What can statistical theory tell us about human cognition?

Dan Navarro
University of Adelaide

COMPUTATIONAL
COGNITIVE
SCIENCE
LABORATORY

THE UNIVERSITY of ADELAIDE

Wouter Voorspoels

Sean Tauber

Amy Perfors

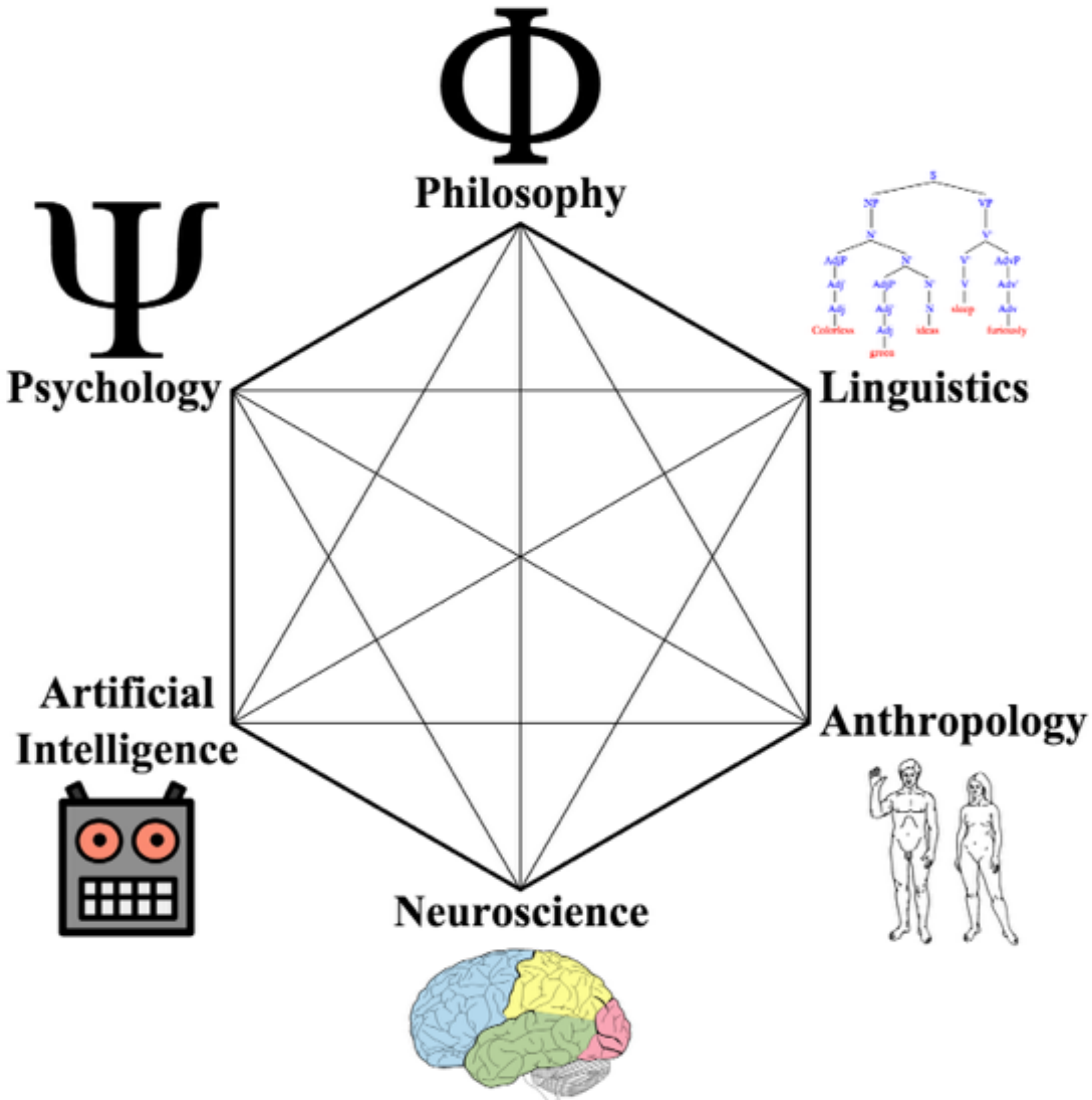
Drew Hendrickson

Simon De Deyne

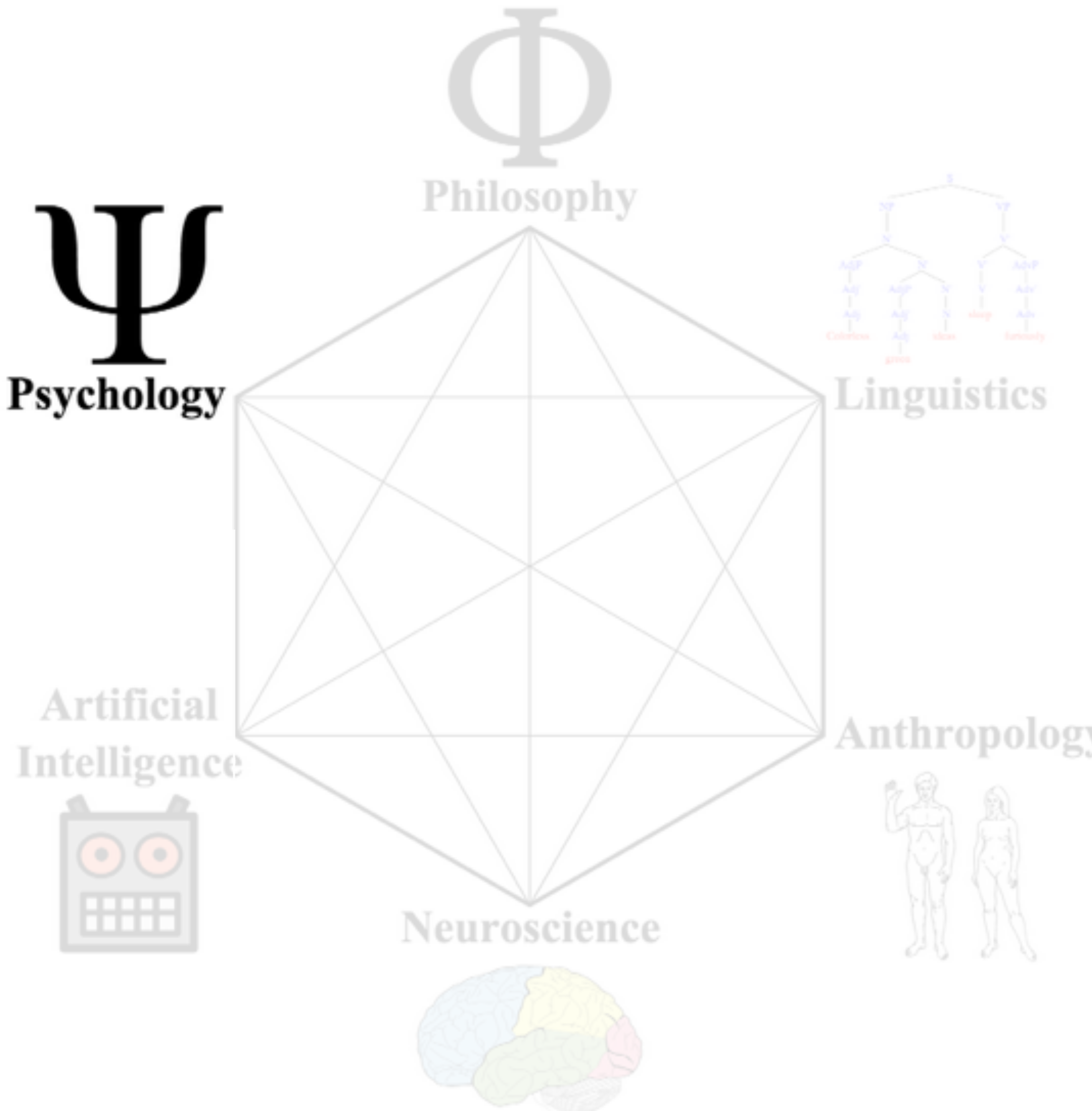
Wai Keen Vong

Keith Ransom

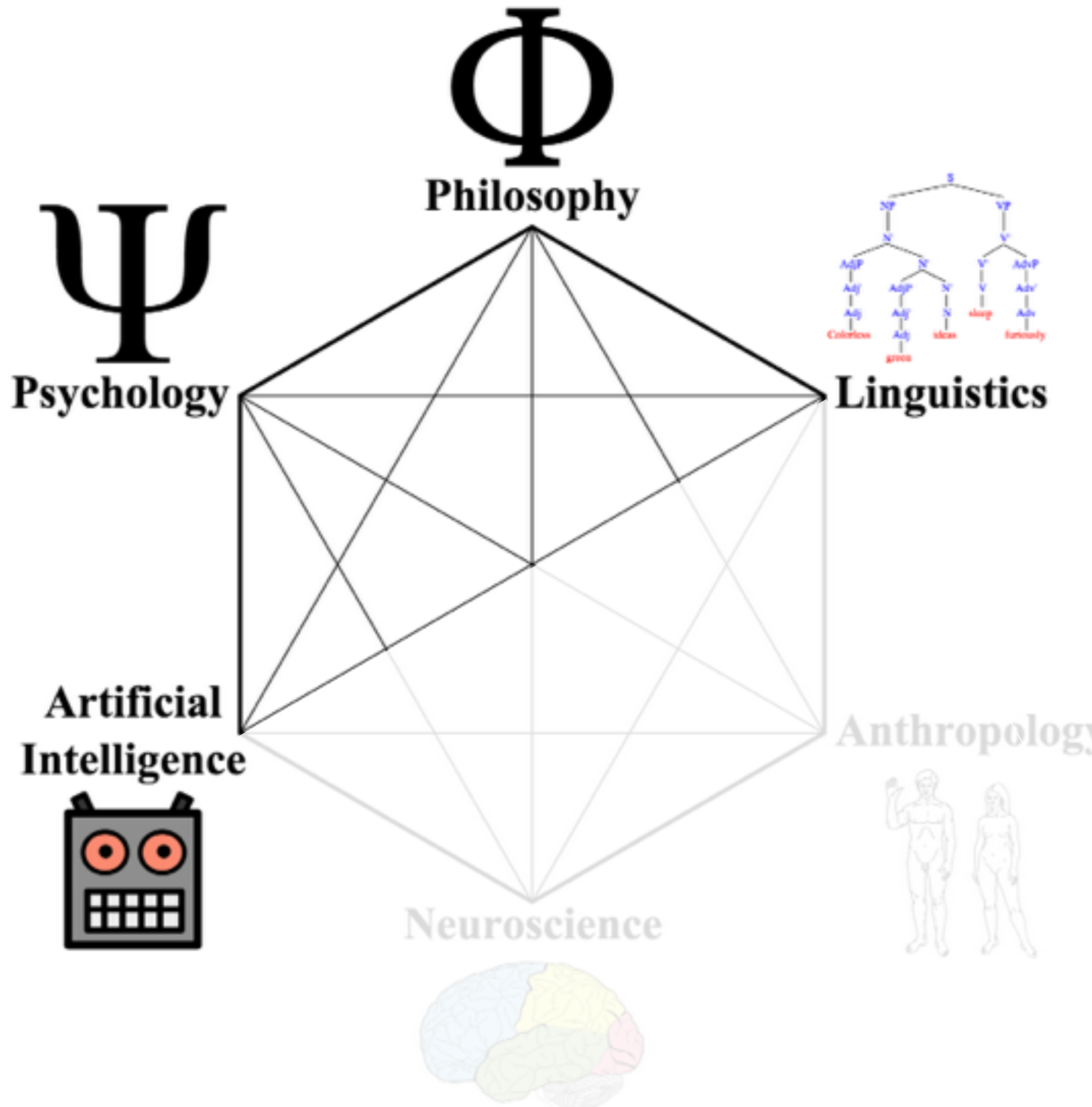
The cognitive science hexagon



We're primarily psychologists…



… but our work draws heavily from machine learning, philosophy and linguistics



<u>Why</u> does human cognition work the way it does?

Overview of the talk

- Inductive reasoning as Bayesian inference
  - Basic motivation
  - Two simple models
  - The role of social inference

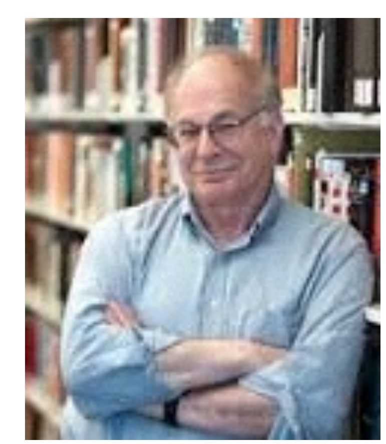
- Cultural evolution of communication systems
- Human decision making as stochastic planning
- And many more…

Part 1.
Inductive reasoning

Linda’s lament

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.

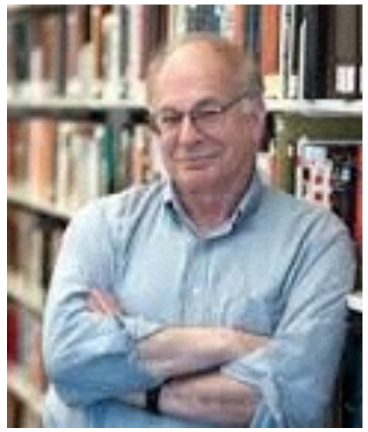
Amos Tversky



Daniel Kahneman

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 bank teller and is active in the feminist movement.

