Deceived by data or savvy about statistics? The impact of sampling on inductive reasoning

Dan Navarro
School of Psychology
University of New South Wales
compcogscisydney.com

I would like to acknowledge this land that we meet on today as the traditional lands for the Kaurna people, and respect their spiritual relationship with their country.

I also acknowledge the Kaurna people as the custodians of the greater Adelaide region and that their cultural and heritage beliefs are still as important to the living Kaurna people today

http://www.statedevelopment.sa.gov.au/upload/aard/welcome/acknowledgement-statements.pdf

Contributors, coauthors, collaborators

Amy Perfors
Keith Ransom
Wouter Voorspoels
Brett Hayes
Steph Banner

Drew Hendrickson
Michelle Keshwa
Sean Tauber
Matthew Welsh
Matt Dry
Michael Lee
Titia Benders
Chris Donkin

Kristy Martire
Ben Newell
Wai Keen Vong
Lauren Kennedy
Steve Langsford
Candy Liu
Anastasia Ejova
Ally Tingey
Rachel Stephens
Gert Storms
Pat Shafto
Baxter Eaves
Charles Kemp
Nancy Briggs

Funding

Australian Government
Australian Research Council

UNSW SYDNEY

THE UNIVERSITY of ADELAIDE

How do people acquire new knowledge?

(categorisation & reasoning)

How do we make choices in an uncertain world?

(judgment & decision making)

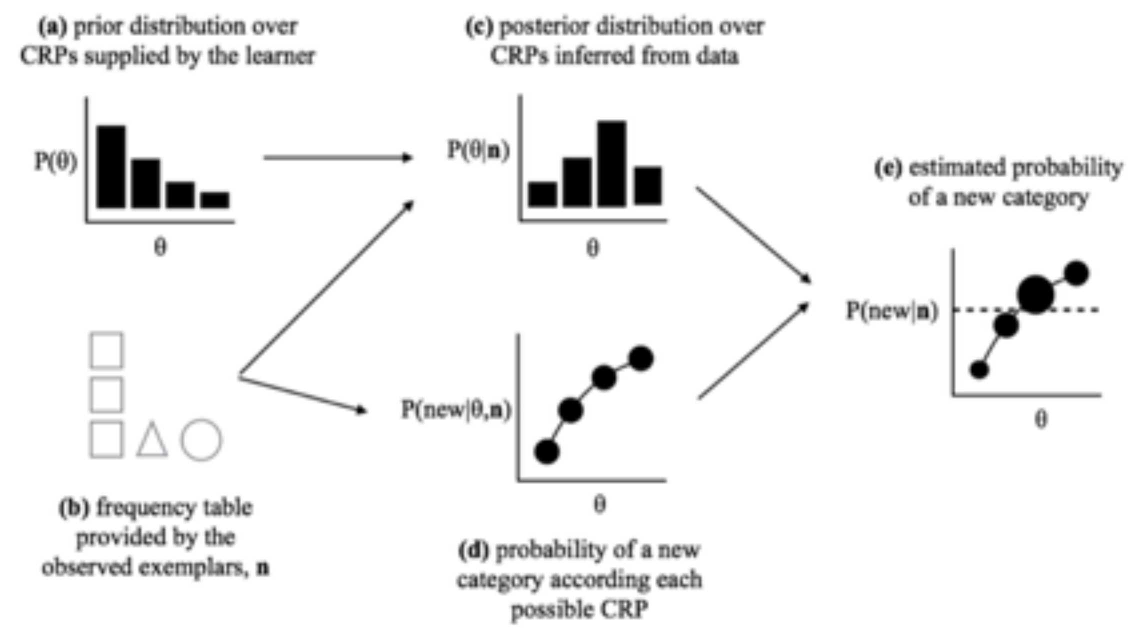
How should psychologists analyse our data?

(math psych & statistics)

How do people acquire new knowledge?

(categorisation & reasoning)

What kind of prior biases shape the acquisition of *new* knowledge?

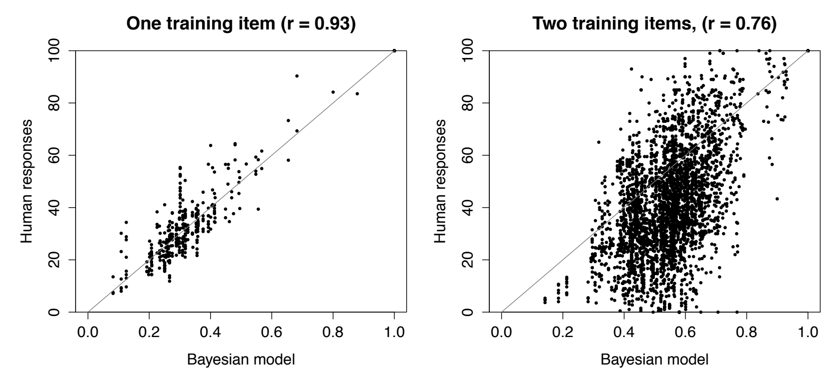


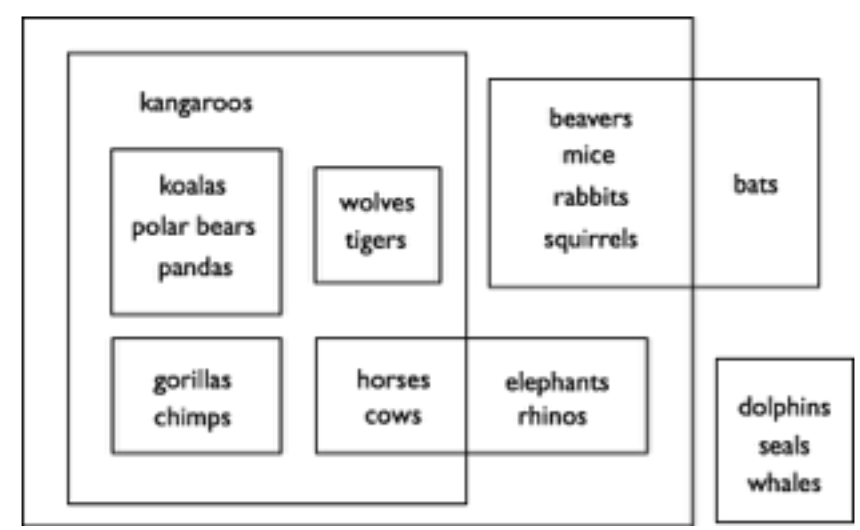
Navarro & Kemp (under revision). *Psych. Review*

How do people acquire new knowledge?

(categorisation & reasoning)

What *old* knowledge do people use to guide inferences?



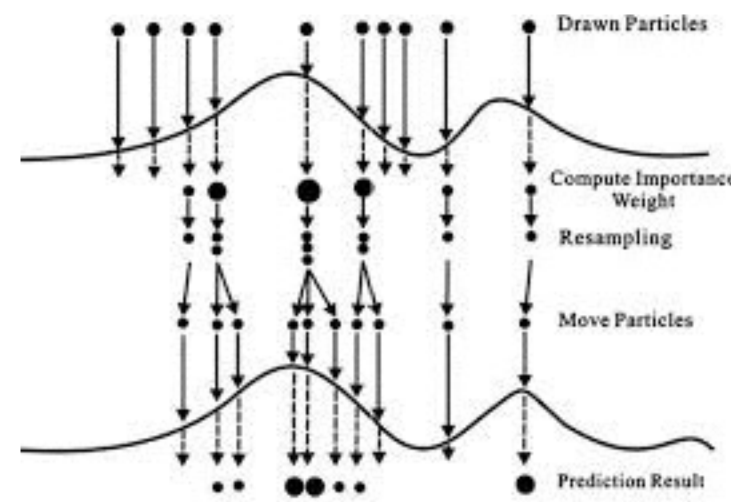


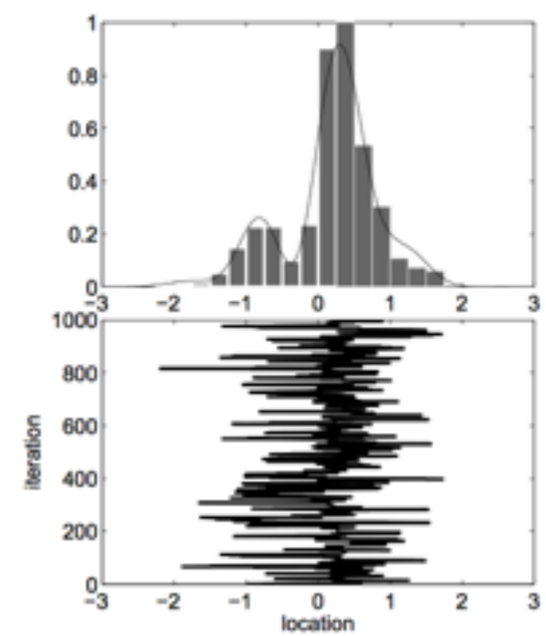
Tauber, Navarro, Perfors & Steyvers (in press). *Psych. Review*

How do people acquire new knowledge?

(categorisation & reasoning)

What computational strategies do people use to simplify complex problems?





Sanborn, Griffiths & Navarro (2010). *Psych. Review*

In the case where $n = 1$ we observe that,

$$\begin{aligned}
\int_{\mathcal{R}} P(x_1, x_1 \in r_t \,|\, r_t = r)\; dr &= \int_0^{z_l} \int_{z_u}^1 P(x_1 \,|\, r_t = [l, u])\; du\; dl \\
&= \int_0^{z_l} \int_{z_u}^1 (u - l)^{-1}\; du\; dl \\
&= \int_0^{z_l} [\ln(u - l)]_{z_u}^1\; dl \\
&= \int_0^{z_l} \ln(1 - l) - \ln(z_u - l)\; dl \\
&= [(l - 1)\ln(1 - l) - l]_0^{z_l} - [(l - z_u)\ln(z_u - l) - l]_0^{z_l} \\
&= ((z_l - 1)\ln(1 - z_l) - z_l) - ((z_l - z_u)\ln(z_u - z_l) - z_l + z_u \ln z_u) \\
&= (z_u - z_l)\ln(z_u - z_l) - (1 - z_l)\ln(1 - z_l) - z_u \ln z_u \qquad (24)
\end{aligned}$$

Applying the same procedure as before yields the expression

$$P(y \in r_t \,|\, x_1, x_1 \in r_t) = \begin{cases} \dfrac{(z_u - y)\ln(z_u - y) - (1 - y)\ln(1 - y) - z_u \ln z_u}{(z_u - z_l)\ln(z_u - z_l) - (1 - z_l)\ln(1 - z_l) - z_u \ln z_u} & \text{if} \quad y < z_l \\ 1 & \text{if} \quad z_l \le y \le z_u \\ \dfrac{(y - z_l)\ln(y - z_l) - (1 - z_l)\ln(1 - z_l) - y \ln y}{(z_u - z_l)\ln(z_u - z_l) - (1 - z_l)\ln(1 - z_l) - z_u \ln z_u} & \text{if} \quad z_u < y \end{cases} \qquad (25)$$

In this case, however, the expression can be further simplified since $z_l = z_u = x_1$:

$$P(y \in r_t \,|\, x_1, x_1 \in r_t) = \begin{cases} \dfrac{(1 - y)\ln(1 - y) + x_1 \ln x_1 - (x - y)\ln(x_1 - y)}{(1 - x_1)\ln(1 - x_1) + x_1 \ln x_1} & \text{if} \quad y < x_1 \\ 1 & \text{if} \quad y = x_1 \\ \dfrac{(1 - x_1)\ln(1 - x_1) + y \ln y - (y - x_1)\ln(y - x_1)}{(1 - x_1)\ln(1 - x_1) + x_1 \ln x_1} & \text{if} \quad x_1 < y \end{cases} \qquad (26)$$

(Obviously, this expression could be derived directly, rather than found as a special case

In the case where $n = 1$ we observe that,

$$\int_{\mathcal{R}} P(x_1, x_1 \in r_t \,|\, r_t = r)\; dr \;=\; \int_0^{z_l} \int_{z_u}^1 P(x_1 \,|\, r_t = [l, u])\; du\; dl$$

$$= \int_0^{z_l} \int_{z_u}^1 (u - l)^{-1}\; du\; dl$$

$$\int_0^{z_l}$$

$$- z_l + z_u \ln z_u)$$

$$z_u \qquad (24)$$

Applying the same proced

$$P(y \in r_t \,|\, x_1, x_1 \in r_t) = \begin{cases} & y < z_l \\ & z_l \le y \le z_u \\ & z_u < y \end{cases} \qquad (25)$$

In this case, however, the

$$P(y \in r_t \,|\, x_1, x_1 \in r_t) = \begin{cases} \dfrac{(1-y)\ln(1-y) + x_1 \ln x_1 - (x-y)\ln(x_1 - y)}{(1-x_1)\ln(1-x_1) + x_1 \ln x_1} & \text{if} \quad y < x_1 \\ 1 & \text{if} \quad y = x_1 \\ \dfrac{(1-x_1)\ln(1-x_1) + y \ln y - (y - x_1)\ln(y - x_1)}{(1-x_1)\ln(1-x_1) + x_1 \ln x_1} & \text{if} \quad x_1 < y \end{cases} \qquad (26)$$

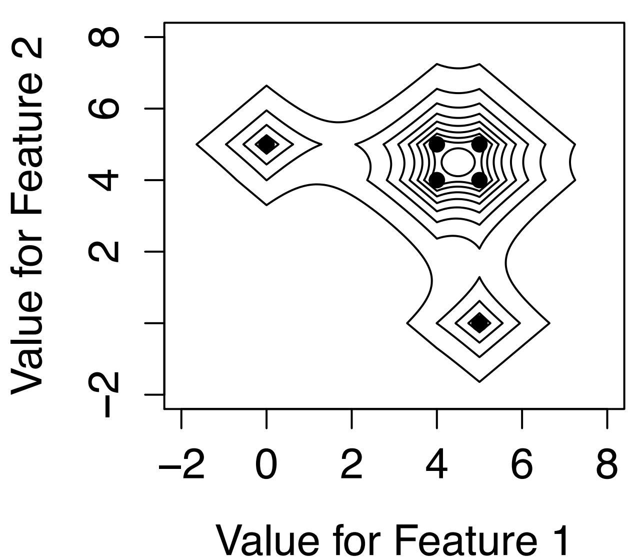
(Obviously, this expression could be derived directly, rather than found as a special case

Why take a computational approach to cognitive science?

Computational models make it easier to be precise about one's theories

??

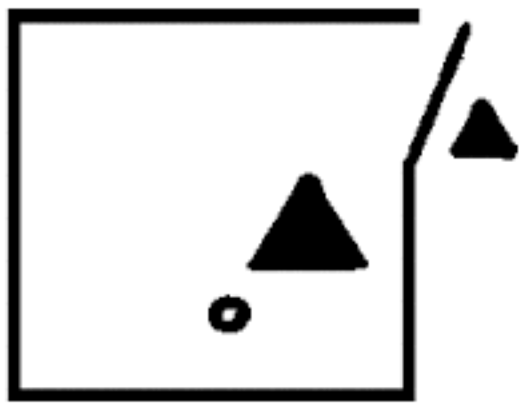
categorisation is sort of related to similarity I guess?



categorisation probability is proportional to the sum of similarities to previous exemplars

Formal descriptions of human inductive biases can improve machine learning

triangles are jerks



inferring intention from actions

"I'm not driving"

understanding the relevance of utterances to context

teapot death star?

constructing categories from instances

Machine agents need to interact with humans, so they need to understand us

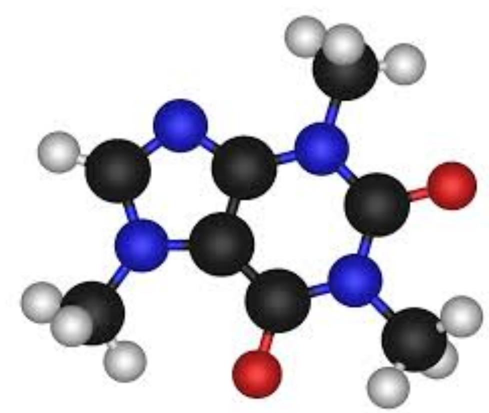
machines need maths to describe how the humans adjust speech patterns when the speech recognition system stuffs up

autonomous vehicles need to understand how human drivers respond to weirdness (e.g., in Sydney)

Conjecture:

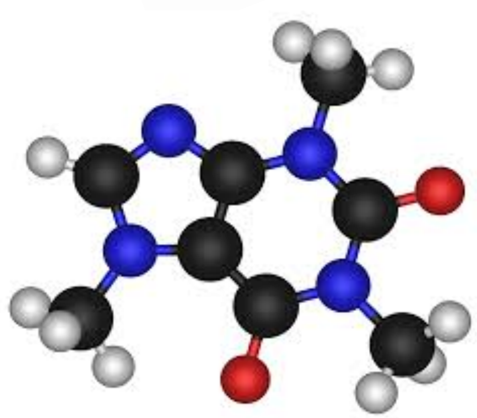
Reasoning is statistical inference

What should we do with
this *sample* of evidence?



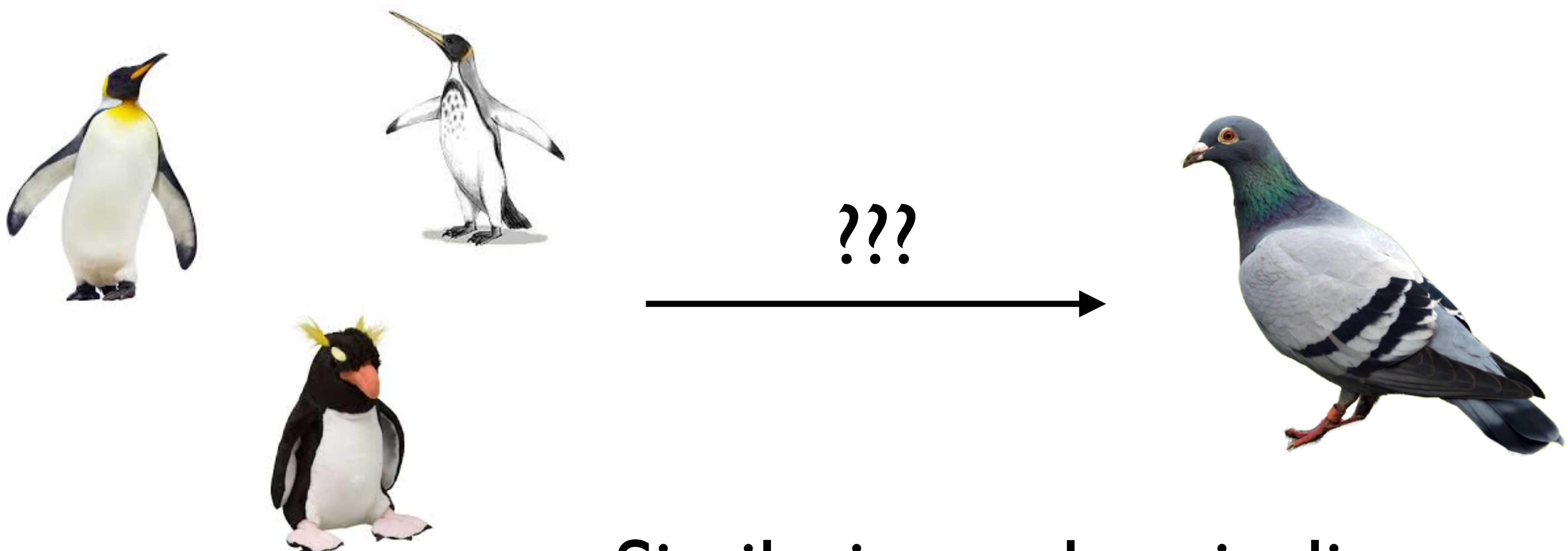
These birds have plaxium blood

The problem of inductive generalisation



?????

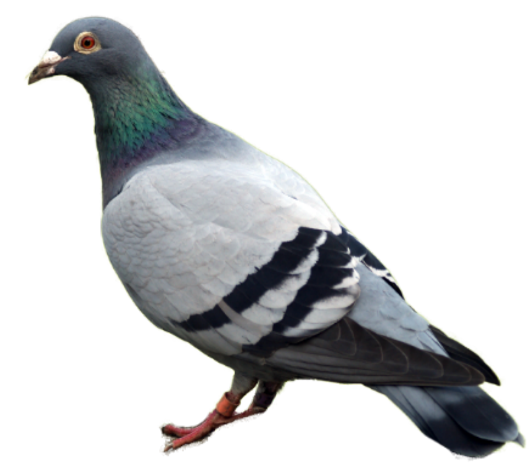
What factors shape our inductive inferences?



Similarity and typicality
of the sample

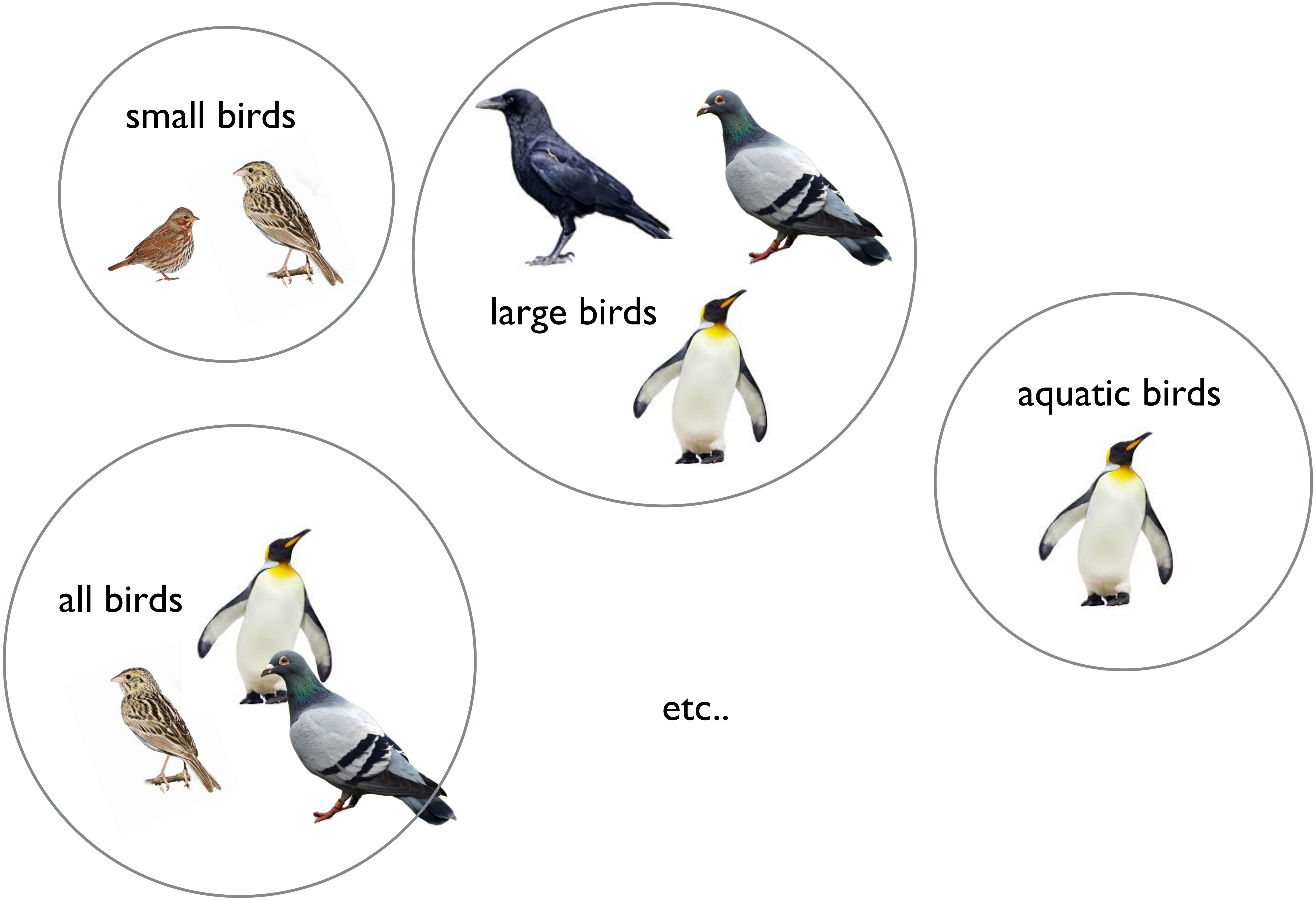
What factors shape our inductive inferences?

???

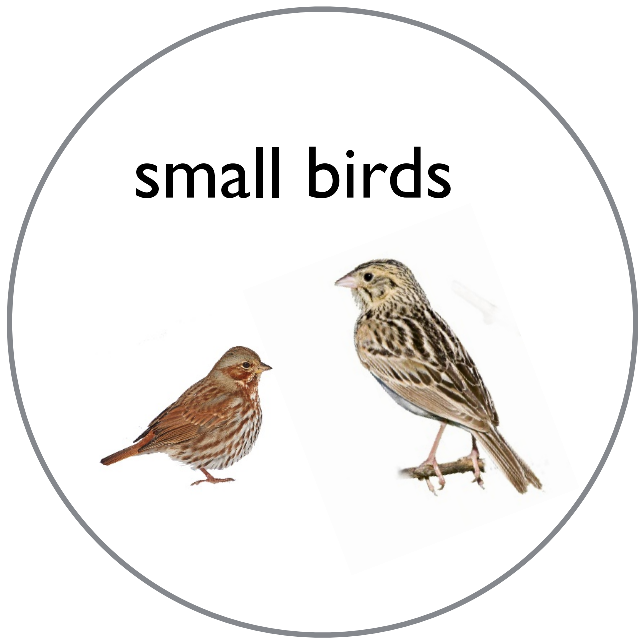


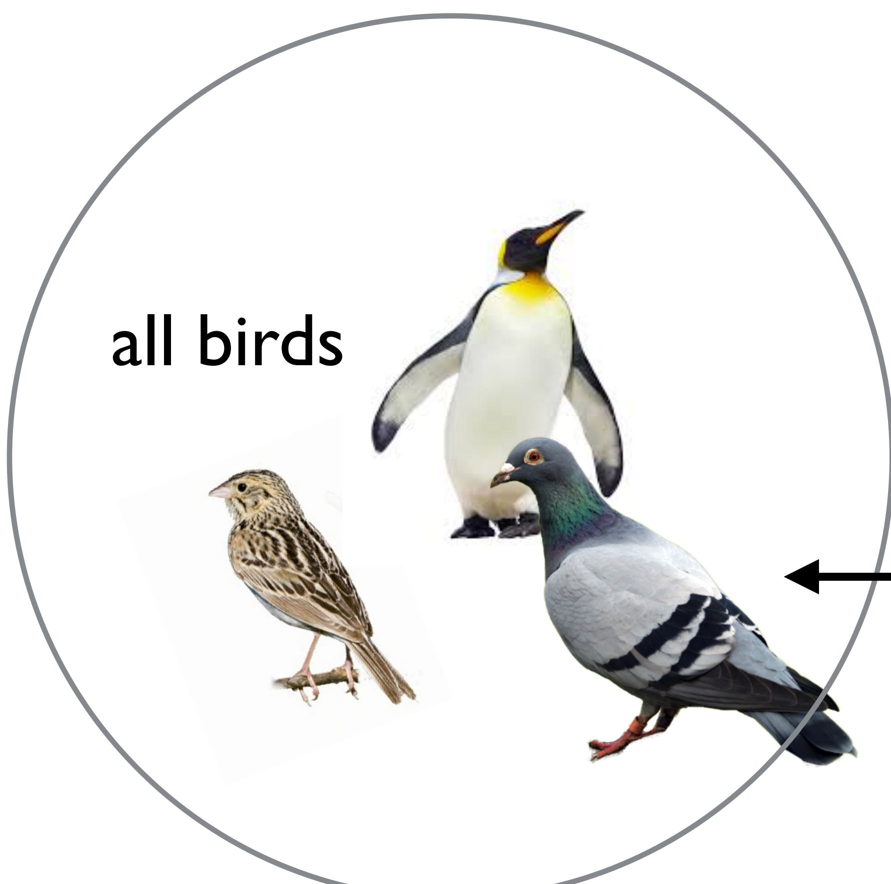
Size and diversity of the sample

Reasoners consider hypotheses



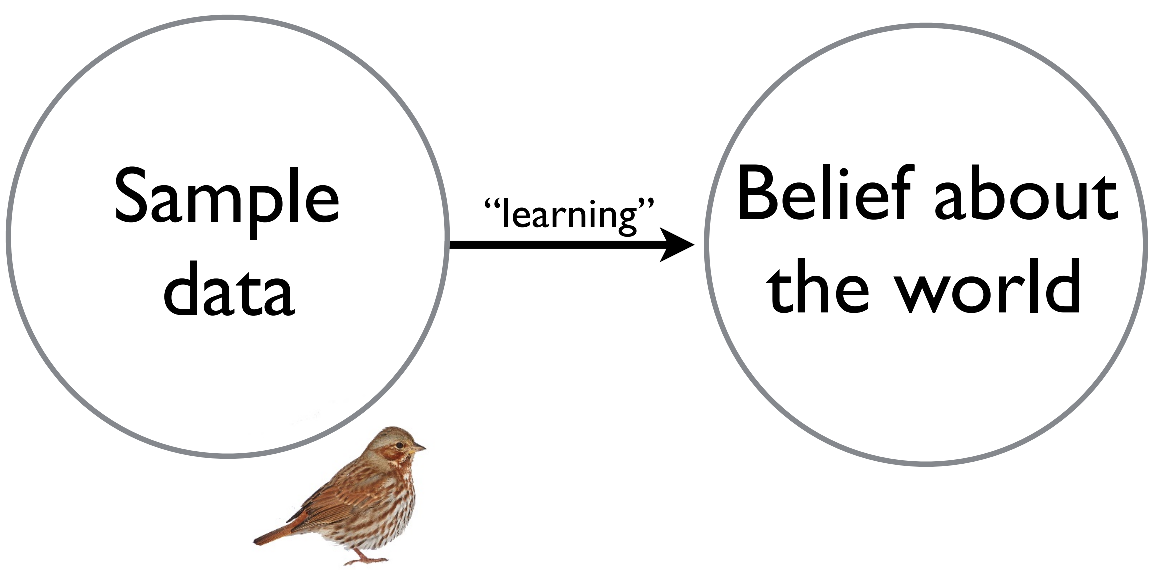
The sample rules out some and not others…





