





How we come to know

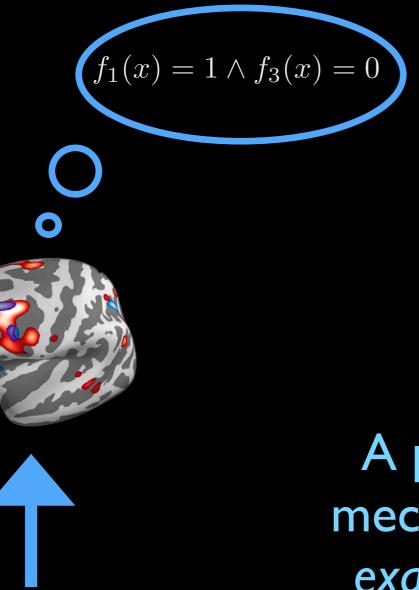
Noah D. Goodman Stanford University

College de France Feb. 11, 2019

The problem of induction



- 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?



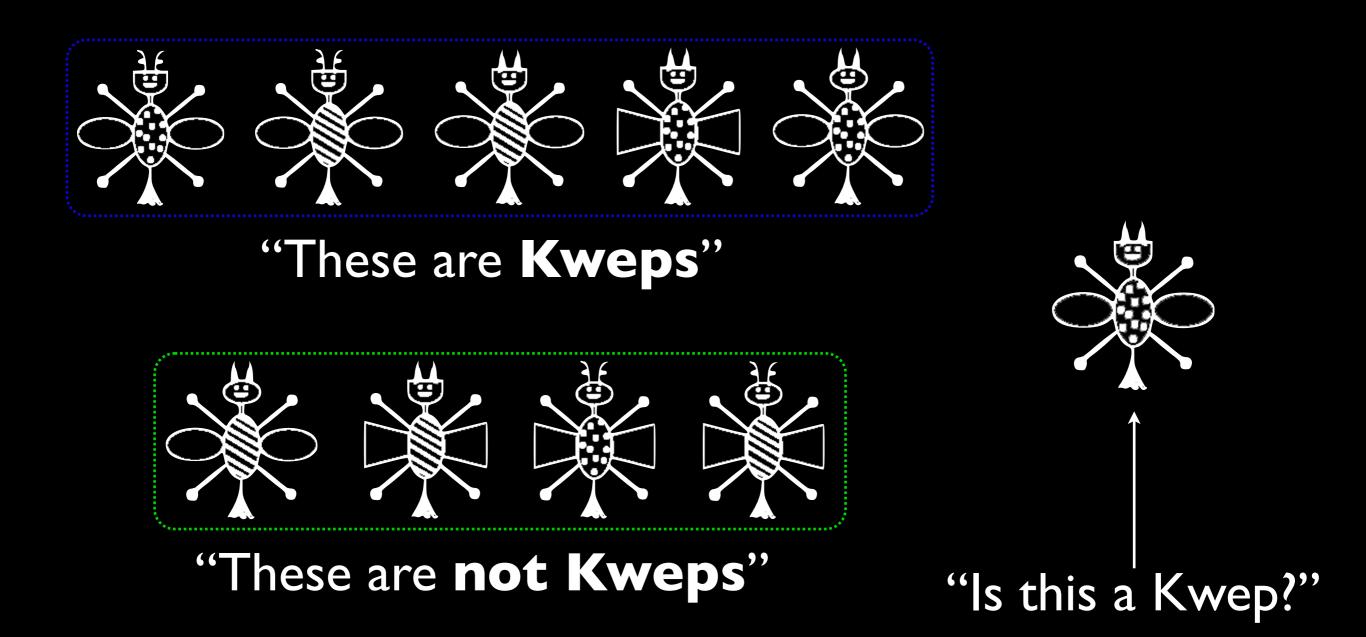
Kwep

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*

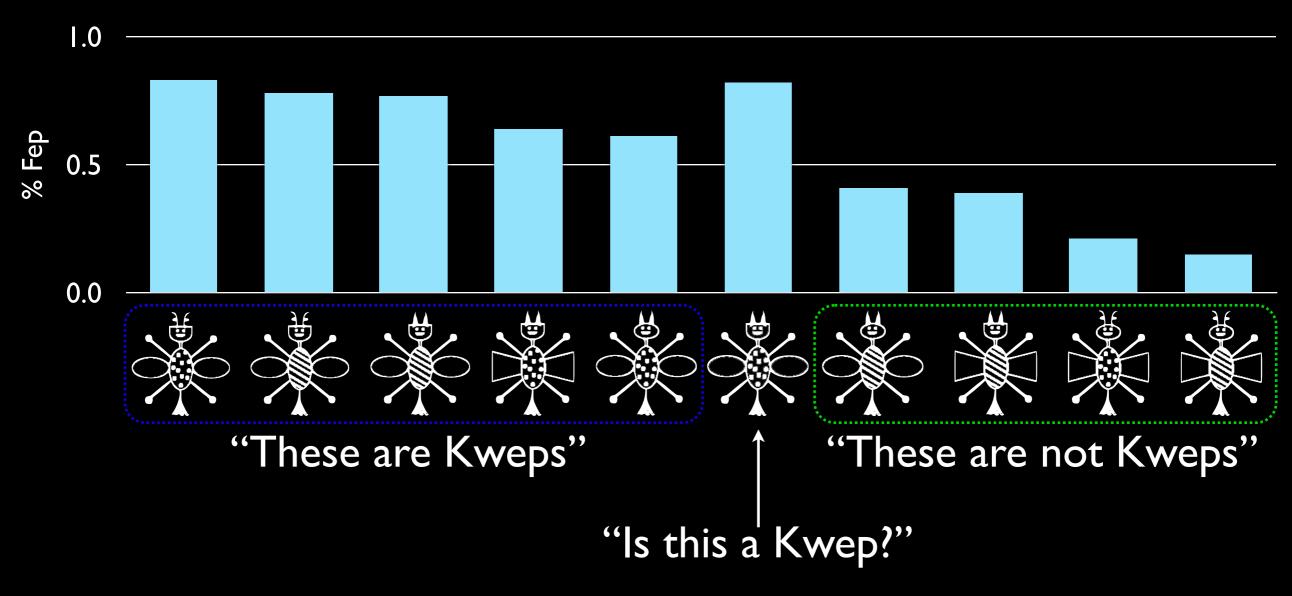
Concept learning

Medin & Schaffer (1978):

