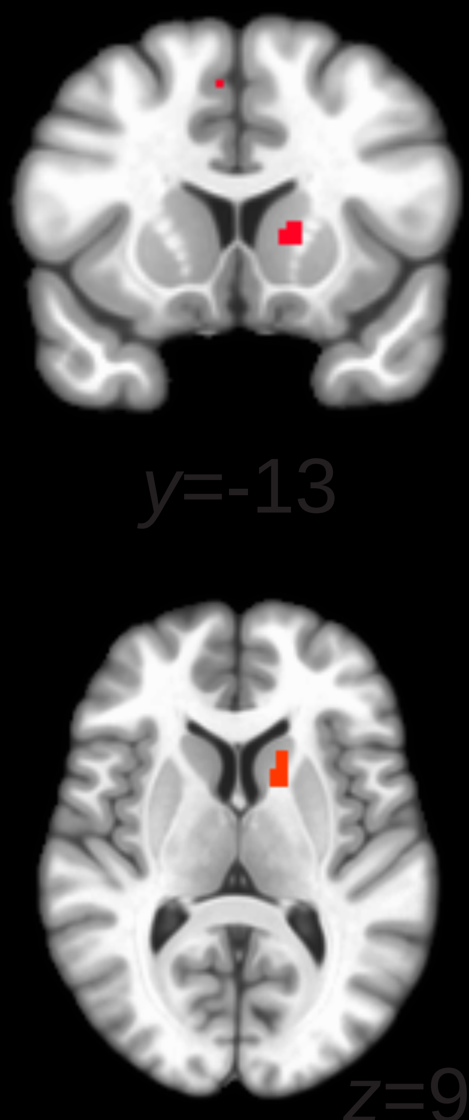


Boolean concepts

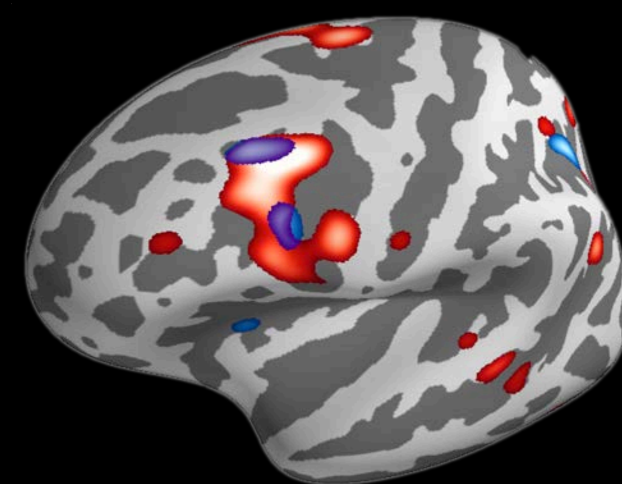


It’s in the brain

RR model surprisal correlates with striatum activity.



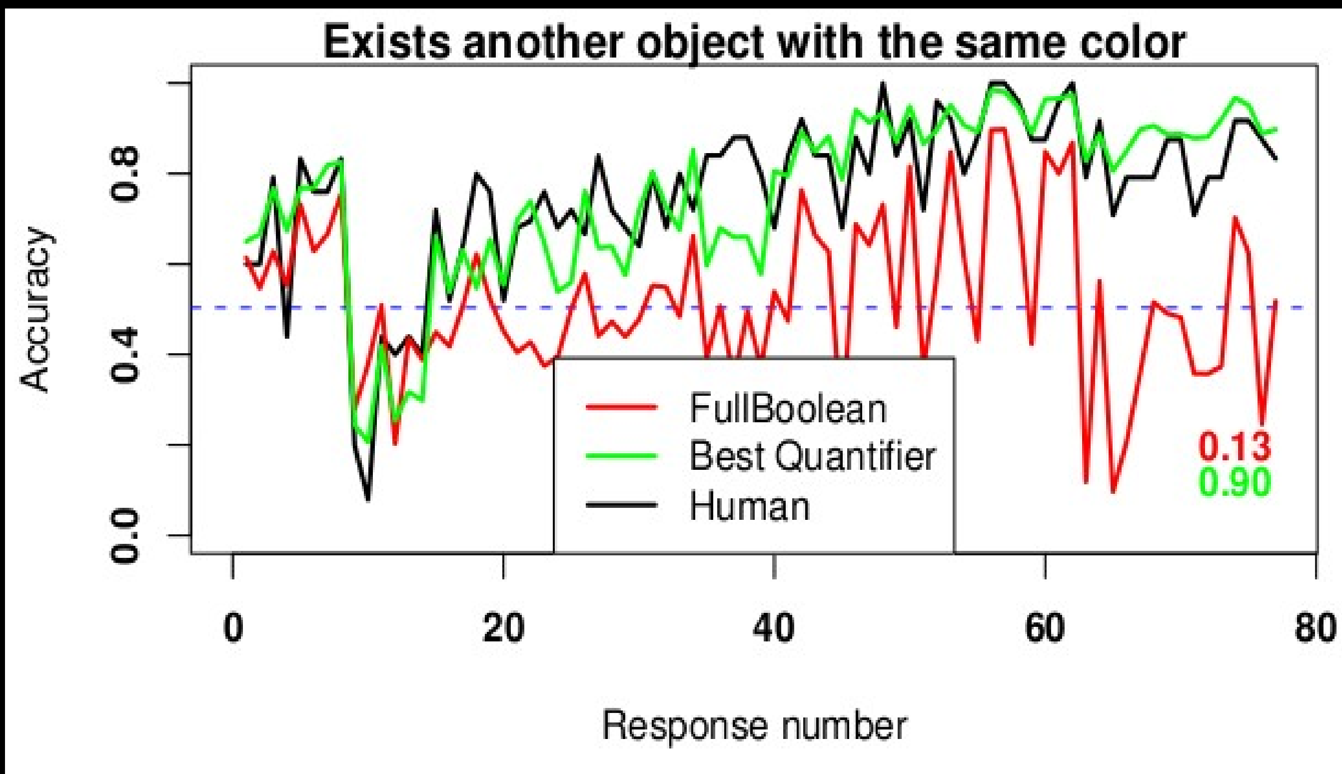
RR model posterior update correlates with DLPFC.

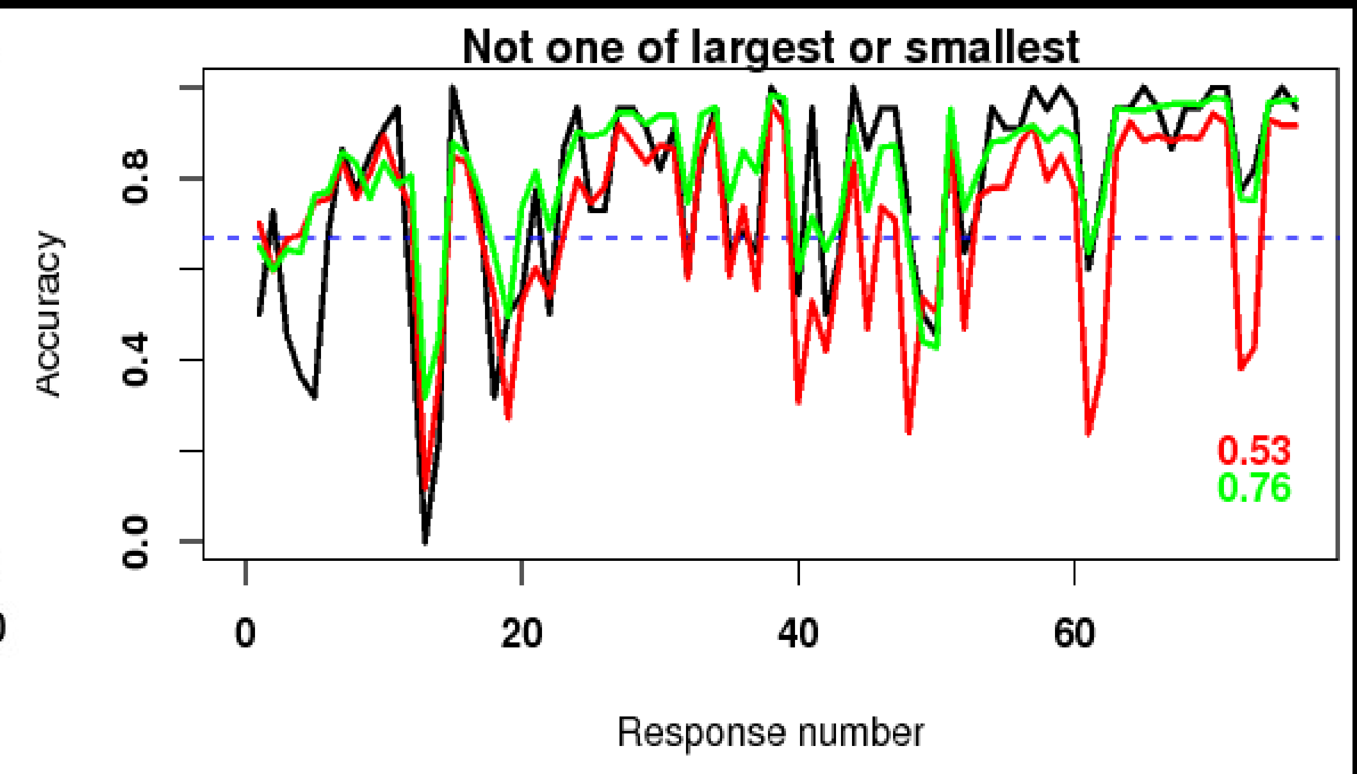


Ballard, Miller, Piantadosi, Goodman, McClure (2018)