People endorse the <u>subset</u> hypothesis as more probable

Which is more probable?

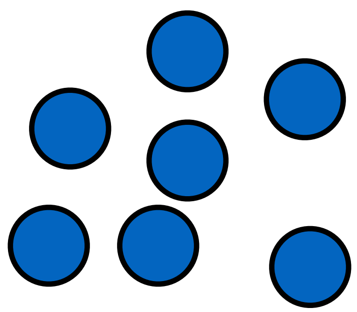
(a) Linda is a bank teller.

(b) Linda is a bank teller and is active in the feminist movement.

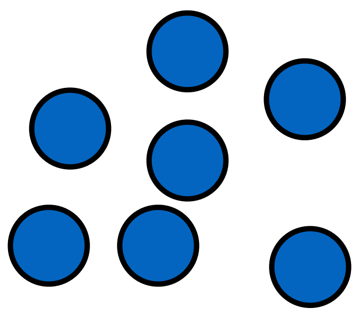
Heh. People are dumb, amiright?

(b) Linda is a bank teller and is active in the feminist movement.

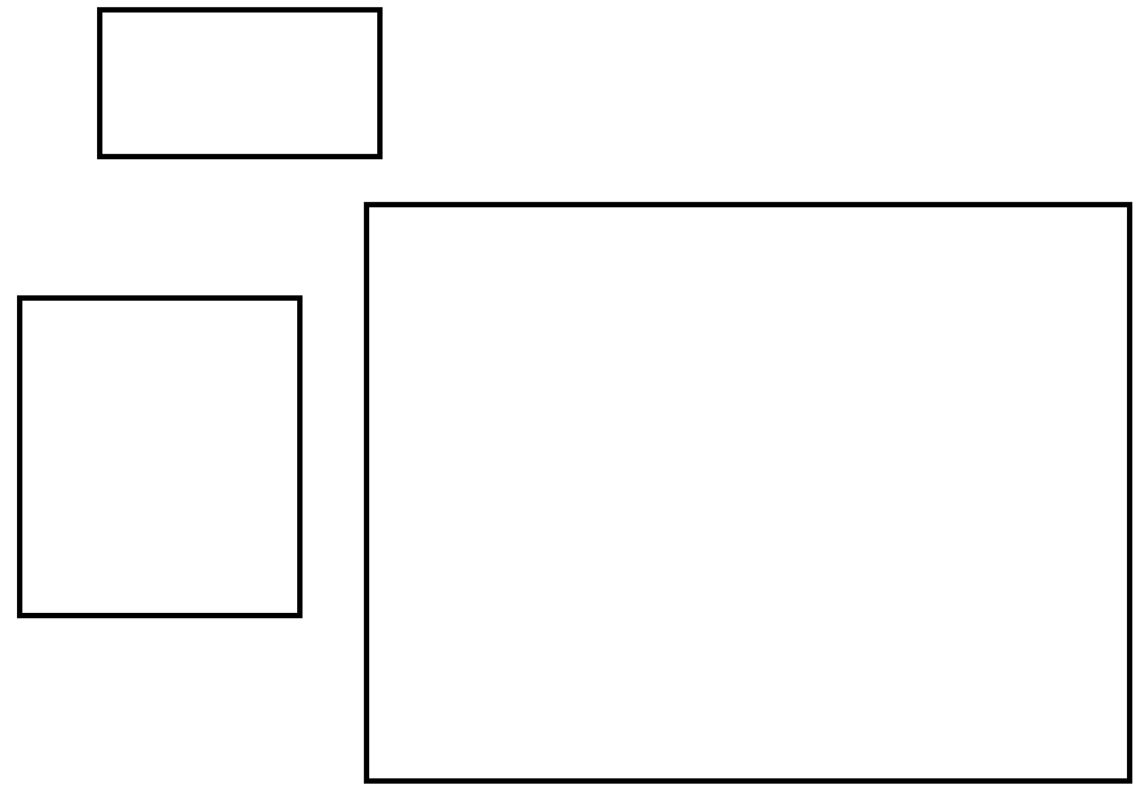
Let’s start over…



Data

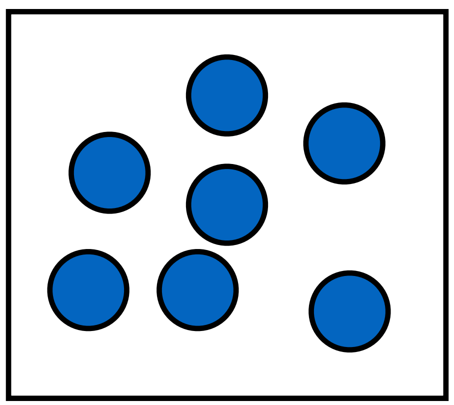


Data



Hypotheses

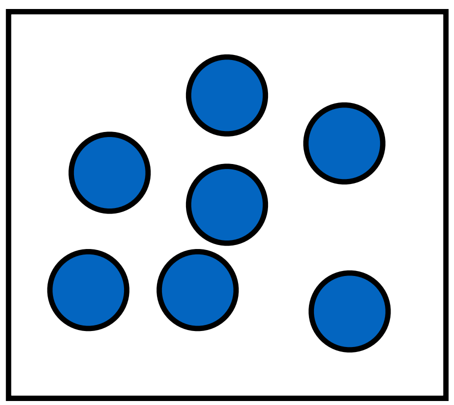
Which theory do <u>you</u> believe in?



A small rectangle that encloses the data?

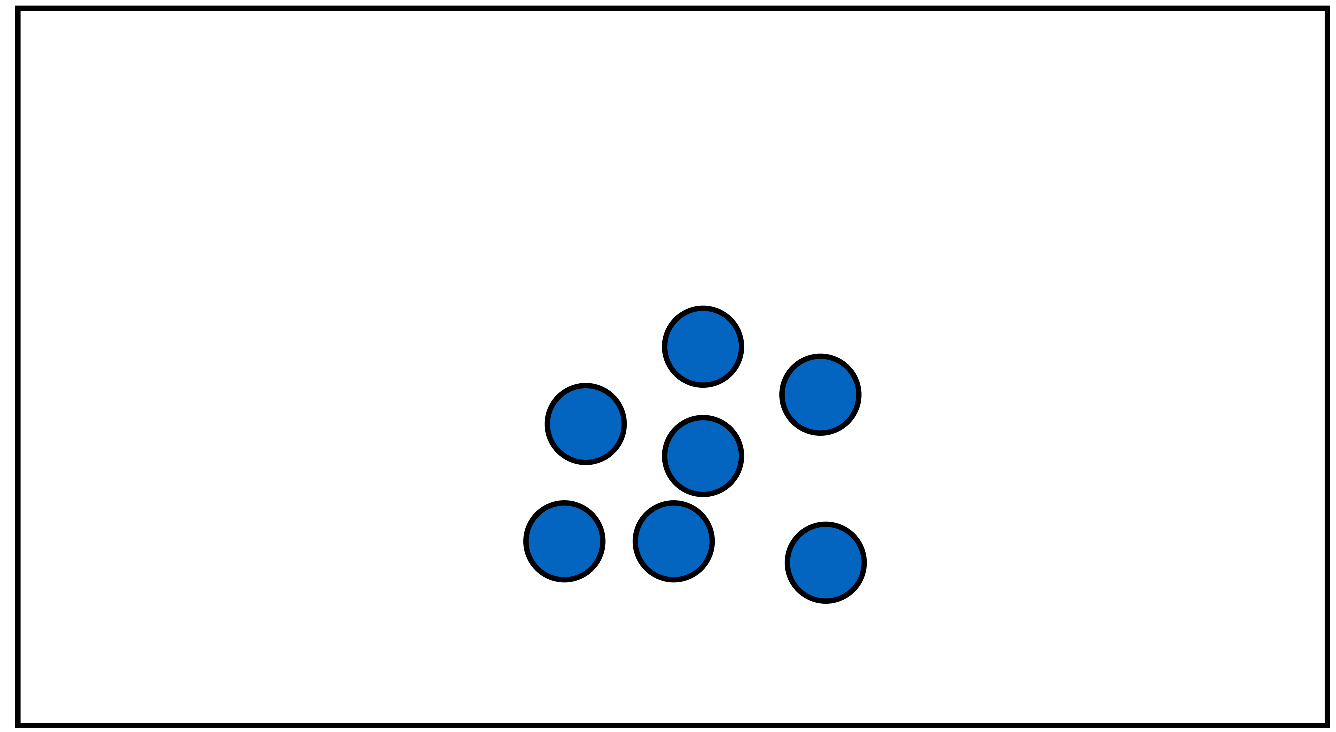
("feminist bank teller")

Which hypothesis do <u>you</u> believe?



A small rectangle that encloses the data?

("feminist bank teller")



Or a big one that strictly includes the small one?

("bank teller")

Hm. Okay, maybe that's a fluke.
Let's try another problem.

Grizzly bears produce hormone TH-L2

Grizzly bears produce hormone TH-L2
Black bears produce hormone TH-L2

Grizzly bears produce hormone TH-L2
Black bears produce hormone TH-L2
Polar bears produce hormone TH-L2

Grizzly bears produce hormone TH-L2
Black bears produce hormone TH-L2
Polar bears produce hormone TH-L2
Sun bears produce hormone TH-L2

Grizzly bears produce hormone TH-L2
Black bears produce hormone TH-L2
Polar bears produce hormone TH-L2
Sun bears produce hormone TH-L2

← Data

Bears?

Mammals?

← Hypotheses

What is the true hypothesis for how Tversky and Kahneman generated the Linda vignette?

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.

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 bank teller and is active in the feminist movement.

Linda is 31 years old, single, outspoken, and very br… sh… di… pa…

W…

feminist movement.

Hm. Maybe people aren't stupid.

Maybe they're correctly inferring that T&K deliberately told a story about a feminist?

The Bayesian heresy

Human learning can be characterised as a form of Bayesian inference

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)}$$

The prior over hypotheses *h* describes the learner's beliefs before any data arise

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

The posterior over hypotheses *h* describes the learner's beliefs <u>after</u> the data have been seen

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

The likelihood of the data under each hypothesis acts as a scoring rule, and guides rational belief revision

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

Likelihoods are <u>theories</u> about how the data came into being.

Rational belief revision depends on how the learner thinks the data were generated

Two ways to tell a story…

Weak sampling:

“present random statements that are true”

Two ways to tell a story…

Weak sampling:

“present random statements that are true”

Strong sampling:

“present random statements that are true and <u>about Linda</u>”

A weakly sampled vignette is stark raving mad

Atoms are smaller than horses. Yellow is not called John. Dogs are not cats. My cat's breath smells like cat food.

A weakly sampled vignette is stark raving mad

Atoms are smaller than horses. Yellow is not called John. Dogs are not cats. My cat's breath smells like cat food.

A strongly sampled vignette reads like a database dump, but at least it's on topic!

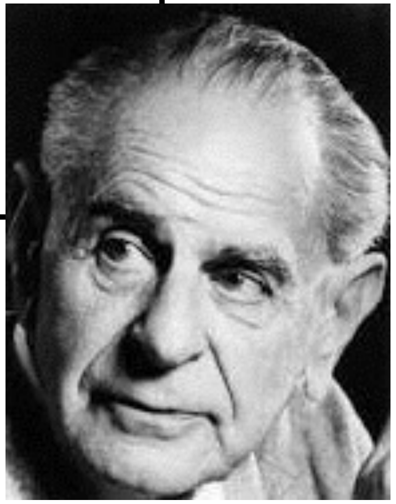
Linda is a person. Linda is an activist. Linda is female. Linda is 31. Linda has a sister.