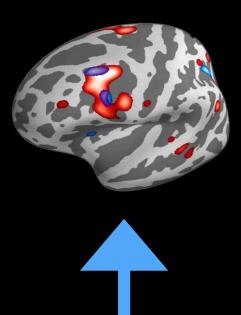


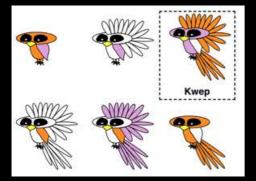
Concept learning

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



- Graded judgements
- Typicality
- Prototype enhancement





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

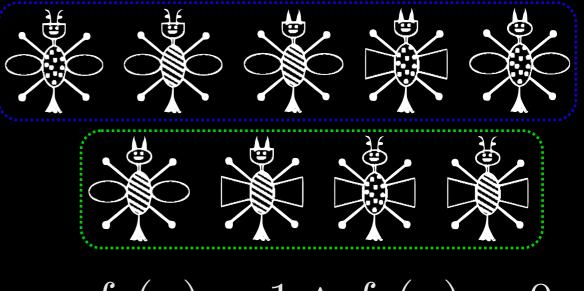
"It's a Kwep if it has flat head and round wings"

"Kwep if *rule*."

rule → feature

rule → rule and rule

rule → rule or rule



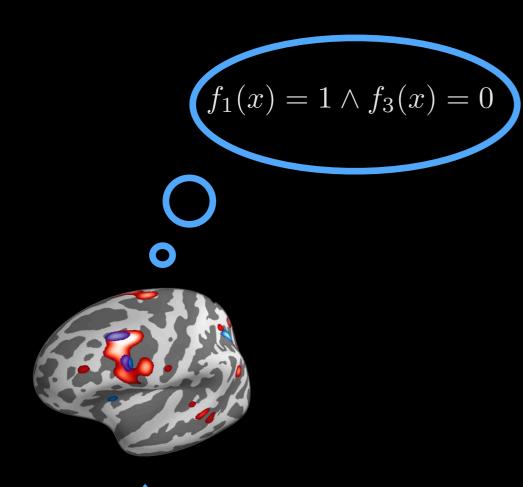
 $f_1(x) = 1 \wedge f_3(x) = 0$

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

Rule prior probability

"Kwep if *rule*." $rule \xrightarrow{50\%}$ feature $rule \xrightarrow{30\%}$ rule and rule $rule \xrightarrow{20\%}$ rule or rule

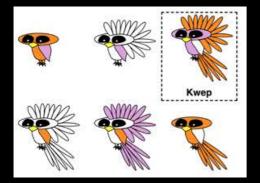
- Assign a probability to each rule-building step (a *probabilistic* grammar).
- The overall probability of a rule is the probability of all choices to make it.
 - Longer rules are less likely a priori.



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

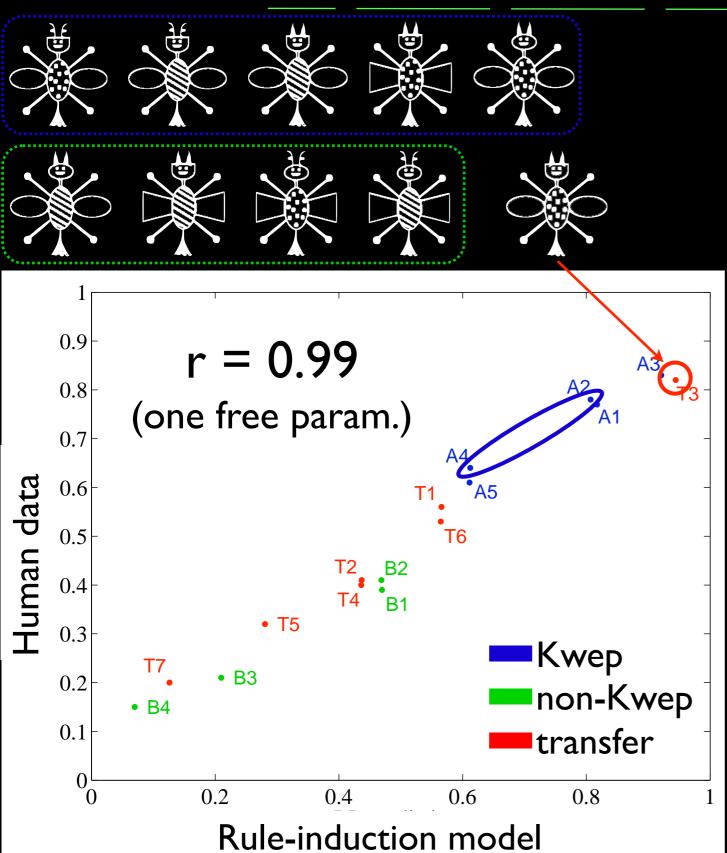
"Kwep if *rule*."
rule → feature
rule → rule and rule
rule → rule or rule

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



Simple noise likelihood: the rule is right with a high probability. $P(\text{Kwep}|\text{rule}(x)) = \epsilon$