Non-Boolean concepts

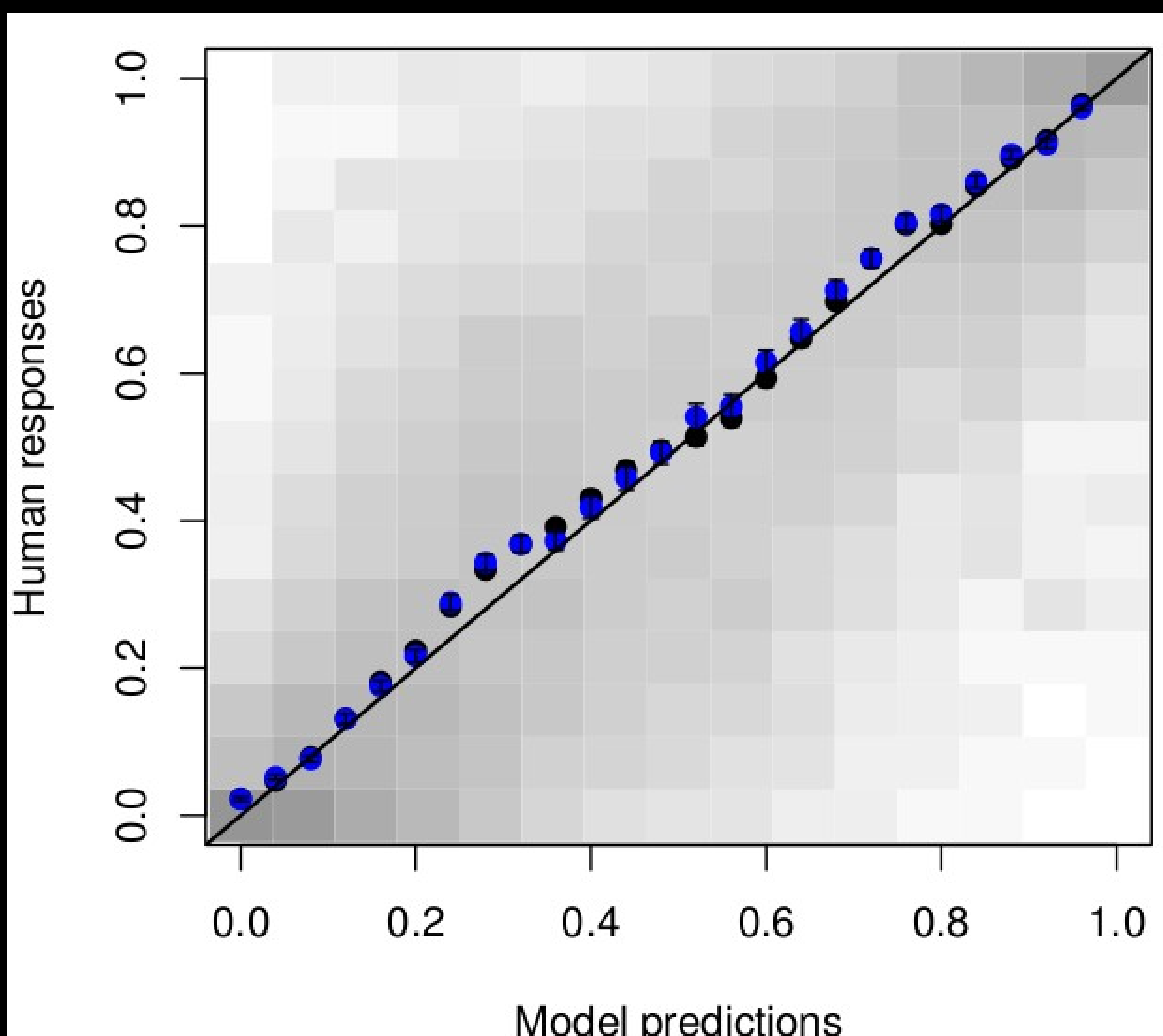
- Experiment included context-dependent (determiner-like) concepts.
- What languages explain inductive bias for these non-boolean concepts?





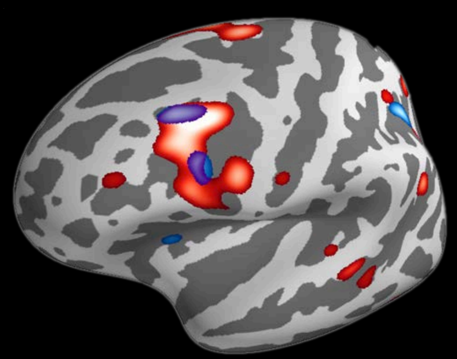
Non-Boolean concepts

- Best hypothesis space is full boolean logic plus quantifiers.

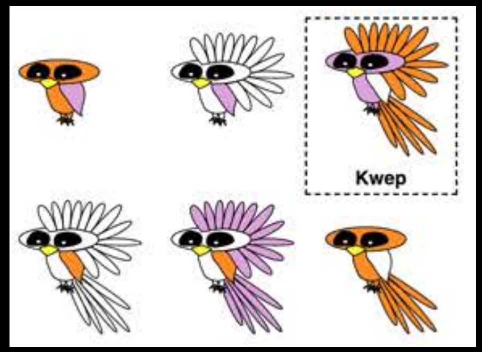


FOL	One-Or-Fewer	Small-Cardinalities	2nd-Ord.-Quan.	H.O. LL
✓	·	·	·	-79279.95
✓	✓	·	·	-79560.90
·	✓	·	·	-79642.46
·	✓	✓	·	-79972.75
✓	✓	·	✓	-80198.75
✓	·	·	✓	-80267.46
✓	·	✓	·	-80285.38
·	✓	·	✓	-80300.00
·	·	✓	·	-80614.35
✓	✓	✓	·	-80942.77
✓	✓	✓	✓	-81138.27
·	✓	✓	✓	-81289.85
✓	·	✓	✓	-81596.68
·	·	✓	✓	-81651.36
FULLBOOLEAN				-81773.43
BICONDITIONAL				-81967.68
SIMPLEBOOLEAN				-82144.71
·	·	·	·	-82219.08
CNF				-82685.21
DNF				-82752.82
·	·	·	✓	-82853.59

$$f_1(x) = 1 \wedge f_3(x) = 0$$

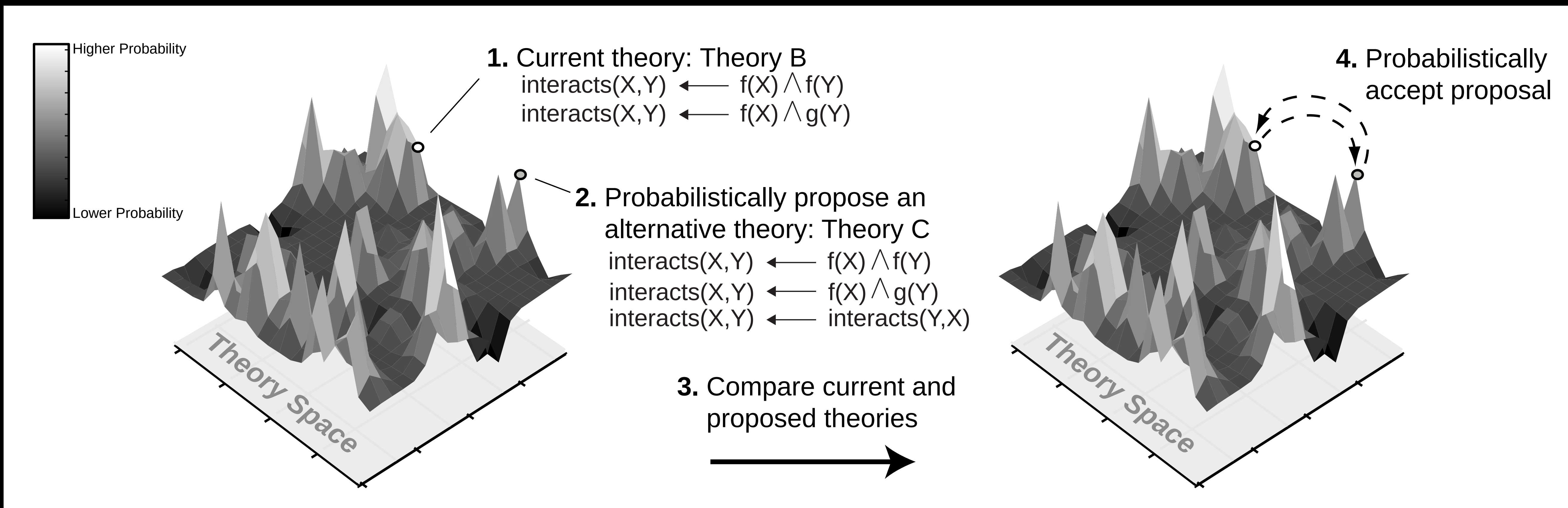


$$P(h|d) \propto P(d|h)P(h)$$



Bayes' rule lets us
compare candidate rules.
How do we find them?