These two kinds of theory lead to <u>very</u> different inductive biases

Weak sampling produces a falsificationist

$$P(x|h) \propto \begin{cases} 1 & \text{if } x \in h \\ 0 & \text{otherwise} \end{cases}$$

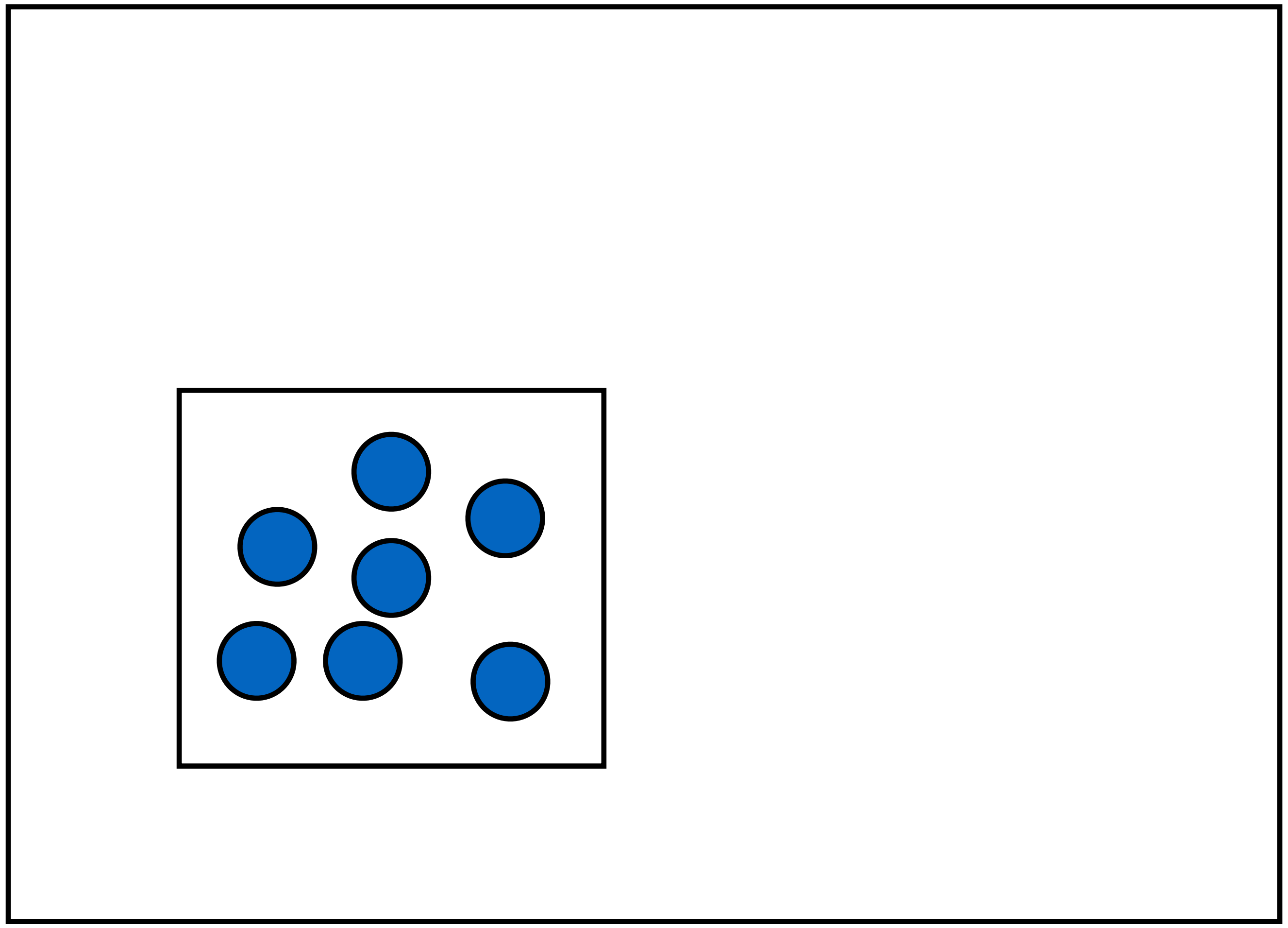
Falsify a hypothesis if it is inconsistent with the facts. Otherwise do nothing.

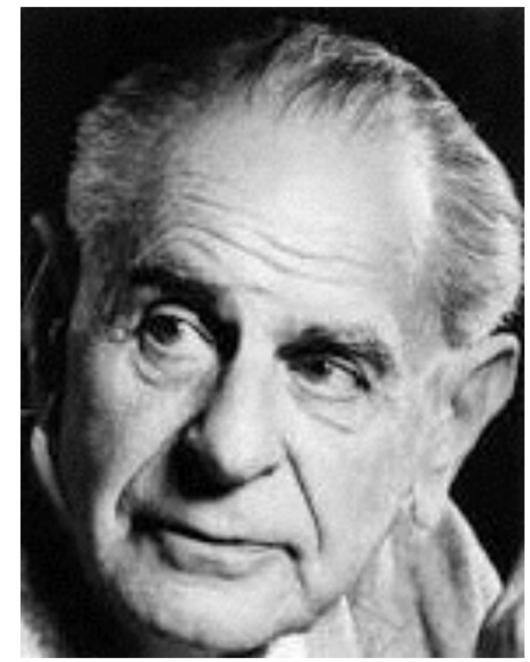


Strong sampling produces an Ockhamist!

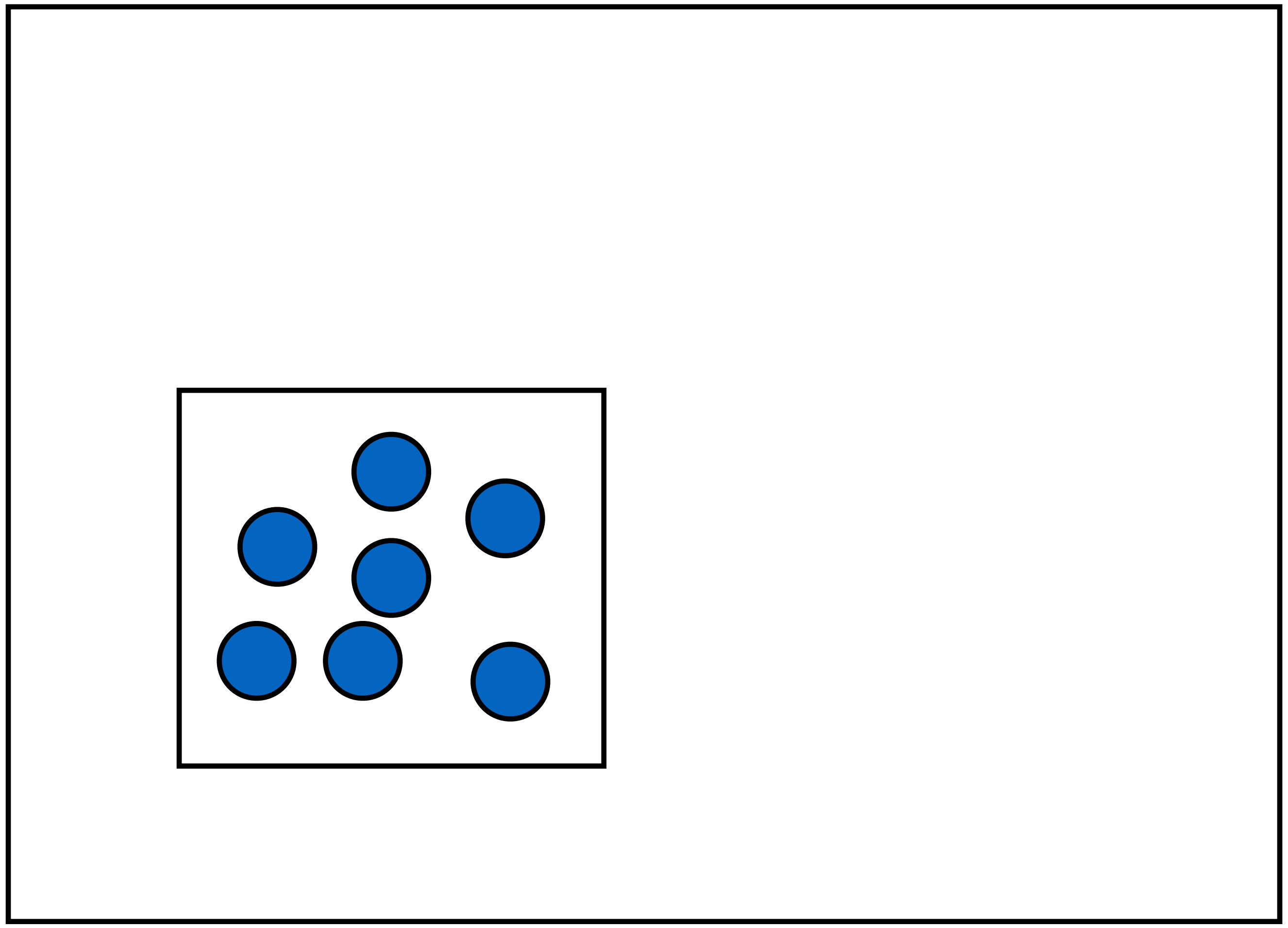
$$P(x|h) = \begin{cases} \frac{1}{|h|} & \text{if } x \in h \\ 0 & \text{otherwise} \end{cases}$$

Prefer the smallest, simplest hypothesis consistent with the data





The data provide <u>no</u> evidence to discriminate between these two hypotheses. Neither can be falsified.



The data provide strong evidence for the small rectangle, because it covers only the observations and no other unobserved possibilities

Humans are Ockhamists by default
(and falsificationists when forced)

Ransom, Perfors & Navarro (under revision)

(1)

Grizzly bears produce hormone TH-L2

Lions produce hormone TH-L2

(2)

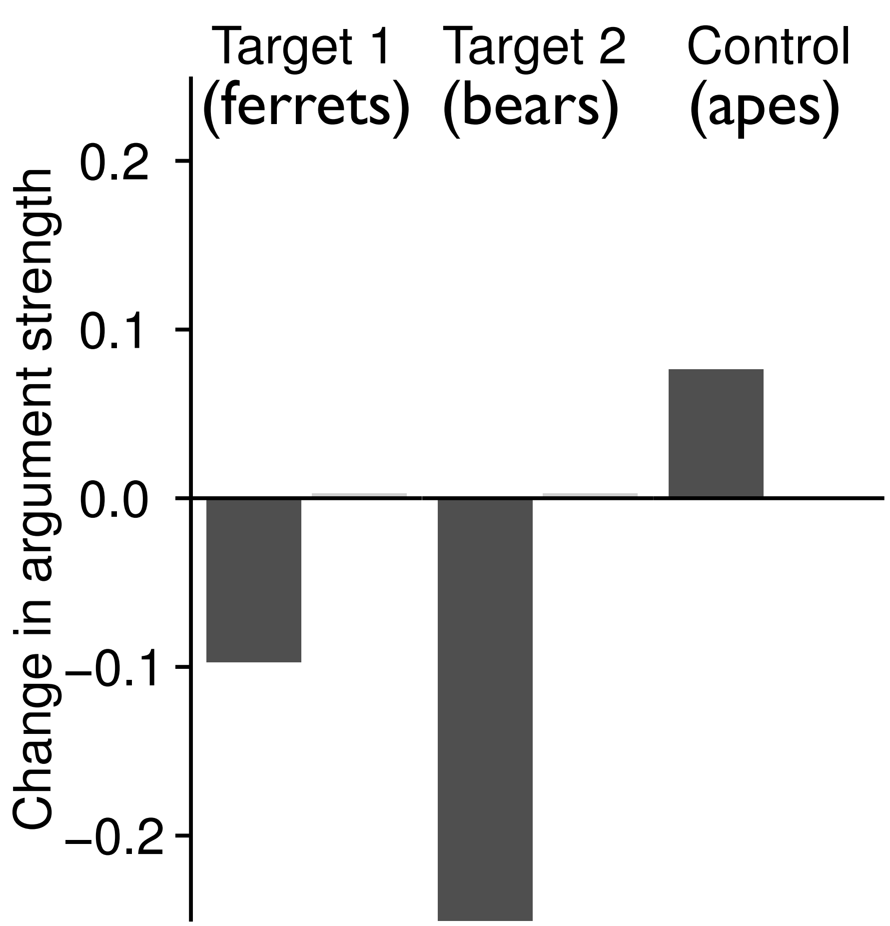
Grizzly bears produce hormone TH-L2

Black bears produce hormone TH-L2

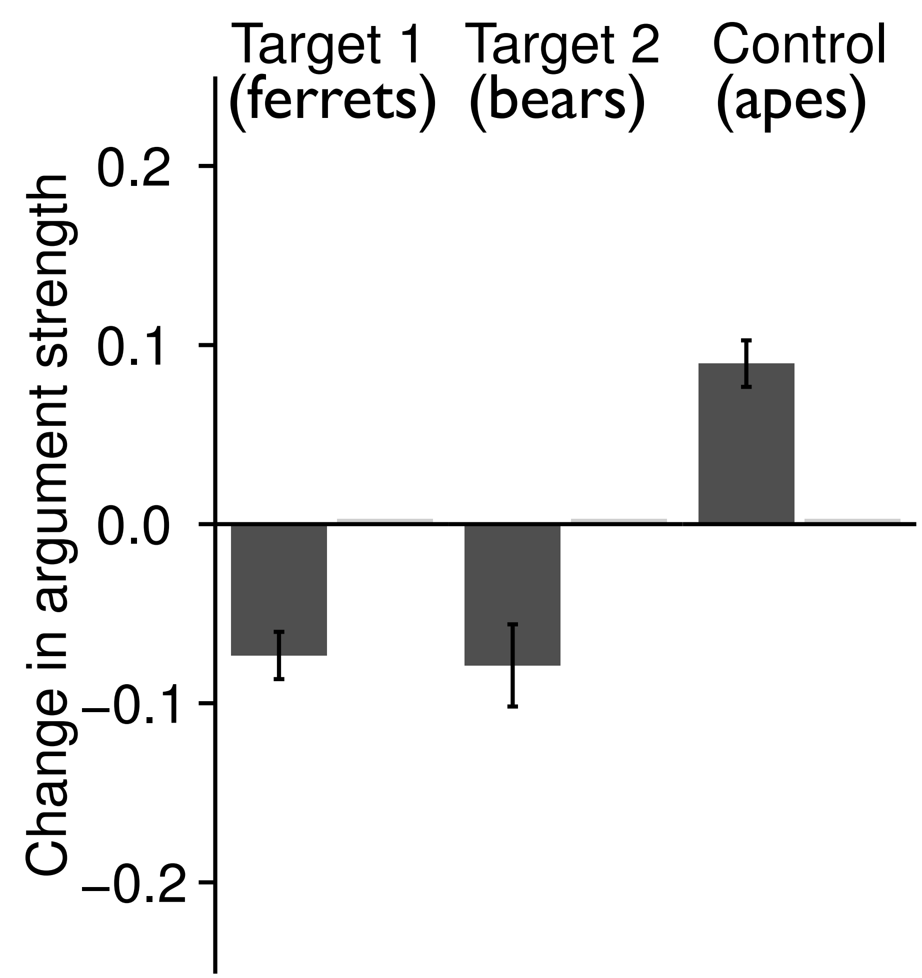
Lions produce hormone TH-L2

Under <u>normal</u> conditions, people match the "Ockhamist" strong sampling model…