Example: concept learning

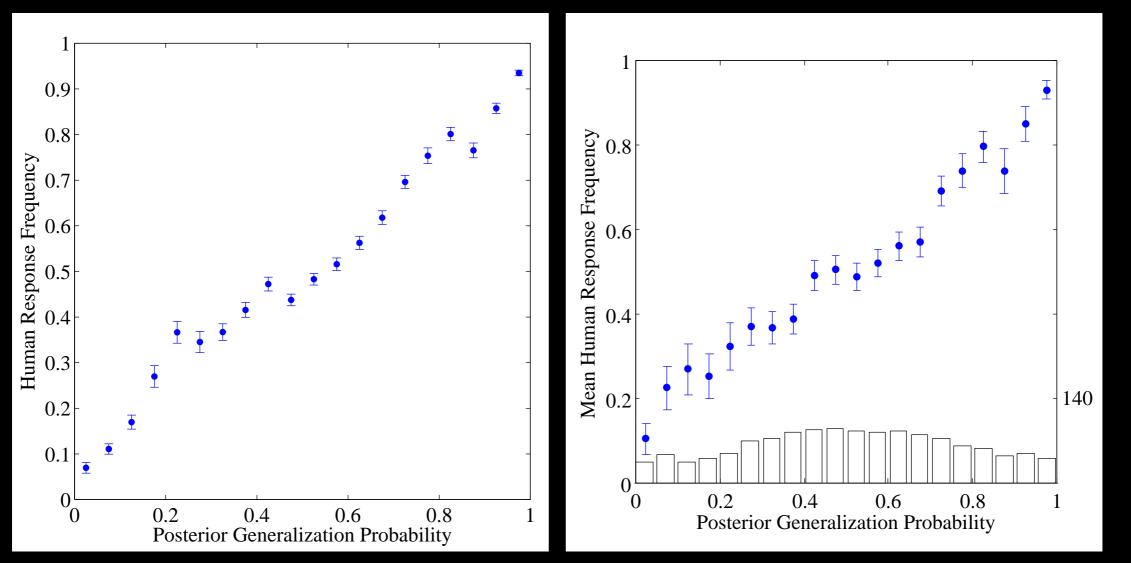


- Graded judgments
- Typicality
- Prototype enhancement

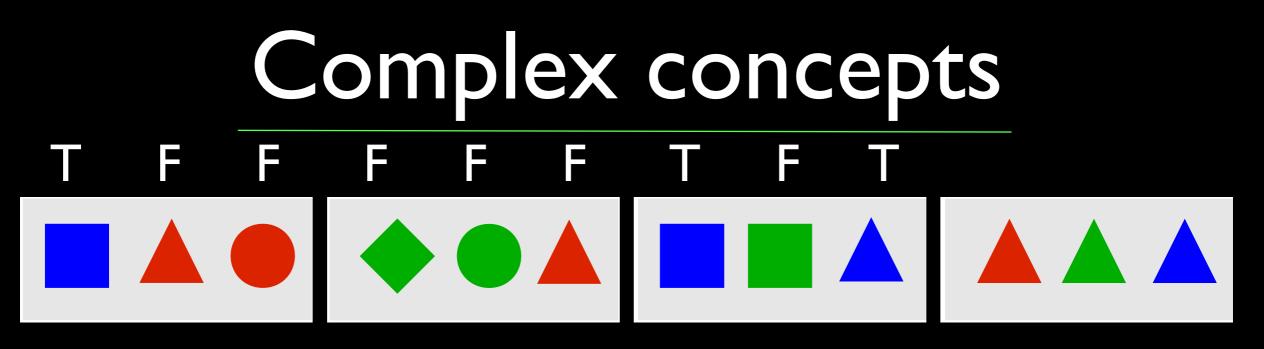
Goodman, et al. (2007, 2008a, 2008b)

Broader test

- 7 Boolean features.
- 43 randomly generated concepts (3-6 pos. + 2 neg. exs)
- I28 judgements (~I22 transfer questions)



Goodman, et al (2008)

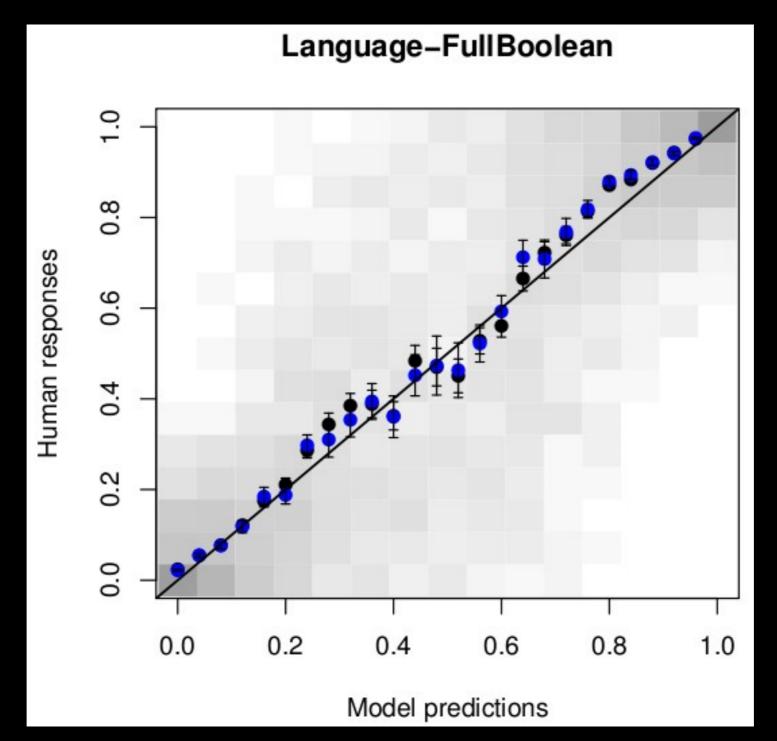


- Big online experiment.
 - I08 concepts,
 - Boolean (circle or red)
 - Context-dependent ("Determiners") (unique largest, exists another with same shape)
 - 2 orders per concept,
 - 1596 participants.

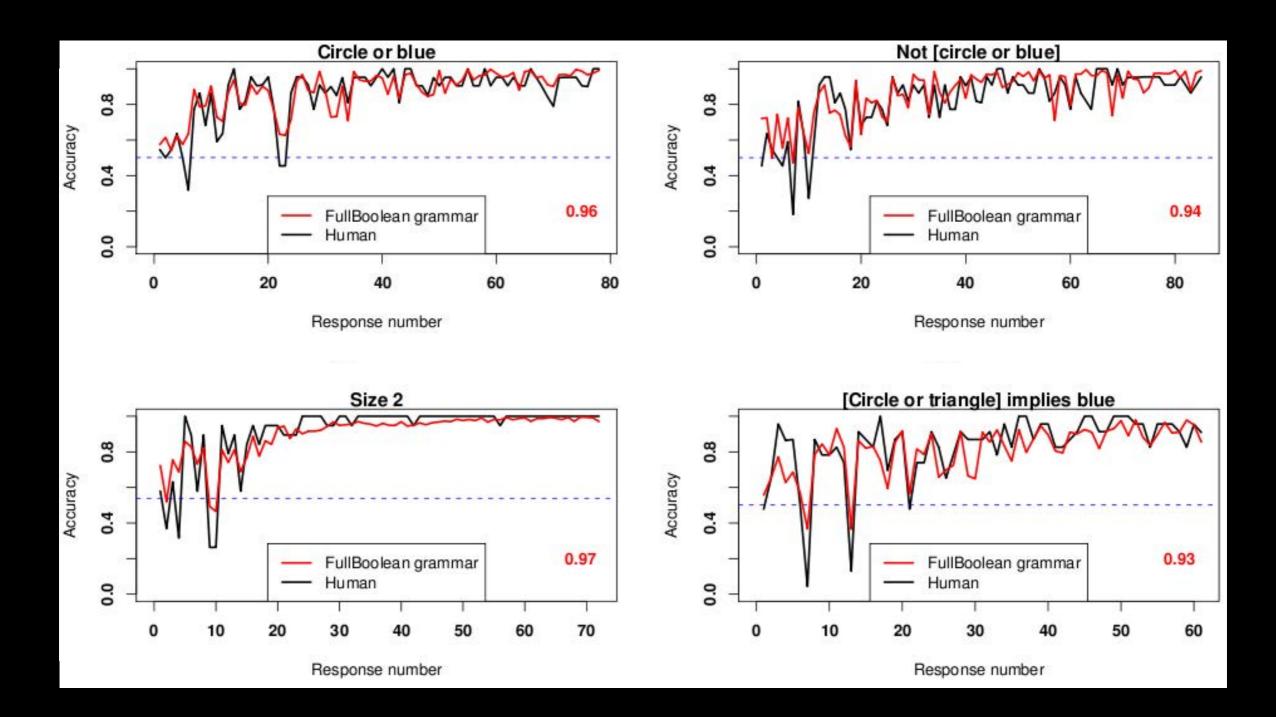
Piantadosi, Goodman, Tenenbaum (2016)