Inductive generalisation is based on hypotheses consistent with the sample

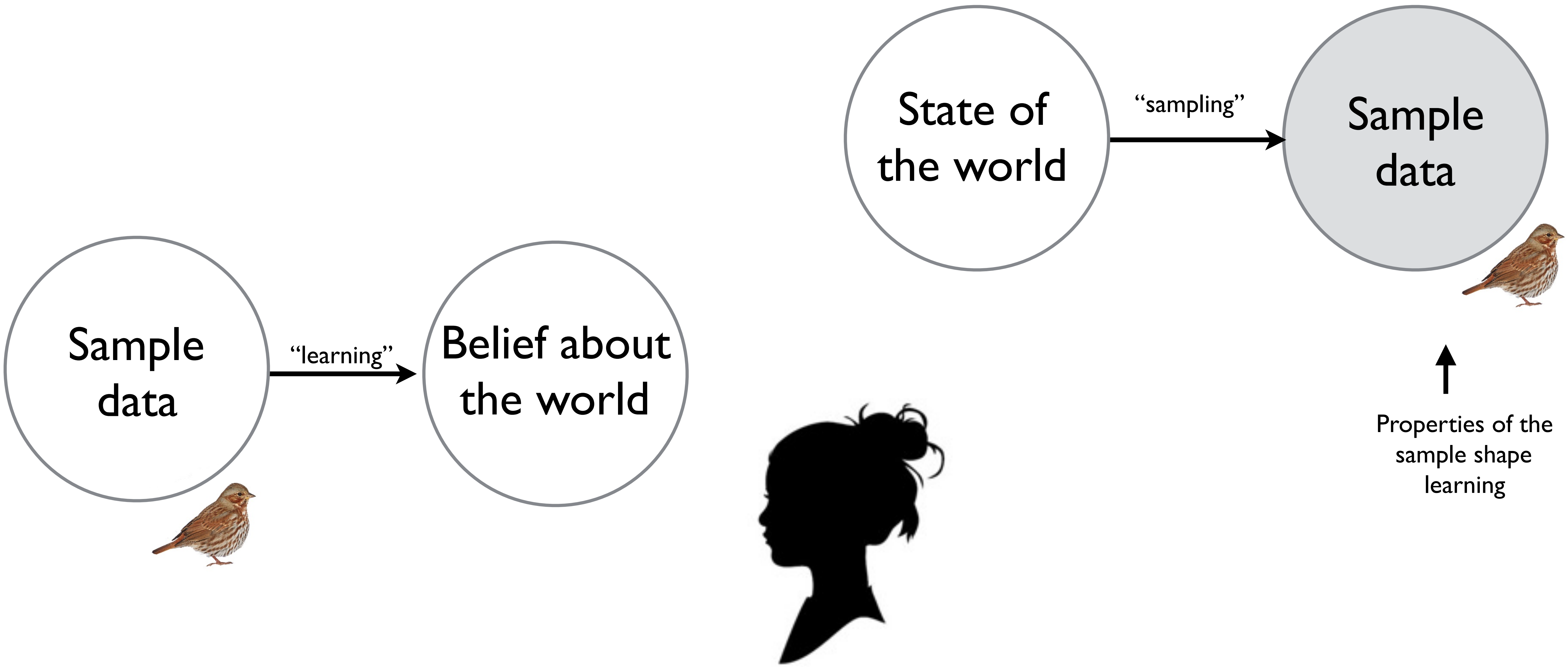
Traditional view of reasoning



Properties of the sample shape learning

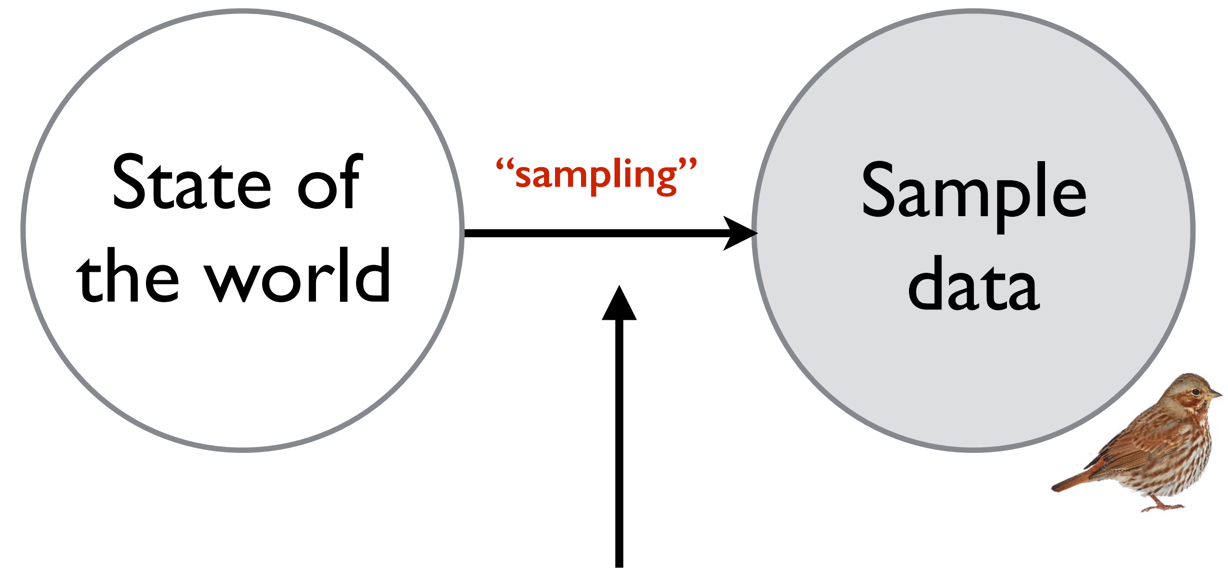
Reasoning as intuitive statistics

$$P(h|d) = \frac{P(d|h)P(h)}{\sum_{h' \in \mathcal{H}} P(d|h')P(h')}$$



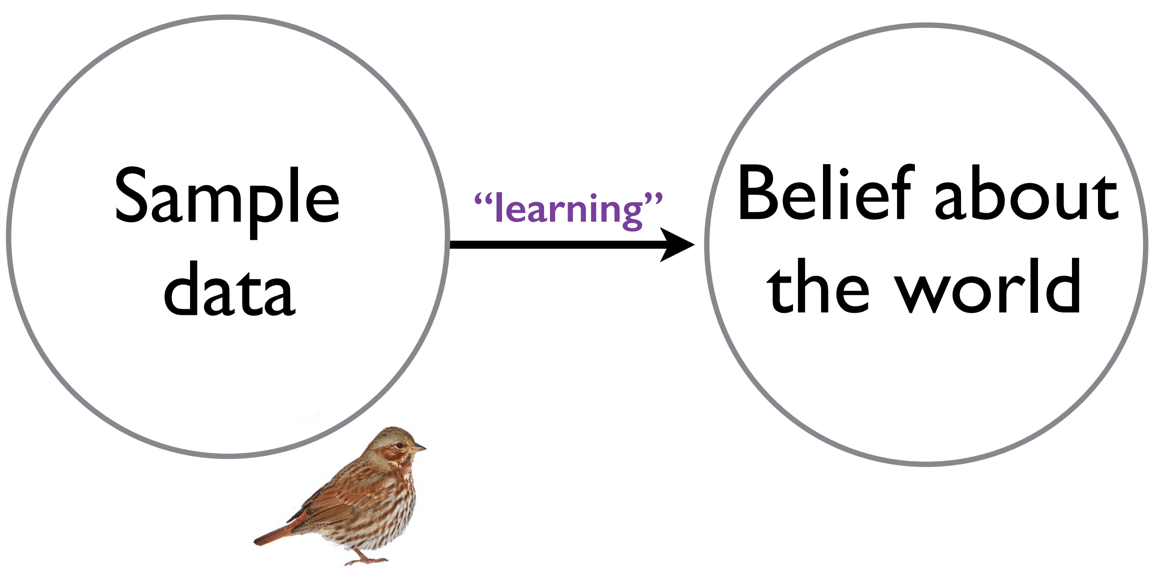
Critical prediction: Learning depends on *sampling*

$$P(h|d) = \frac{P(d|h)P(h)}{\sum_{h' \in \mathcal{H}} P(d|h')P(h')}$$



The evidentiary value of the sample depends on how the learner thinks it was generated, or how it came to their attention

Properties of the sample shape learning



Epistemic vigilance: Statistical reasoning about untrustworthy data

These birds have plaxium blood

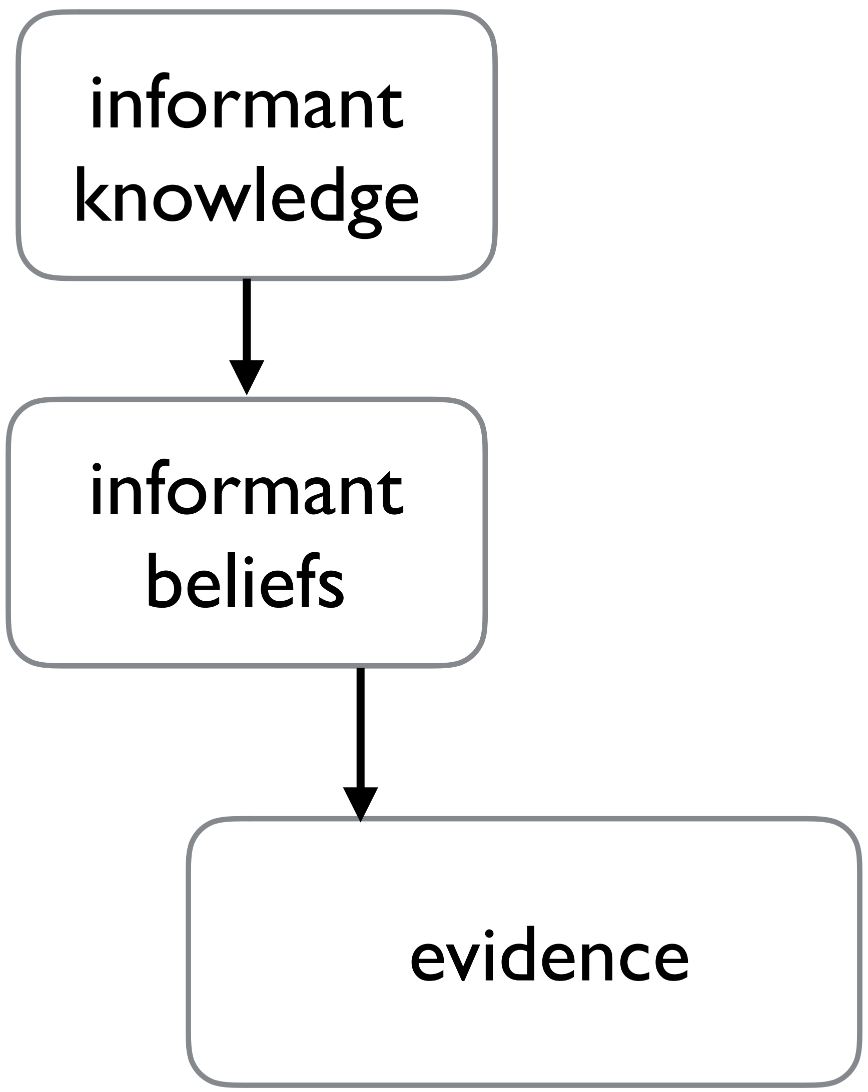
Does this bird have plaxium blood?

↑

"It's all made up" is *absolutely* a legitimate sampling assumption

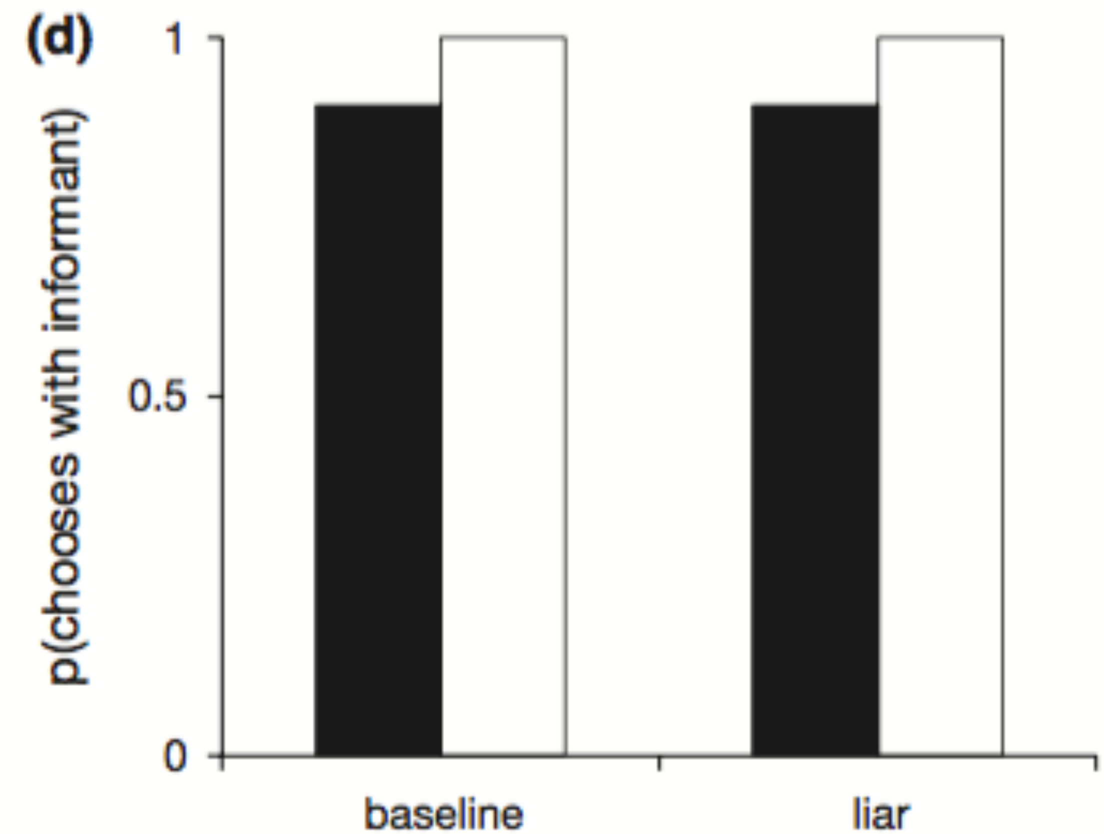
Does this bird have plaxium blood?

The price of inductive freedom is epistemic vigilance



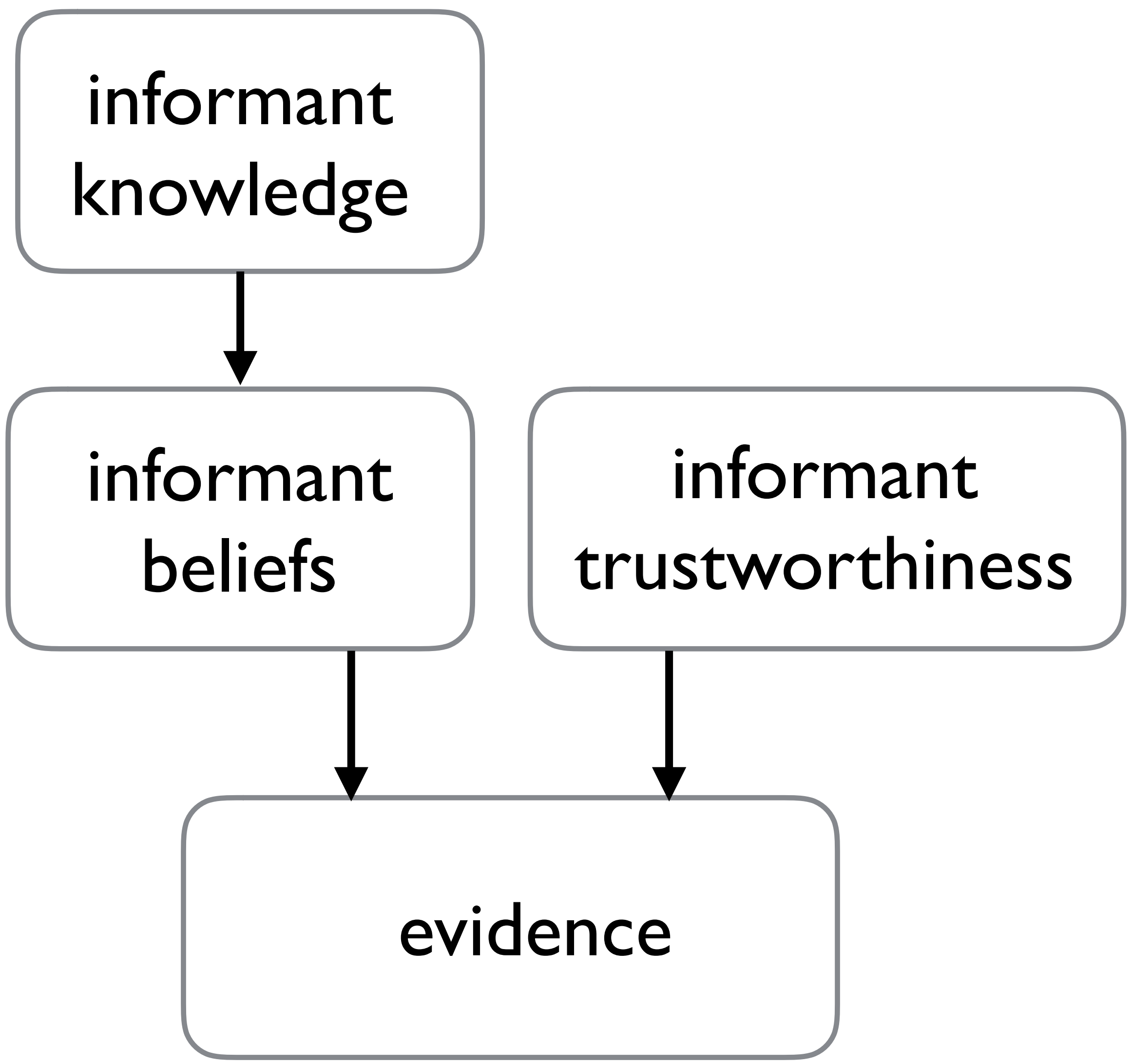
Shafto, Eaves, Navarro & Perfors (2012) *Developmental Science*

Three year olds are easily deceived…



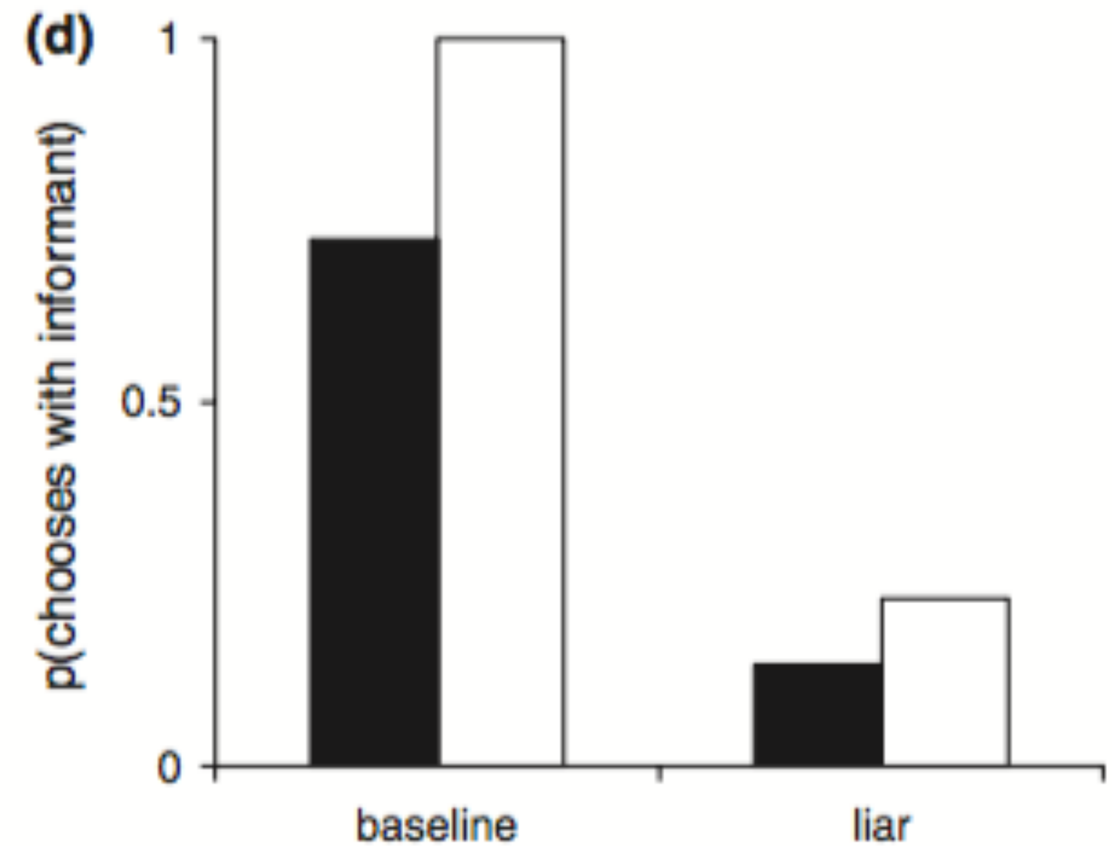
Mascaro & Sperber (2009)

The price of inductive freedom is epistemic vigilance



Shafto, Eaves, Navarro & Perfors (2012) *Developmental Science*

… but four year olds are savvy statisticians



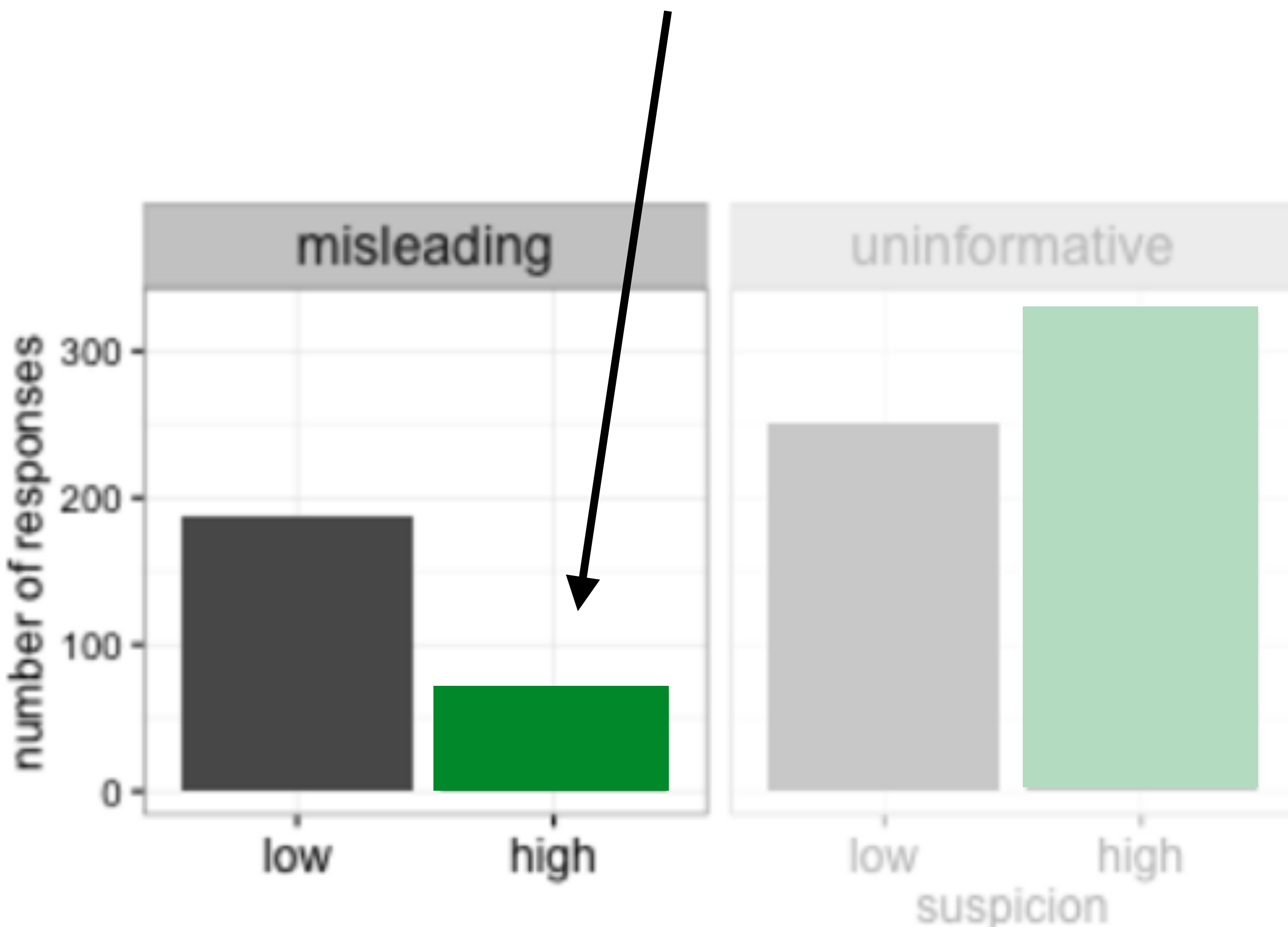
Mascaro & Sperber (2009)

Why epistemic vigilance?

People will try to "mislead with a half truth" if the listener is naive…

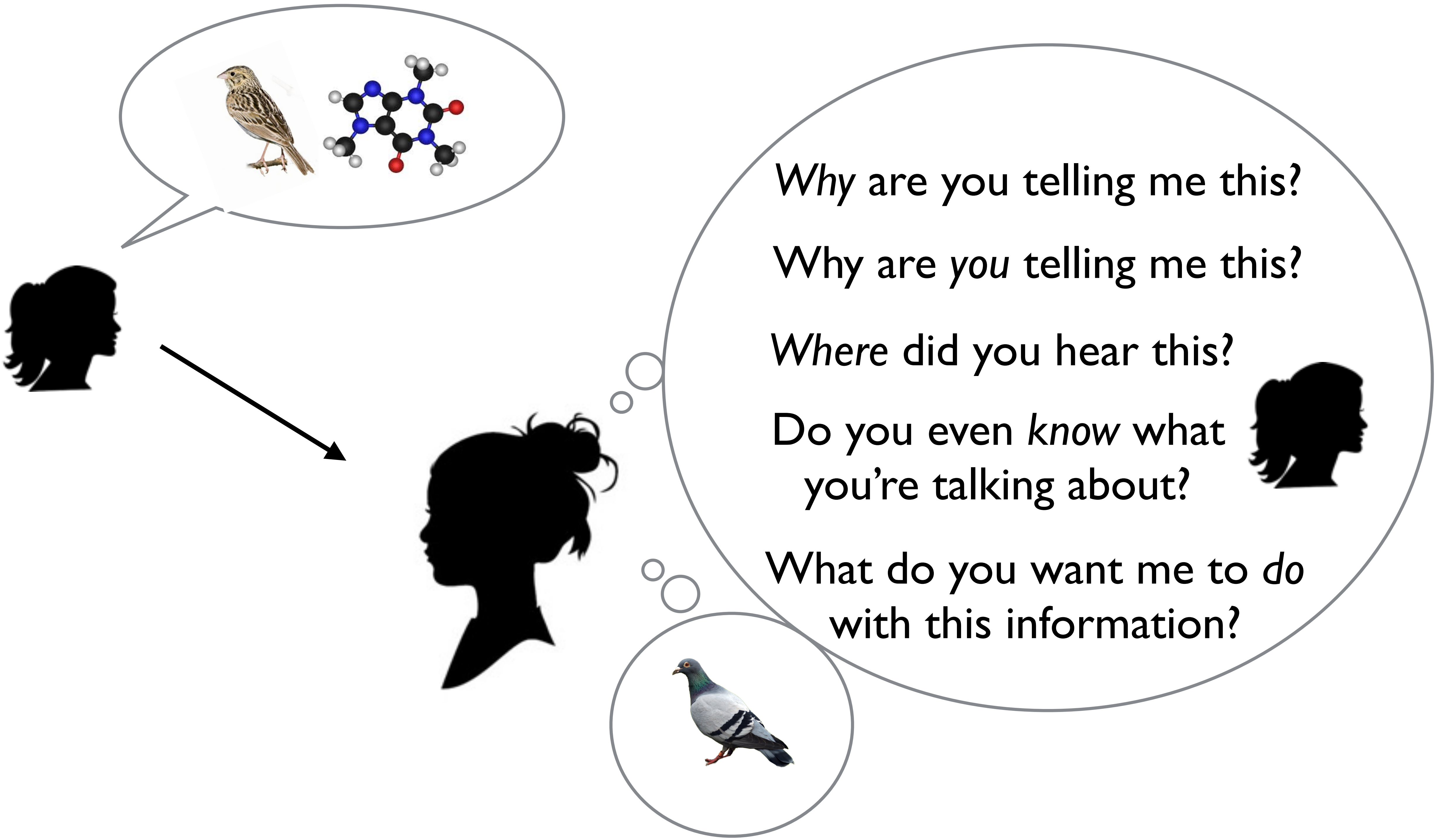
Ransom, Voorspoels, Perfors & Navarro (submitted)

They rarely try this when the listener is suspicious!