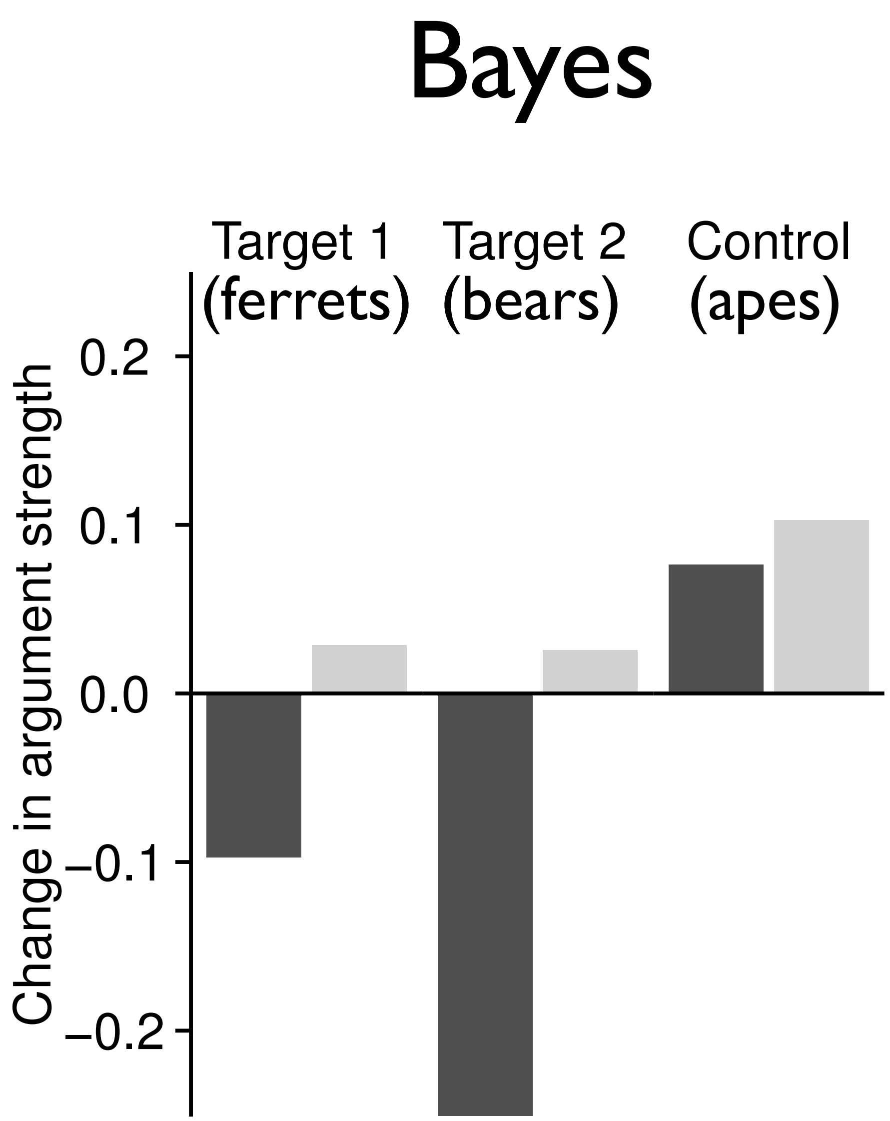
Bayes

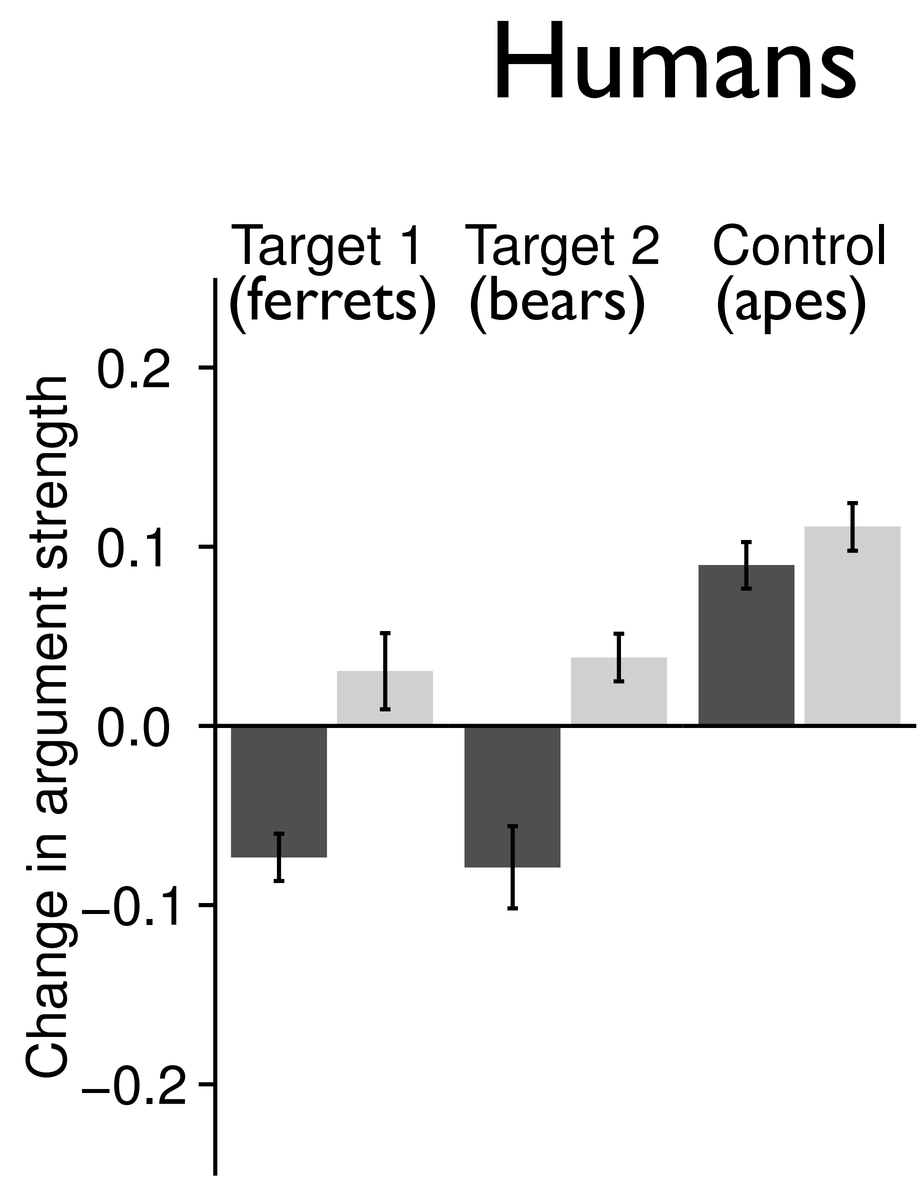


Humans



But if you rig the experiment so the facts are "random truths" they switch to a falsificationist weak sampling logic





- Inductive reasoning is not just about the evidence that facts provide for a conclusion, it’s also about <u>how</u> you think those facts were put together

- Bayesian models explain the reversal as a shift in the sampling assumption

How to take a hint

(Rational reasoning by <u>social</u> agents has a rather different inductive logic)

Voorspoels, Storms, Navarro & Perfors (under review)

Weak sampling is <u>really</u> stupid

Electrons are smaller than horses. Yellow is not called John. Dogs are not cats. My cat's breath smells like cat food.

Strong sampling is less stupid, but it's still stupid

Linda is a person. Linda is an activist. Linda is female. Linda is 31. Linda has a sister.

Weak sampling is <u>really</u> stupid

Electrons are smaller than horses. Yellow is not called John. Dogs are not cats. My cat's breath smells like cat food.

Strong sampling is less stupid, but it's still stupid

Linda is a person. Linda is an activist. Linda is female. Linda is 31. Linda has a sister.

But <u>real</u> story telling is designed to <u>communicate</u> an idea

Linda is a 31 year old woman with a strong commitment to social justice and a history of activism.

“Stories” are told (and “arguments” made) by intelligent agents who wants to <u>shape your beliefs</u>

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

The data x selected by the <u>communicator</u>…

… is chosen to maximise the <u>learner's</u> posterior degree of belief in hypothesis h