Boolean concepts

Learning Boolean concepts, model performance on Boolean concepts:

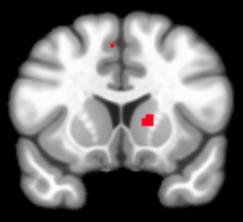


Boolean concepts



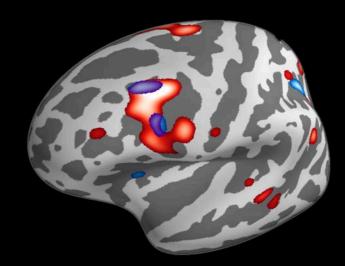
It's in the brain

RR model surprisal correlates with striatum activity.





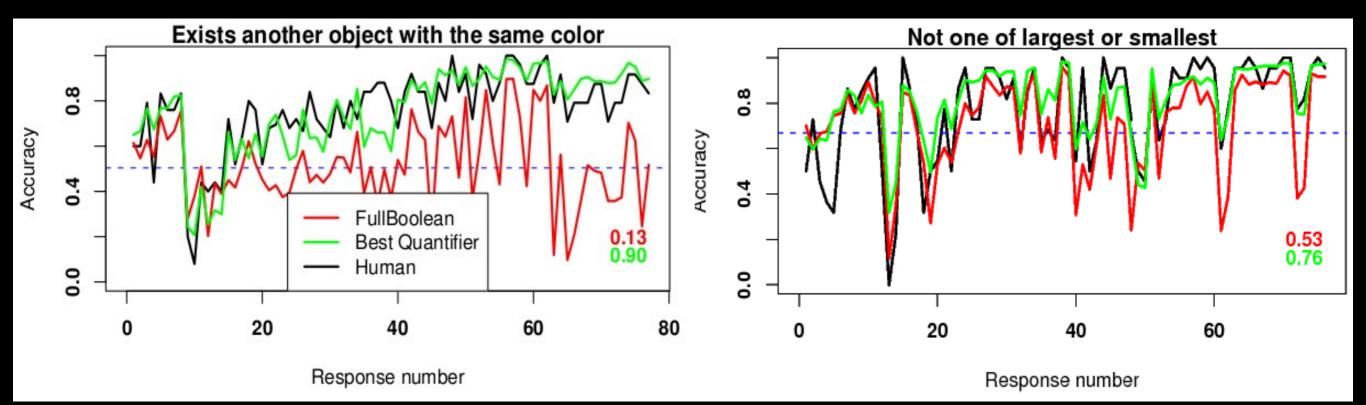
RR model posterior update correlates with DLPFC.



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

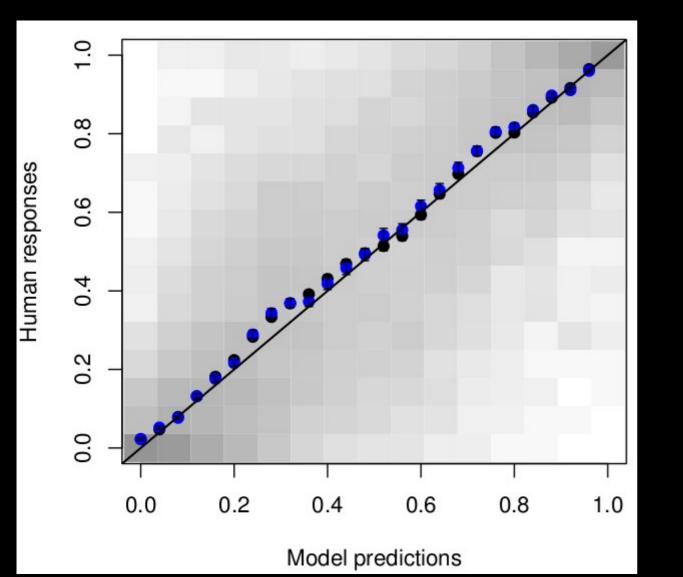
Non-Boolean concepts

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



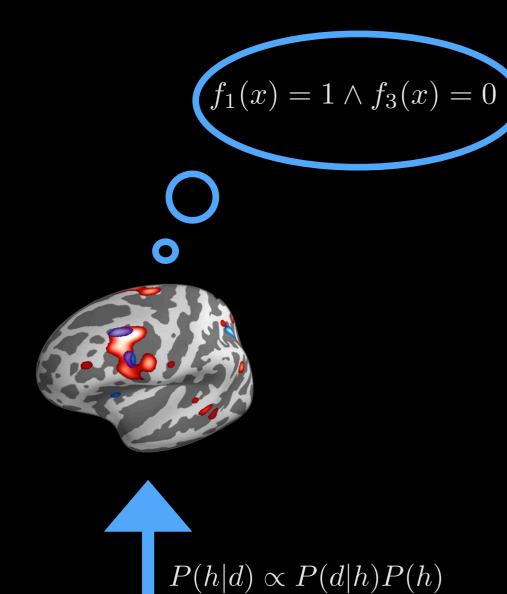
Non-Boolean concepts

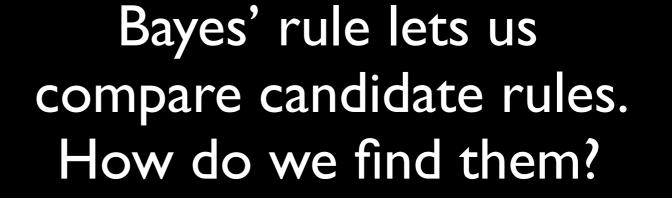
Best hypothesis space is full boolean logic plus quantifiers.