Ransom, Voorspoels, Perfors & Navarro (submitted)

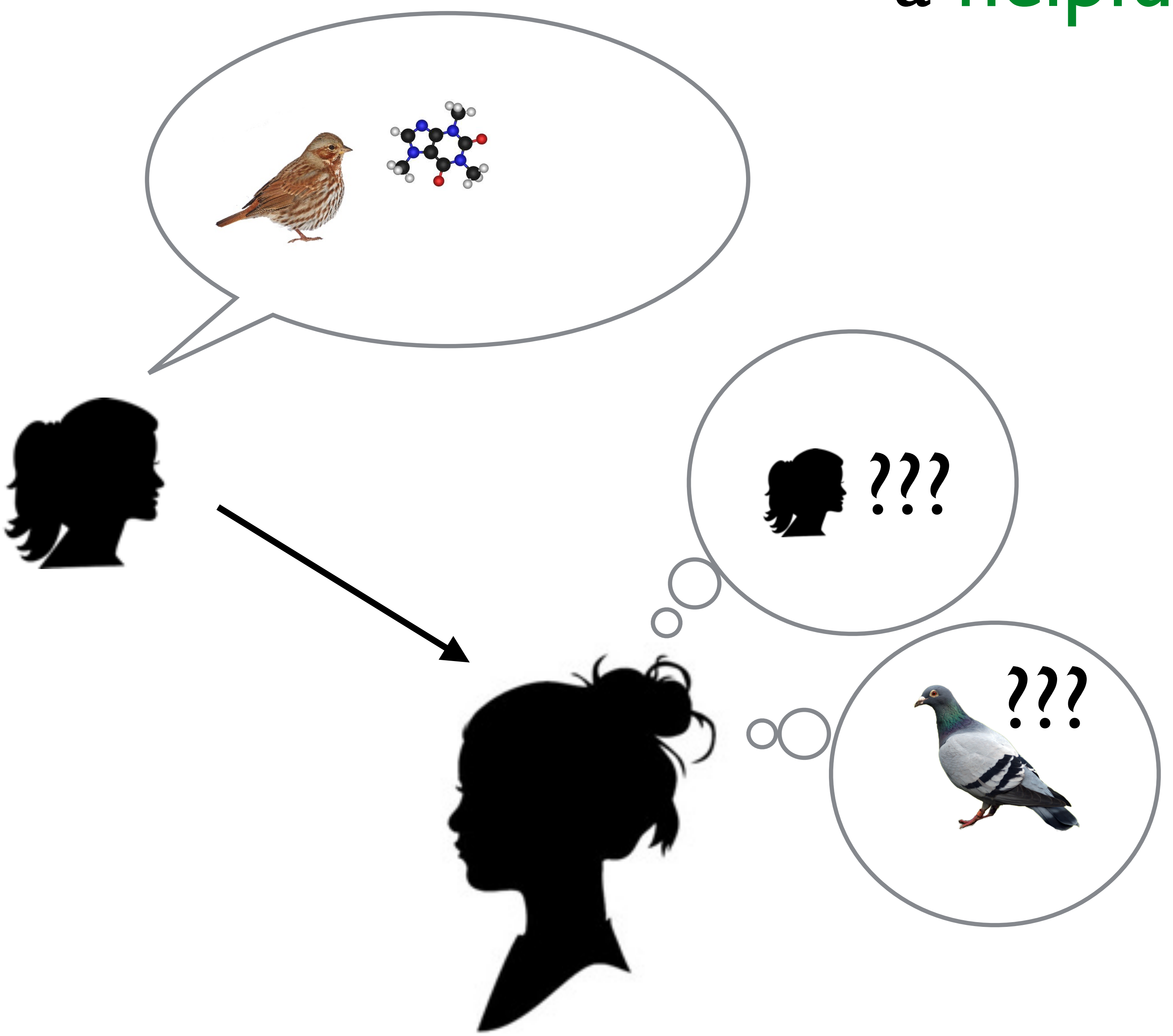
Everyday reasoning *about the world* is intertwined with *social reasoning* about other people



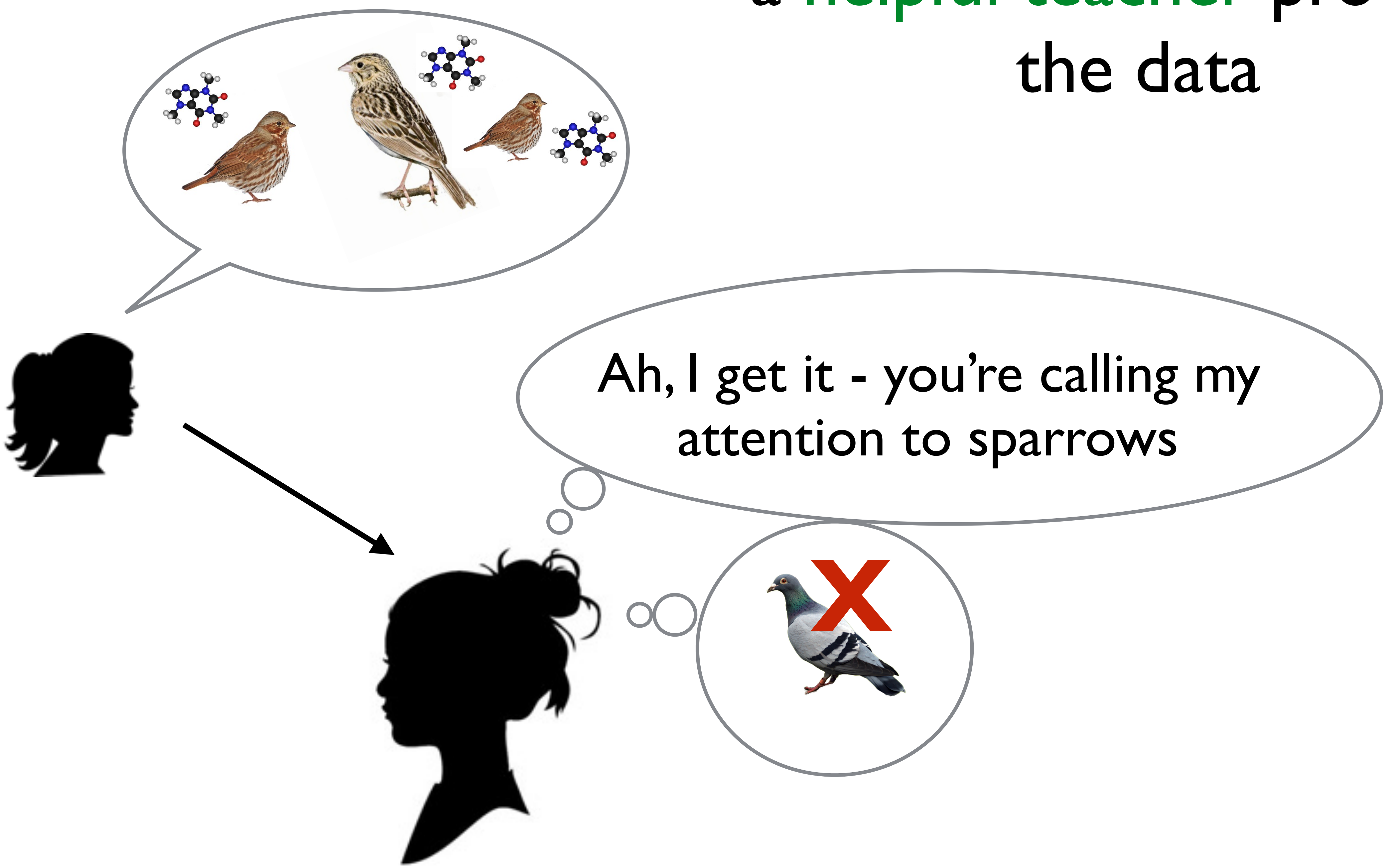
What does all this buy us?
Taking a hint from a helpful teacher

Ransom, Perfors & Navarro
(2016). *Cognitive Science*

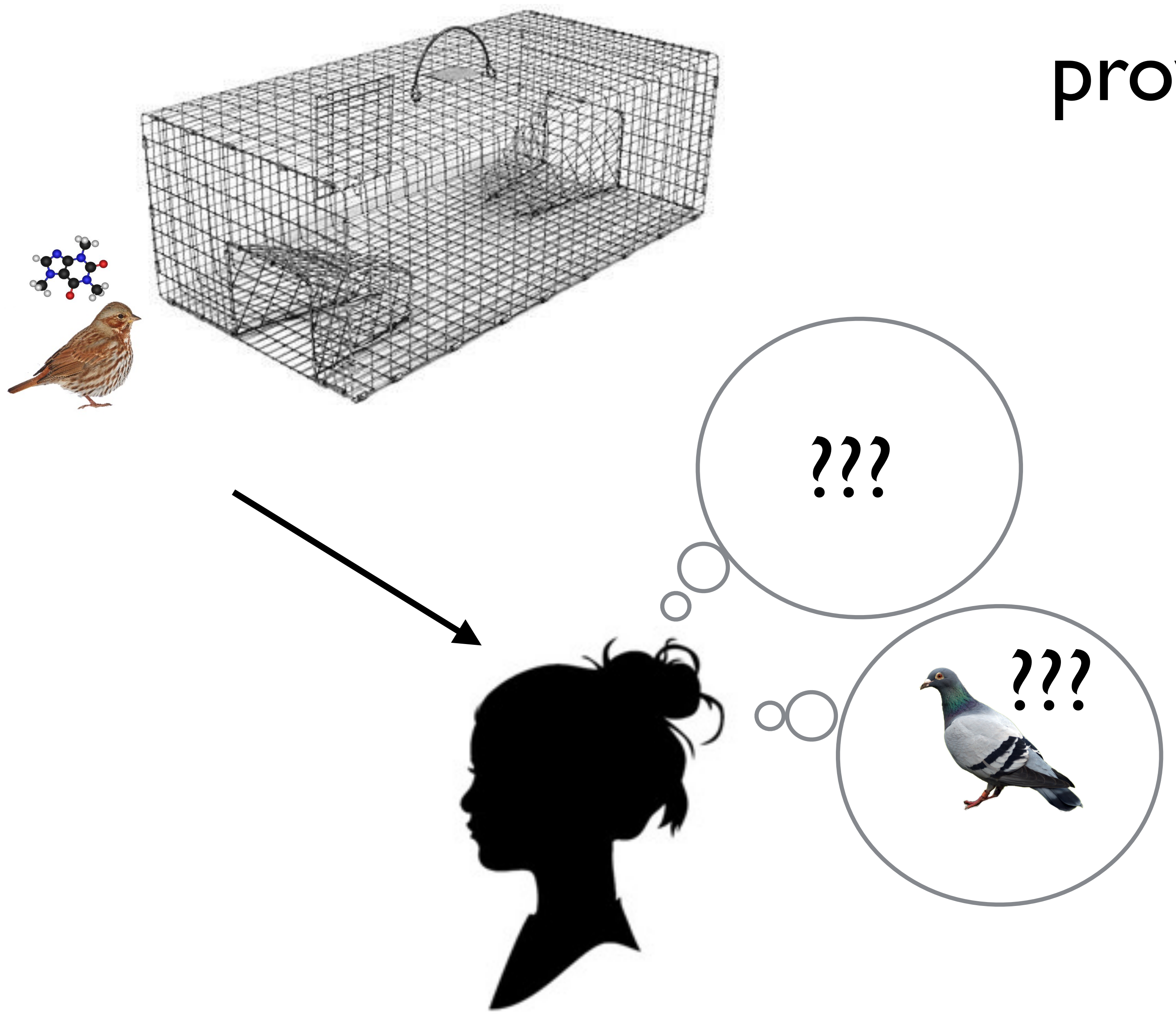
Inductive reasoning when a helpful teacher provides the data



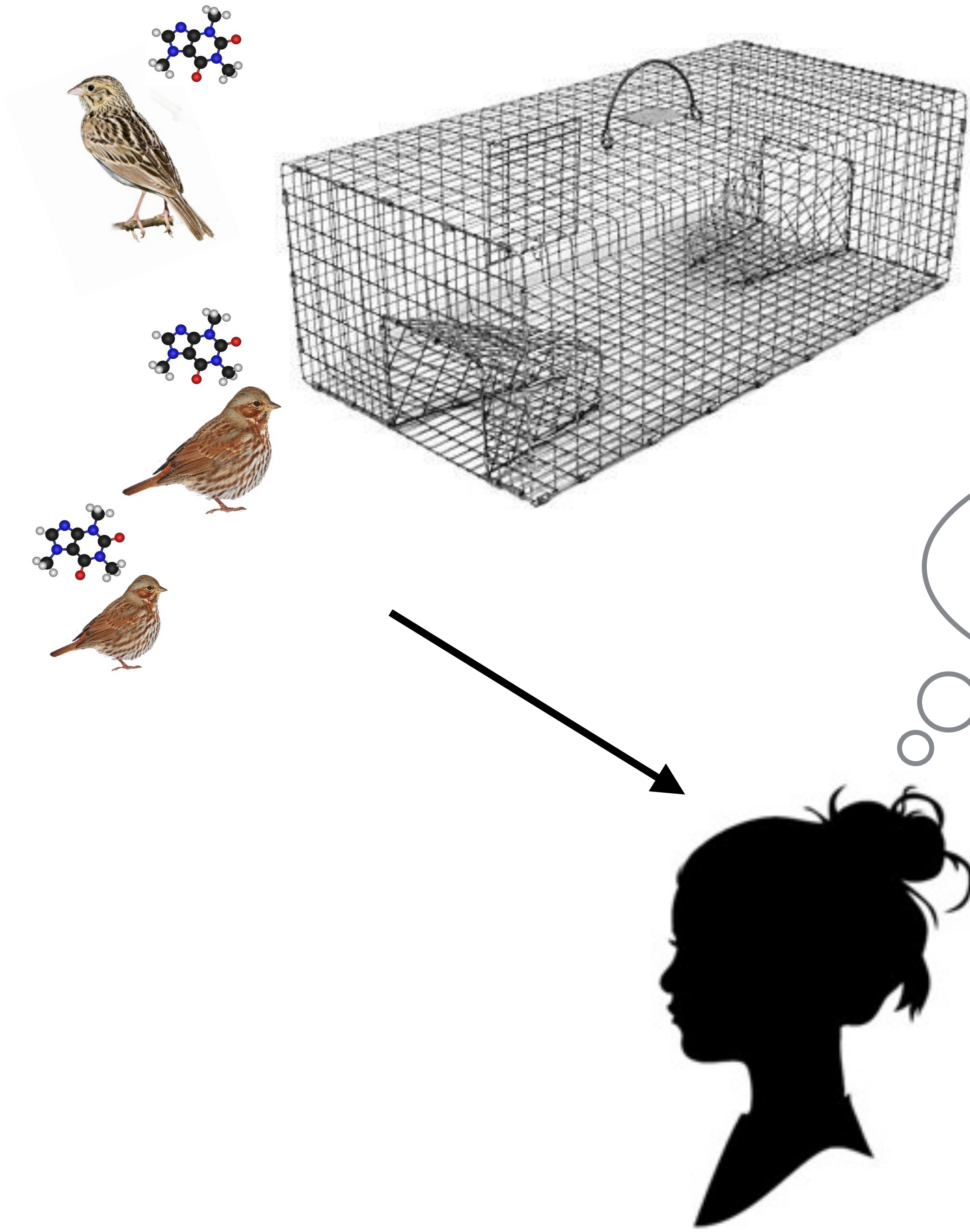
Inductive reasoning when a helpful teacher provides the data



Inductive reasoning when an indifferent world provides the data



Inductive reasoning when an indifferent world provides the data



Sampling mechanism:

Random:

"select items at random"

Helpful:

"select items to efficiently communicate an idea"

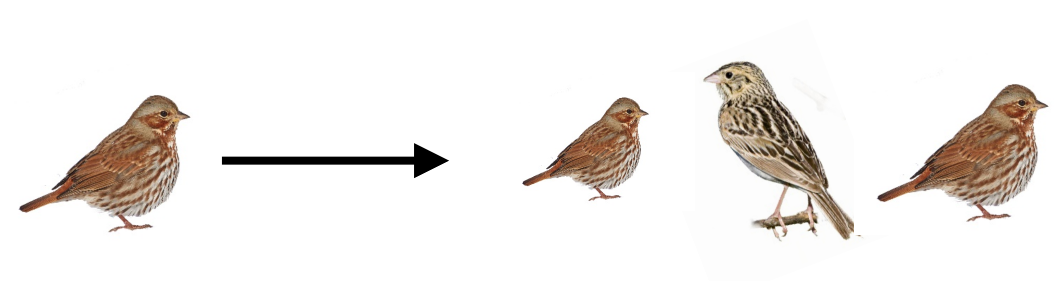
Prediction:

Random:

Adding positive instances has minimal effect if they're too similar to things I already know about

Helpful:

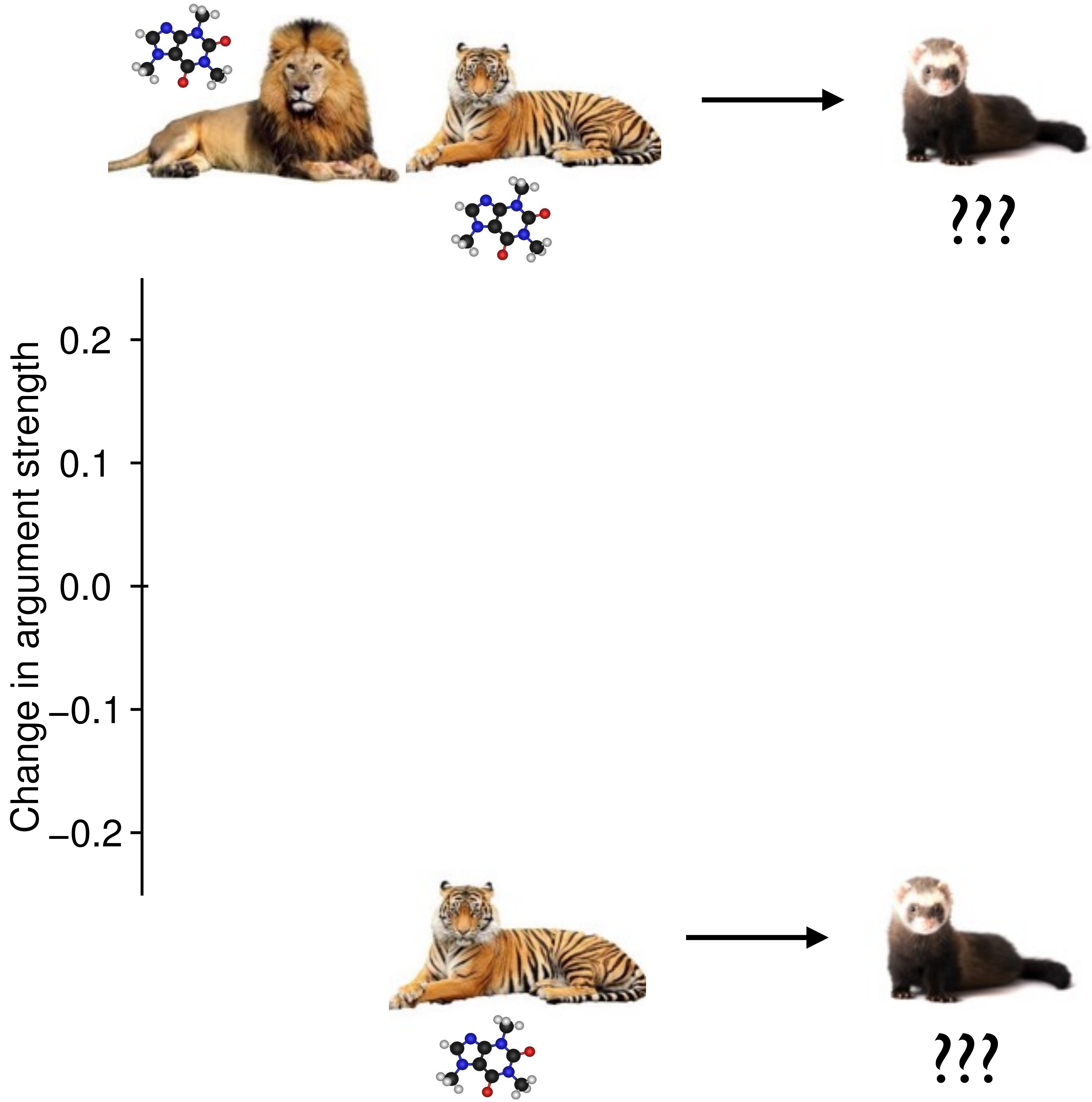
Adding positive instances from the same category conveys *intent*, and drives attention to that category



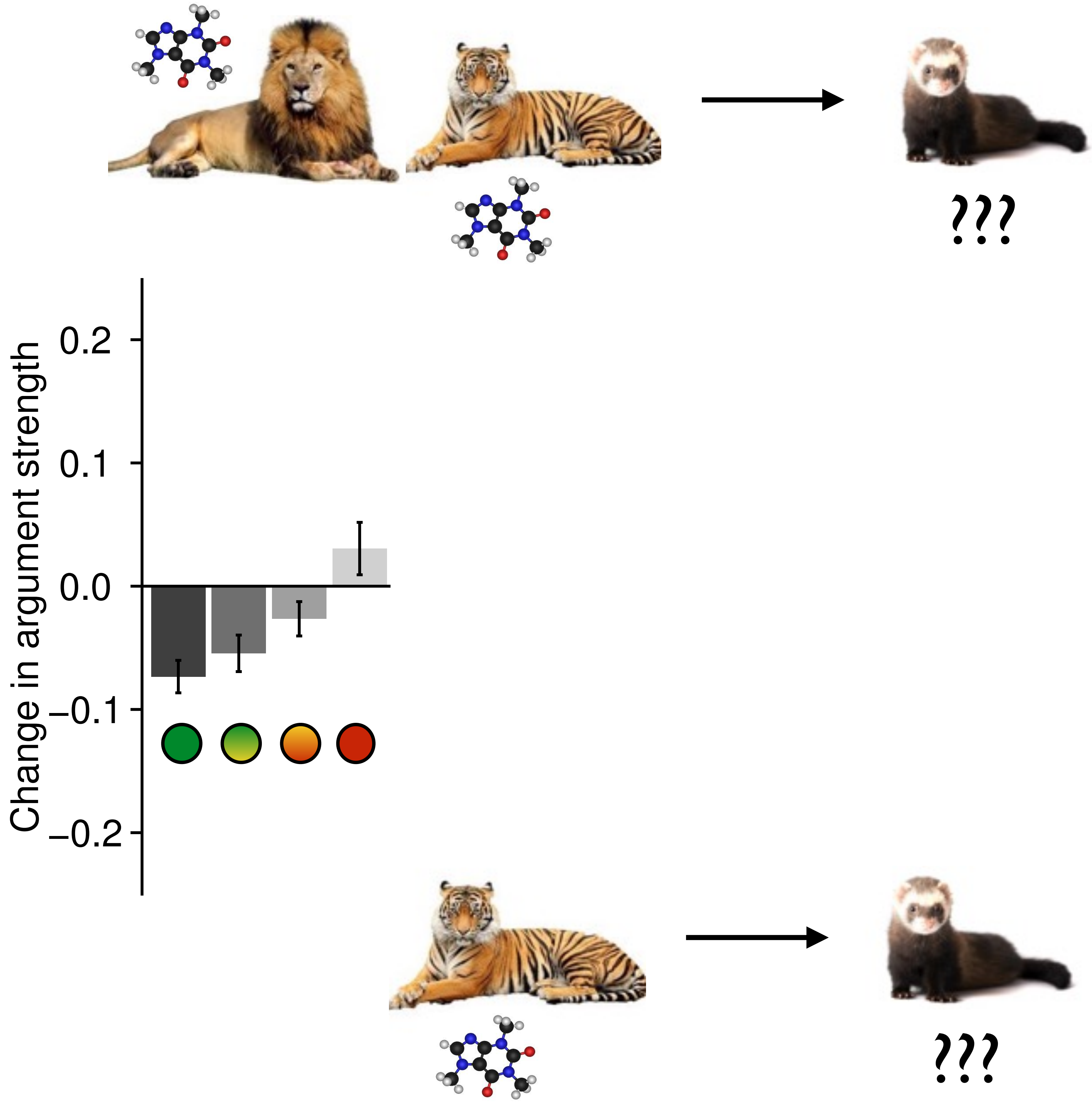
Previous experience?
(filler trials)

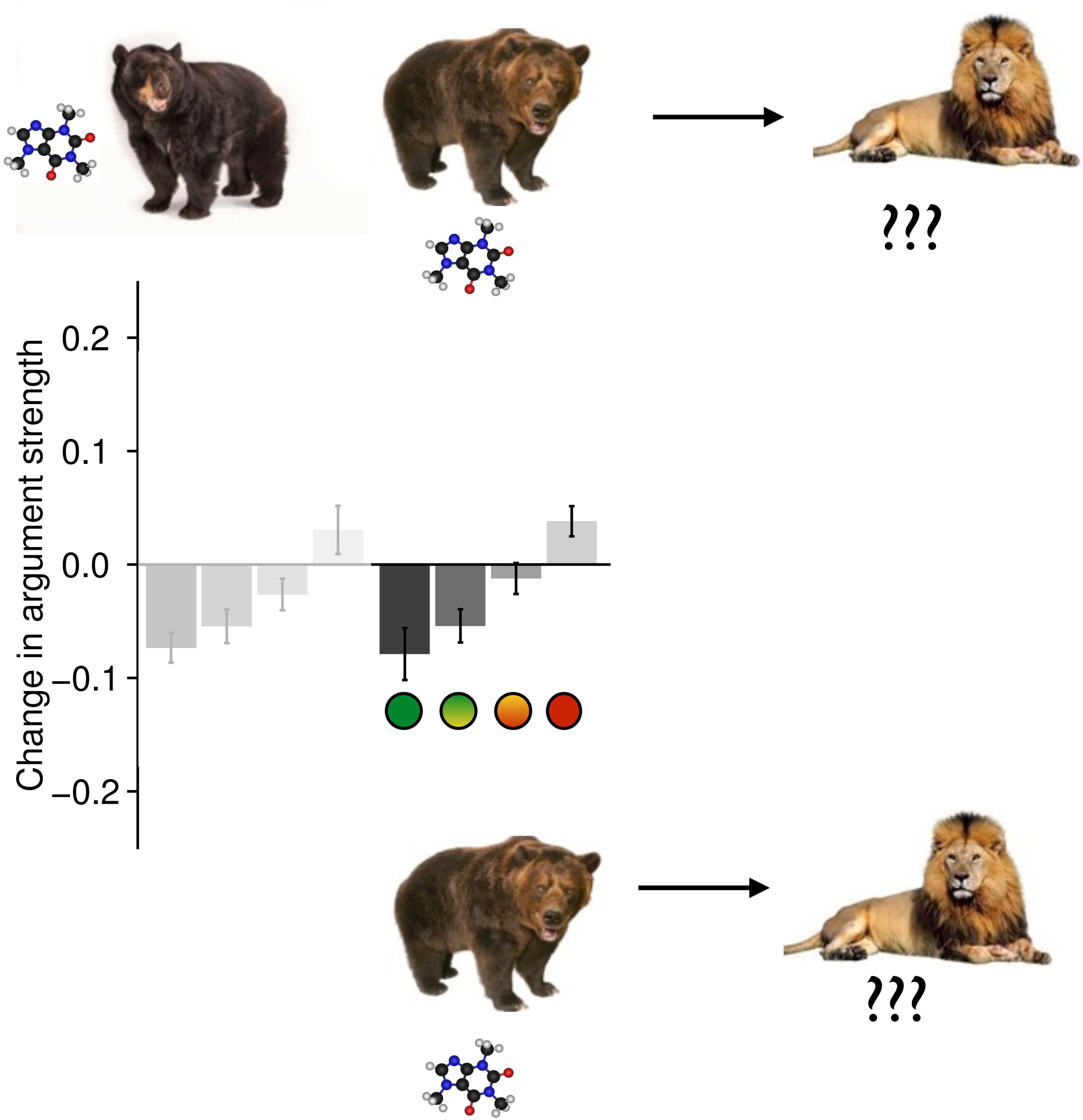
Cover story?