Avalanche

Bach

Metallica

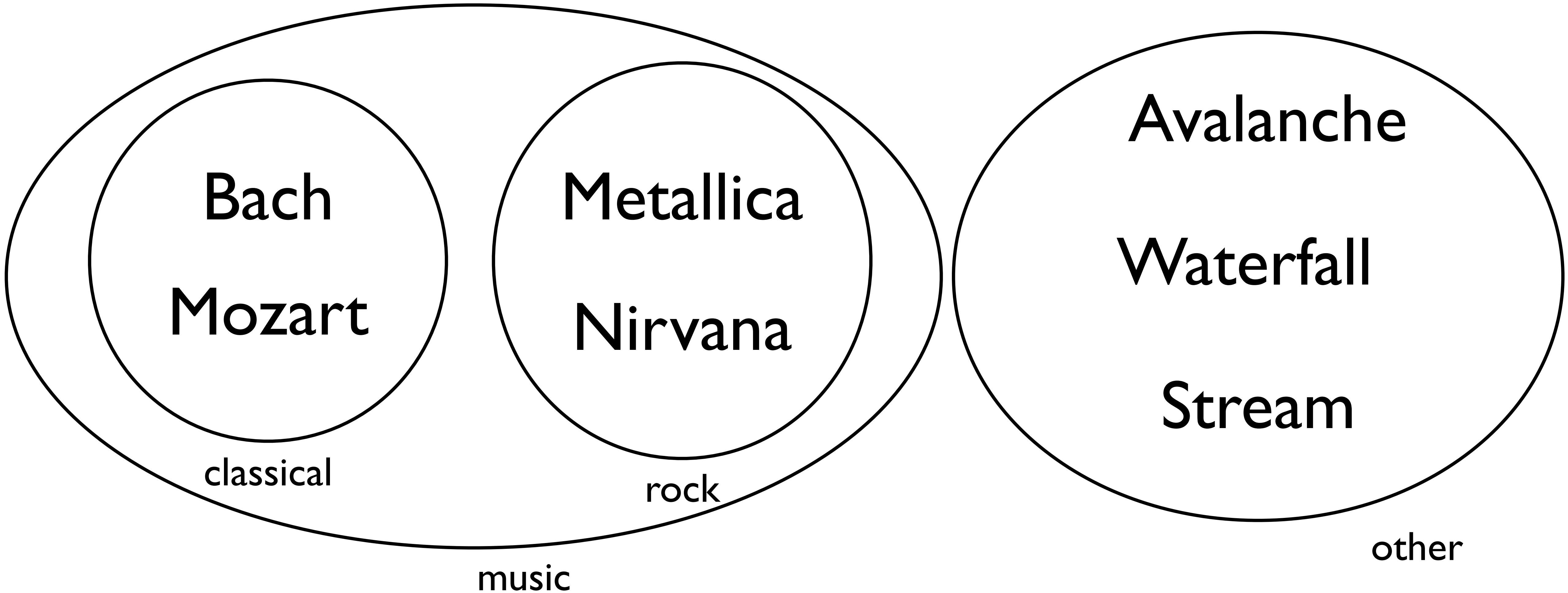
Waterfall

Mozart

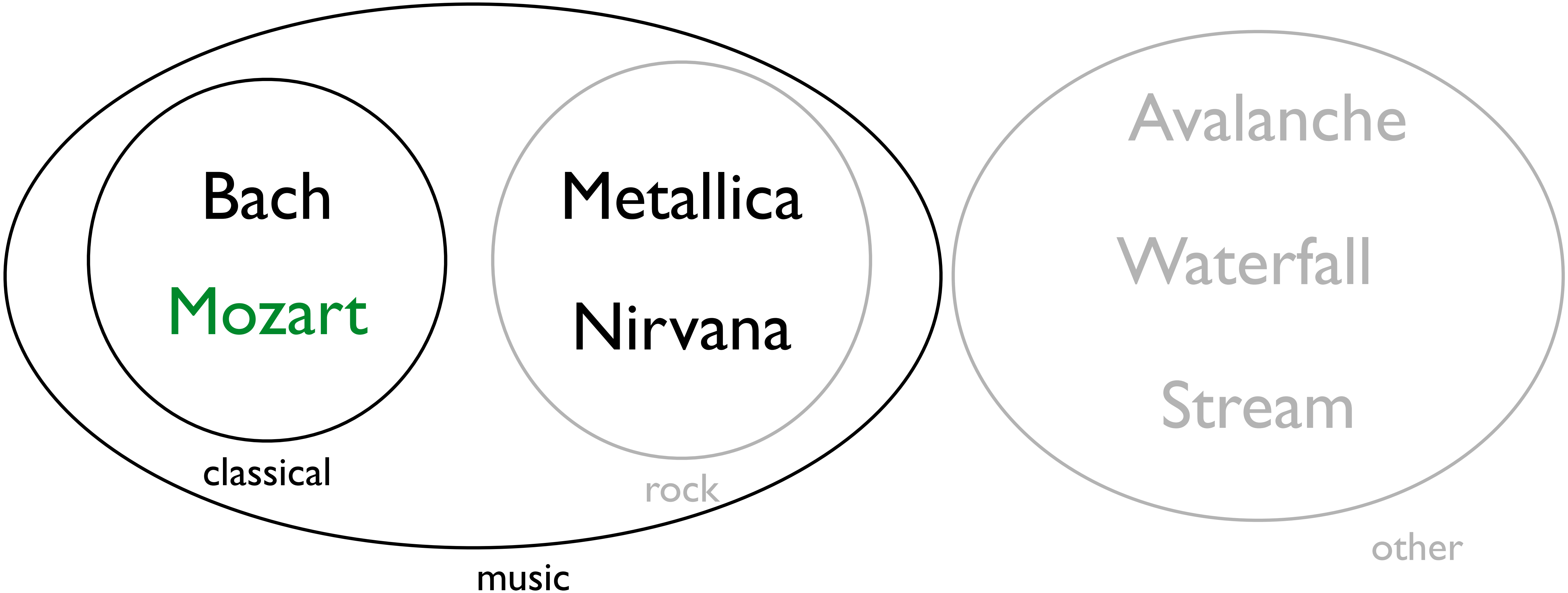
Nirvana

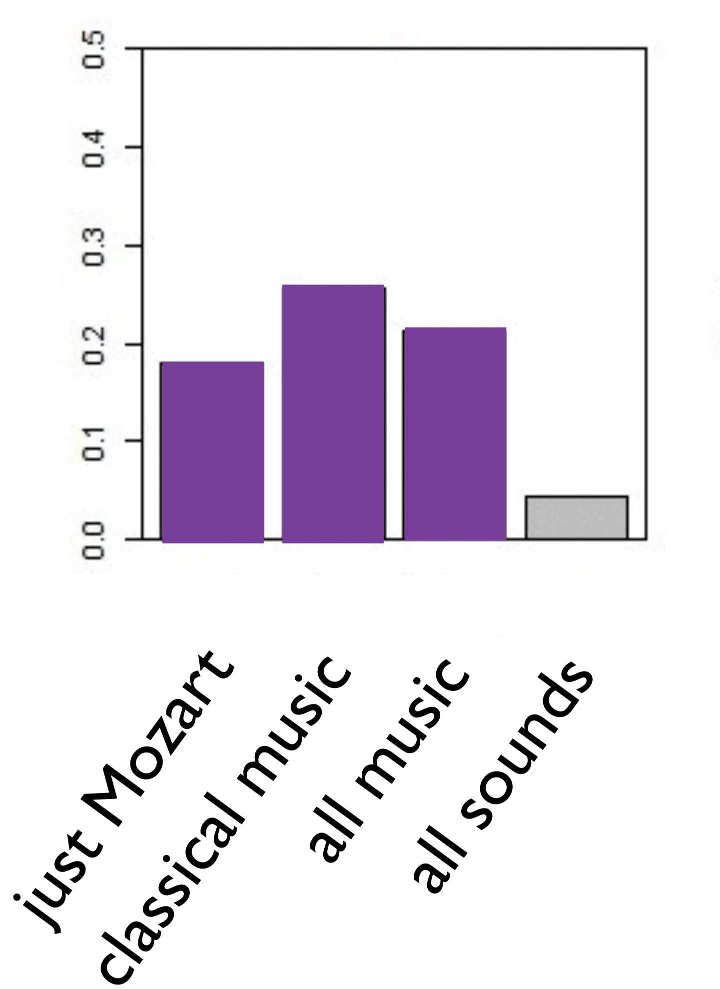
Stream

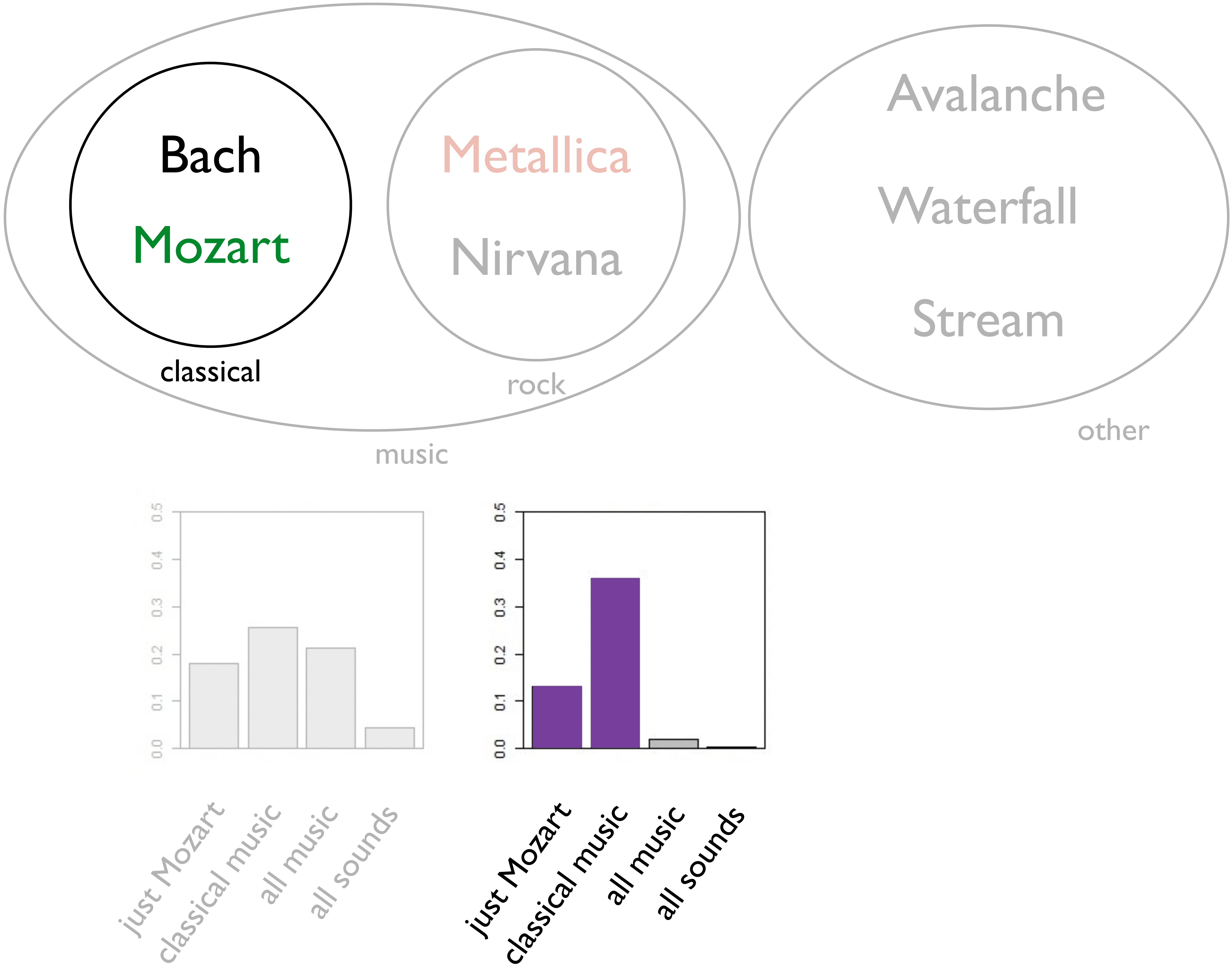
Which of these “produce alpha waves” in the brain?

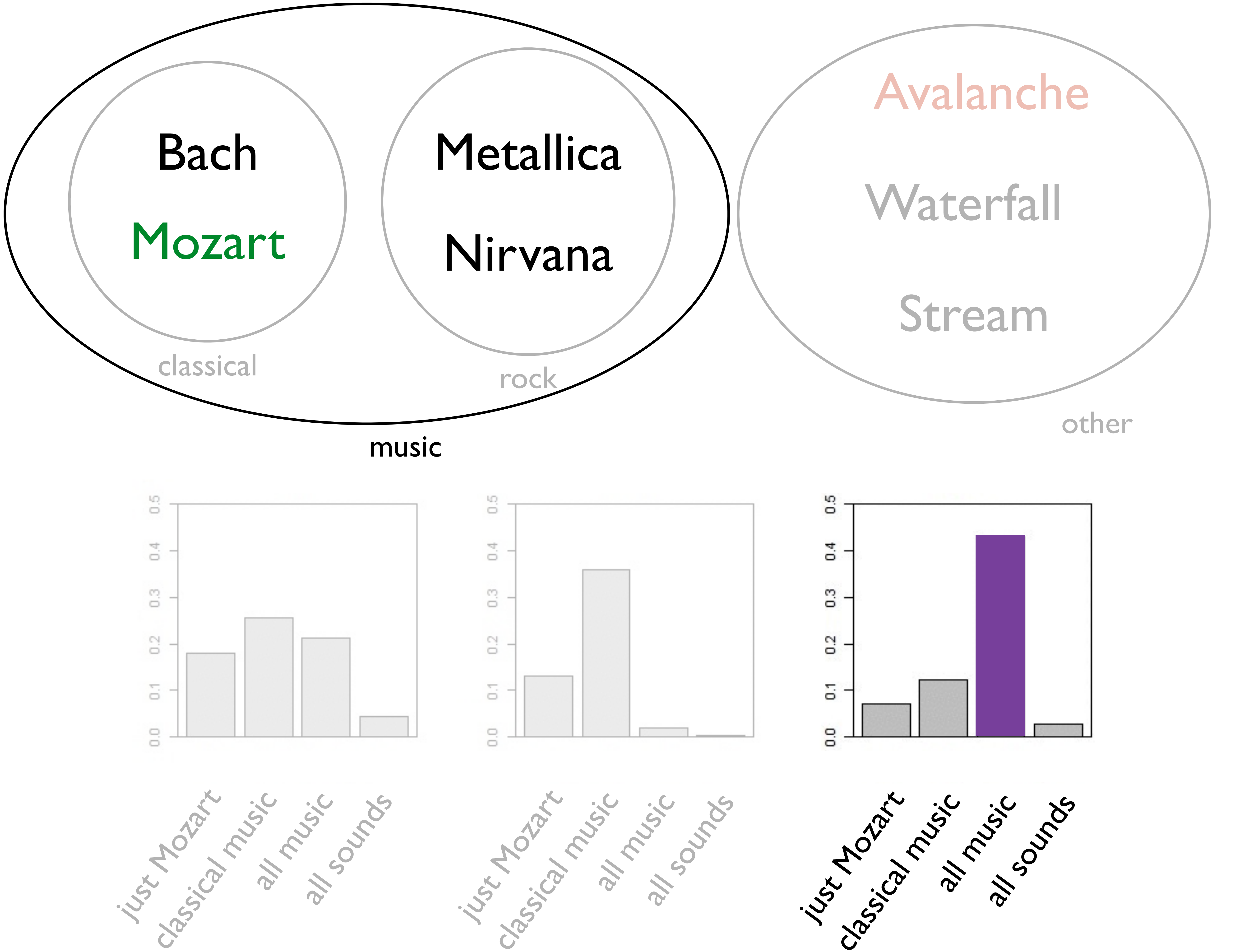


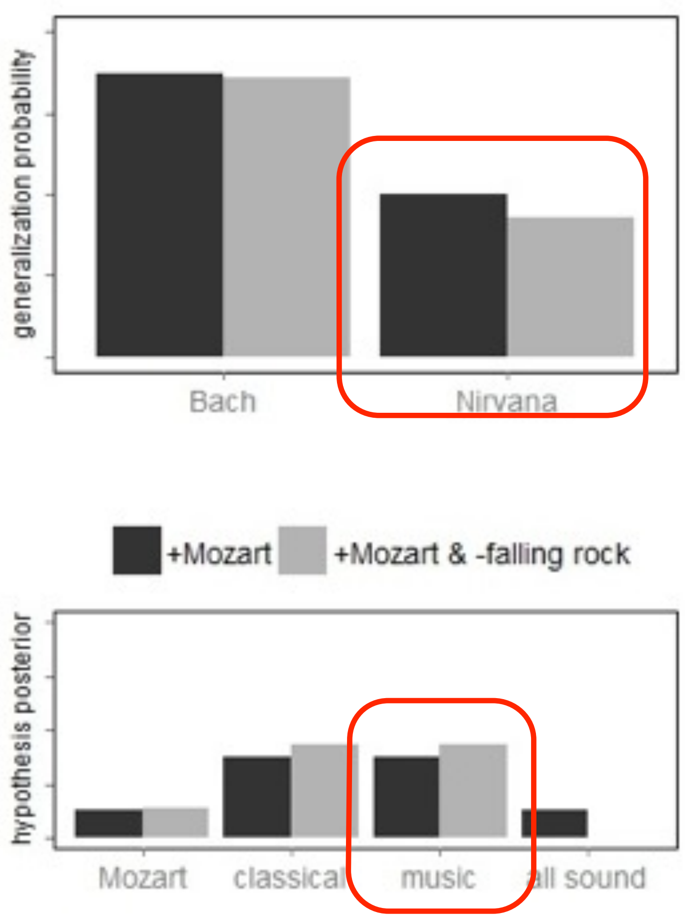
Some plausible categories that might describe the extension of this new property