Finding rules?

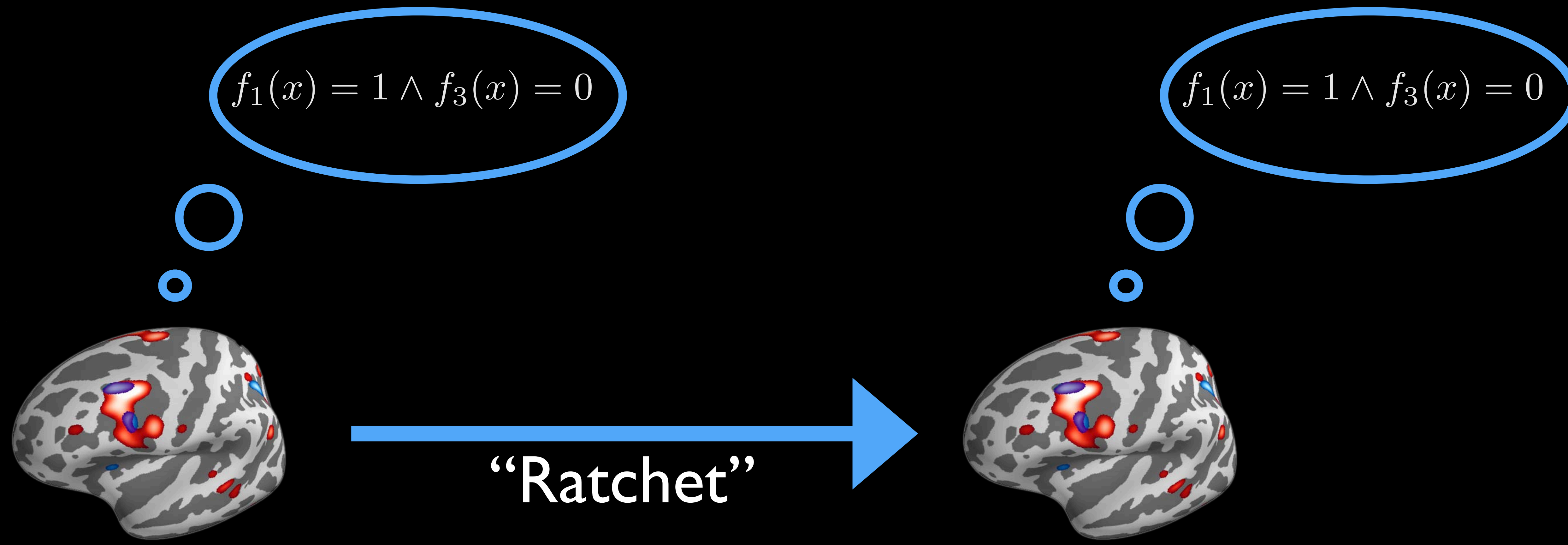


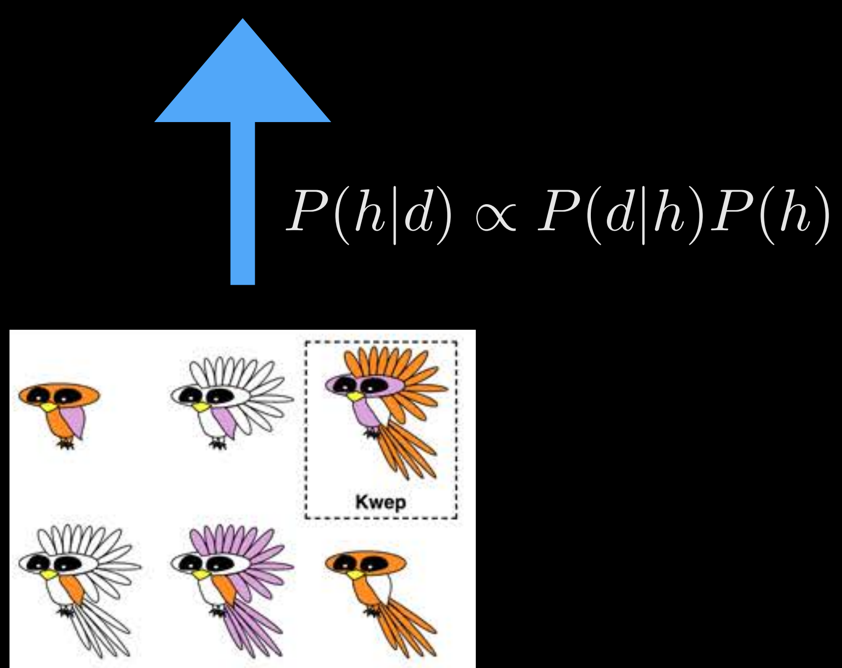
- Random search works for simple concepts…

Ullman, Goodman, Tenenbaum (2012)

The problem of induction

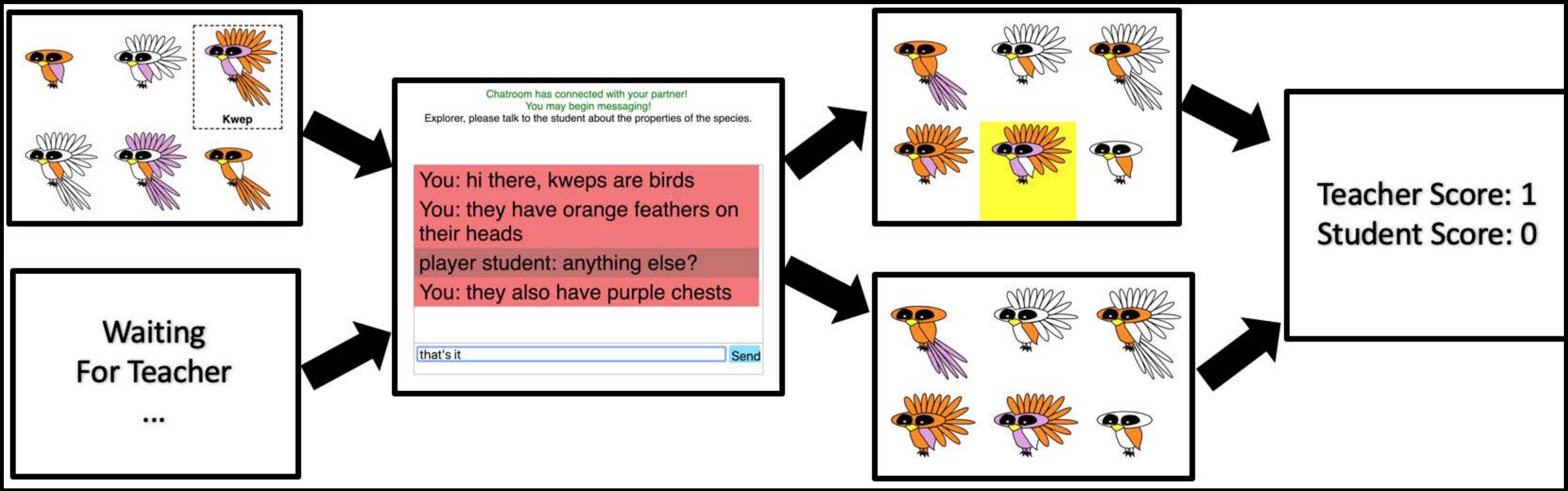
- Concept learning quickly gets hard for people…
  - How do we learn many complex concepts with many features from lots of data?
- A solution: amplify limited individual learning by accumulation over generations — the "cultural ratchet" (Tomasello, 1999).





This requires **faithful transmission** of knowledge, and it has to be easier than directly learning from examples.