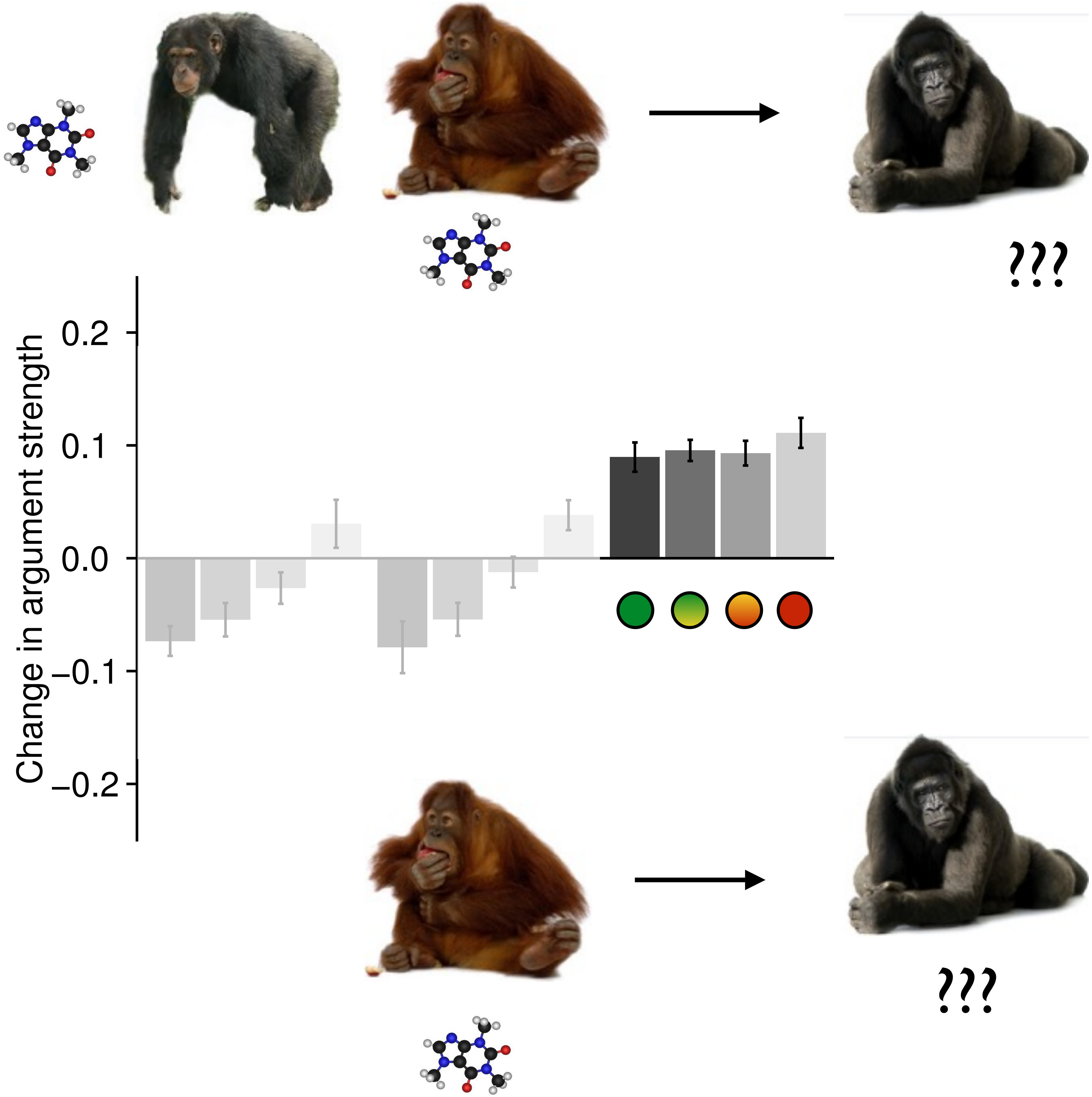
	Helpful cover story, filler trials imply helpful	
	Neutral cover story, filler trials imply helpful	Neutral cover story, filler trials imply random
		Random cover story, filler trials imply random



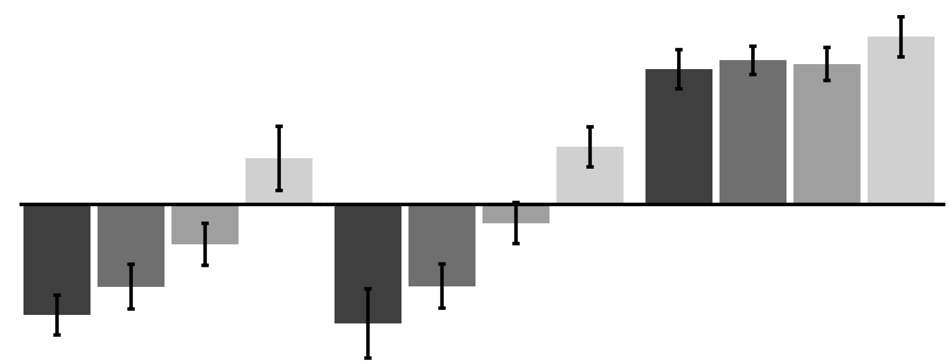
???
Change in argument strength
0.2
0.1
0.0
−0.1
−0.2
???


???
Change in argument strength
0.2
0.1
0.0
−0.1
−0.2
???


???
Change in argument strength
0.2
0.1
0.0
−0.1
−0.2
???

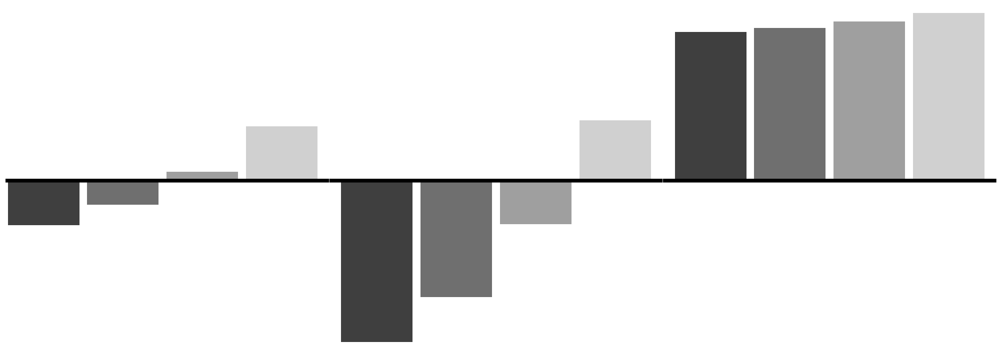

???
Change in argument strength
0.2
0.1
0.0
−0.1
−0.2
???

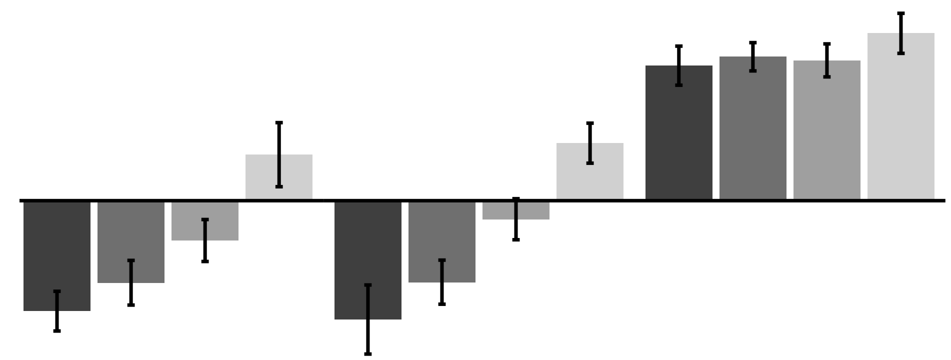
Humans



Bayes

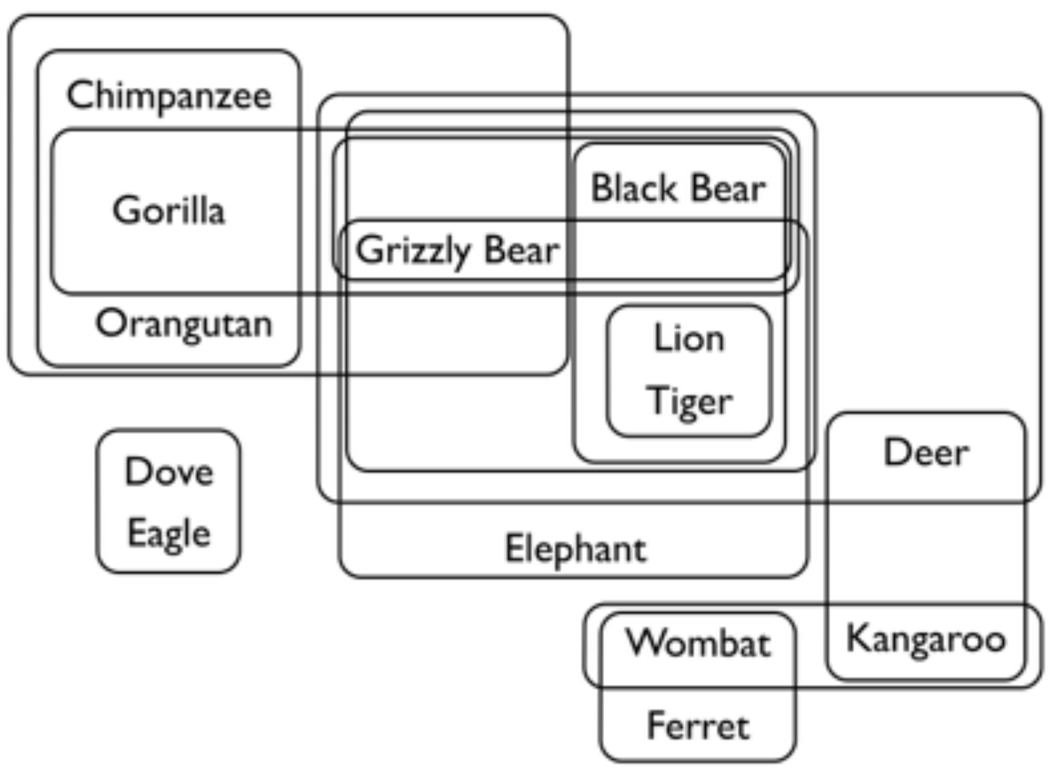
Humans





Knowledge about animal categories (*theory of the world*) creates structural differences between the different arguments

The sampling model (*theory of the context*) describes how "adding more data" can have different effects across conditions and arguments



Taking a hint from a helpful teacher… with negative evidence

Voorspoels, Navarro, Perfors, Ransom & Storms (2015). *Cognitive Psychology*

You want to infer whether <u>all ravens are black</u>.
Which of these observations is more helpful?

(raven, black)

(¬black, ¬raven)

Paradox of the raven: see Hempel (1945), Good (1960), etc

Positive evidence

Negative evidence

(raven, black)

(¬black, ¬raven)

Positive evidence

Mozart produces alpha waves in the brain

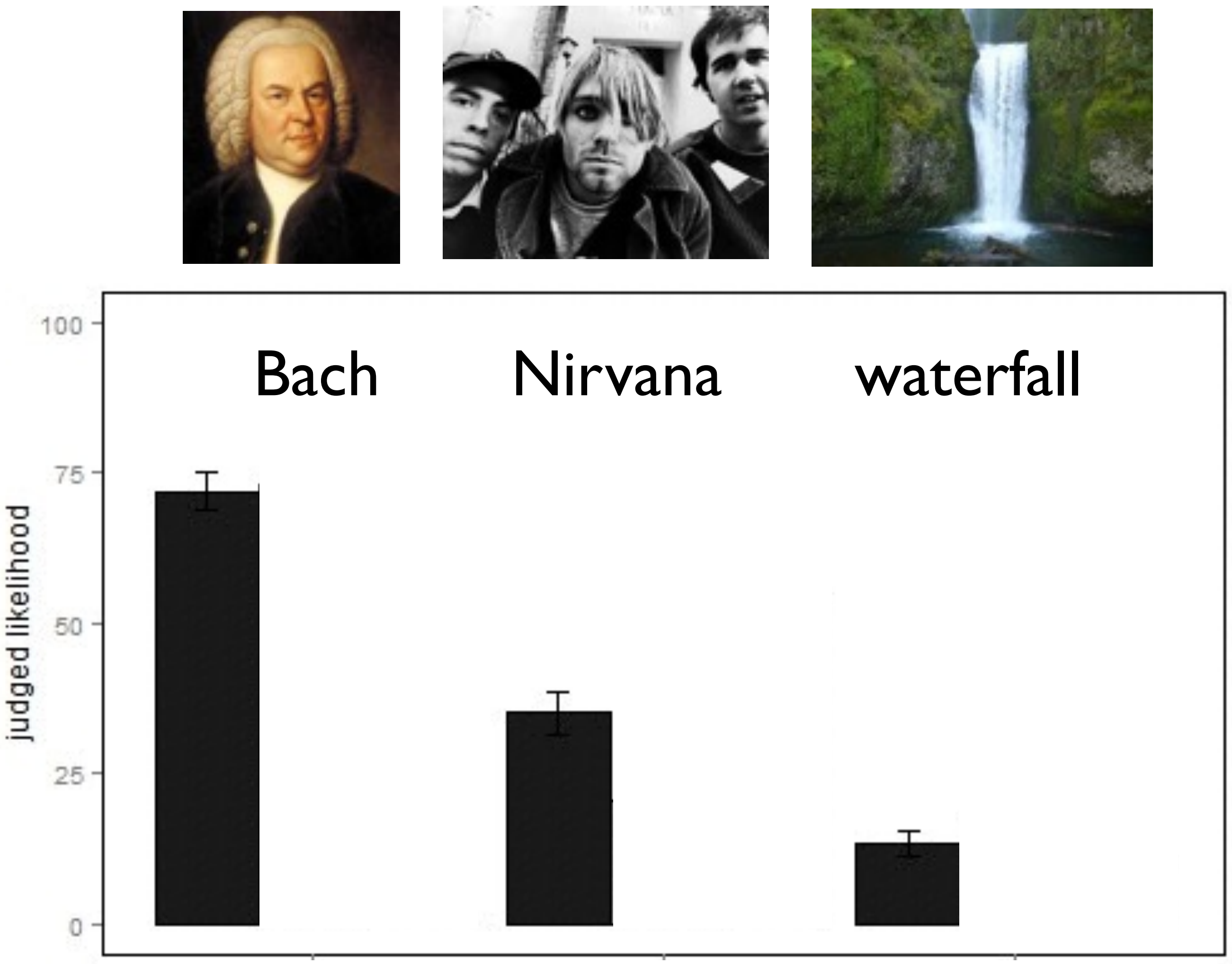
Negative evidence

The sound of a falling rock does not

Example stimuli only - the real experiments used *many* variations

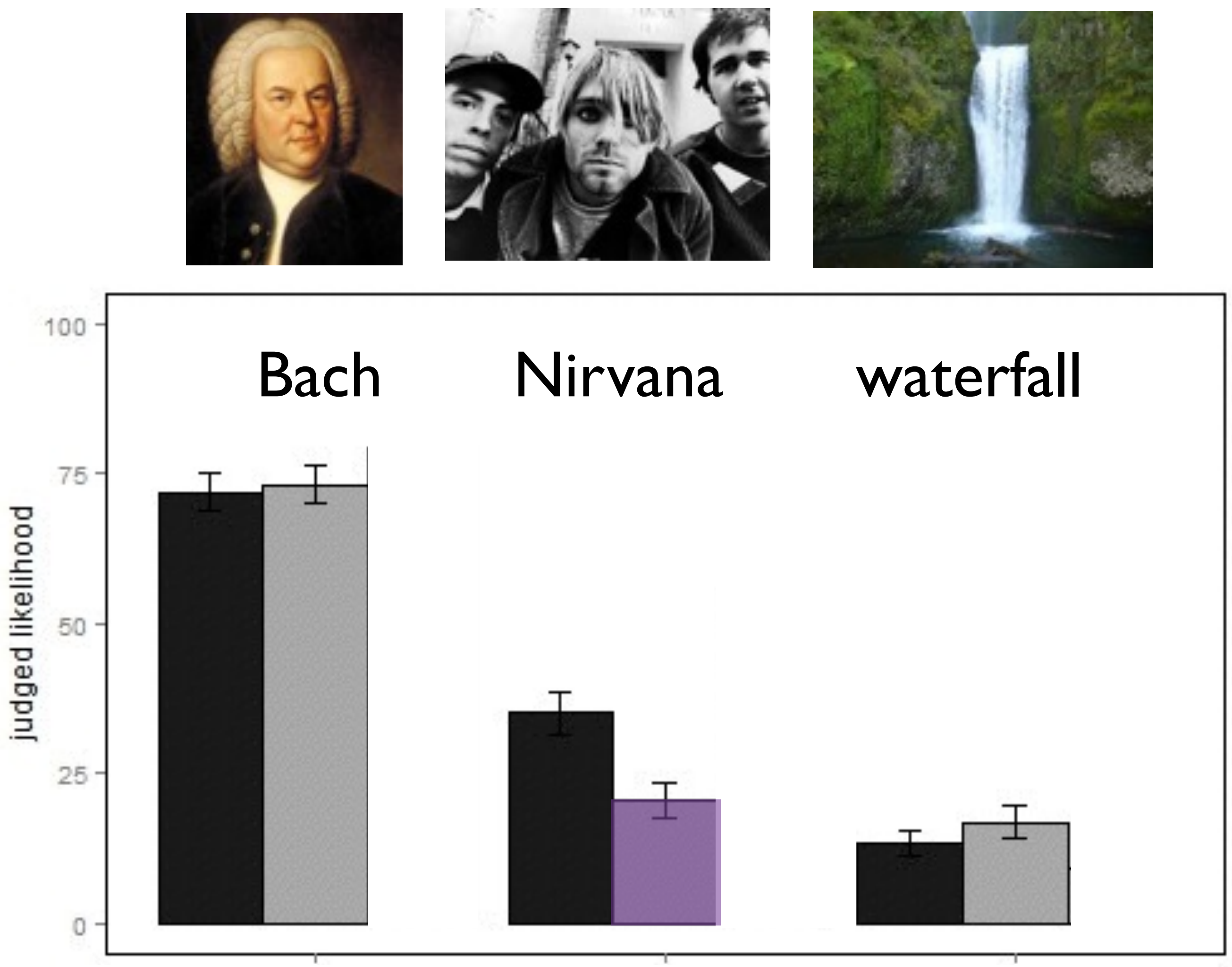
Okay, we start by telling people that Mozart does produce alpha waves…

+Mozart



+Mozart

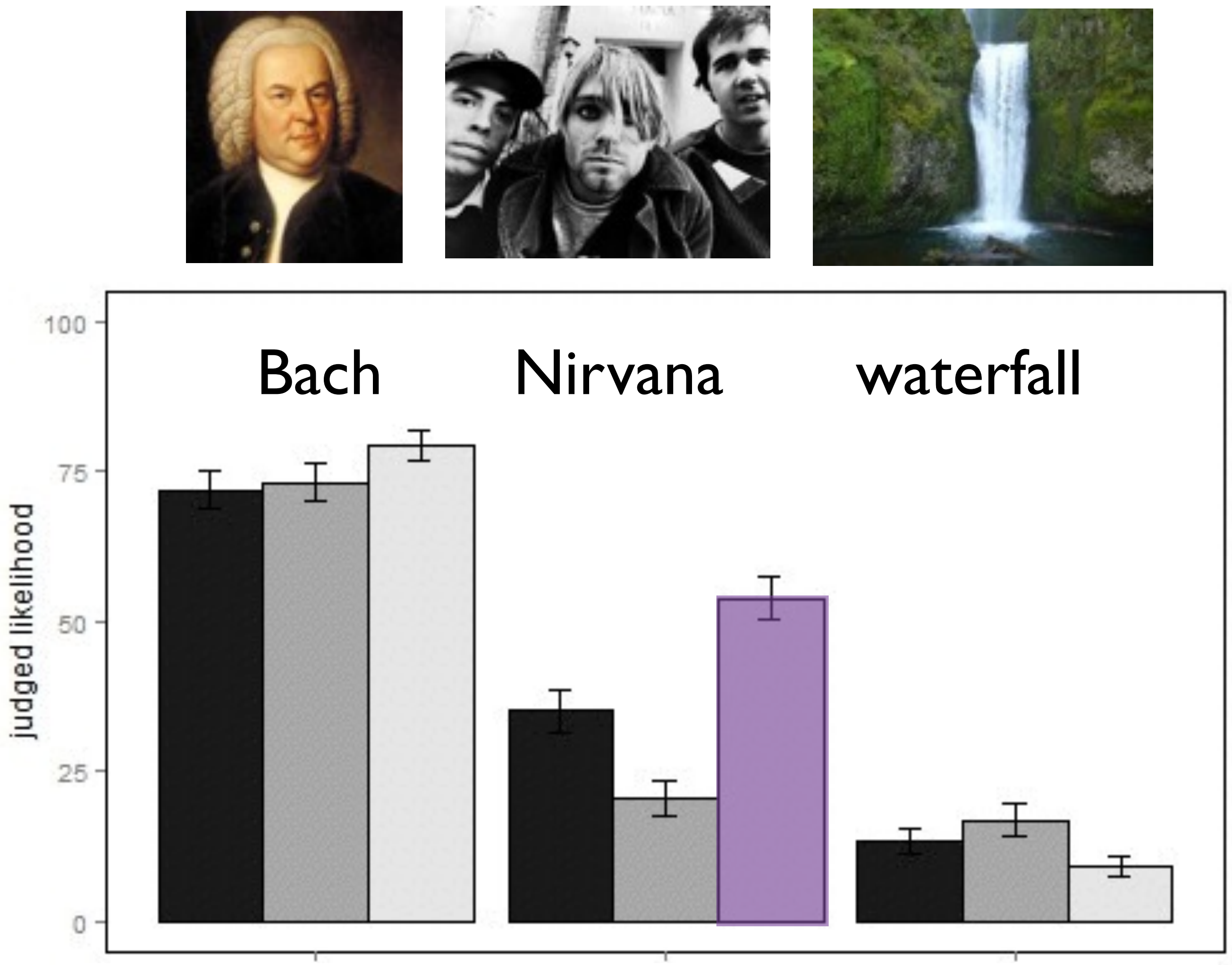
… and they reason sensibly



+Mozart

-Metallica

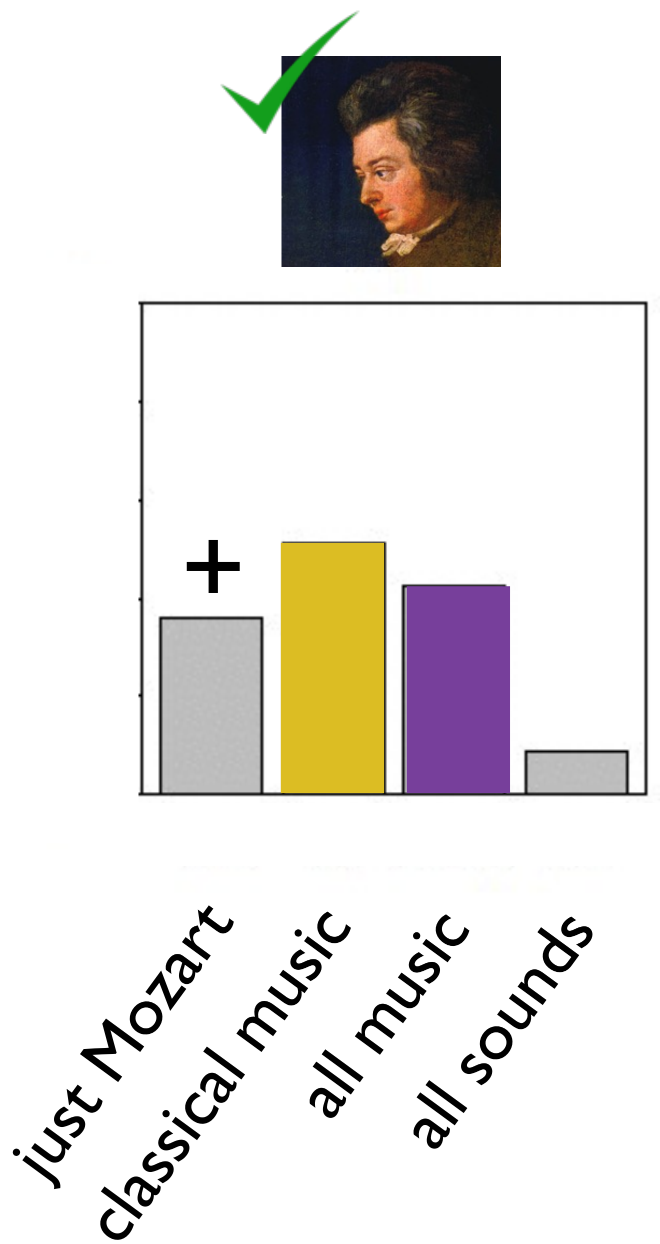
Adding Metallica as a negative example has a modest, sensible effect on inferences about Nirvana



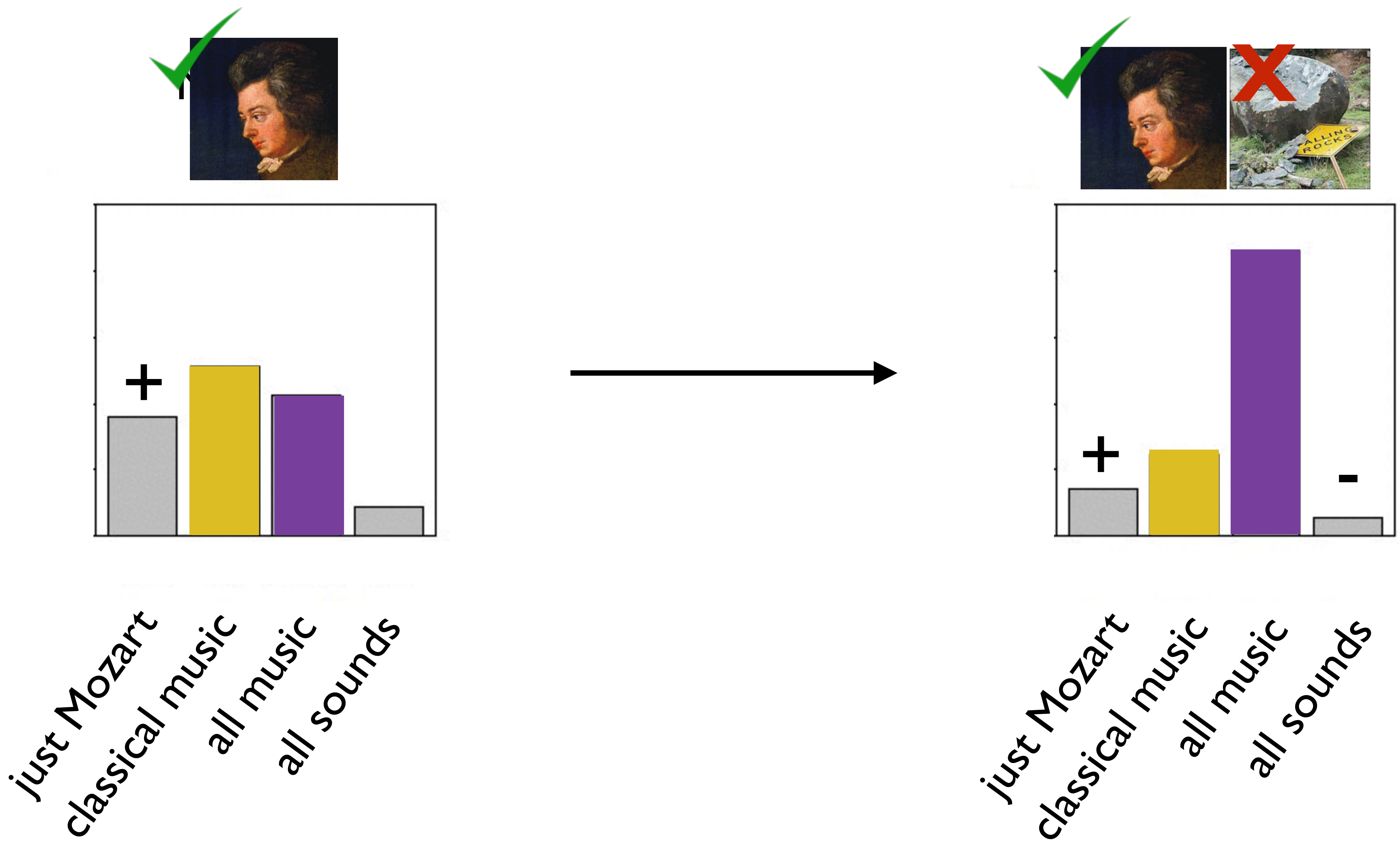
Um.