Bach
Mozart
classical
Metallica
Nirvana
rock
music
Avalanche
Waterfall
Stream
other


0.0
0.1
0.2
0.3
0.4
0.5
just Mozart
classical music
all music
all sounds

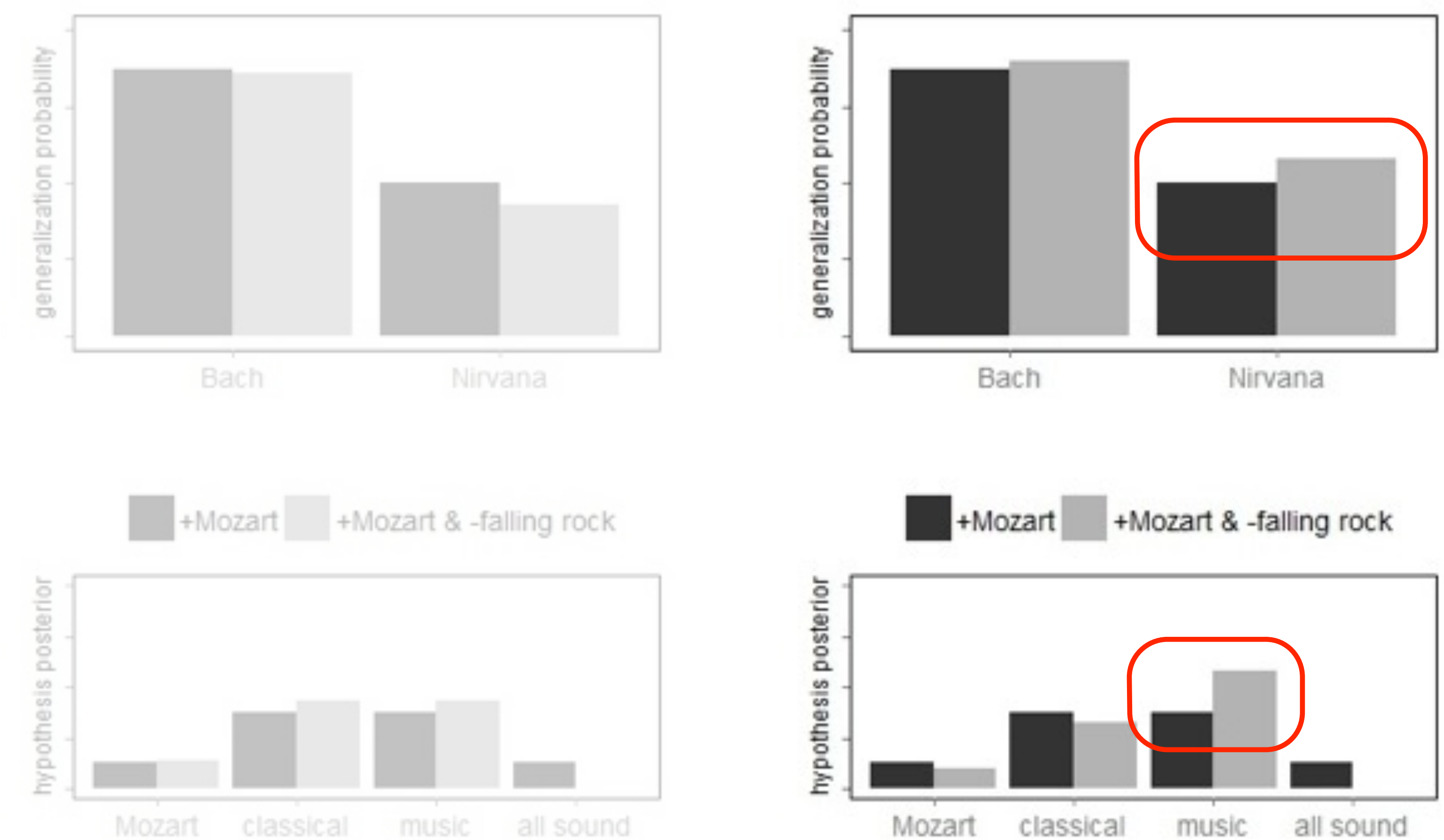

Bach
Mozart
classical
Metallica
Nirvana
rock
music
Avalanche
Waterfall
Stream
other
0.0
0.1
0.2
0.3
0.4
0.5
just Mozart
classical music
all music
all sounds
0.0
0.1
0.2
0.3
0.4
0.5
just Mozart
classical music
all music
all sounds


Bach
Mozart
classical
Metallica
Nirvana
rock
music
Avalanche
Waterfall
Stream
other
0.0
0.1
0.2
0.3
0.4
0.5
just Mozart
classical music
all music
all sounds
just Mozart
classical music
all music
all sounds
just Mozart
classical music
all music
all sounds



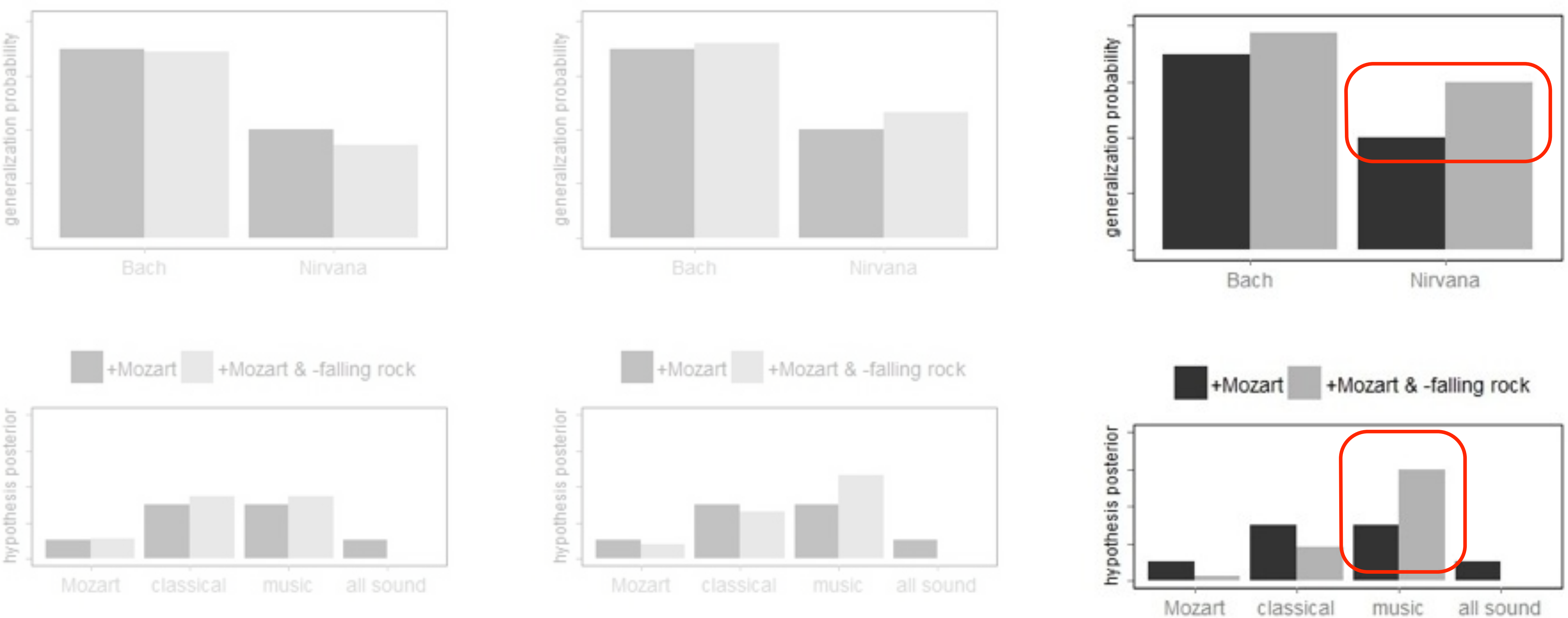
Weak sampling doesn't predict this effect in any version of our experiments

Strong sampling does, but only barely.



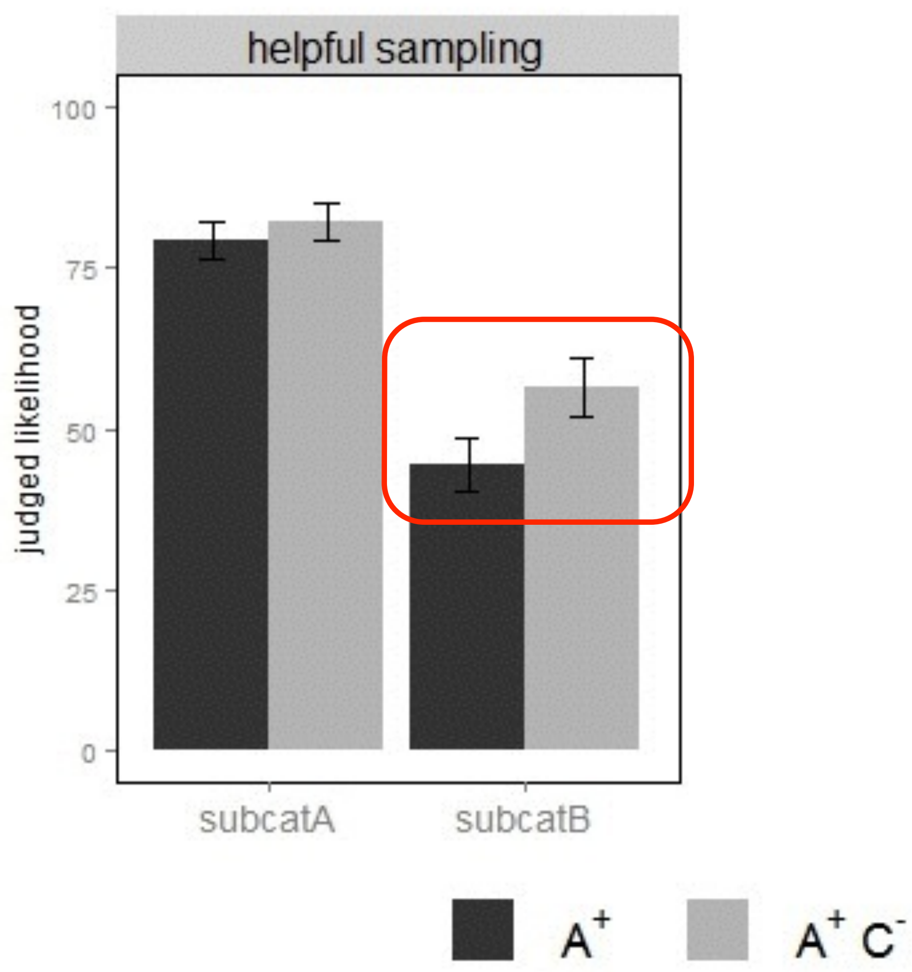
(The effect sizes are too small and highly dependent on how you “fiddle” with parameter settings)

The result emerges naturally within a communicative sampling framework

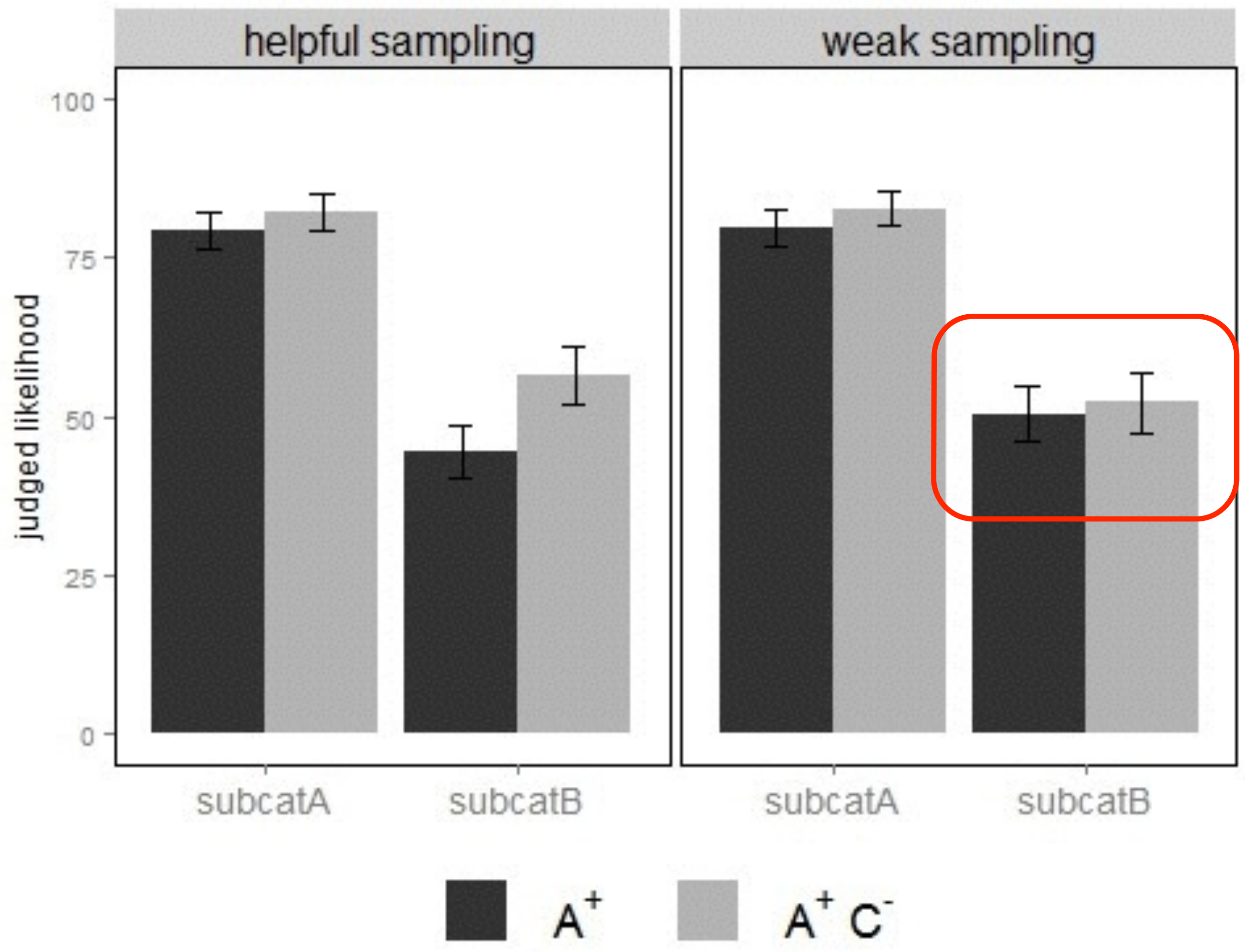


More experiments!