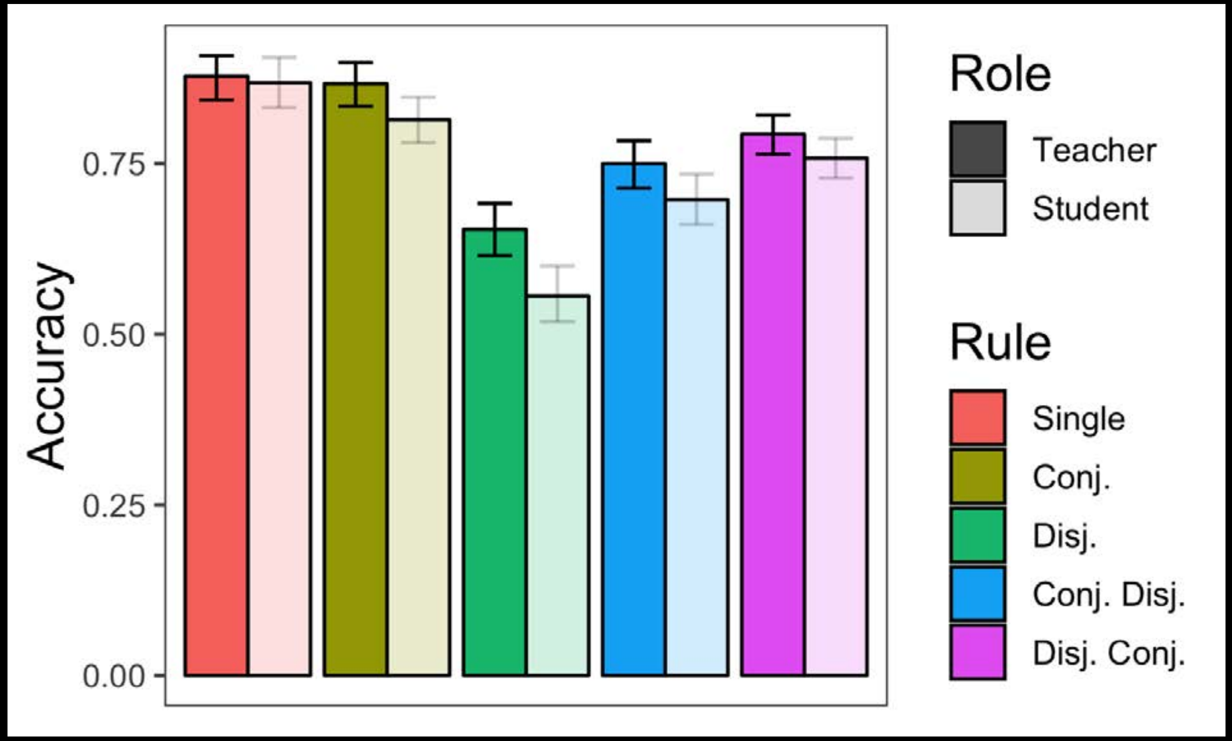
Learning from language



- A minimal paradigm to compare concept learning from observing examples and from linguistic communication.

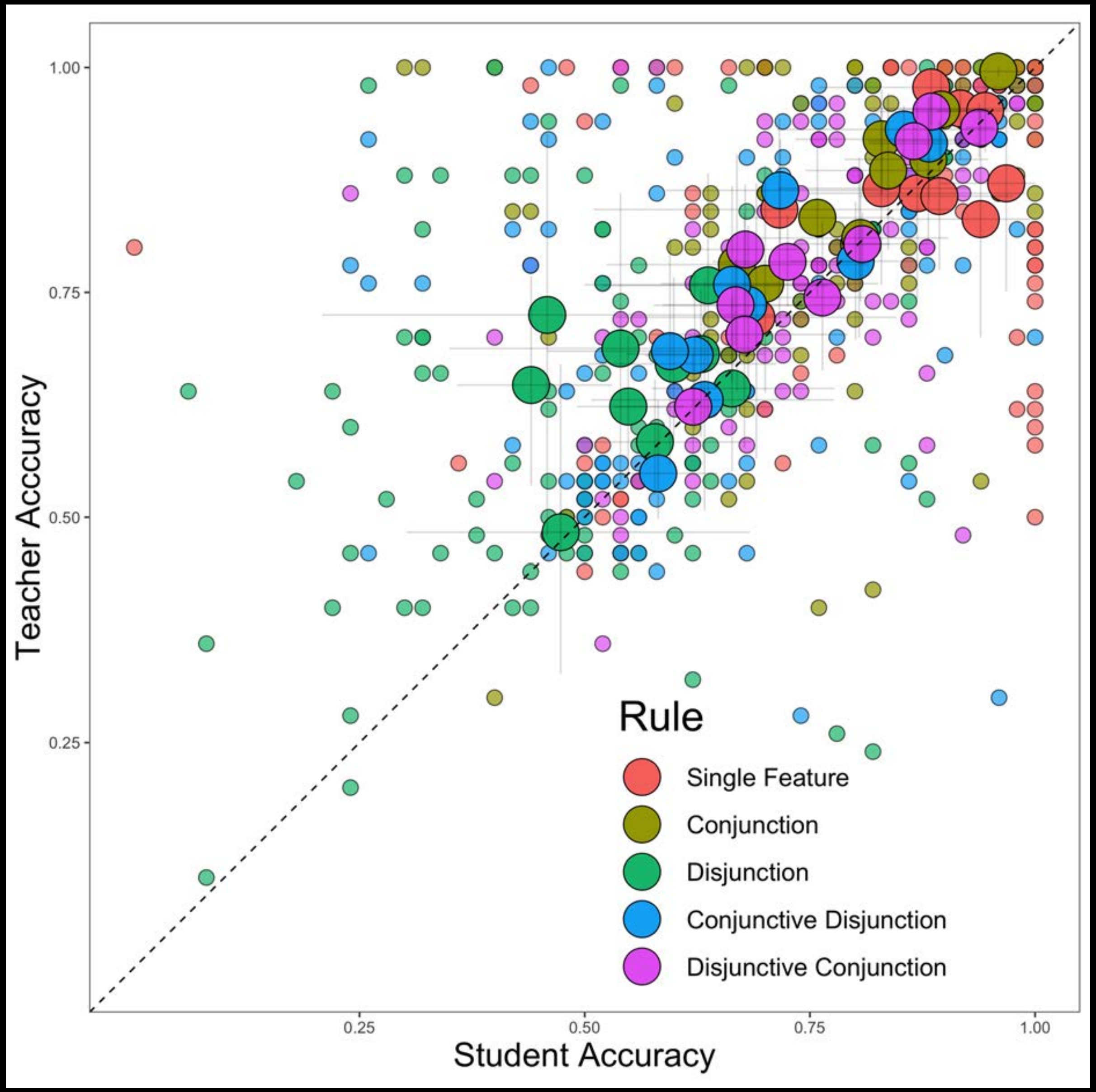
Chopra, Tessler, Goodman (subm)

Results



- Language is *sufficient:*
  - Students who learn from language perform only slightly worse than their teacher.
  - (Approx. 5% lower accuracy for students, by Bayesian mixed effects model.)

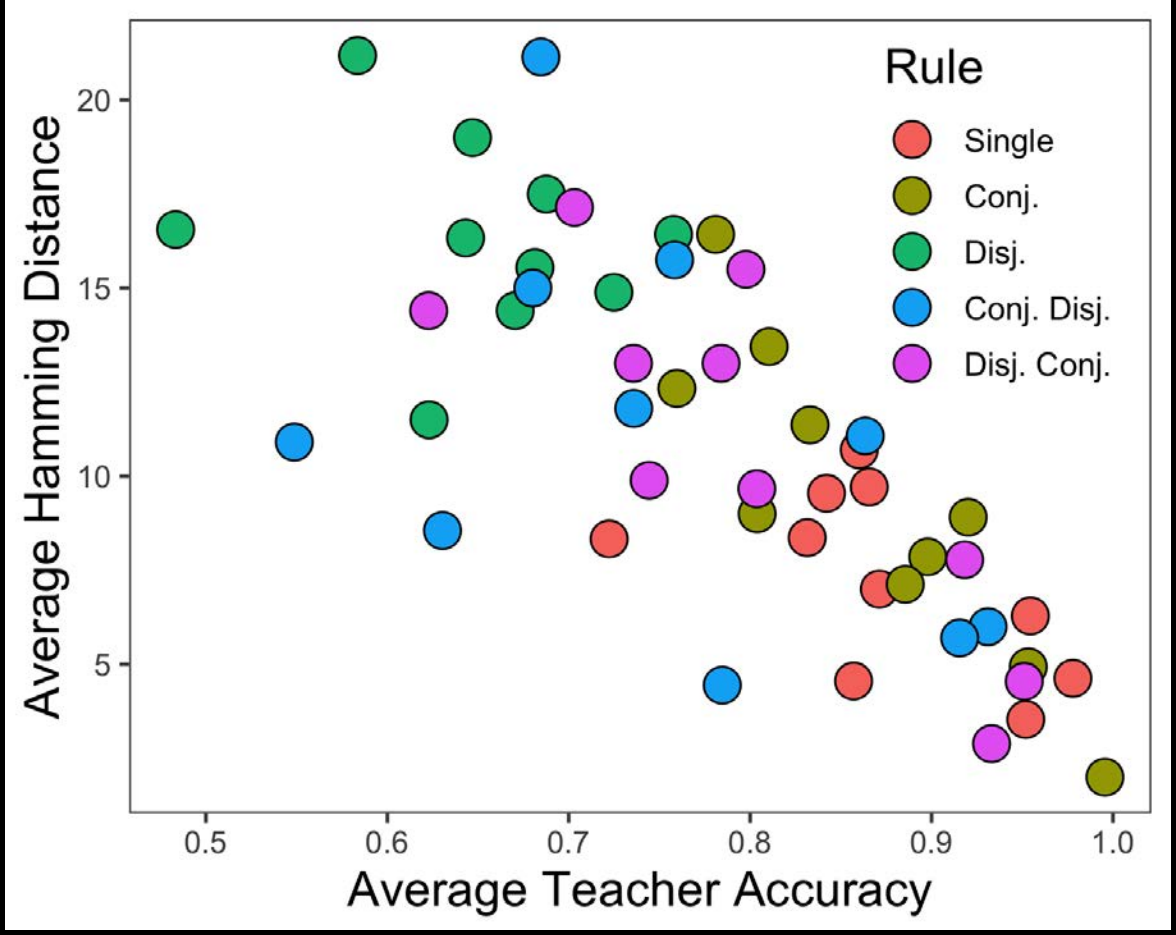
Results



- Teacher accuracy predicts student accuracy.