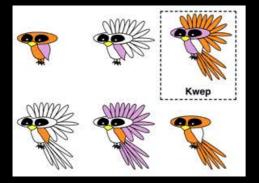


FOL	One-Or-Fewer	Small-Cardinalities	2nd-OrdQuan.	H.O. LL
\checkmark				-79279.95
\checkmark	\checkmark			-79560.90
	\checkmark			-79642.46
	\checkmark	\checkmark		-79972.75
\checkmark	\checkmark		\checkmark	-80198.75
\checkmark			\checkmark	-80267.46
\checkmark	•	\checkmark	•	-80285.38
•	\checkmark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

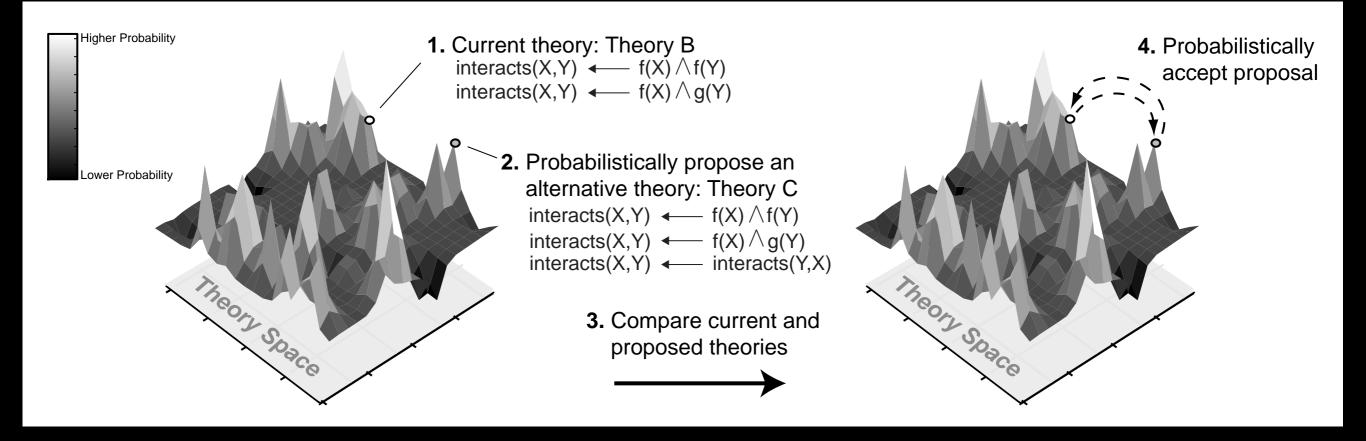
 $\tilde{\mathbf{x}}$







Finding rules?

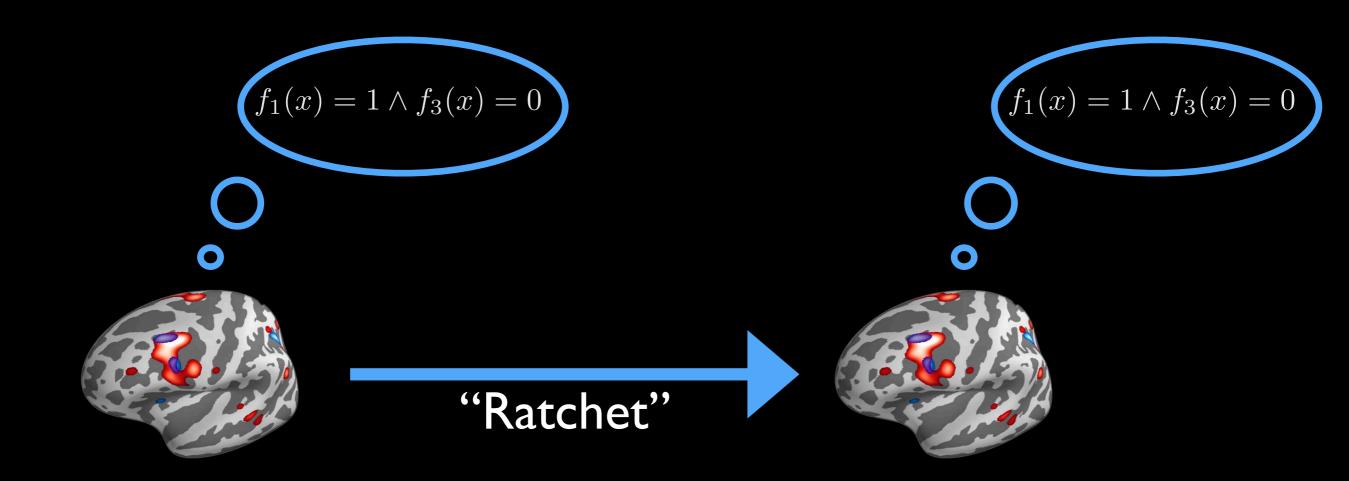


Random search works for simple concepts...

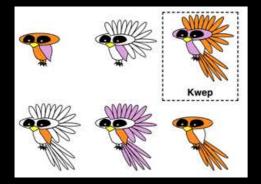
Ullman, Goodman, Tenenbaum (2012)

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).