When the negative evidence is described as a "helpful hint" the effect replicates…



… but when construed as a “random true fact” about the world, the effect vanishes

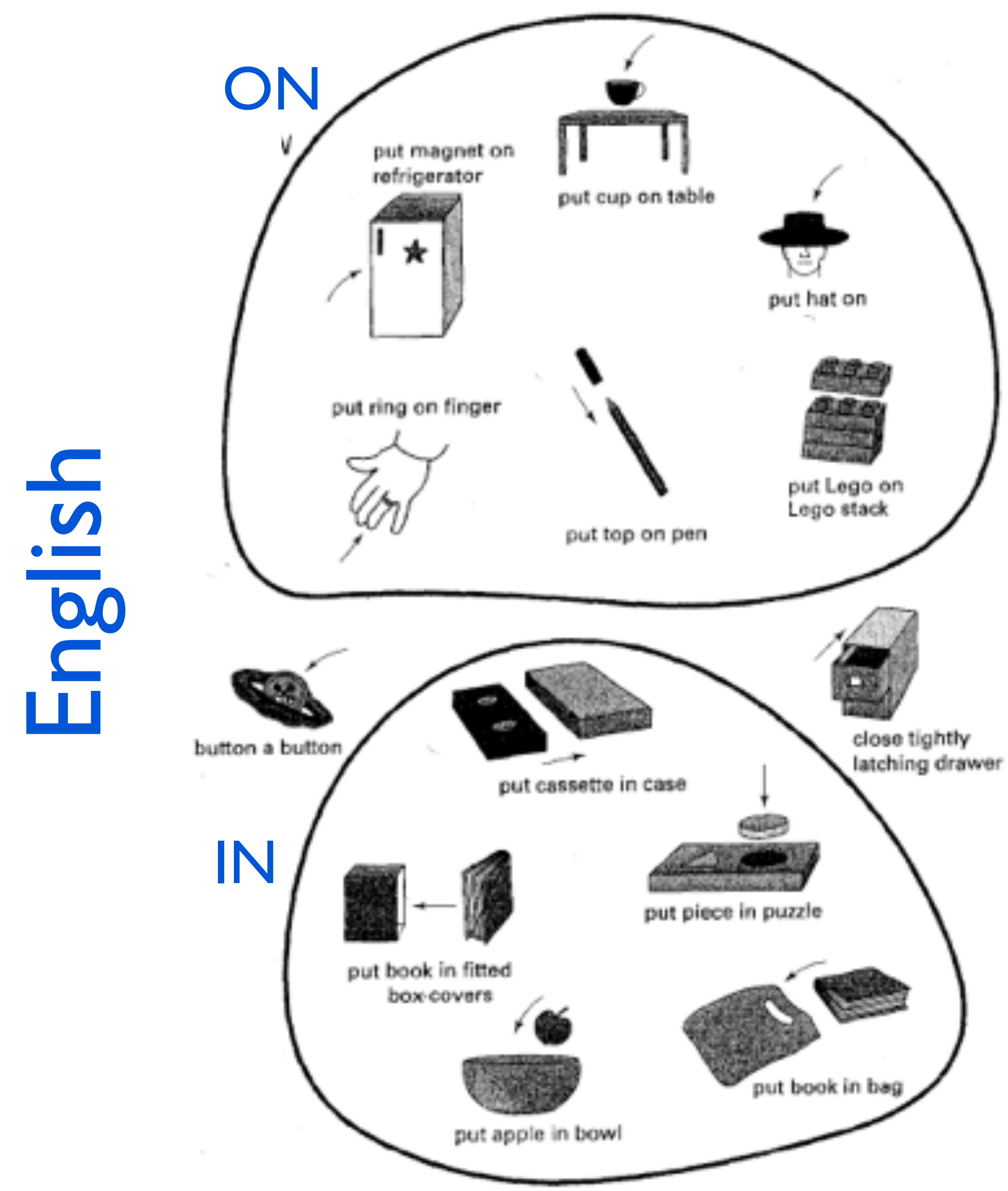


Part II.
Everything else

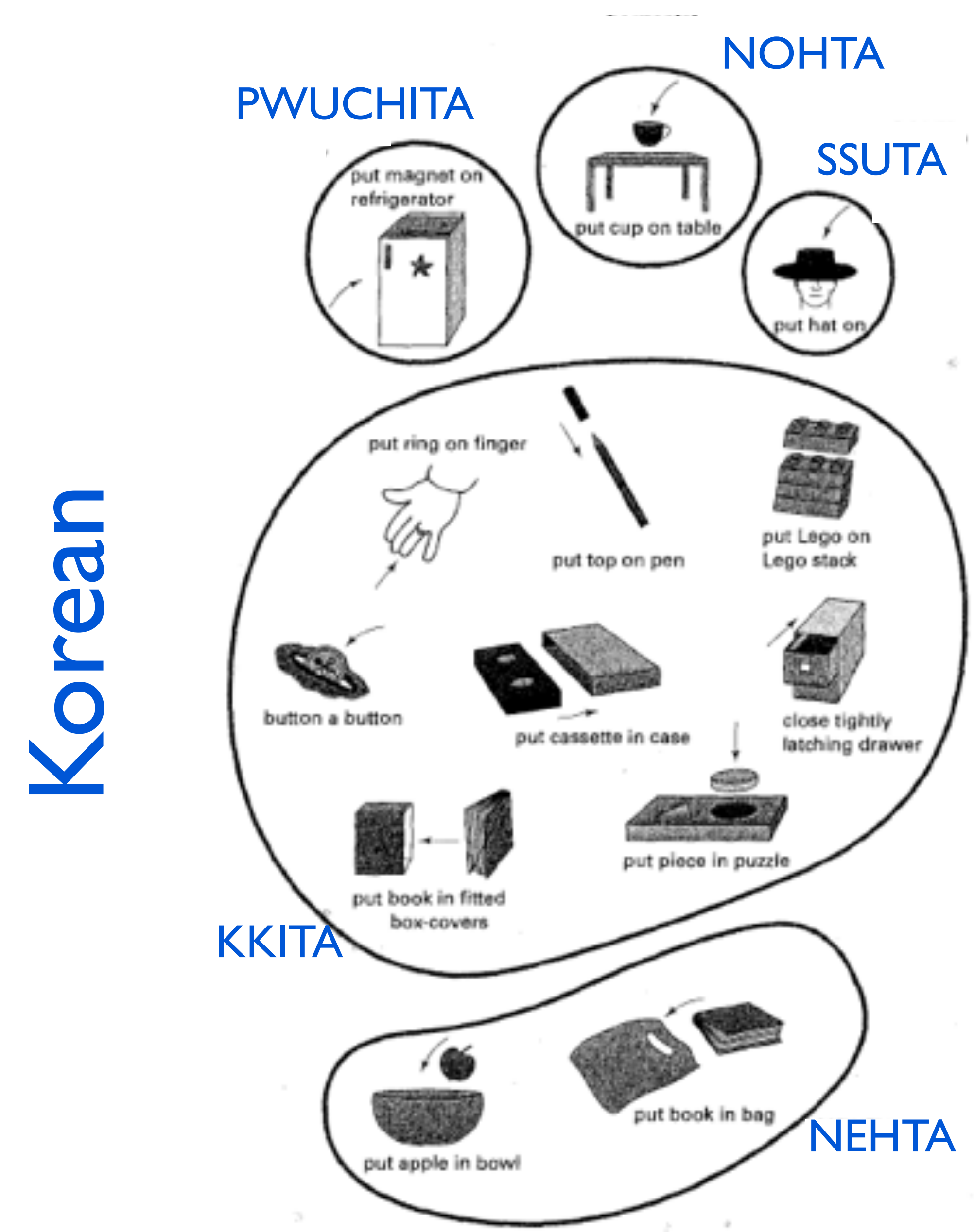
On the (cultural) evolution of communicative codes

Perfors & Navarro (2014)