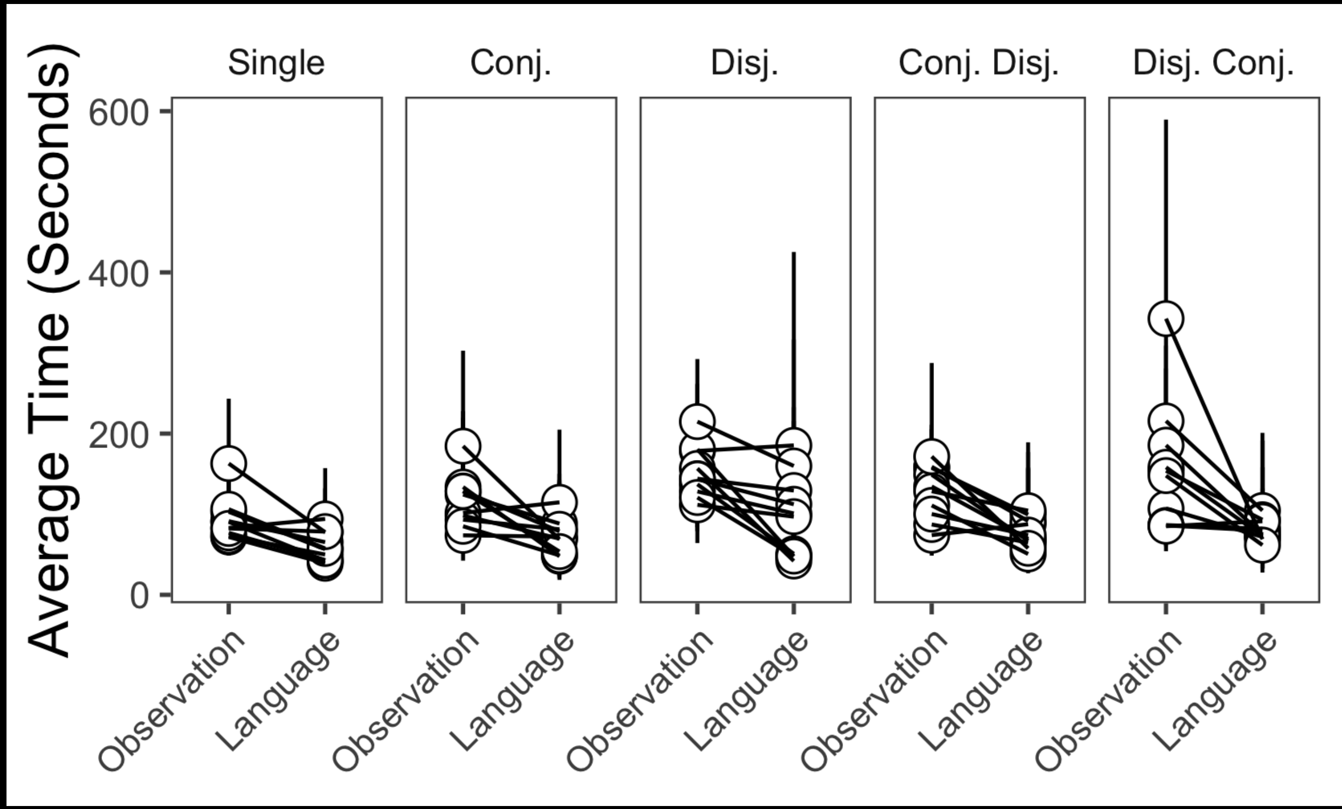
Results

- Individual students make the same mistakes as their teachers (hamming distance lower than permutation baseline).



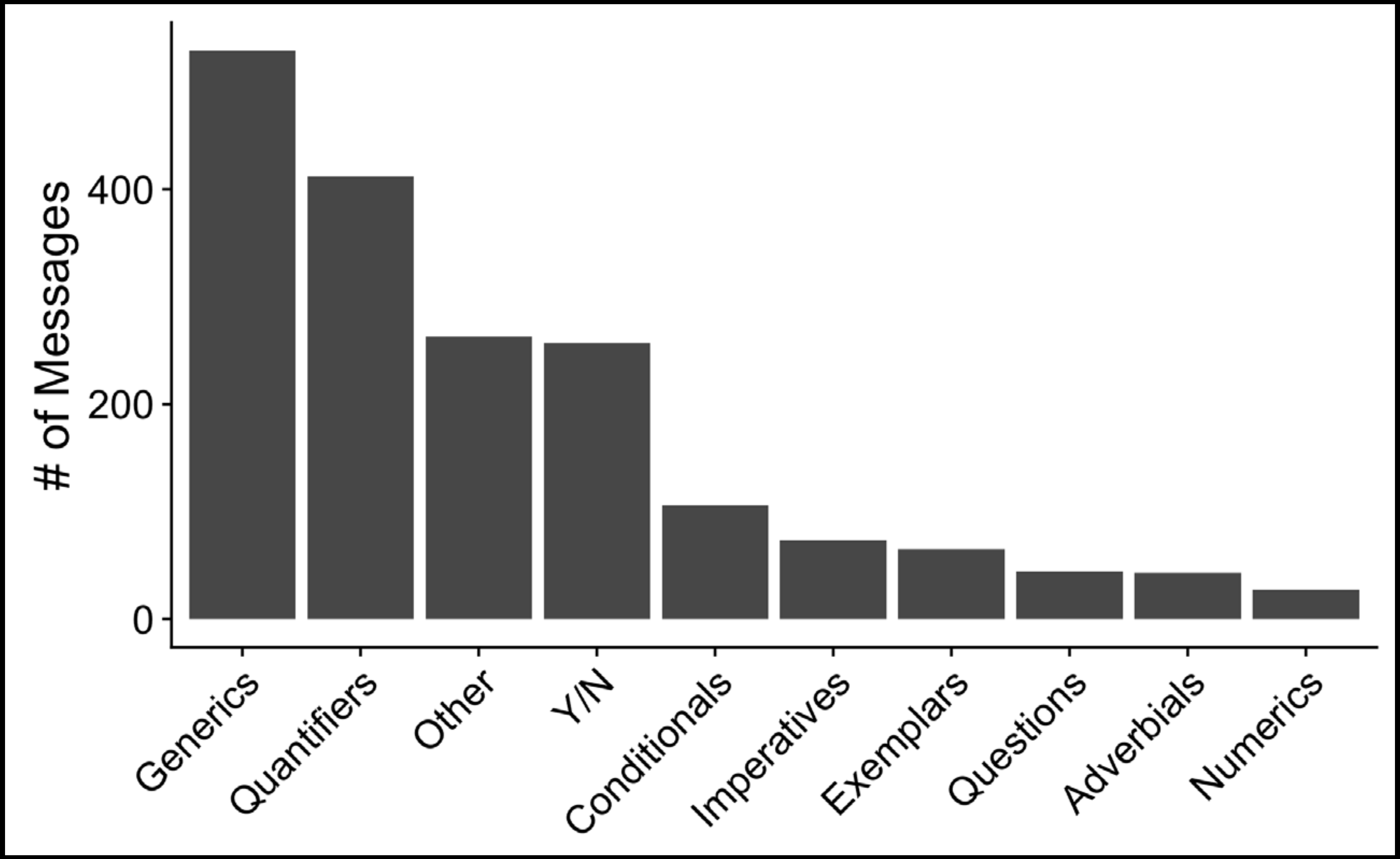
Average of 2.4 more different answers from a student to a different teacher, in same concept.

Results



- Language is *efficient*.
  - Participants spent longer learning from examples than from language. (Both were freely determined by participants.)