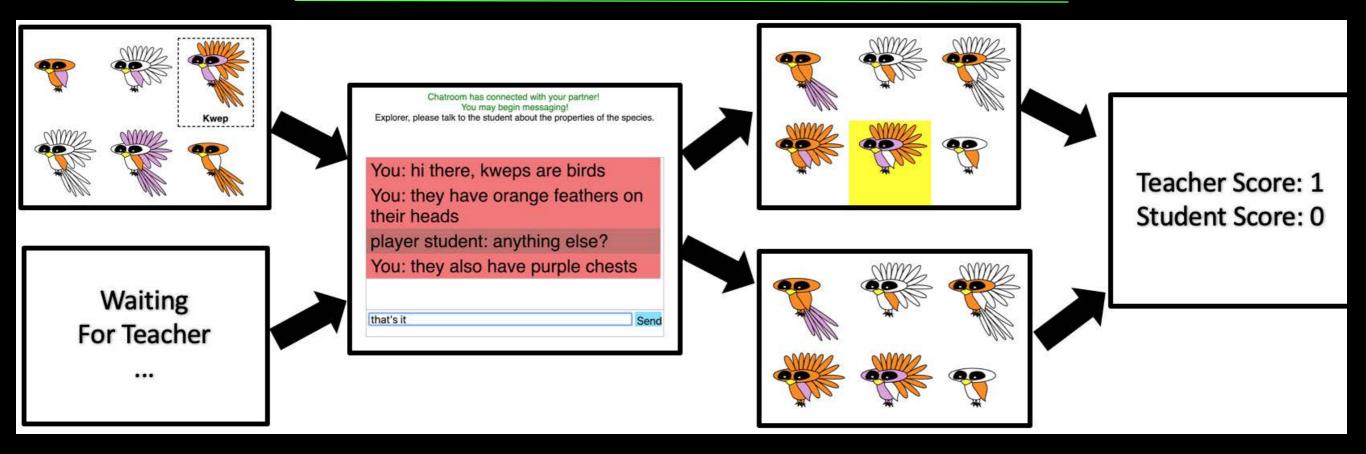
 $P(h|d) \propto P(d|h)P(h)$



This requires **faithful transmission** of

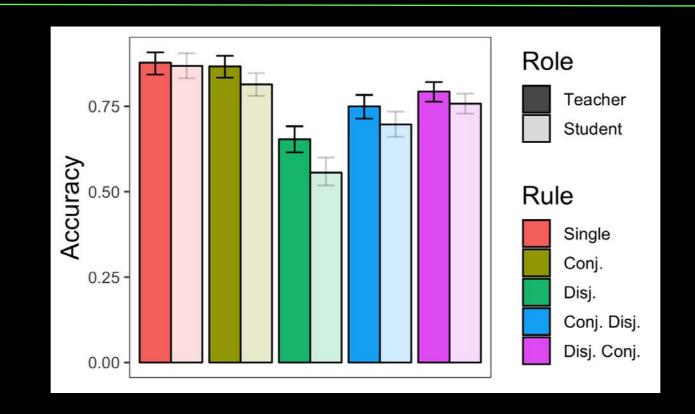
knowledge, and it has to be easier than directly learning from examples.

Learning from language

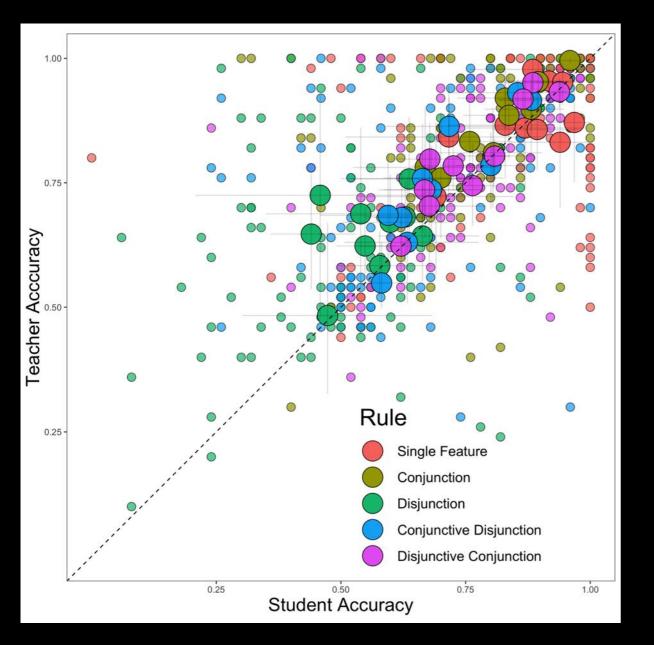


 A minimal paradigm to compare concept learning from observing examples and from linguistic communication.

Chopra, Tessler, Goodman (subm)

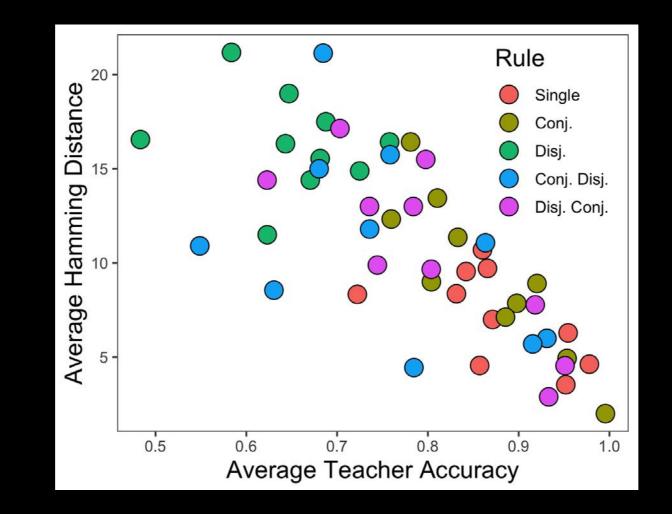


- Language is sufficient:
 - Students who learn from language perform only slightly worse than their teacher.
 - (Approx. 5% lower accuracy for students, by Bayesian mixed effects model.)

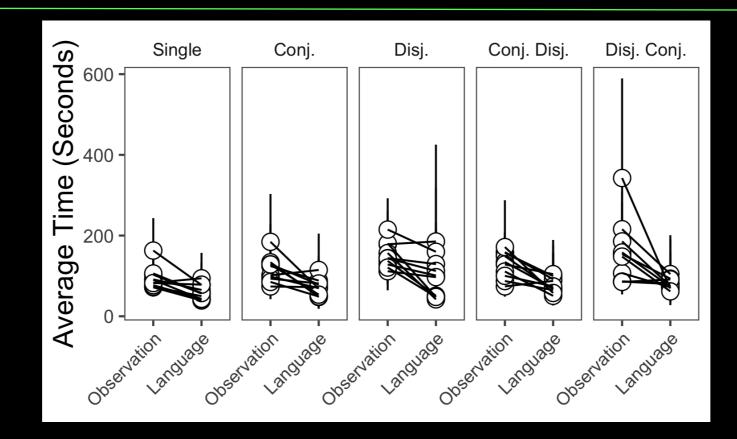


 Teacher accuracy predicts student accuracy.

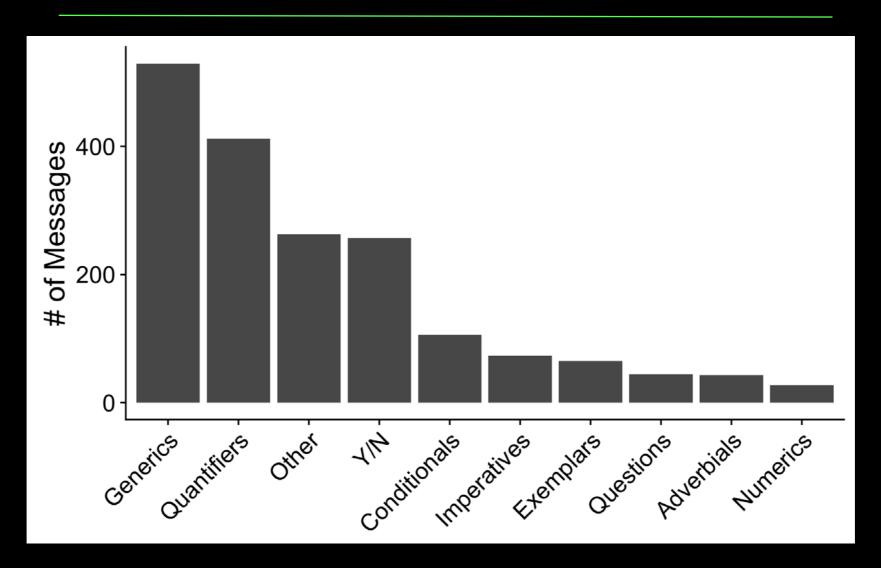
 Individual students make the same mistakes as their teachers (hamming distance lower than permutation baseline).