+Mozart

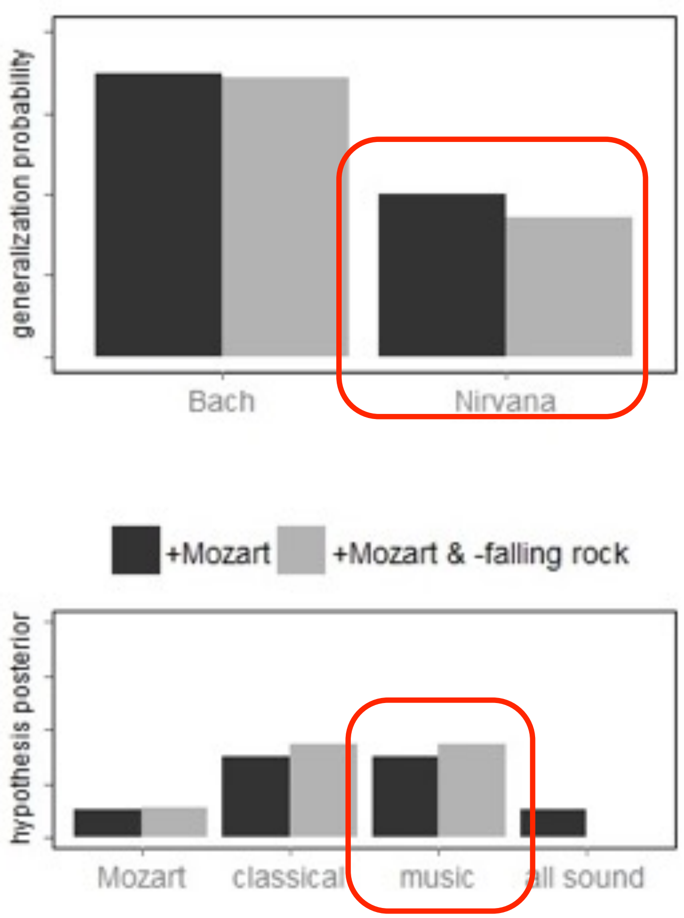
-Falling rock


+
just Mozart
classical music
all music
all sounds

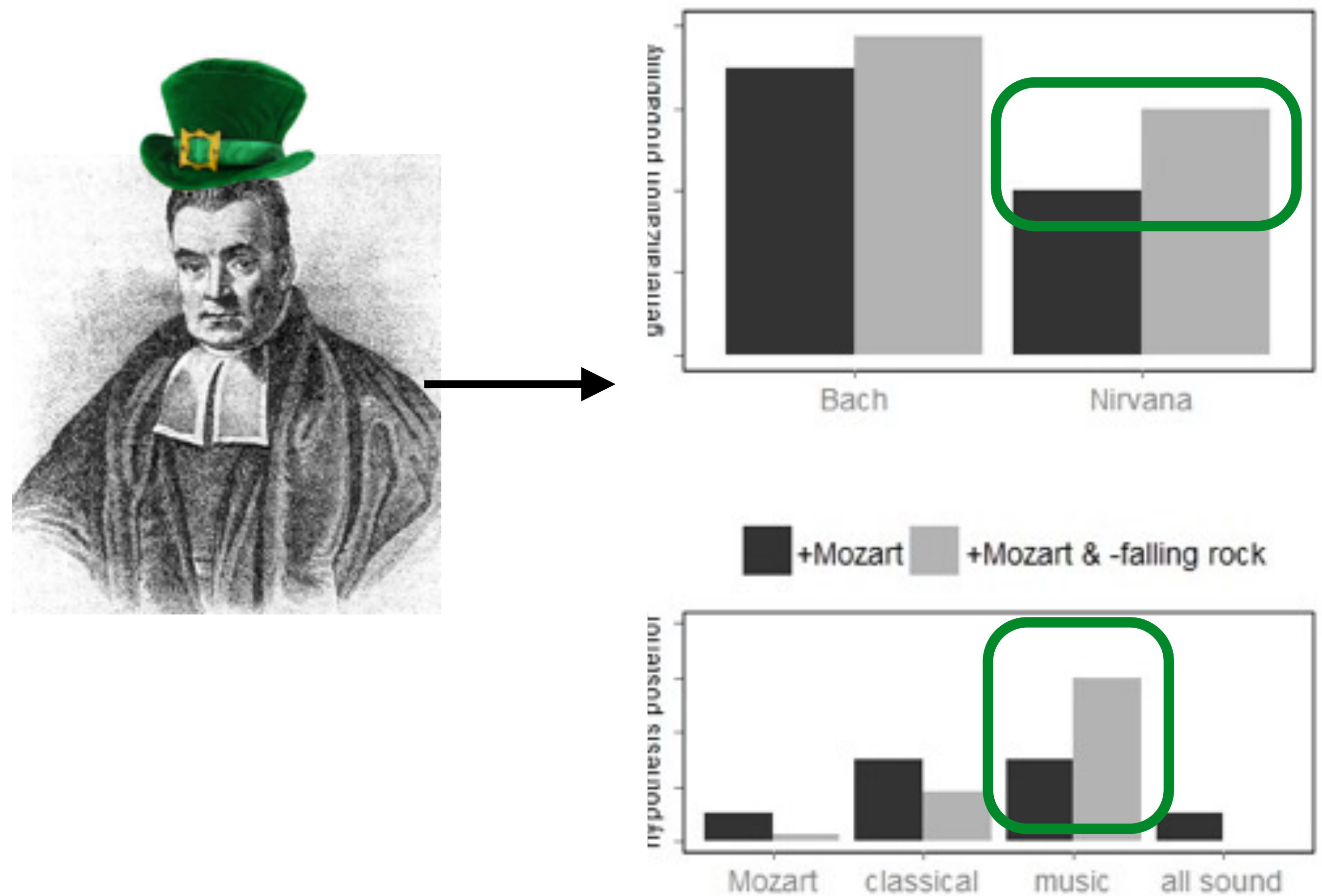
Negative evidence is interpreted as marking the category boundary



Bayesian reasoners with a random sampling assumption do *not* produce the effect



Bayesian reasoners with a helpful sampling assumption *do* produce the effect



What does it mean to be "helpful" anyway?

$$P(x|h) \propto P(h|x)^{\alpha}$$

The data x sampled by the communicator…

… is designed to maximise the learner's degree of belief in hypothesis h

Mozart but ***not*** rocks.
Wink wink

Gotcha!

Prediction:

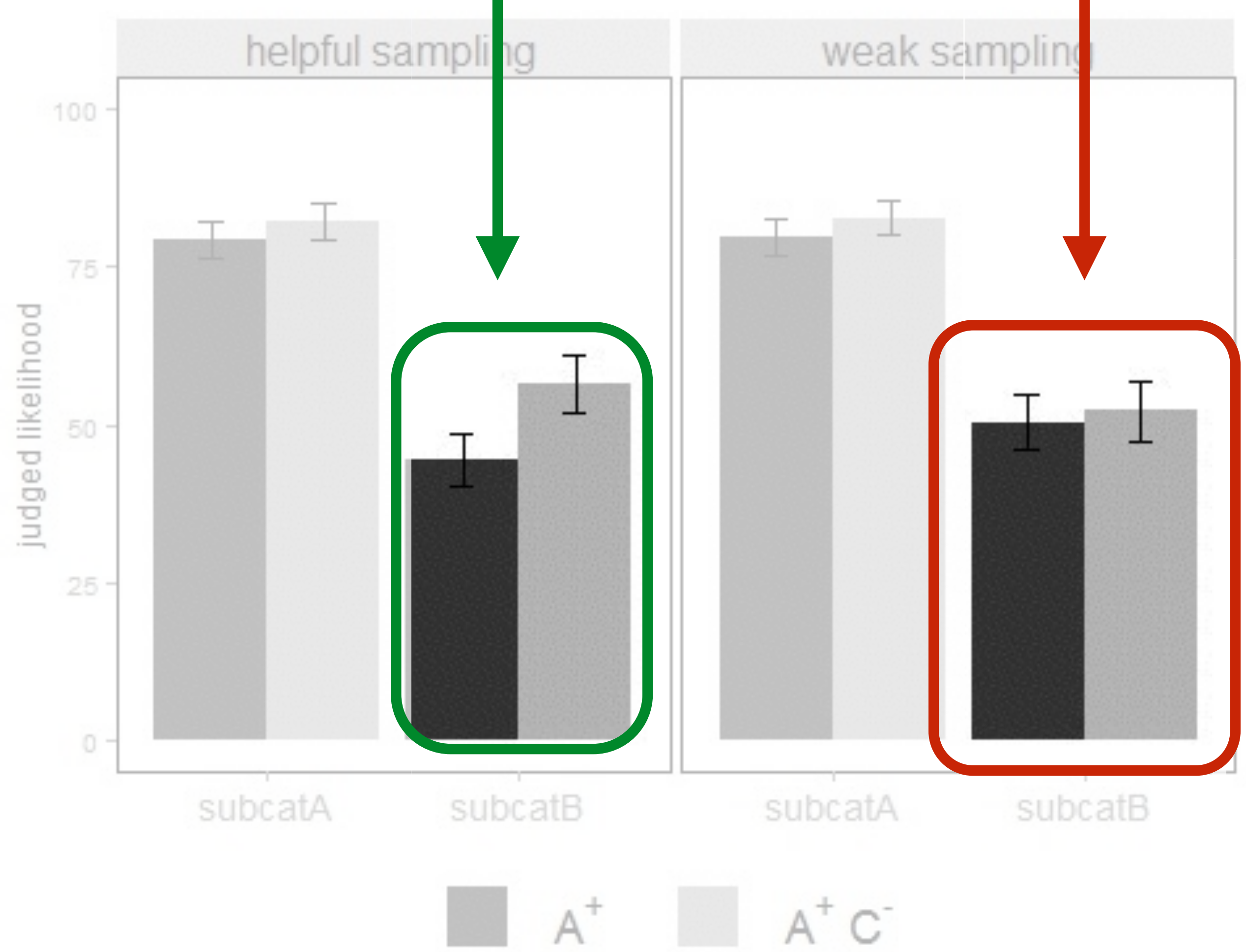
If the negative evidence is perceived as a helpful hint we should continue to get the effect

If it is construed as an arbitrary fact, the effect should vanish

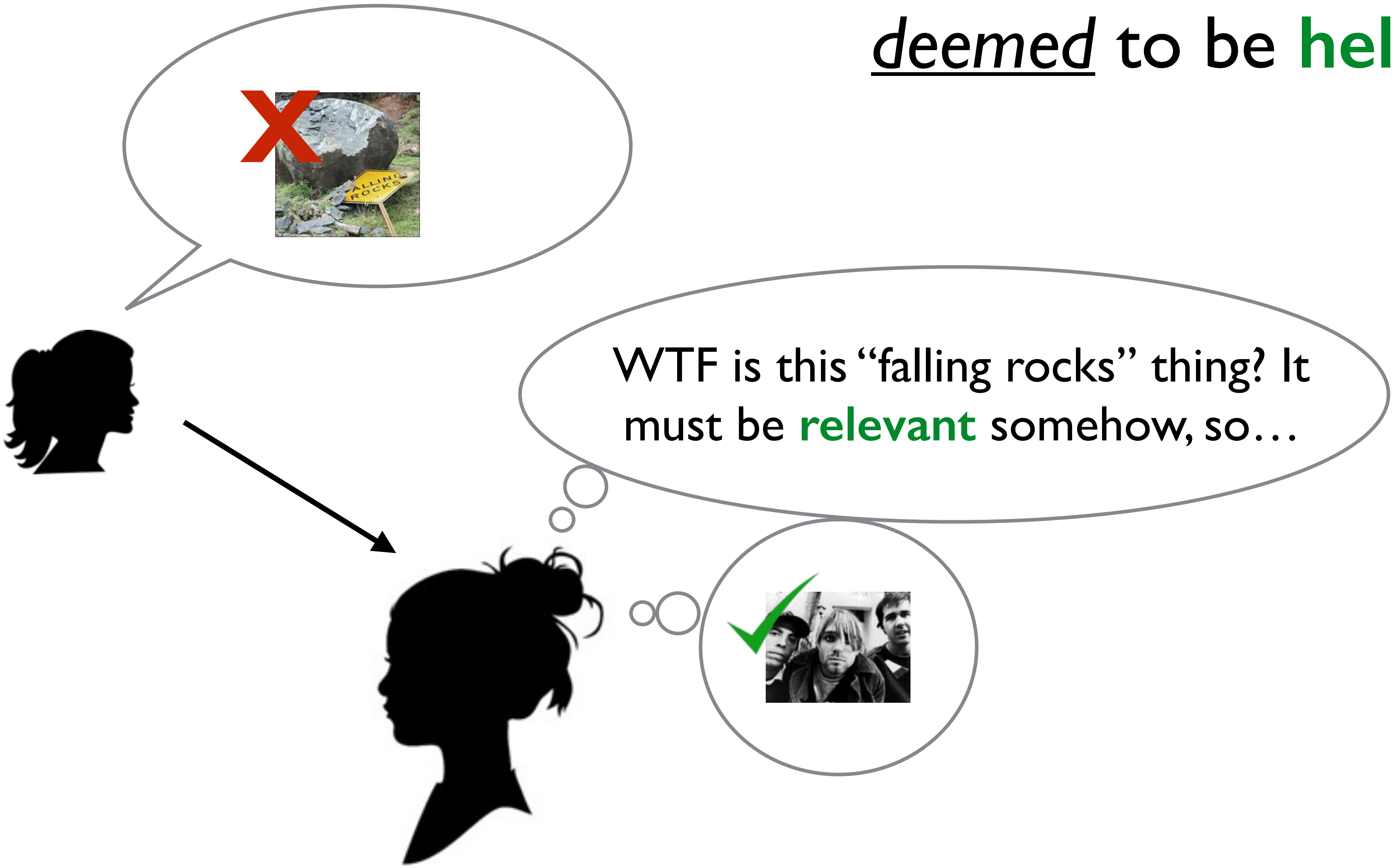
Here's the experimental results:

Hint

Arbitrary



Superficially useless information can have a huge effect when it is *deemed* to be **helpful**



Taking the *wrong* hint when your teacher is a jerk

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

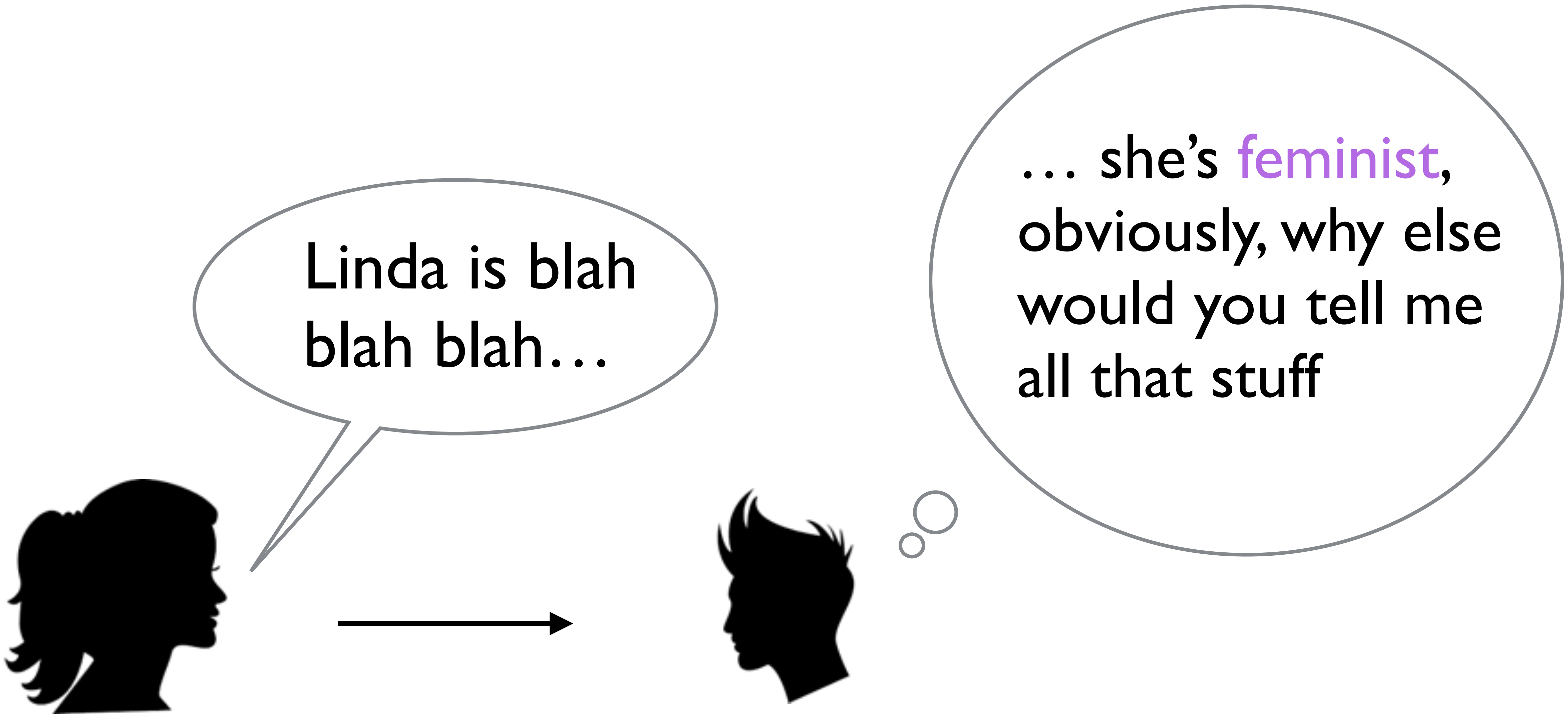
Which is more probable?

(a) Linda is a bank teller

(b) Linda is a feminist bank teller

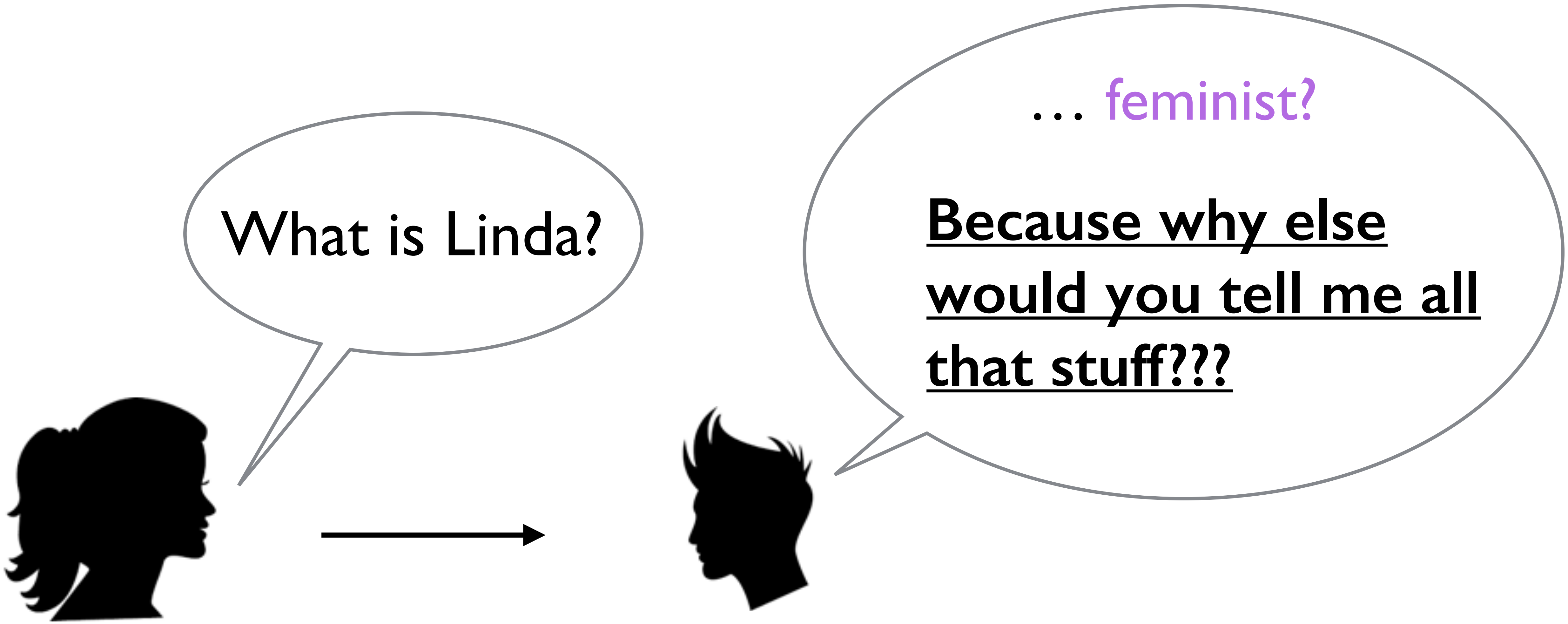
Kahneman & Tversky (1983)

The social/pragmatic account

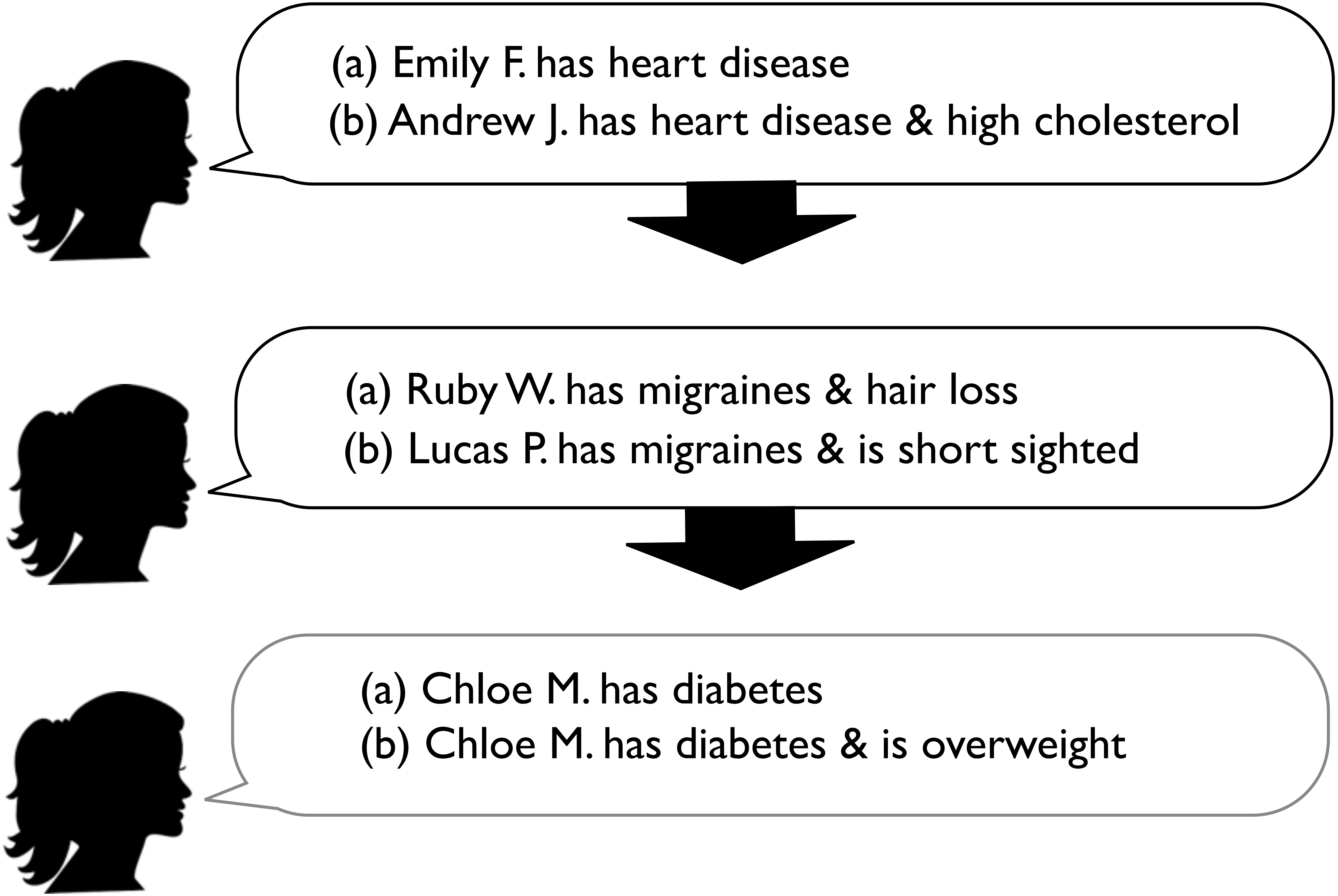


Hertwig & Gigerenzer (1999)

The social/pragmatic account

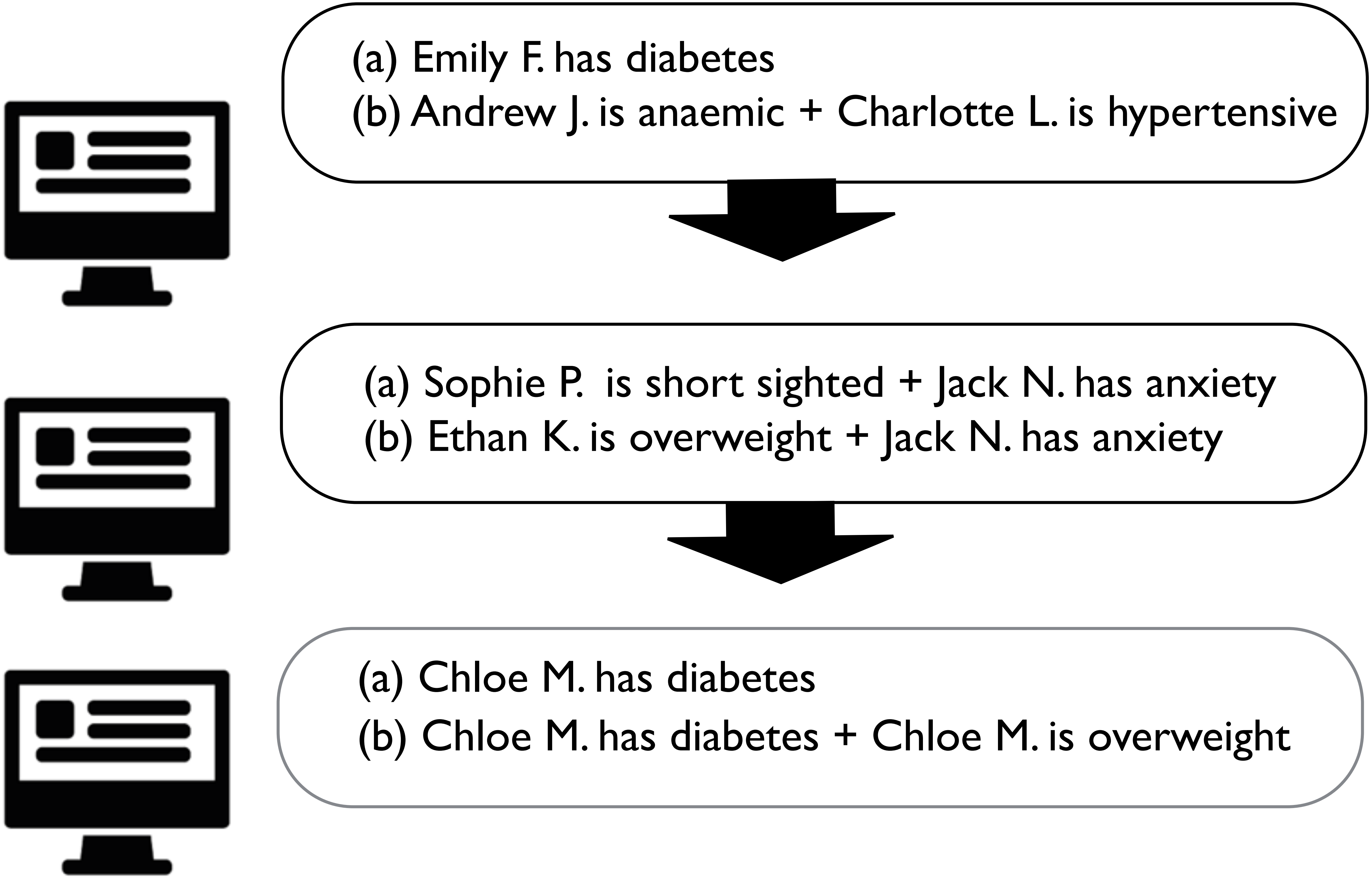


Hertwig & Gigerenzer (1999)