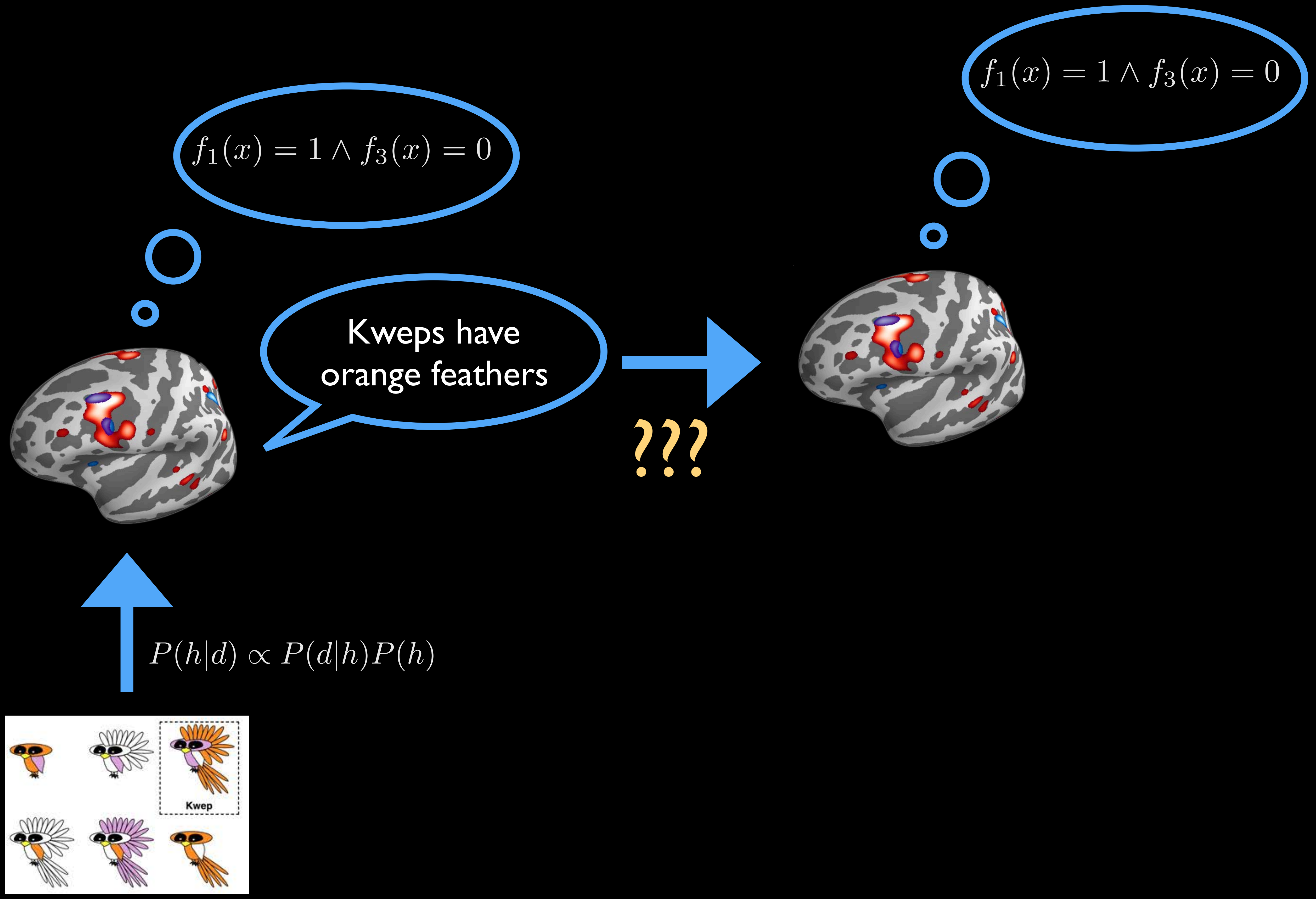
Results



- It's the language of generalization that matters.
  - Most messages use generics or quantifiers.

Hypothesis

- Claim: The cultural ratchet arises specifically out of the ability of language to convey generalizations through generics and quantifiers.

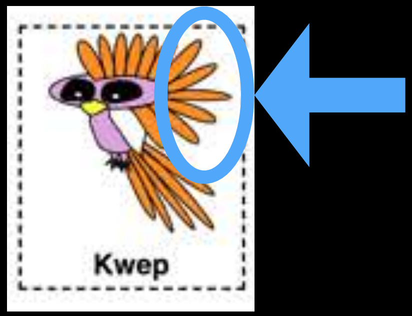

$f_1(x) = 1 \wedge f_3(x) = 0$
Kweps have orange feathers
???
$f_1(x) = 1 \wedge f_3(x) = 0$
$P(h|d) \propto P(d|h)P(h)$
Kwep

Generics

“Wugs have red legs.”

“Wugs have broken wings.”

- Two ideas of how generics work:
  - they provide a minimal example,
  - they have social force — they’re *intended* examples.



“Mosquitos cary malaria.”

“Birds lay eggs.”

“Birds are female.”

Formalizing generics

- Let r be the probability of feature F for objects of category C.
  - Generics provide a minimal example. By Bayes rule:

$$P_{L_0}(r|\text{“Cs F”}) \propto r \cdot P(r)$$

  - Generics have social force — they're *intended* examples (Cf RSA models, Goodman and Frank, 2016):

$$P_{L_1}(r|\text{“Cs F”}) \propto P_S(\text{“Cs F”}|r)P(r)$$

$$P_S(\text{“Cs F”}|r) \propto P_{L_0}(r|\text{“Cs F”})$$

Prior elicitation

Category elicitation

supplied to participants

- Kangaroos
- Robins
- Sharks
- Mosquitos
- Ducks
- Ticks

participants generate animal kinds

Continue

Prevalence elicitation

For each kind of animal, what percentage of the species do you think

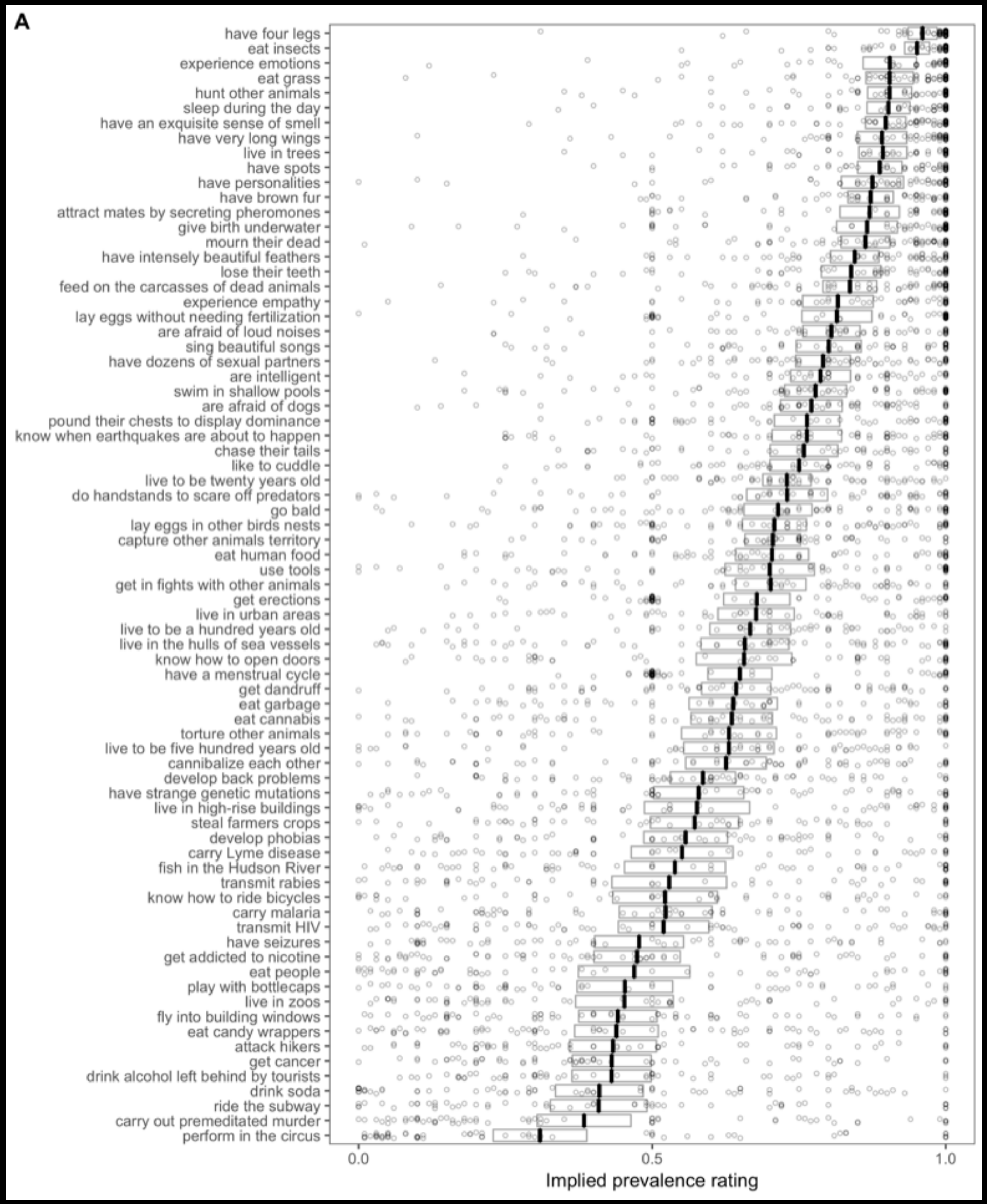
	carry malaria
Kangaroos	0 %
Robins	0 %
Sharks	0 %
Mosquitos	10 %
Ducks	0 %
Ticks	3 %
dogs	1 %
cats	1 %
geese	0 %
monkeys	1 %
falcons	0 %