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



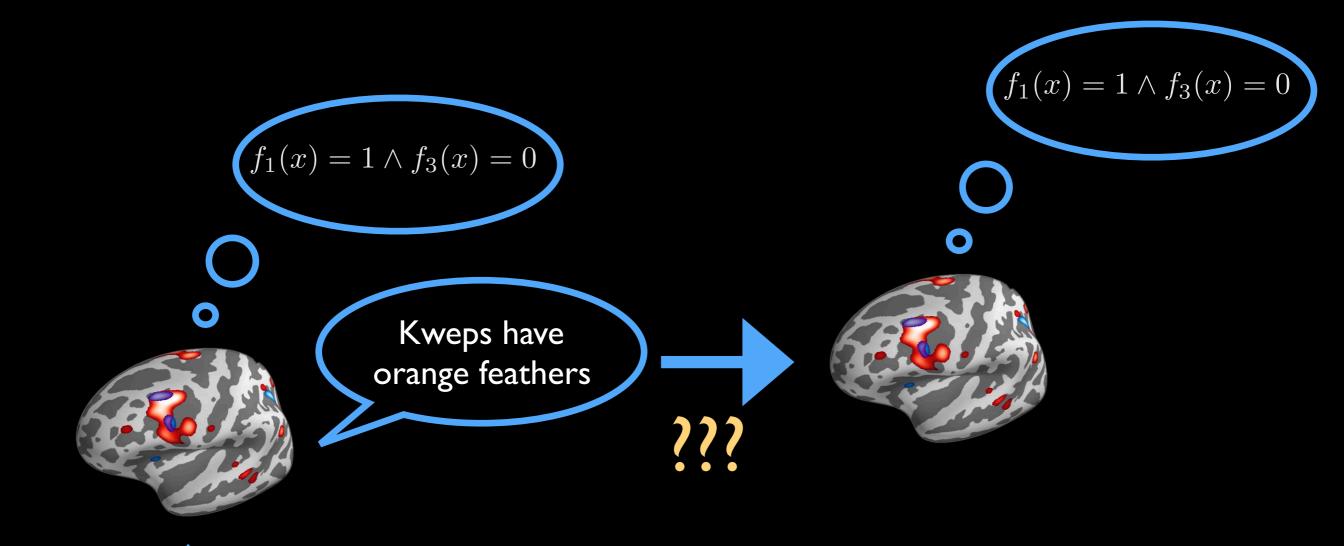
- Language is efficient.
 - Participants spent longer learning from examples than from language. (Both were freely determined by participants.)



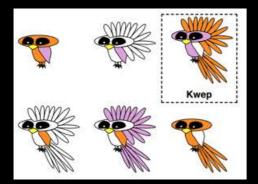
- It's the language of generalization that matters.
 - Most messages use generics or quantifiers.

Hypothesis

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



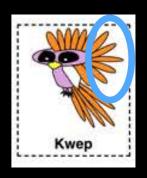
 $P(h|d) \propto P(d|h)P(h)$

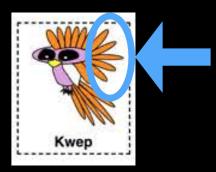


Generics

"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:

• 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

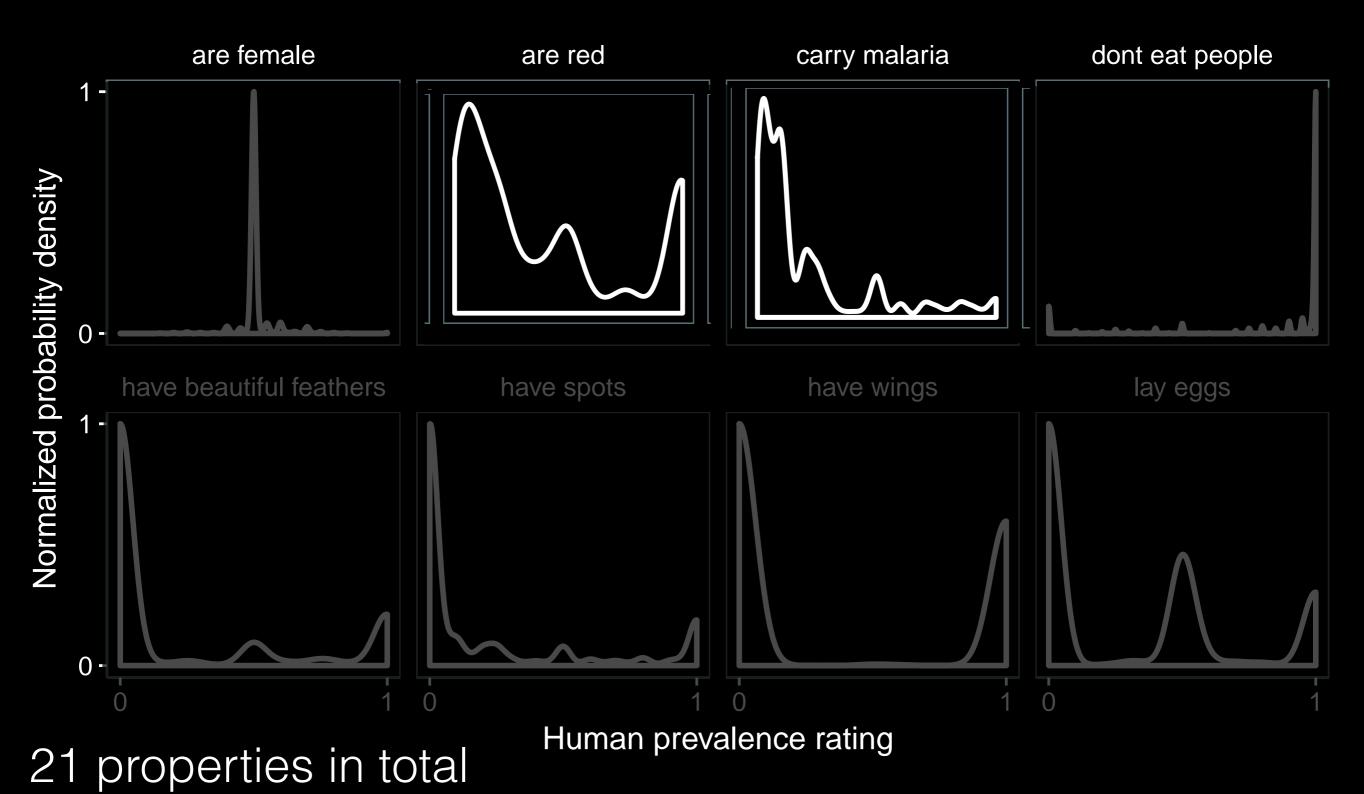
Category elicitation

Prevalence elicitation

For each kind of animal, what percentage of the species do you think Kangaroos Robins carry malaria Sharks supplied to % 0 Kangaroos participants Mosquitos % 0 Robins % Sharks 0 Ducks % Mosquitos 10 Ticks % 0 Ducks % 3 Ticks % 1 dogs participants % 1 cats generate % 0 animal kinds geese % monkeys 1 % 0 falcons Continue