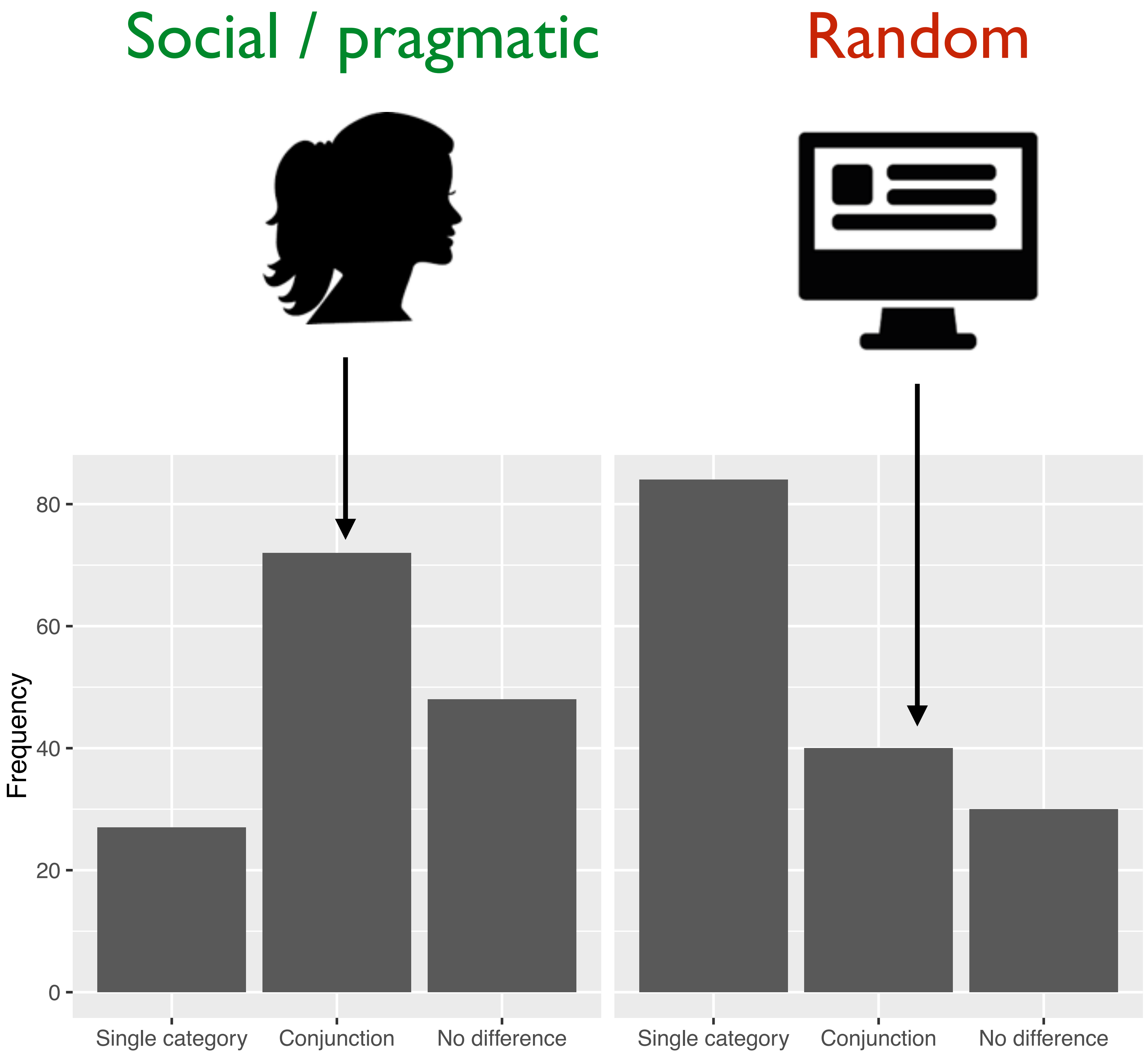
Social / pragmatic context

… with Michelle Keshwa



Random / disconnected fact condition

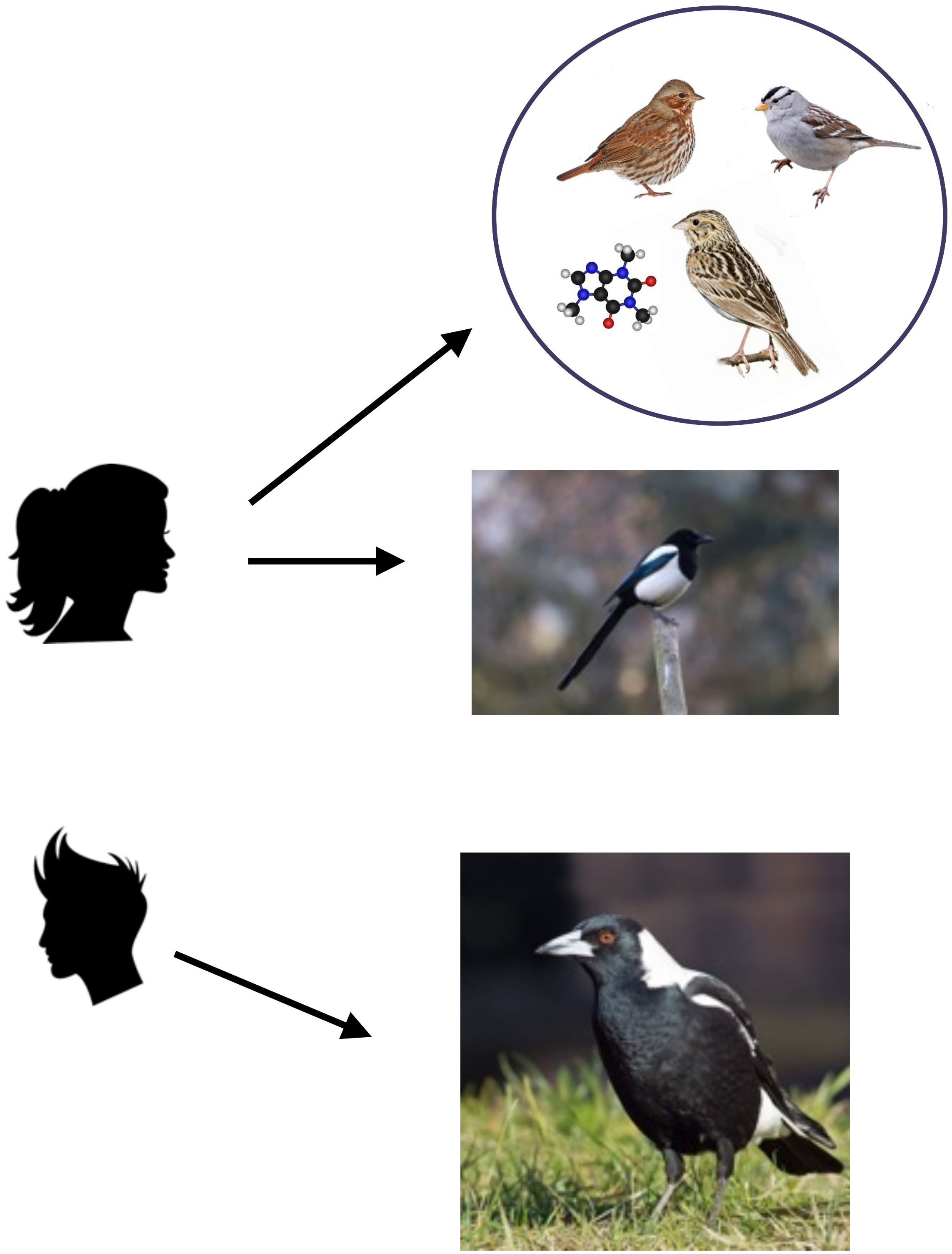
… with Michelle Keshwa



… with Michelle Keshwa

Sampling shapes reasoning even *without* a helpful (or deceitful) human involved

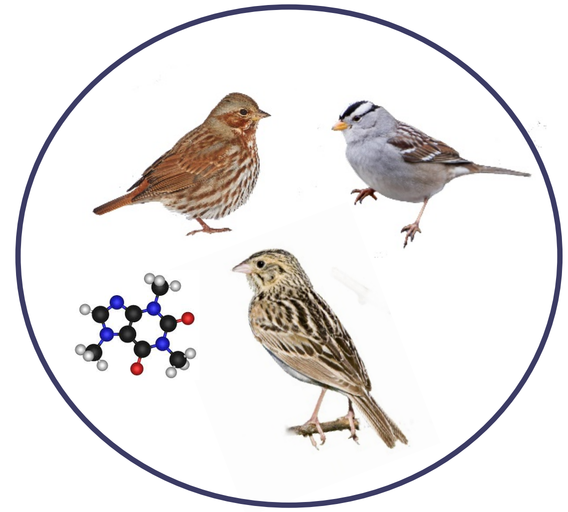
Sampling by different people



This problem can be solved using social cognition

Maybe this is all social reasoning?

Sampling across spatial locations

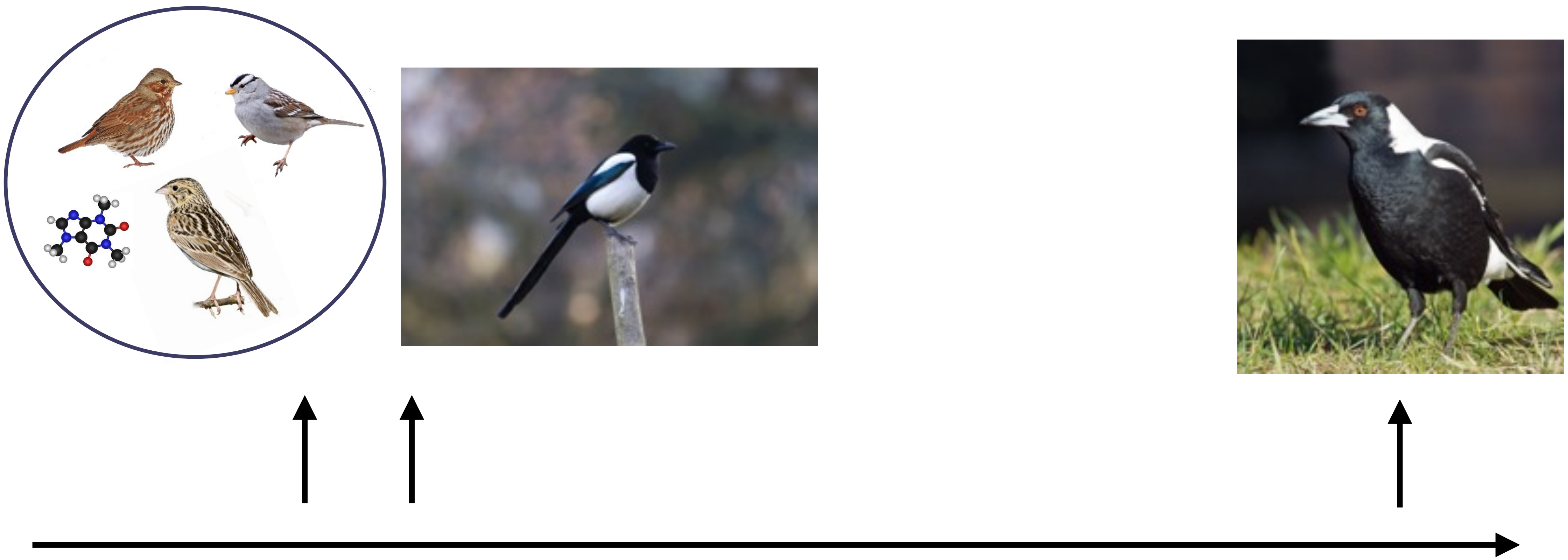


Eurasian magpie

Australian magpie

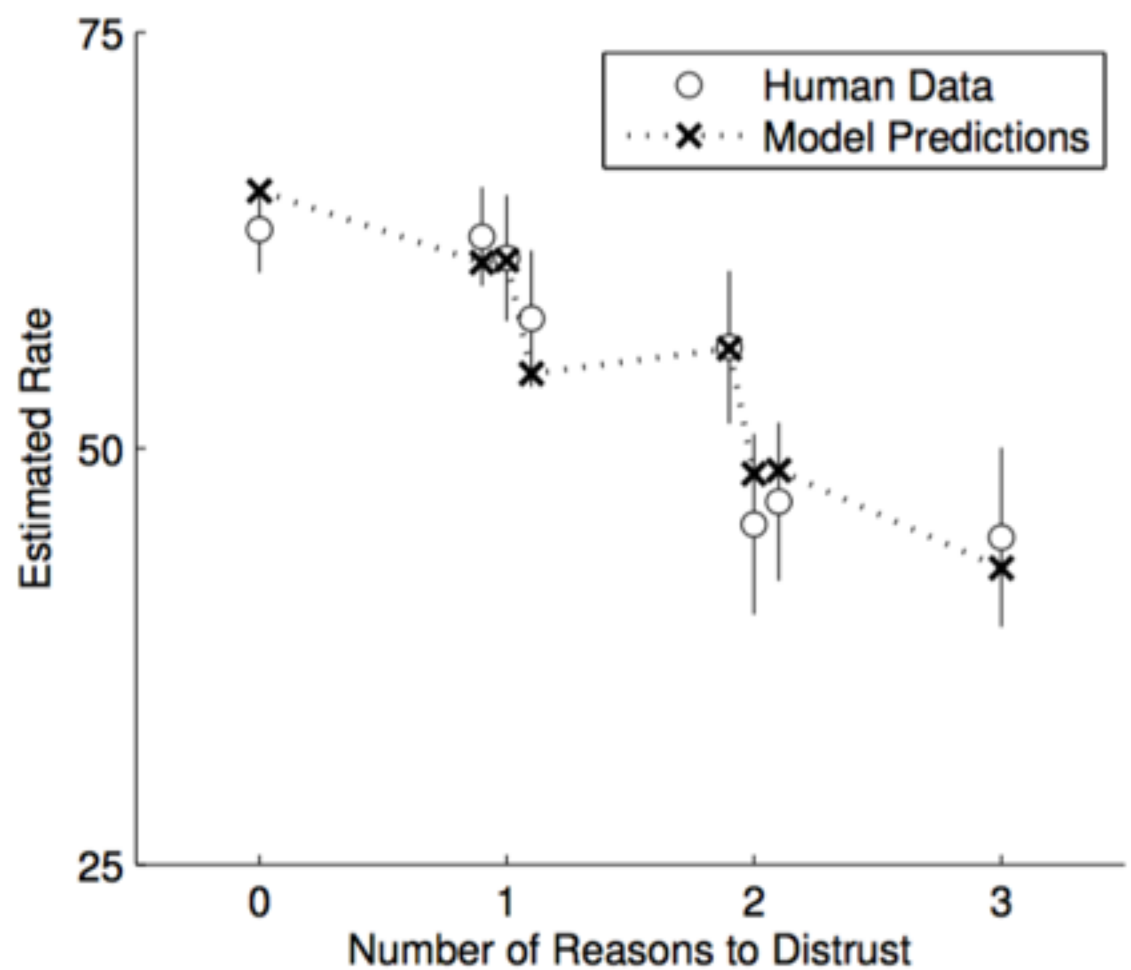
This is not social cognition!

Sampling across time



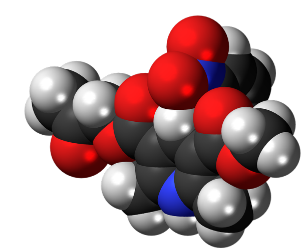
This is not social cognition!

You are currently classifying predators according to whether they pose a threat to humans. *Your team*, working at *this location* *recently* collected 200 observations and found that 50 (25%) of them met this criterion. This week, you have made another 4 observations, of which 3 (75%) met the above criterion. What proportion of predators in the area do you estimate pose a threat to humans?



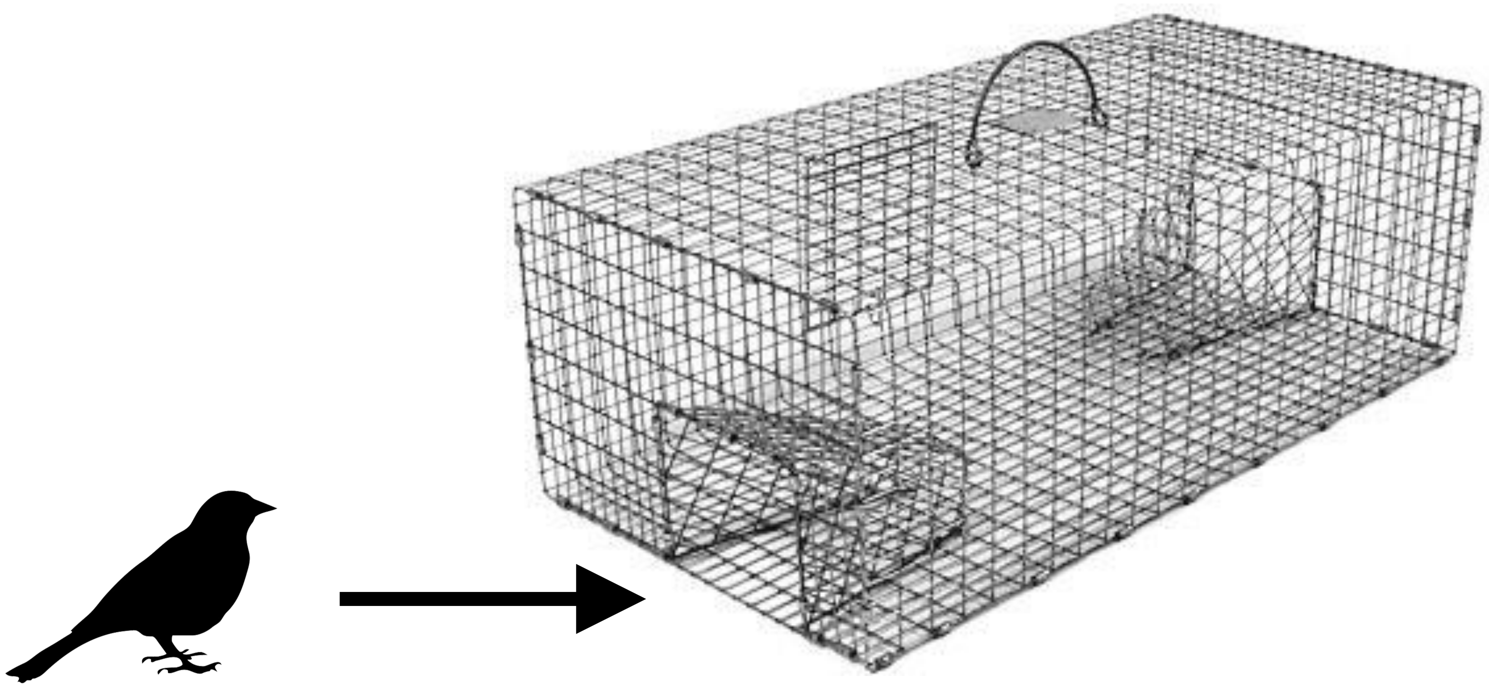
Welsh & Navarro (2012). *Organisational Behavior and Human Decision Processes*

Let’s make this a little more sneaky…



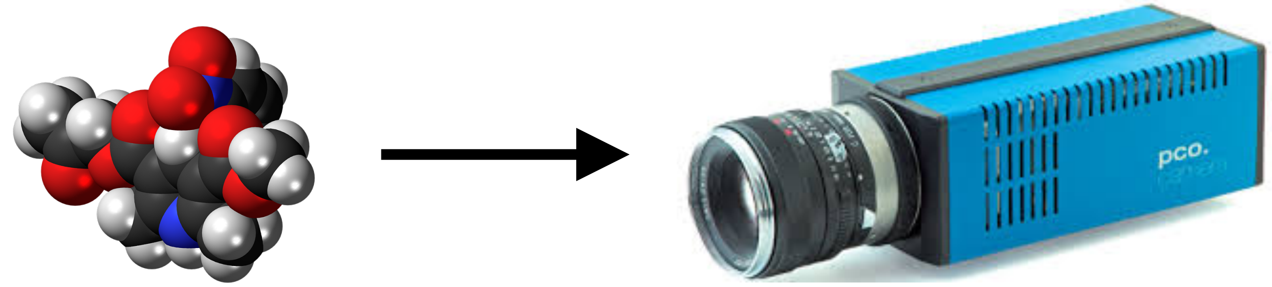
20 *small* birds with *plaxium* blood (SP+)

Hayes, Banner & Navarro (under review)



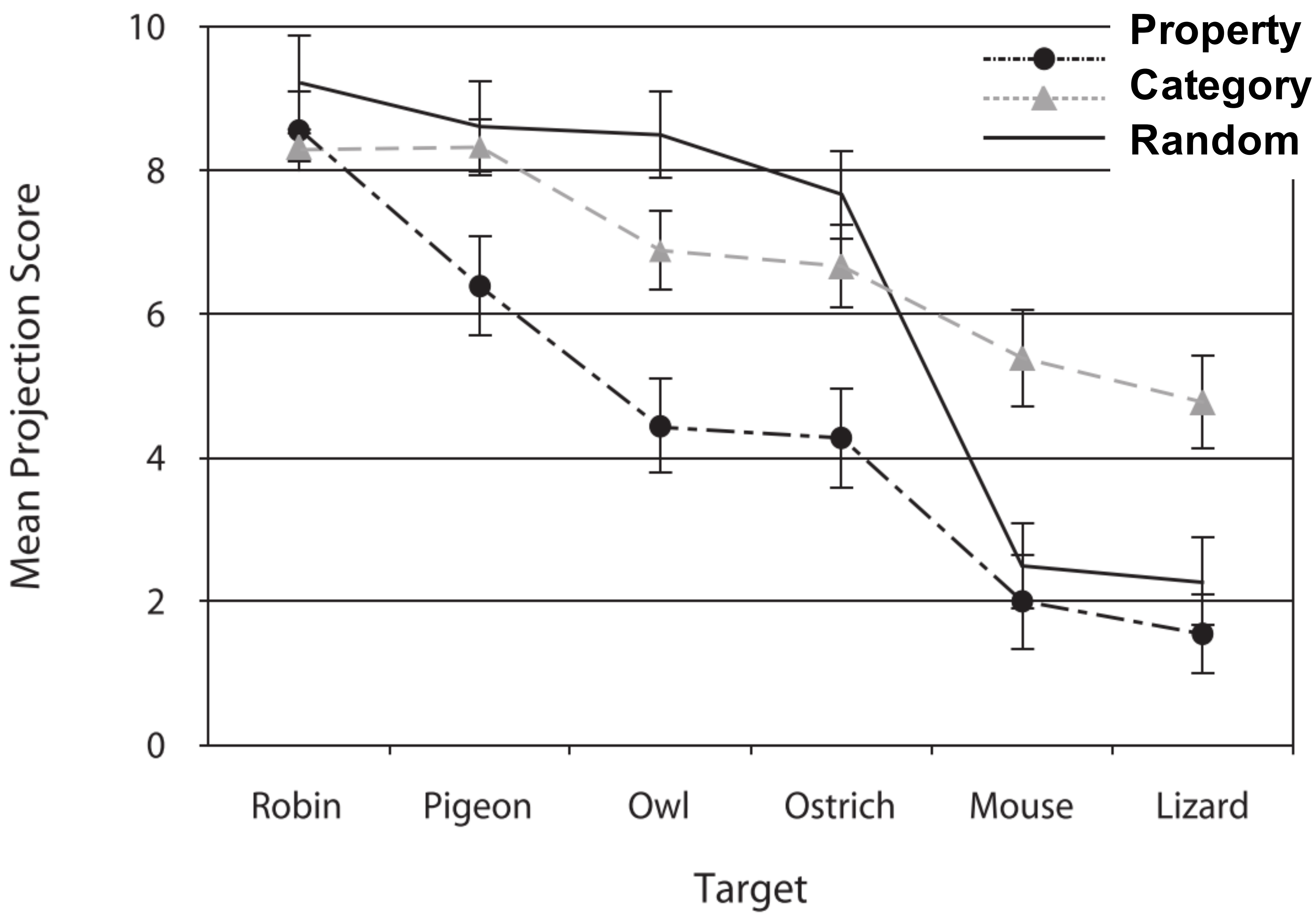
Category sampling: select items based on category membership (i.e. small birds)

Hayes, Banner & Navarro (under review)



Property sampling: select items based on possession of the property (i.e. plaxium blood)

Hayes, Banner & Navarro (under review)

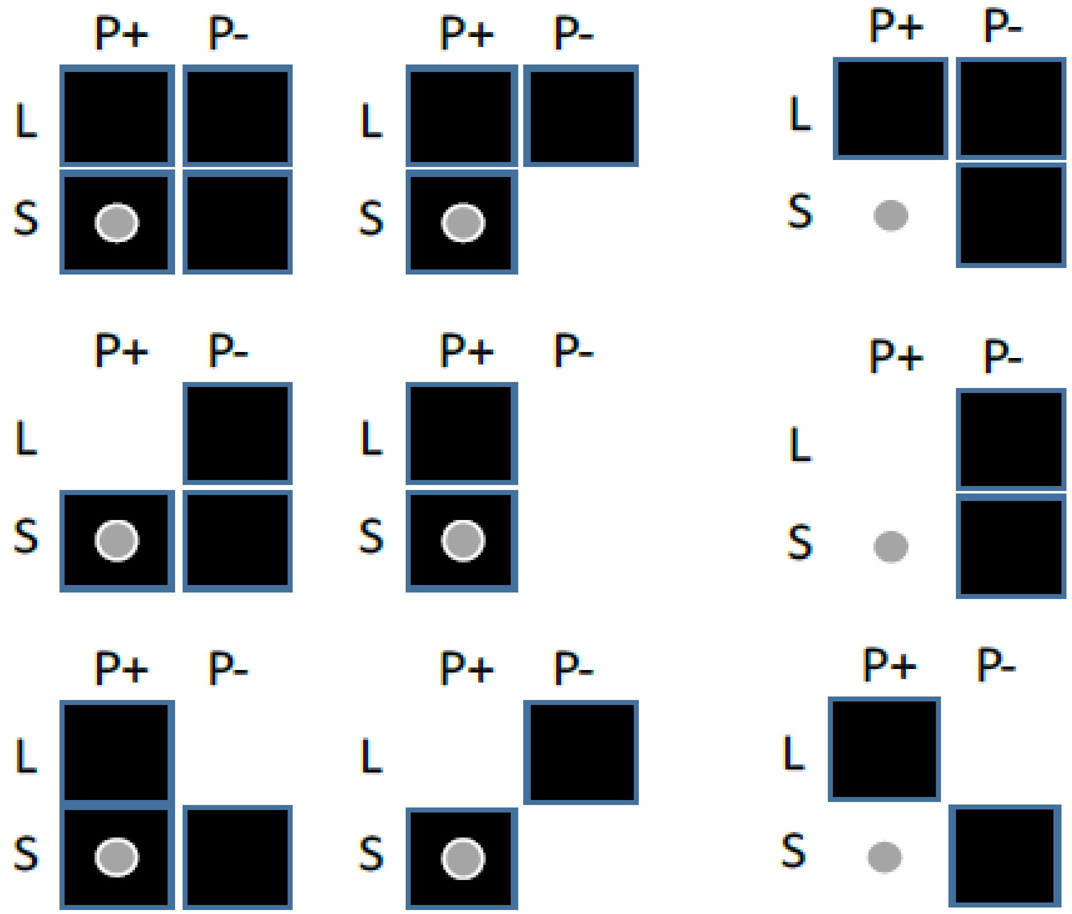


category

property

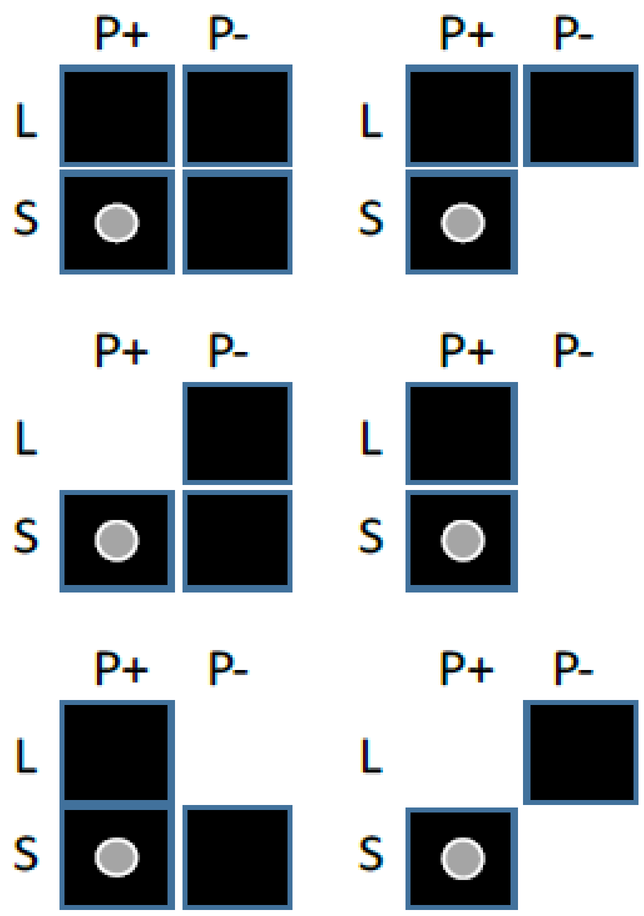
Lawson & Kalish (2009)

Hypotheses a reasoner might consider



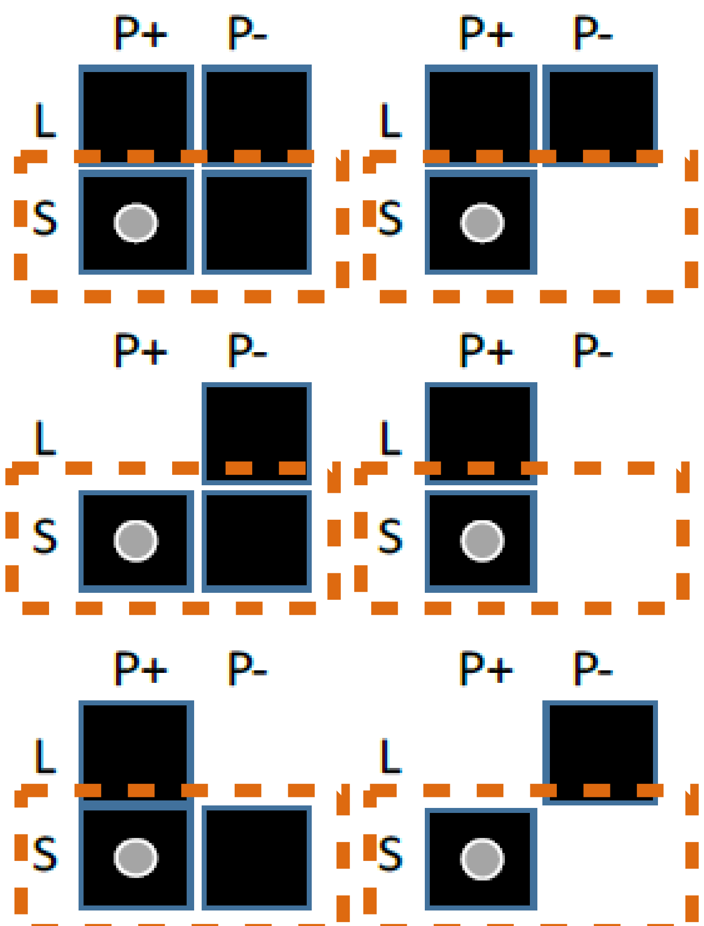
Hayes, Banner & Navarro (under review)