n = 60 from Amazon’s Mechanical Turk

filtering 0% responses

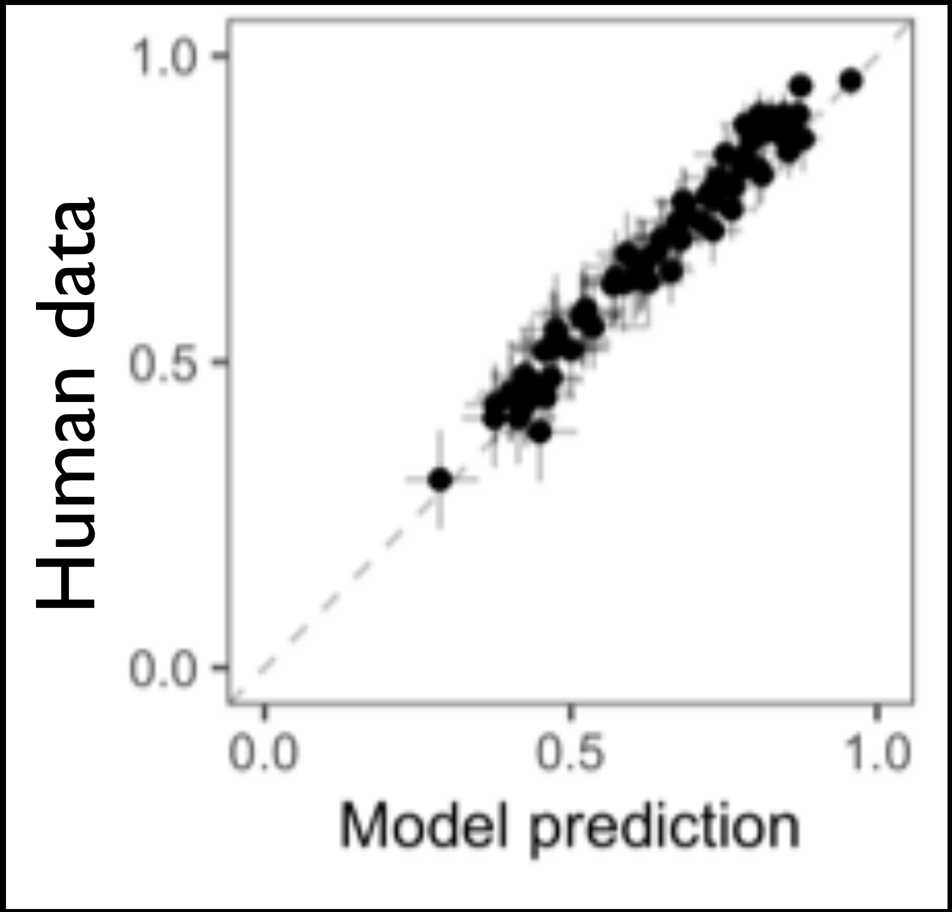
21 properties in total

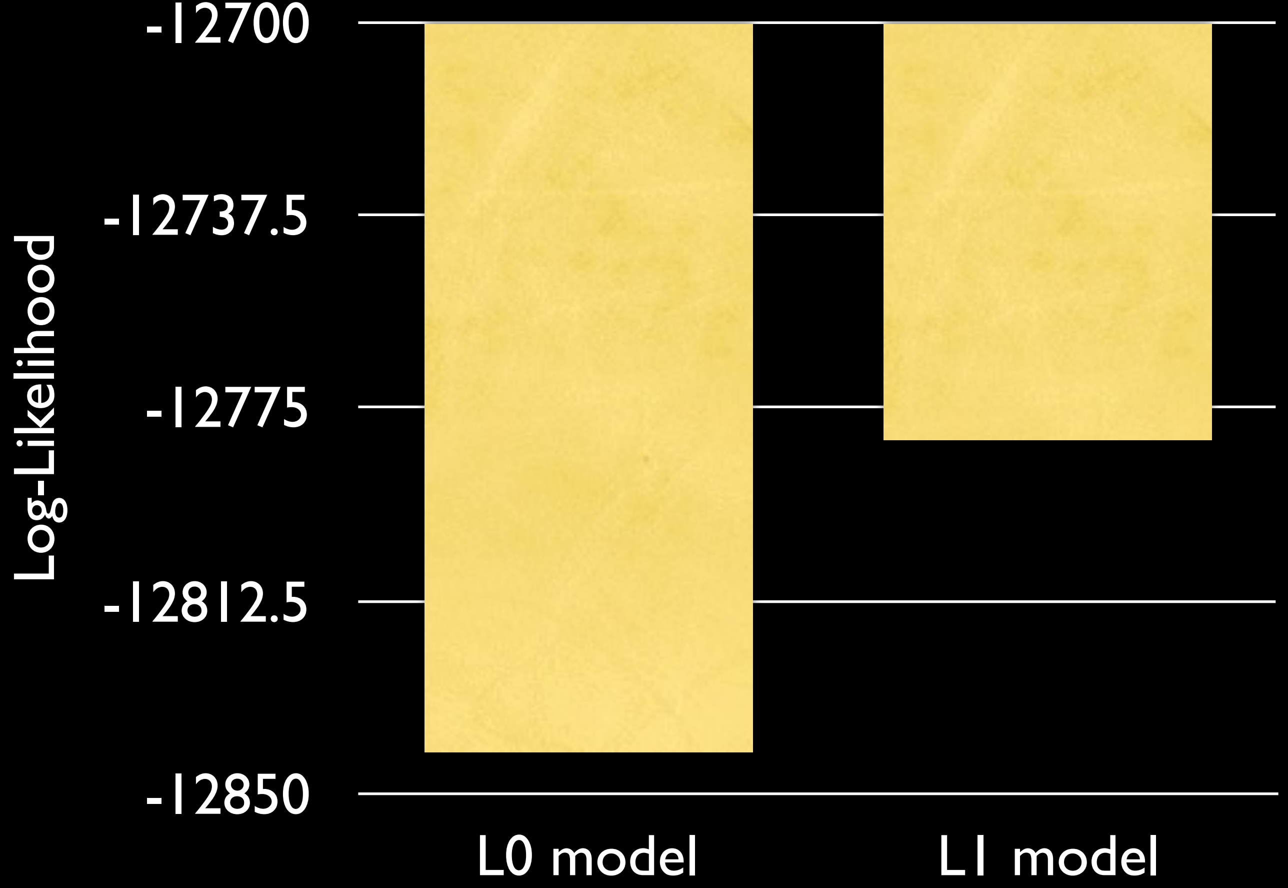
Interpretation data



Tessler and Goodman (in prep)