n = 60 from Amazon's Mechanical Turk

filtering 0% responses



Interpretation data

have four legs eat insects experience emotions o 000 eat grass · hunt other animals bao-89 000088 10-98**06** live in trees 0 0 0 have spots . 08000 have personalities have brown fur A0.00 000000 mourn their dead · 60000 0 0 0 00 0.80 008-8-8 lose their teeth 0 . . . 000 Bg00.00 -080 080 8 0 08000 0000 0000 00 00 o 0 Ro R 80 8 8 00 00 00.0 000 00 0000 are intelligent 00 00 ില്ല ାଲ୍ଲ 08-00 are afraid of dogs 0 0 000 08 00805 0 000 ○8 µ 0 0 0 chase their tails like to cuddle · 8008 000 0 0800 080 go bald 0 0 8 ~0 0B~B 0 0 - 0000000000 00 0.0 00. Bo80 08 eat human food 000 o 0 use tools ode 080-800 get erections -0.0008 live in urban areas RO. 0 0 00 000 0 086 5 get dandruff eat garbage eat cannabis 00 800 0 0 0 8 00 0 60 00 000 a0 0 8 08 0 0008 000 00 00 00 00 00 0.0 8000 0 0 000 8 8 00 80 0 8 2000 0 0000 000 0 0.0.08 0 00000 080 000 0 0 0.0 08 8 0 000 0 8 -0 B steal farmers crops -0.0 0 0 00 00 develop phobias o 8⁰0 0 00 0 800800 8 0 0 carry Lyme disease 0 0 00 00 00 00 00 00 00 00 0 0 0000 0 0000 8 0 00 000 ്റ്റ 880 000 0 000 0 00 0 0 80 transmit rabies 8 0 0 0 8 0 000 0.0 ñ 00 08.80 80 808 carry malaria 0 000,00 0 0 00 0 00 0 0 8 0 000 transmit HIV 00.-8 00 00 0 00 0 000 0 have seizures · 000 00 08 0 0 0 0 00 0 00, 0 eat people -8000 0 0 0 0 0 0 00 00 co 00 00 play with bottlecaps 00 8 0⁰808 00 live in zoos 88 8o 000 800 0 000 0,0000 0 eat candy wrappers 680 00000000 attack hikers 000 8000 8 00 08000 0 00 get cancer 000,80 80 08 008 0 800 00 o o^ooo o 80 drink soda 000<u>90</u>00000 00 ride the subway oo oogo @**⊙**oo<mark>⊙</mark>o∂o∂ o 0008 80808- 0 00 0.0 0.5 1.0

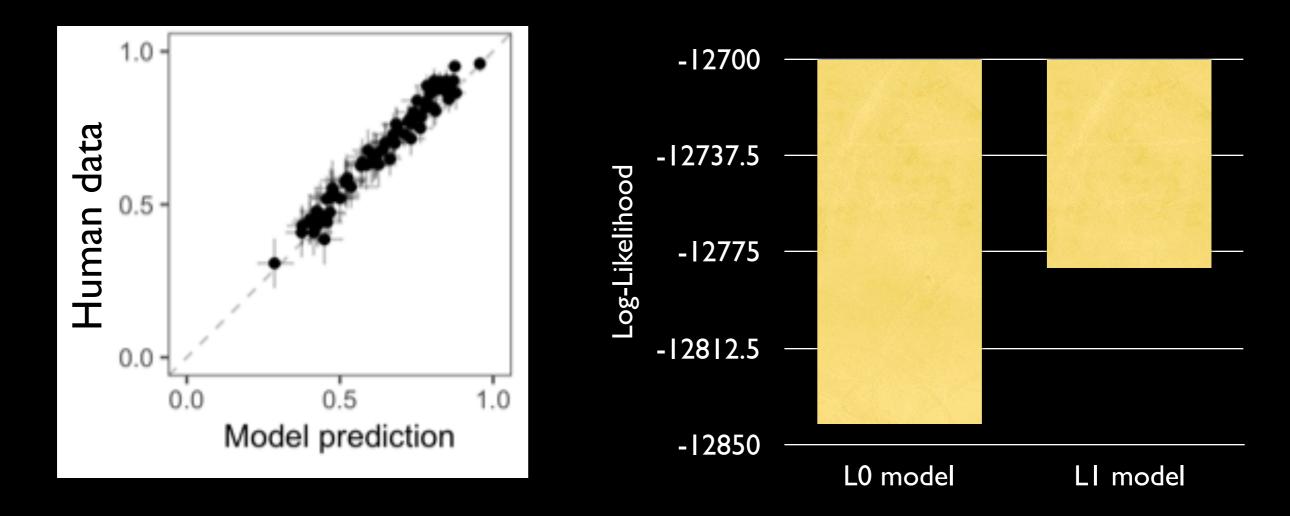
Implied prevalence rating

sleep during the day have an exquisite sense of smell have very long wings attract mates by secreting pheromones give birth underwater have intensely beautiful feathers feed on the carcasses of dead animals experience empathy lay eggs without needing fertilization are afraid of loud noises sing beautiful songs have dozens of sexual partners swim in shallow pools pound their chests to display dominance know when earthquakes are about to happen. live to be twenty years old do handstands to scare off predators lay eggs in other birds nests capture other animals territory get in fights with other animals live to be a hundred years old live in the hulls of sea vessels. know how to open doors have a menstrual cycle torture other animals live to be five hundred years old cannibalize each other develop back problems have strange genetic mutations live in high-rise buildings fish in the Hudson River know how to ride bicycles get addicted to nicotine fly into building windows drink alcohol left behind by tourists carry out premeditated murder perform in the circus -

Α

Tessler and Goodman (in prep)

Model comparison



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

Endorsement data + model