Hypotheses consistent with the data



Hayes, Banner & Navarro (under review)

Category sampling

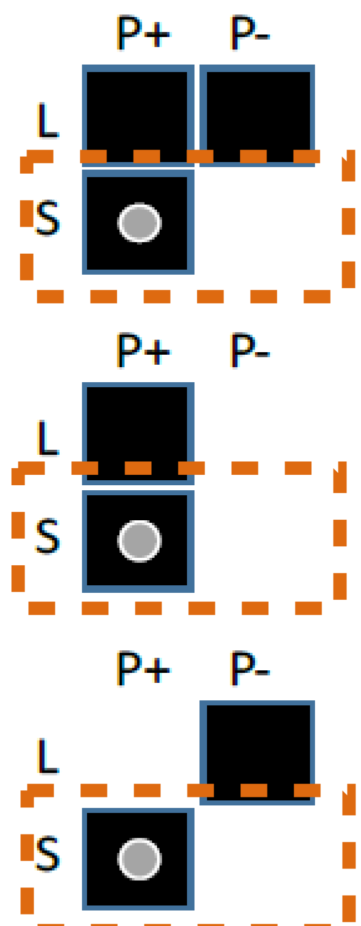


Frame explains absence of LP+ and LP-

Hypothesis must account for absence of SP-

Hayes, Banner & Navarro (under review)

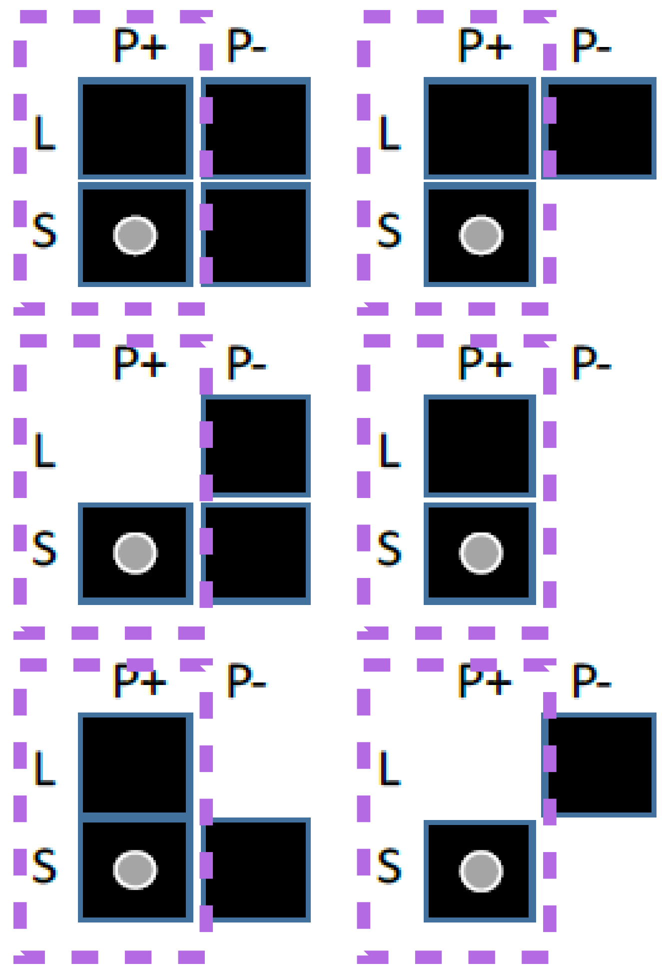
Category sampling



2 of 3 hypotheses allow LP+
… so generalisation to large birds is *very plausible*

Hayes, Banner & Navarro (under review)

Property sampling

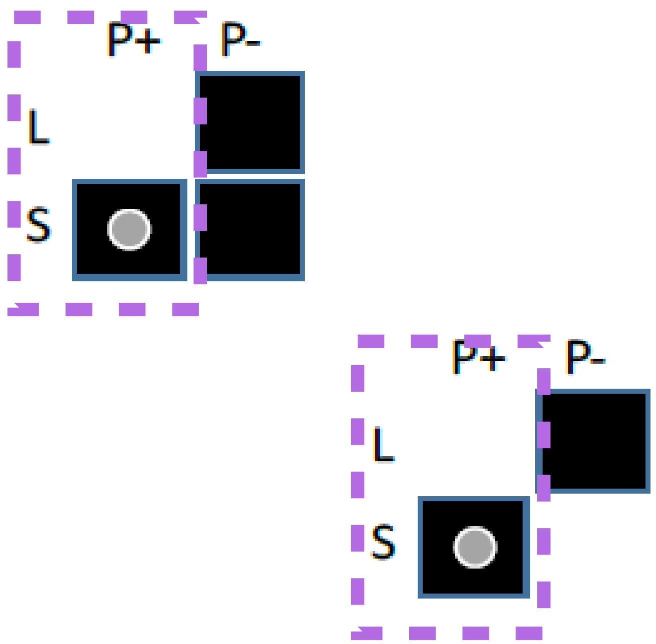


Frame explains absence of SP- and LP-

Hypothesis must account for absence of LP+

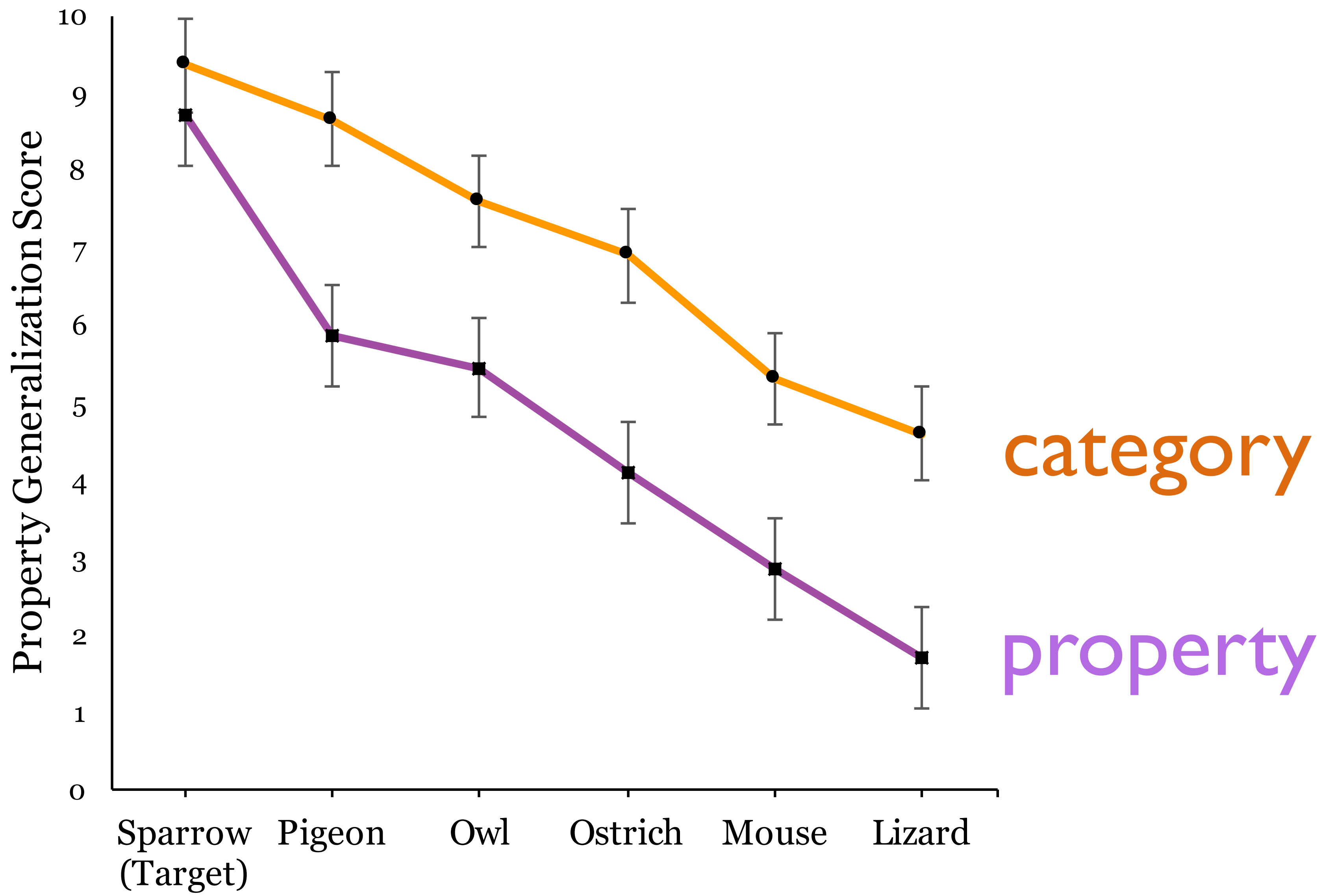
Hayes, Banner & Navarro (under review)

Property sampling



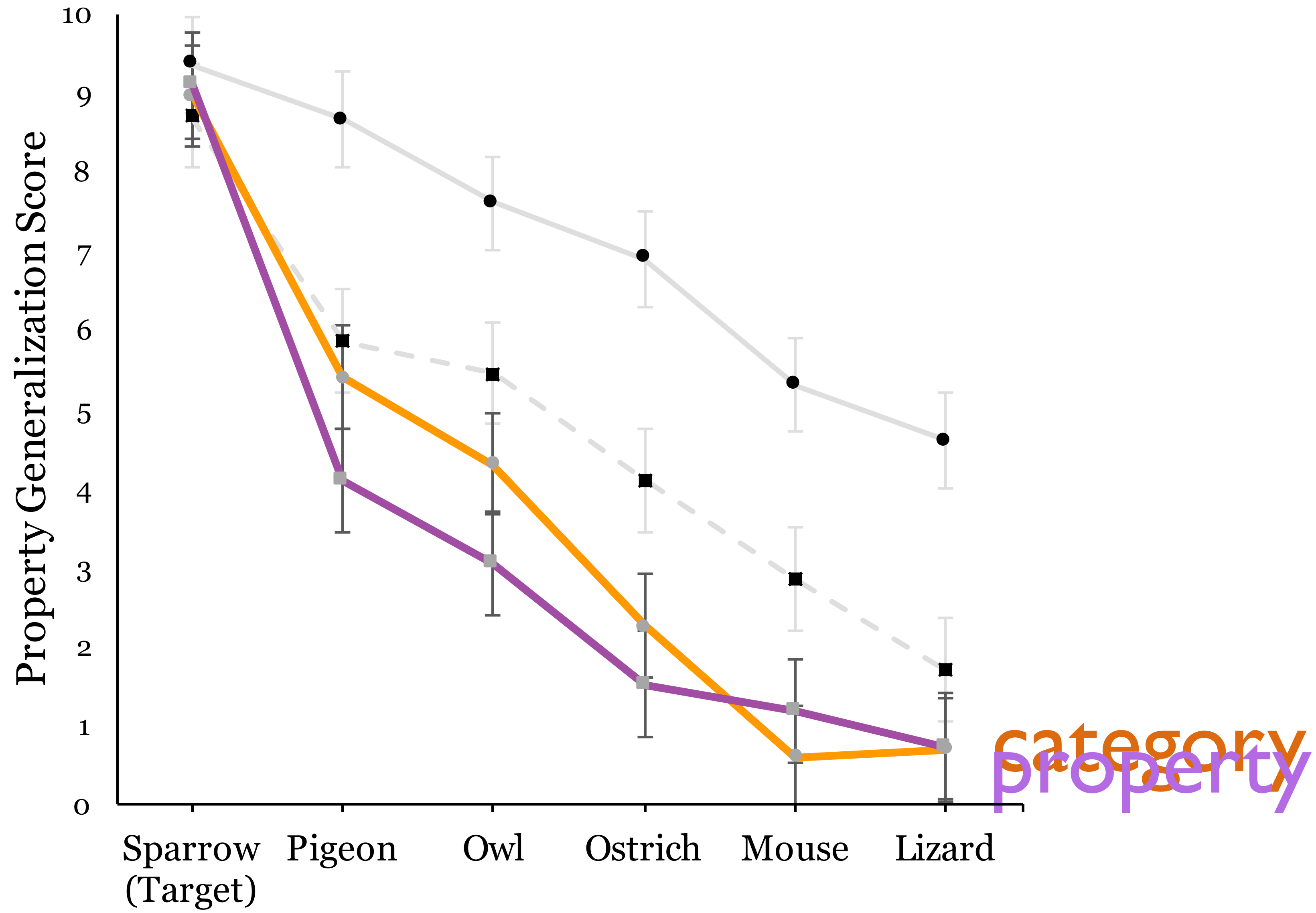
No remaining hypotheses allow LP+… so generalisation to large birds is *very implausible*

Replication of L&K 2009



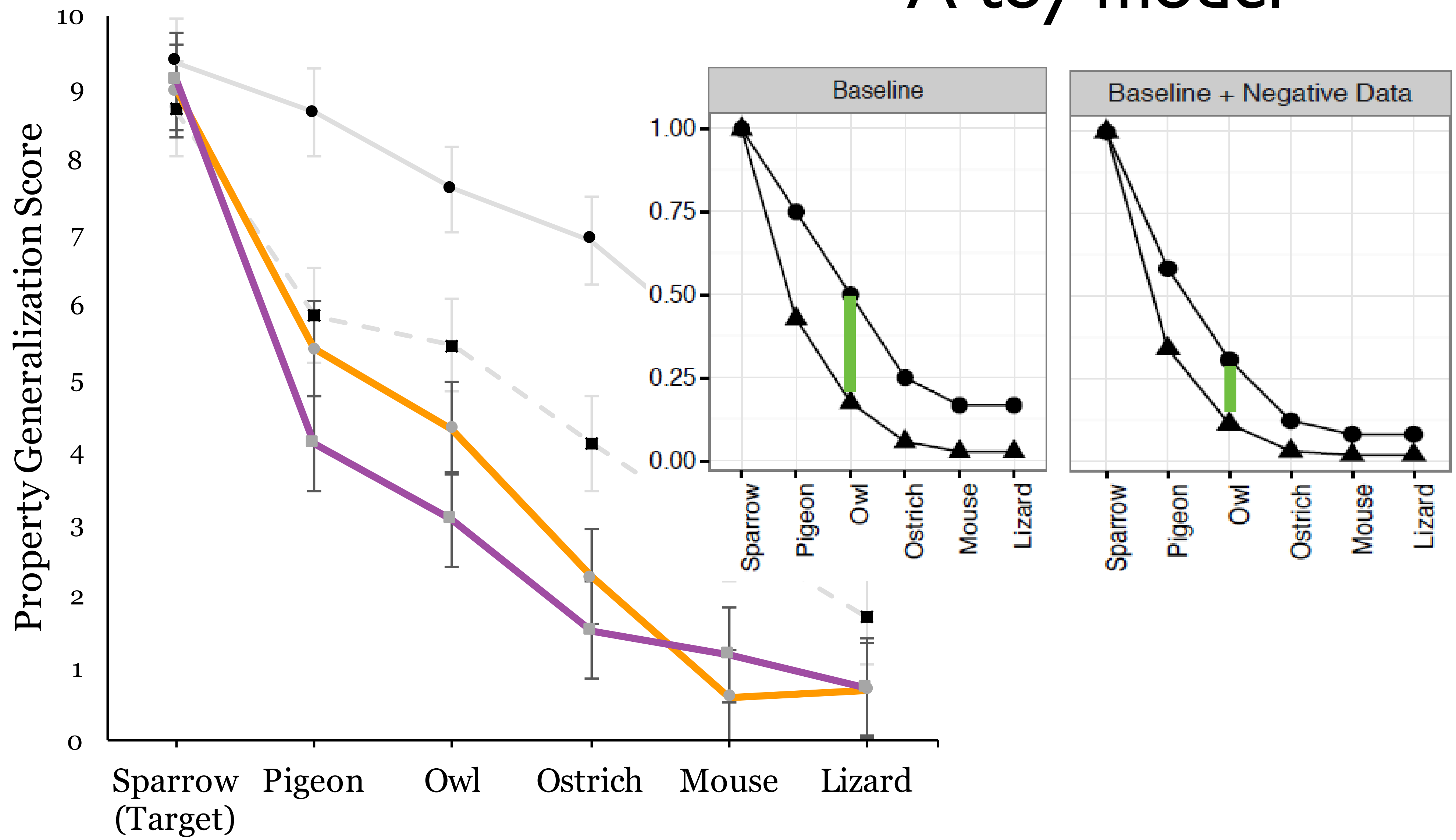
Hayes, Banner & Navarro (under review)

Explicit negative evidence (actual LP-) attenuates value of *implicit* negative evidence (no LP+)



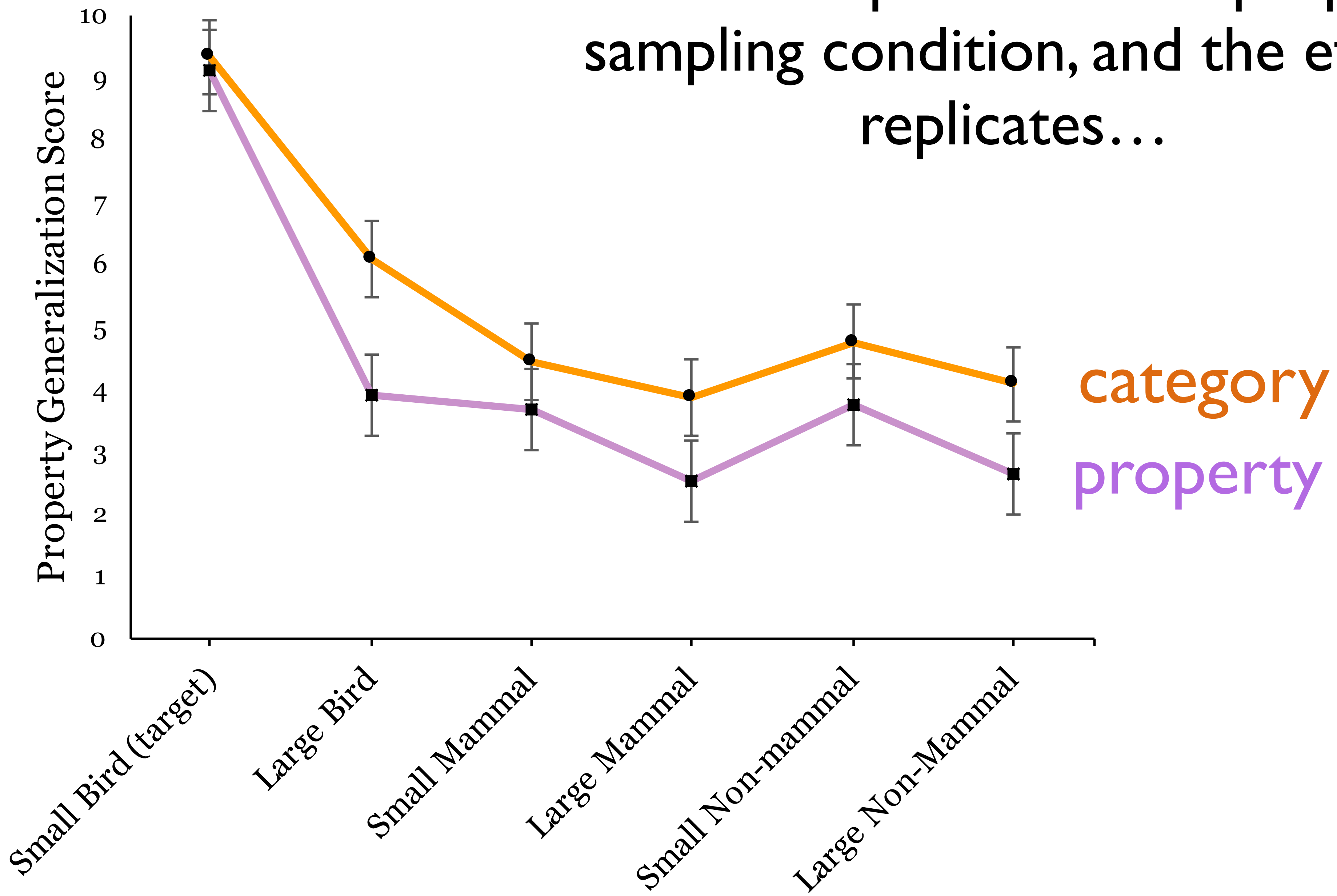
Hayes, Banner & Navarro (under review)

A toy model



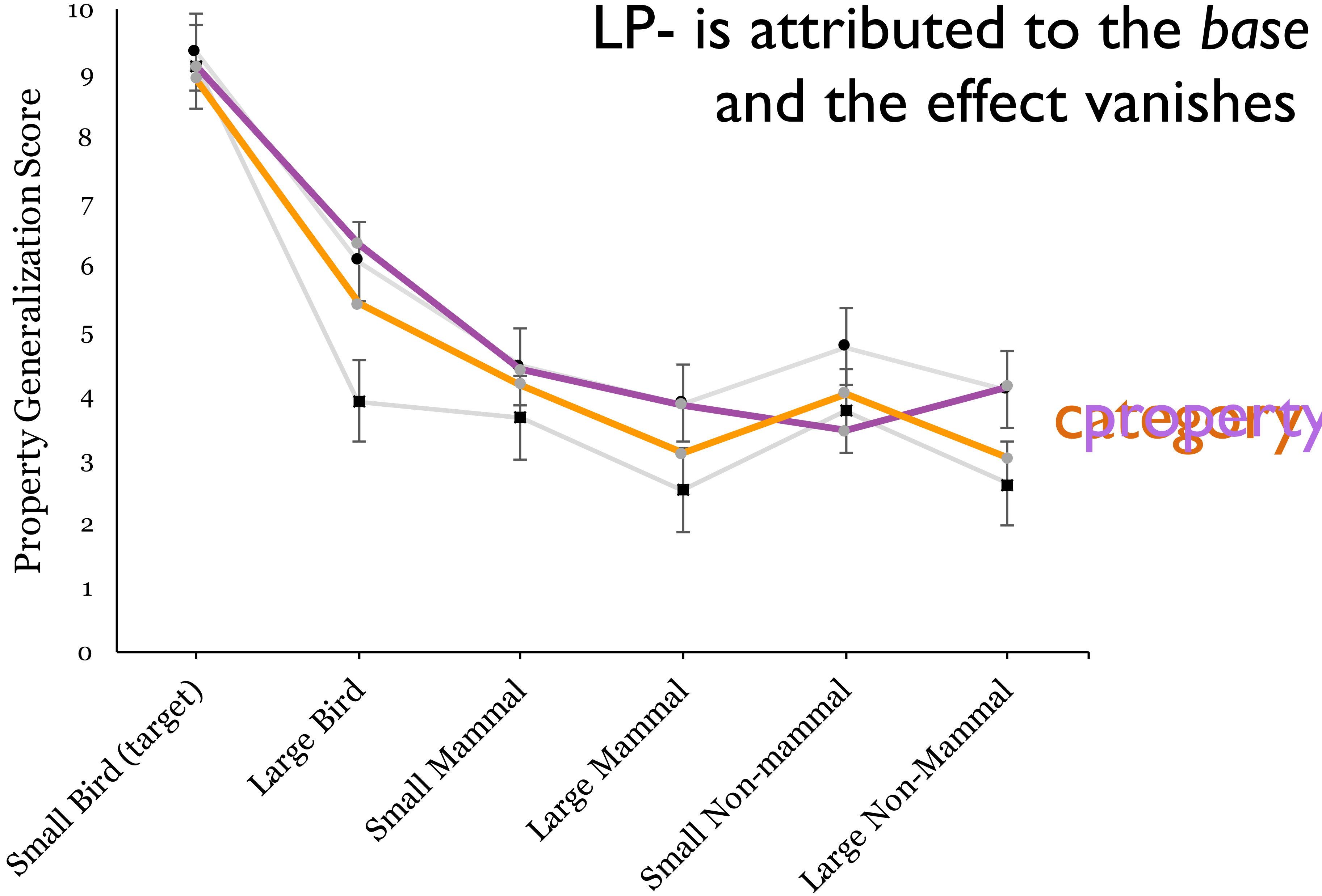
Hayes, Banner & Navarro (under review)

If we tell people large birds are common, then the absence of LP+ remains suspicious in the property sampling condition, and the effect replicates…



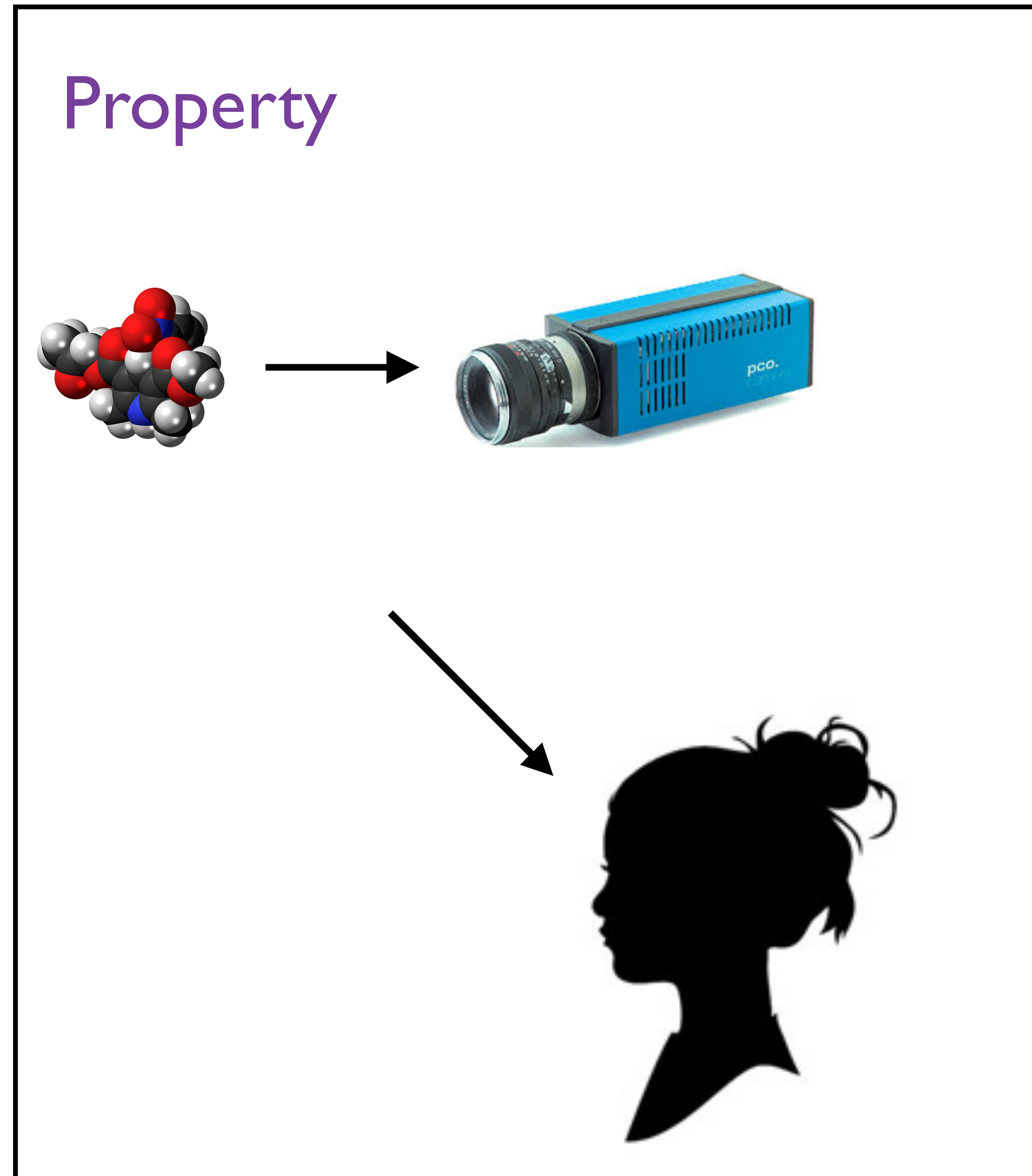
Hayes, Banner & Navarro (under review)

But if we tell people large birds are rare, then the absence of LP+ and LP- is attributed to the *base rate* and the effect vanishes



Hayes, Banner & Navarro (under review)

People pay attention to *mechanistic* constraints on sampling processes (not just social cues), and this shapes our reasoning in a sensible way