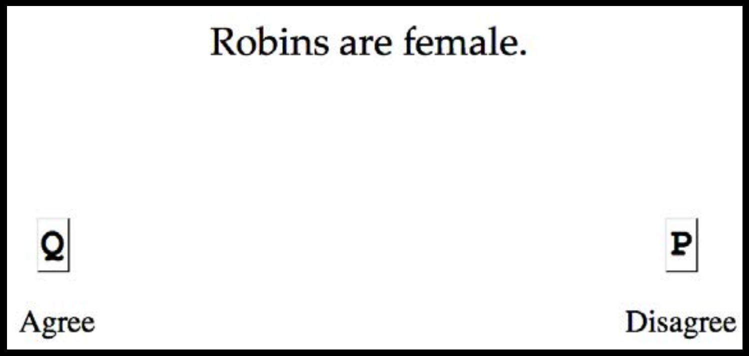
Model comparison

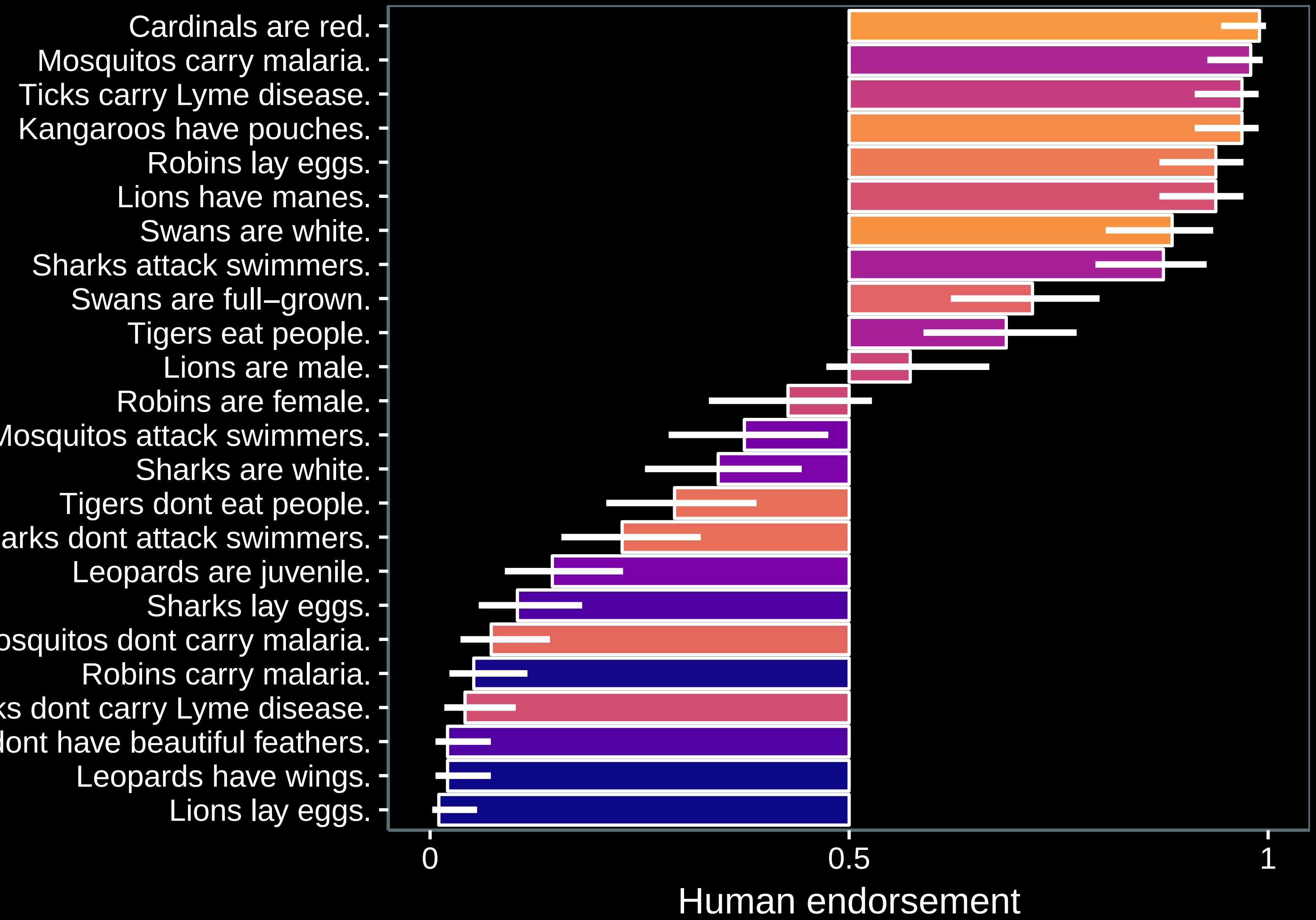


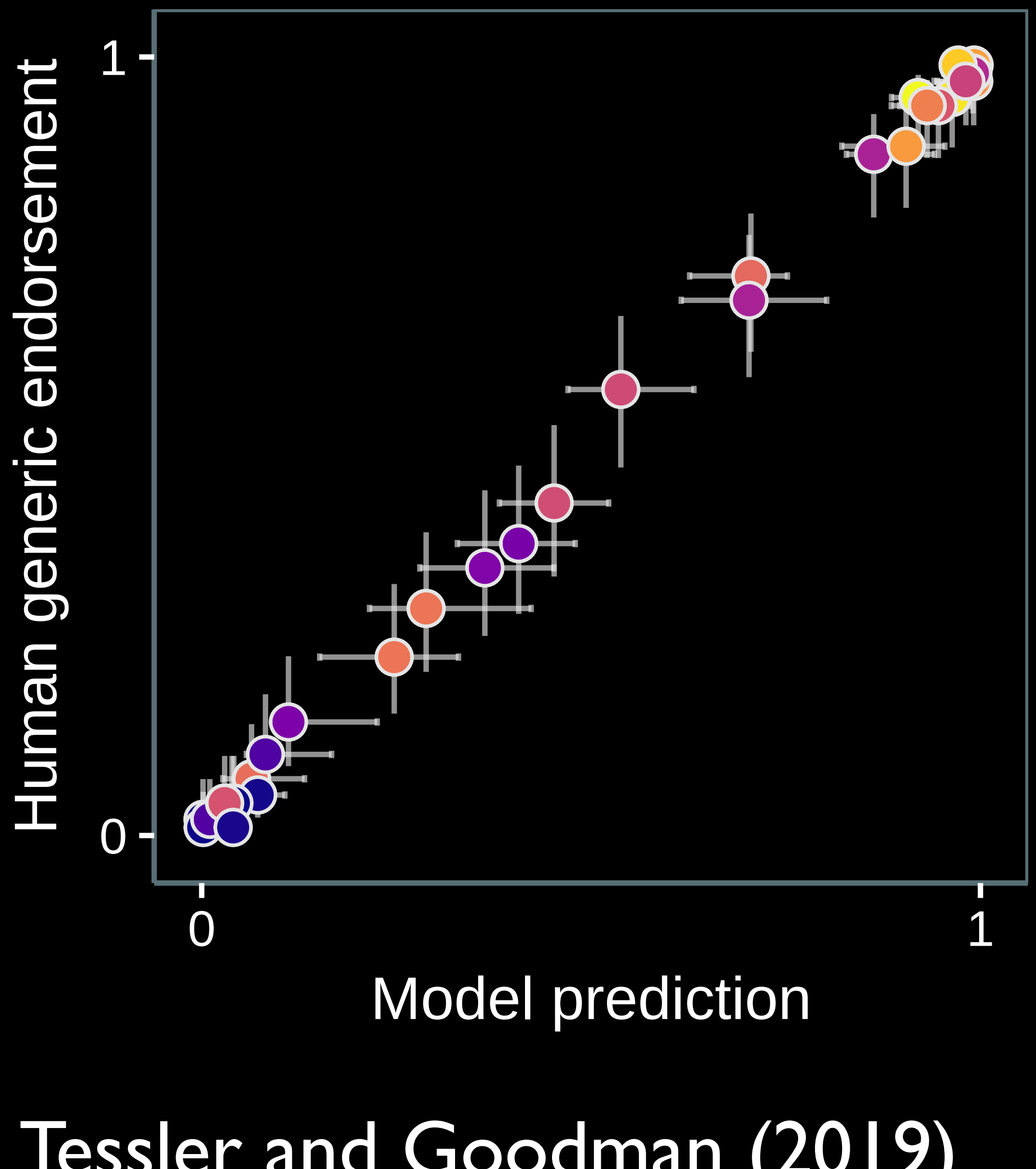


- The social model better accounts for the data: generics are *intended* minimal examples.

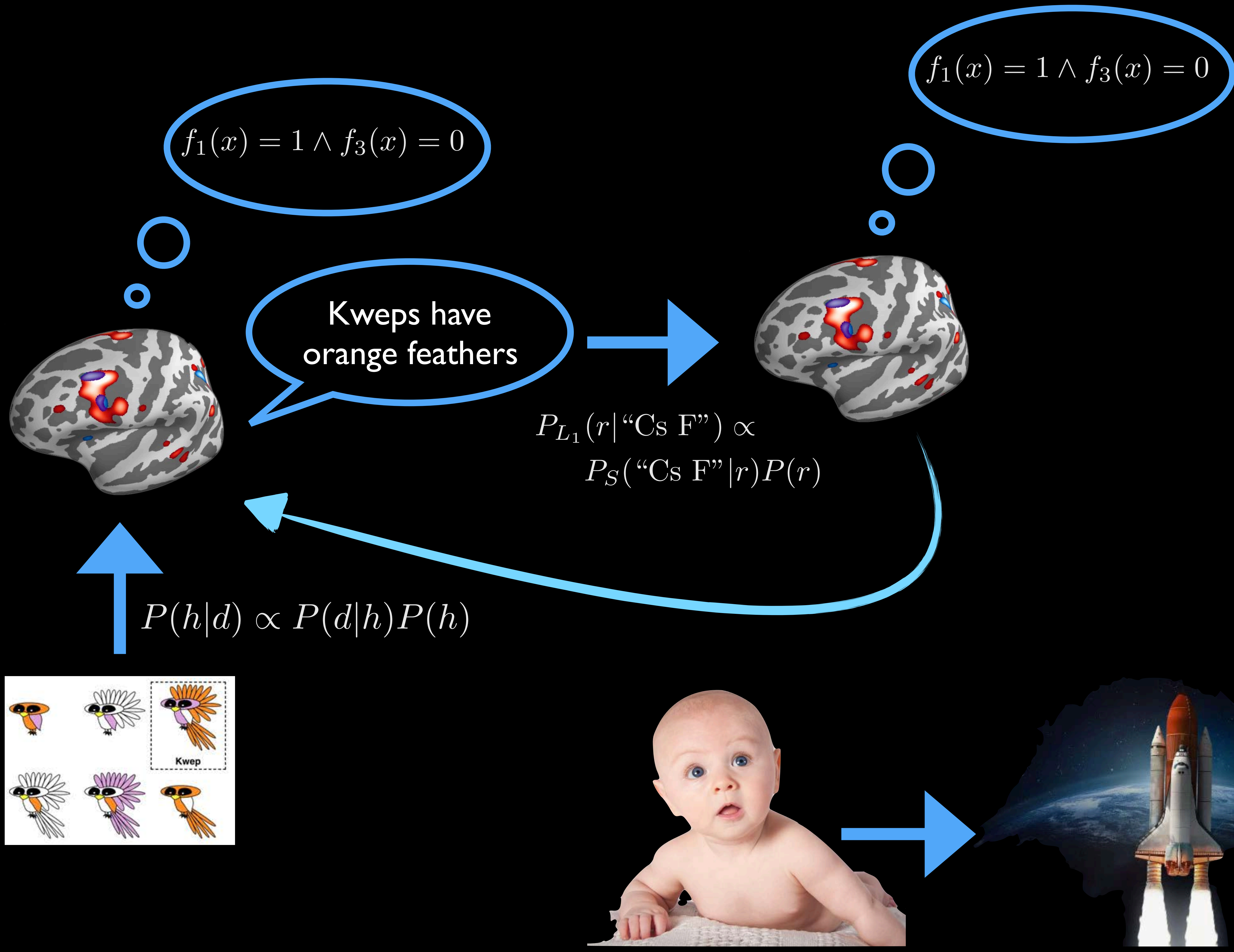
Endorsement data + model







Tessler and Goodman (2019)


$f_1(x) = 1 \wedge f_3(x) = 0$
$f_1(x) = 1 \wedge f_3(x) = 0$
Kweps have orange feathers
$P_{L_1}(r|\text{“Cs F”}) \propto P_S(\text{“Cs F”}|r)P(r)$
$P(h|d) \propto P(d|h)P(h)$
Kwep