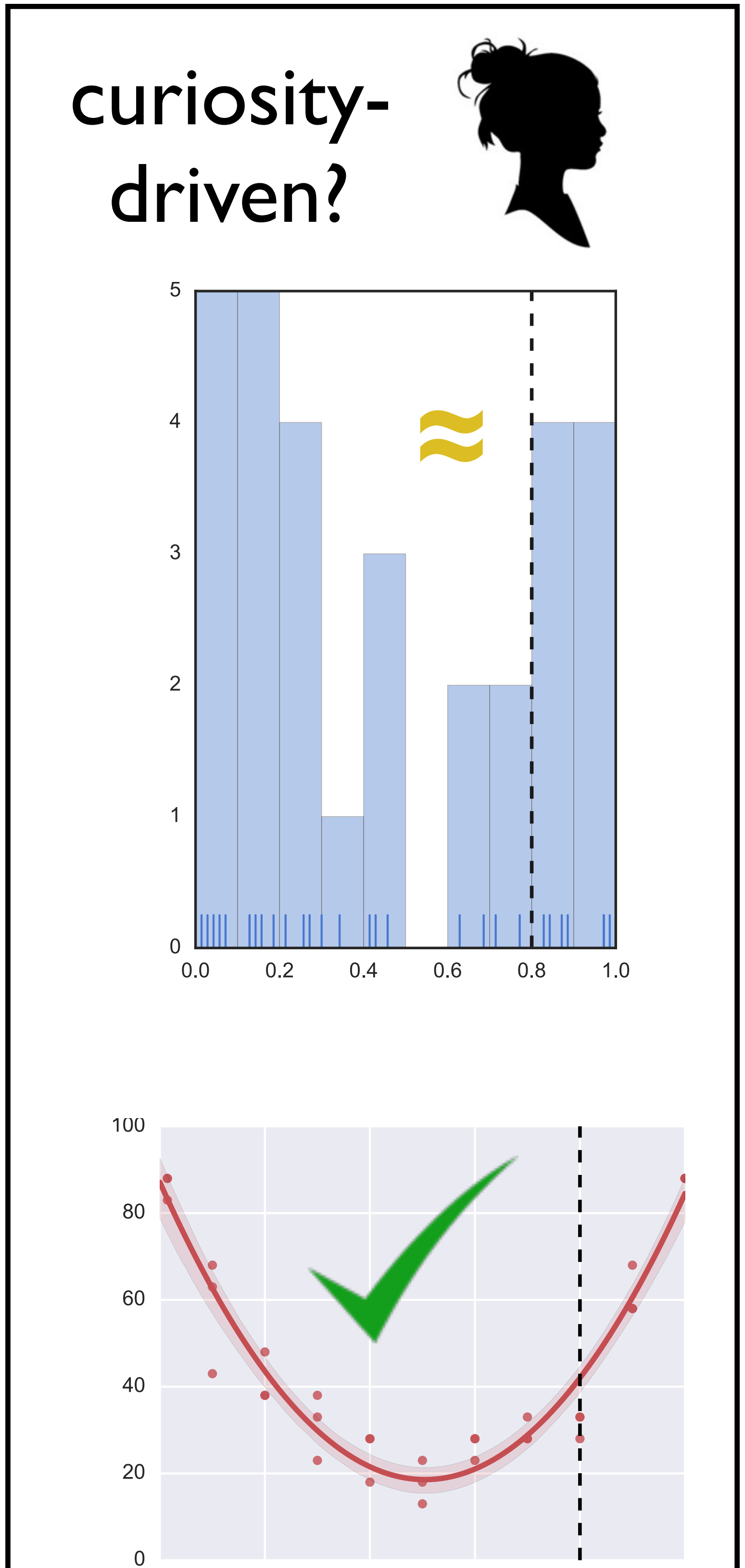
Hayes, Banner & Navarro (under review)

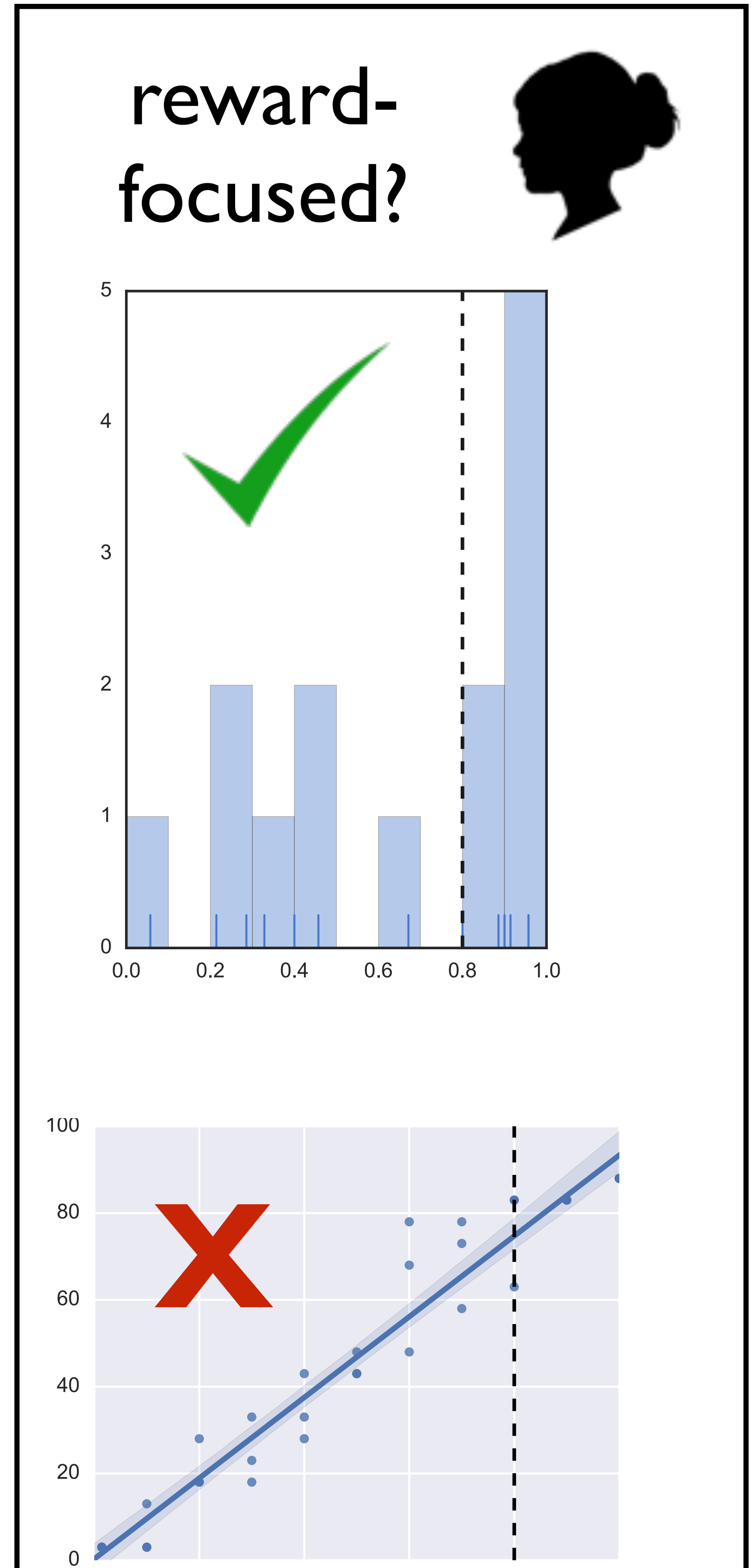
Extensions?

Choice: What drives people's active sampling?

instrumental learning task

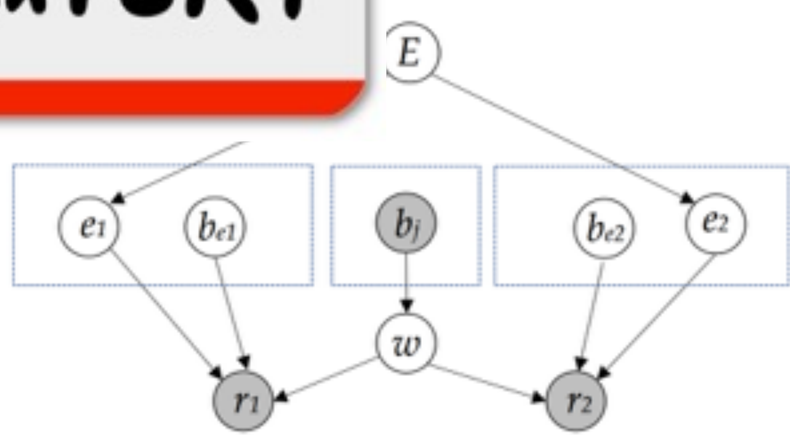
transfer task





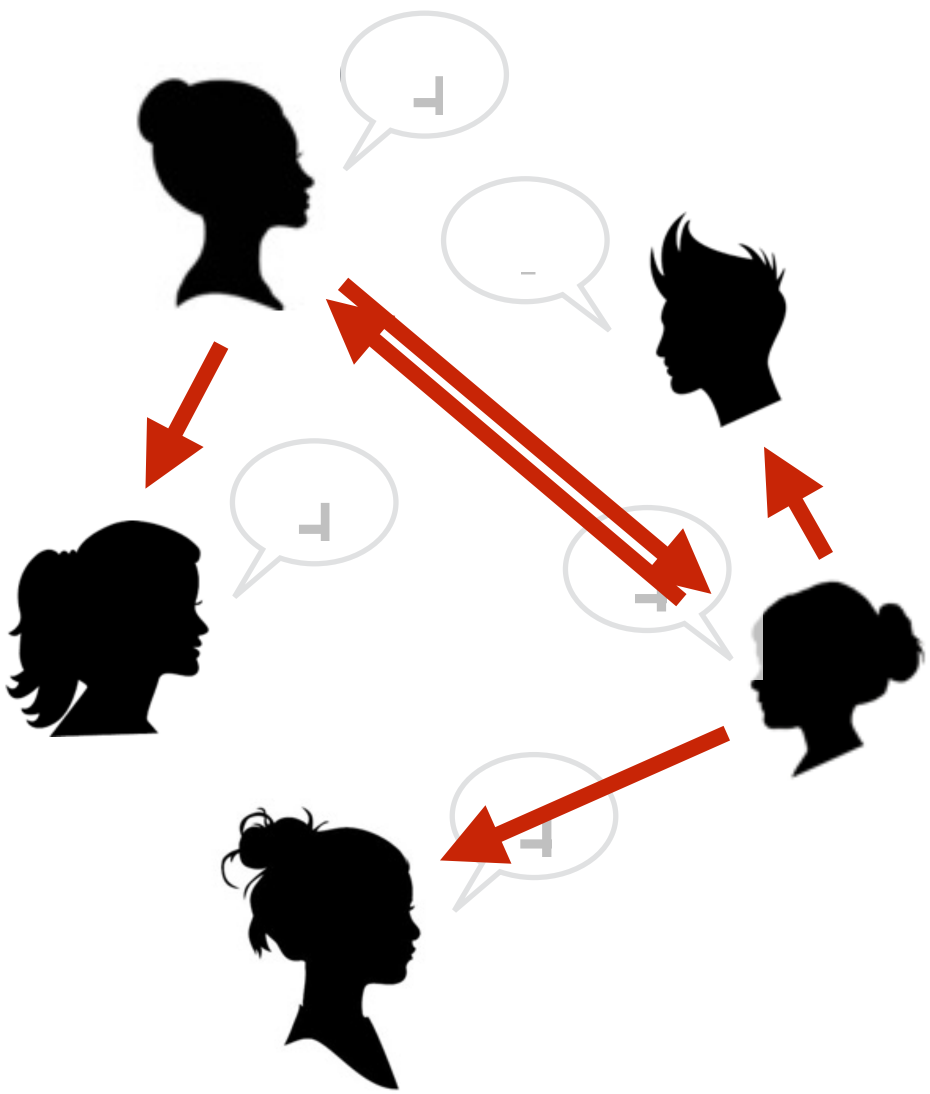
with Sean Tauber and Ben Newell

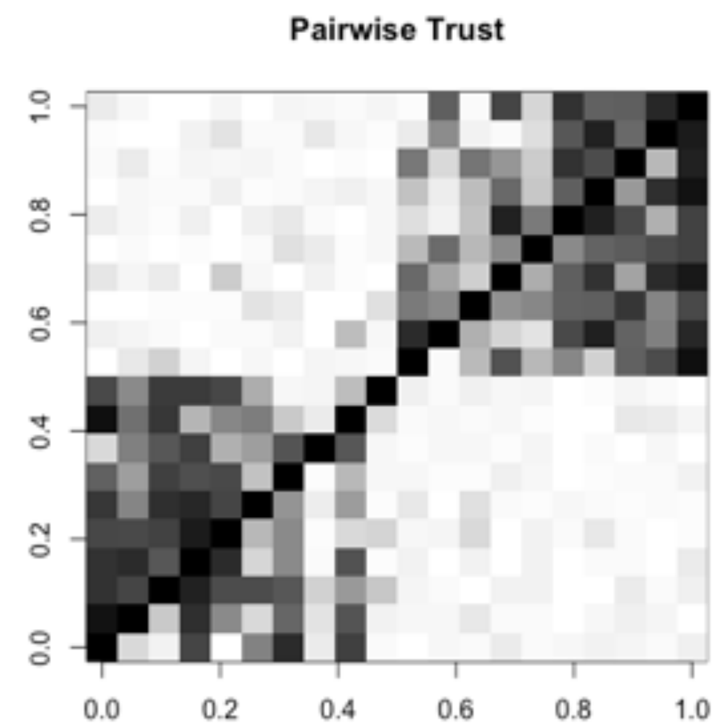
Law: Evidence sampling and expertise in the courtroom



with Kristy Martire and Gary Edmond

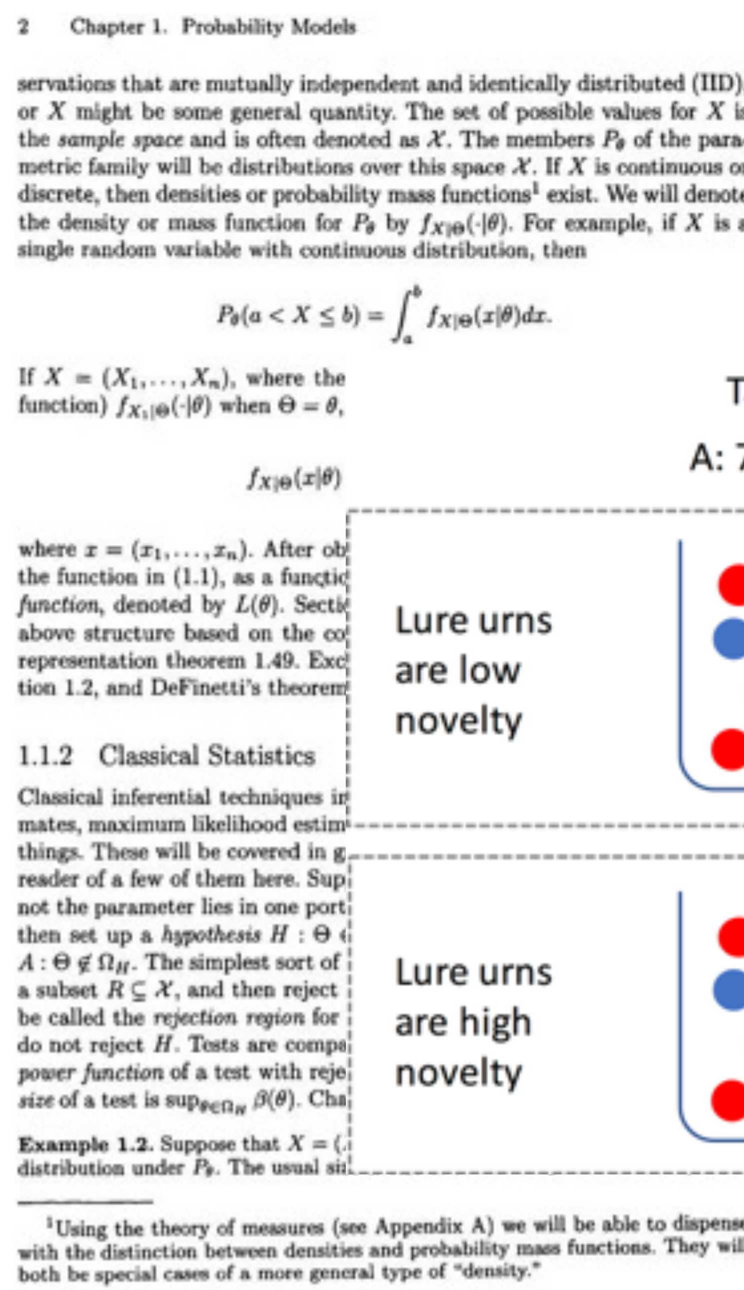
Society: Trust-based sampling via self-organising social networks (fake news…)





with Amy Perfors

Development: Exploratory versus goal-directed sampling by preschoolers



2 Chapter 1. Probability Models

servations that are mutually independent and identically distributed (IID), or X might be some general quantity. The set of possible values for X is the *sample space* and is often denoted as $\mathcal{X}$. The members P_θ of the parametric family will be distributions over this space $\mathcal{X}$. If X is continuous or discrete, then densities or probability mass functions[1] exist. We will denote the density or mass function for P_θ by $f_{X|\Theta}(\cdot|\theta)$. For example, if X is a single random variable with continuous distribution, then

$$P_\theta(a < X \leq b) = \int_a^b f_{X|\Theta}(x|\theta)dx.$$

If $X = (X_1, \dots, X_n)$, where the
function) $f_{X_1|\Theta}(\cdot|\theta)$ when $\Theta = \theta$,

$$f_{X|\Theta}(x|\theta)$$

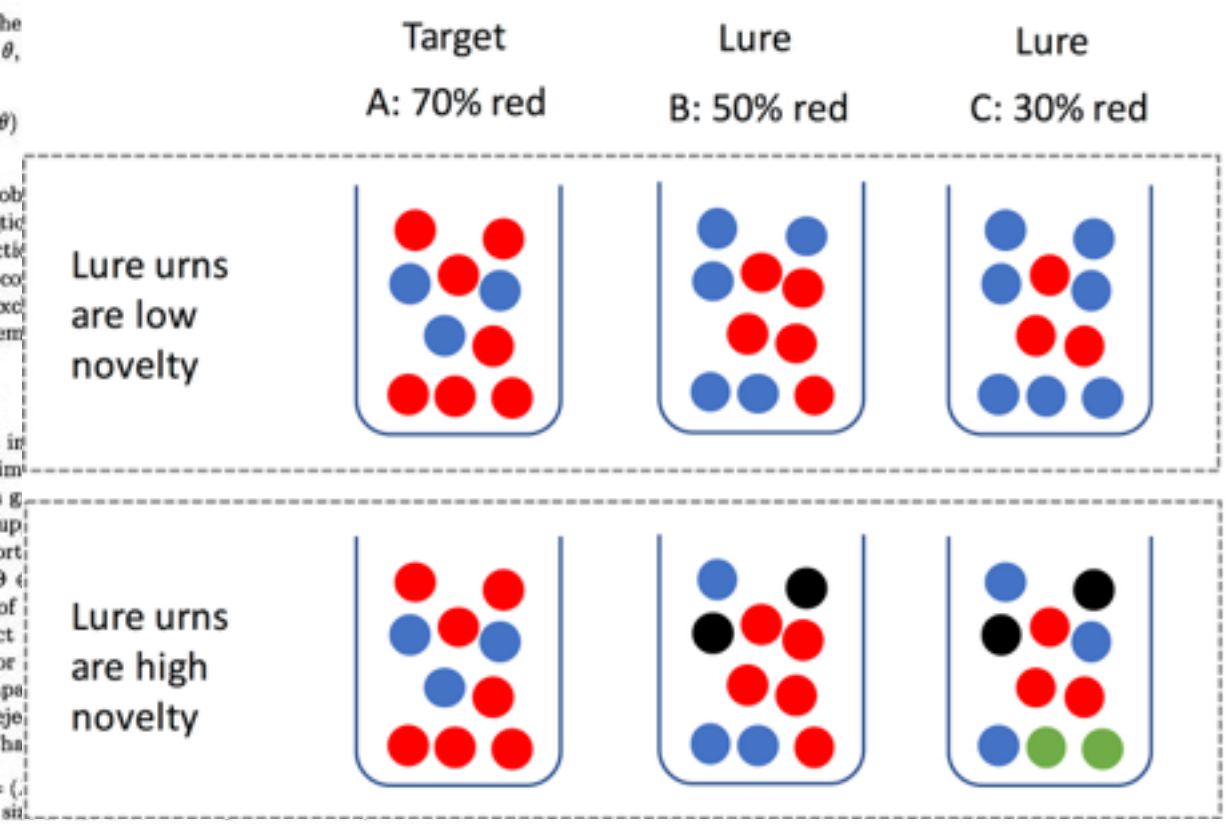
where $x = (x_1, \dots, x_n)$. After ob
the function in (1.1), as a functio
function, denoted by $L(\theta)$. Secti
above structure based on the co
representation theorem 1.49. Exc
tion 1.2, and DeFinetti's theorem

1.1.2 Classical Statistics

Classical inferential techniques in
mates, maximum likelihood estim
things. These will be covered in g
reader of a few of them here. Sup
not the parameter lies in one port
then set up a *hypothesis* $H : \Theta$
$A : \Theta \notin \Omega_H$. The simplest sort of
a subset $R \subseteq \mathcal{X}$, and then reject
be called the *rejection region* for
do not reject H. Tests are compa
power function of a test with reje
size of a test is $\sup_{\theta \in \Omega_H} \beta(\theta)$. Cha

Example 1.2. Suppose that $X = ($
distribution under P_θ. The usual si

[1] Using the theory of measures (see Appendix A) we will be able to dispense with the distinction between densities and probability mass functions. They will both be special cases of a more general type of "density."



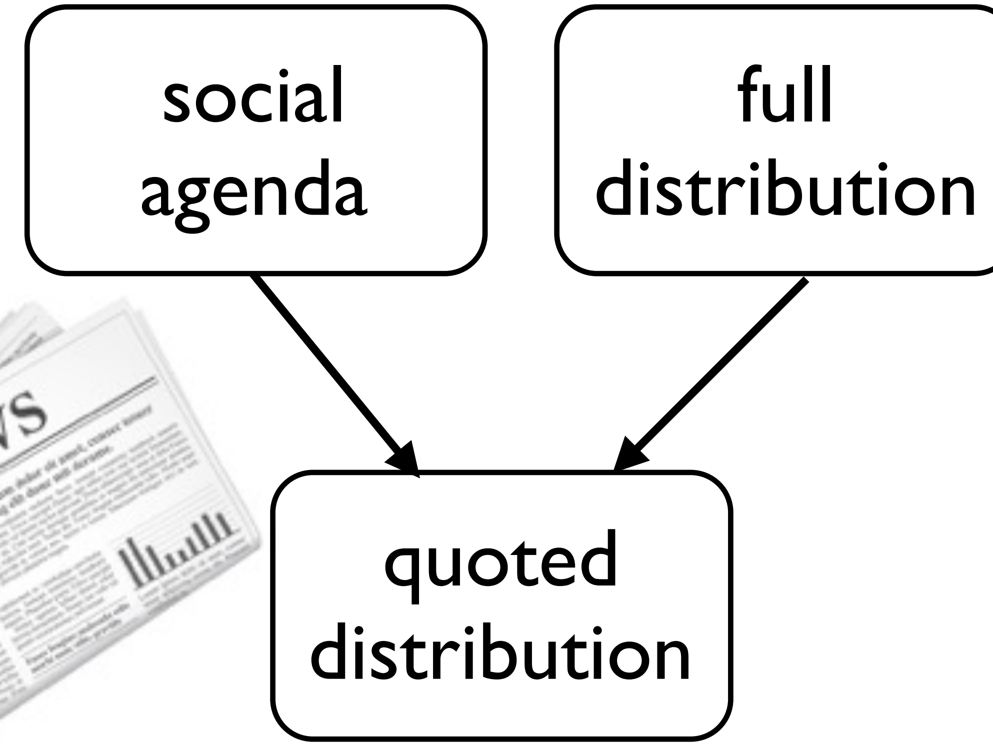
with Candy Liu, Jenny Richmond & Amy Perfors

Wrap-up:

On the origins of data and the rationality of human reasoning

People are smart. Limited, but smart.

"*Common sense*" reasoning is infuriatingly cunning, and requires people to learn from complex data sources (e.g., other people)



We need to disentangle facts from agendas

with Amy Perfors and Pat Shafto

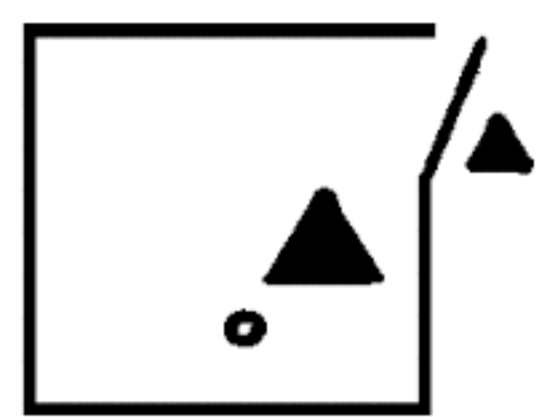
We need to detect trickery

too many collaborators to list

We need to know when to reject the rules we're given

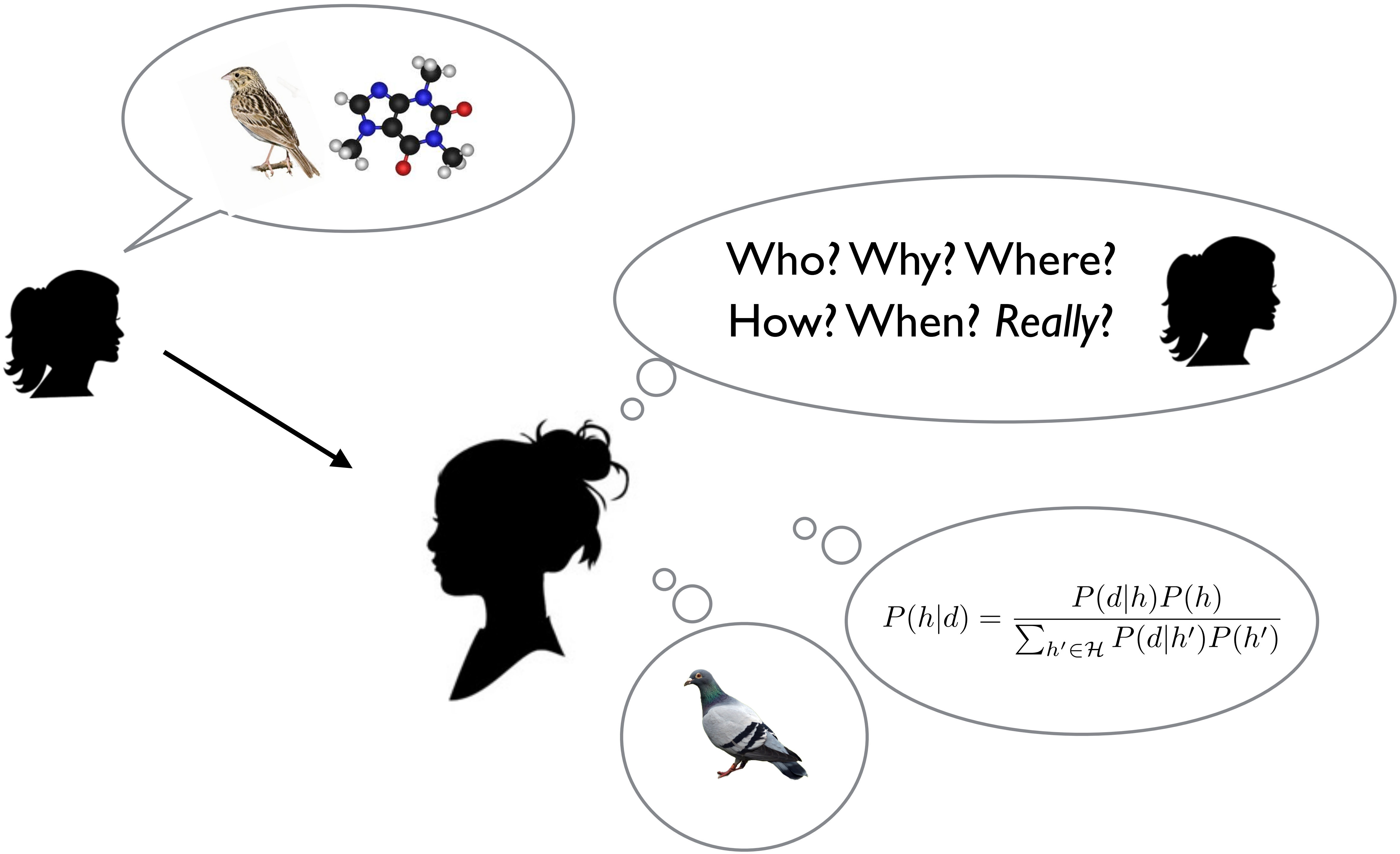
with Charles Kemp

We need to read the intention of potentially malicious agents



too many collaborators to list

Common sense reasoning requires uncommonly rich statistical models



Thanks!