Lexical categories are organised differently across languages

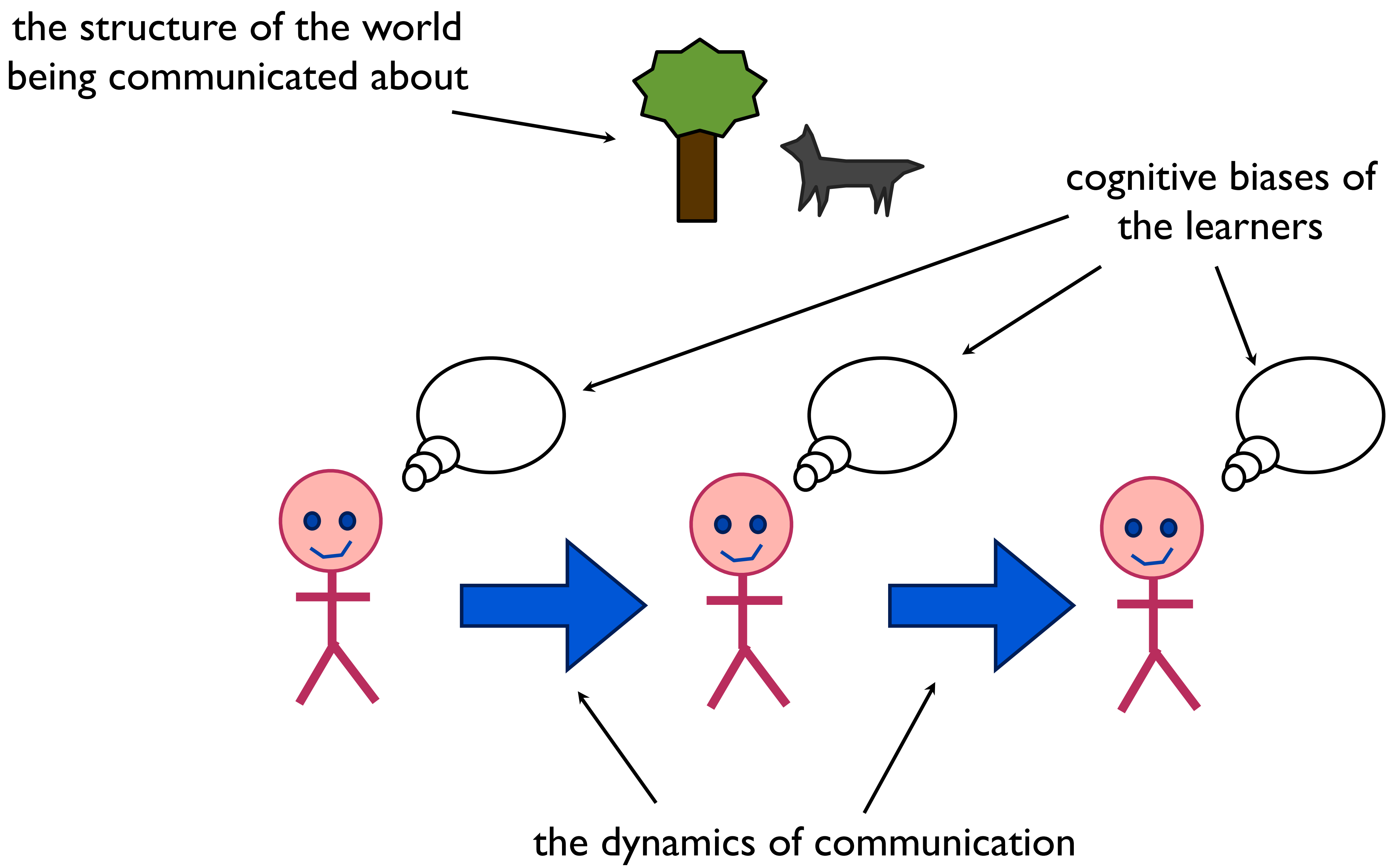


a. English

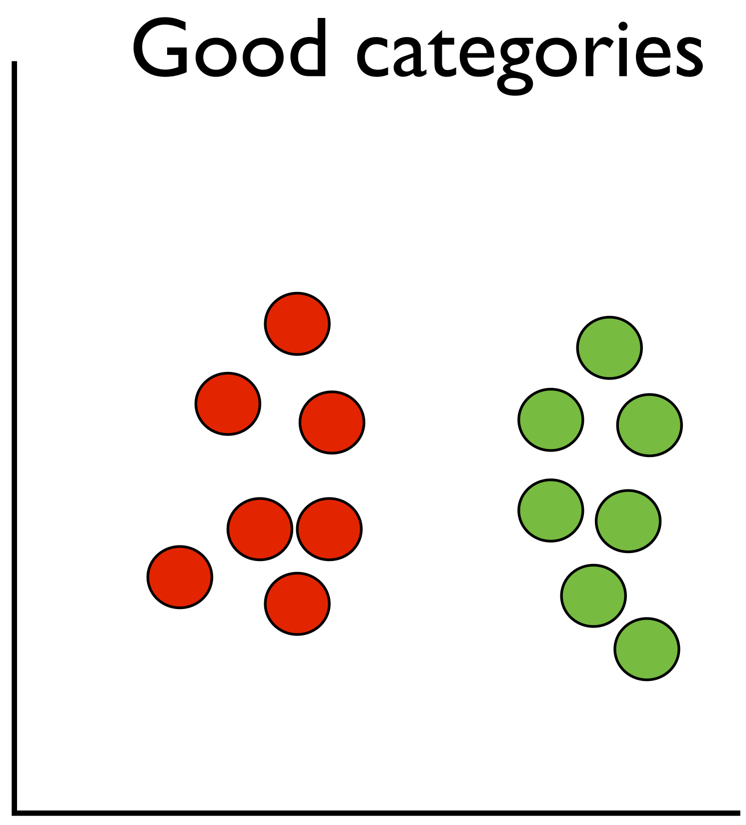


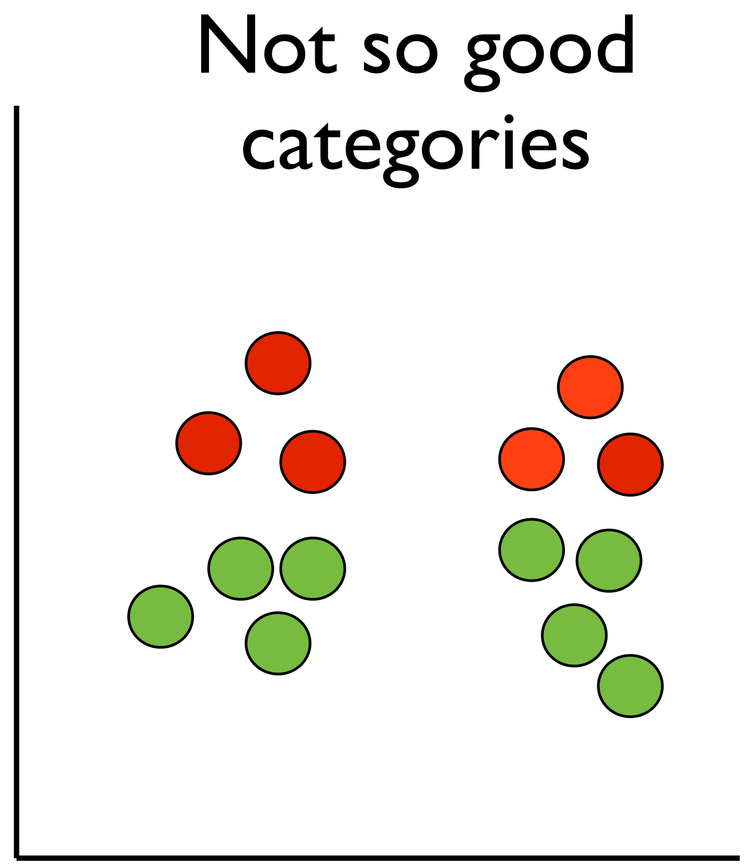
b. Korean

How do these naming systems evolve?

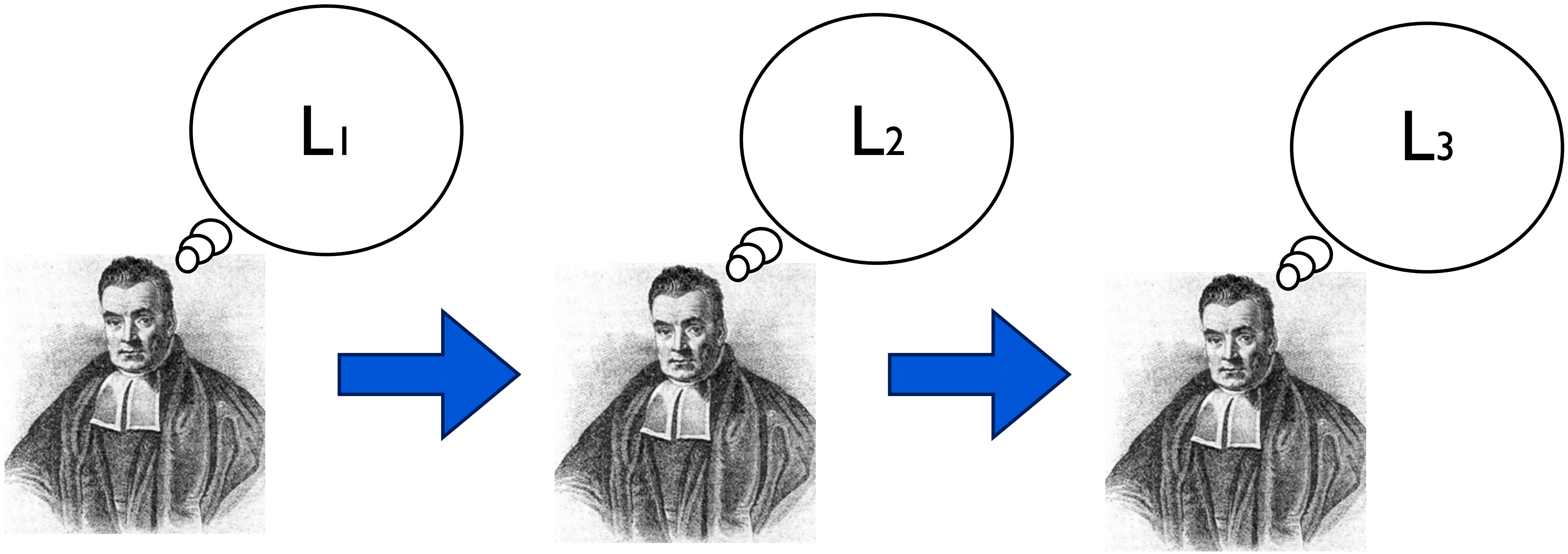


The usual idea in categorisation: stimulus structure (i.e., the world) shapes inferences





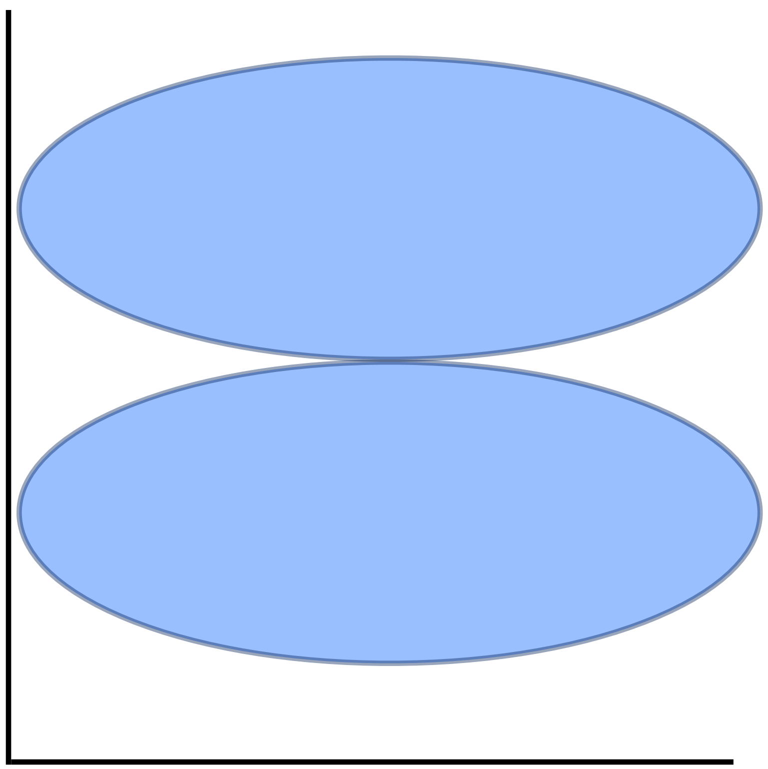
And yet… the current "hot topic" in language evolution says otherwise



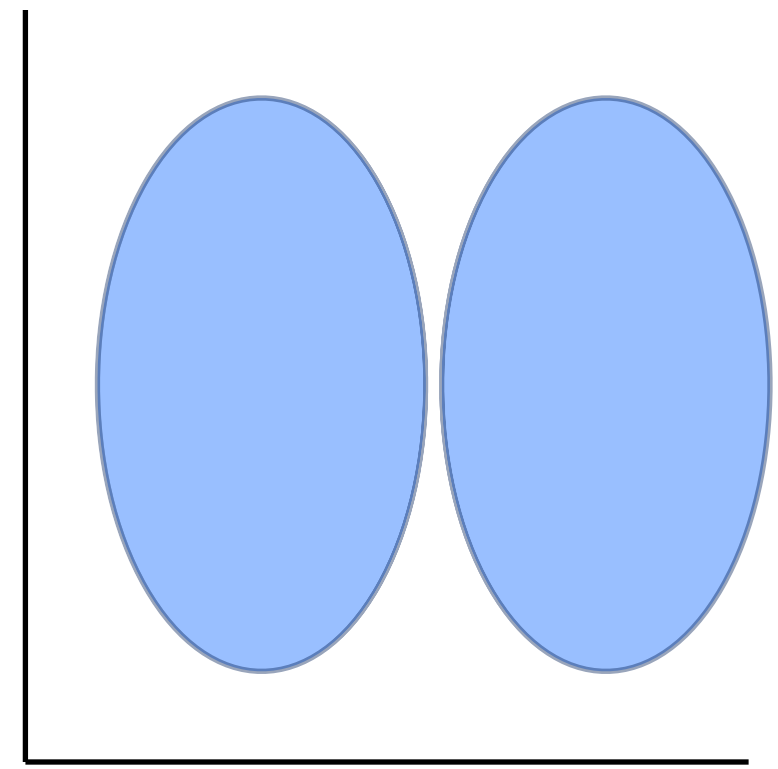
A sequence of Bayesian learners, each learning from the language output of the last one, and then generating the input for the next one... **converges to the learner's prior**.

Griffiths & Kalish (2005, 2007)

This is just <u>bizarre</u>. Here is a prior…

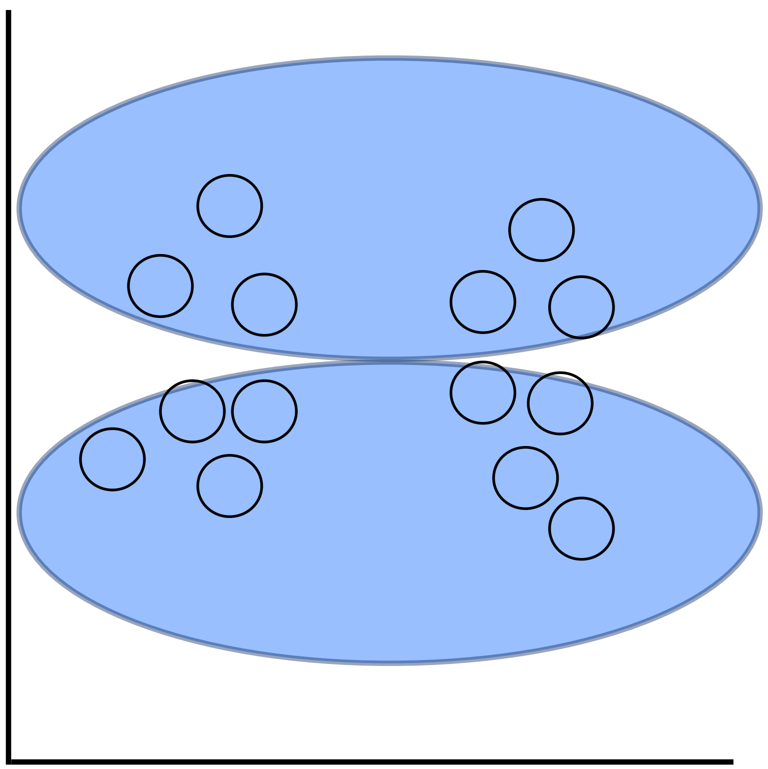


High probability

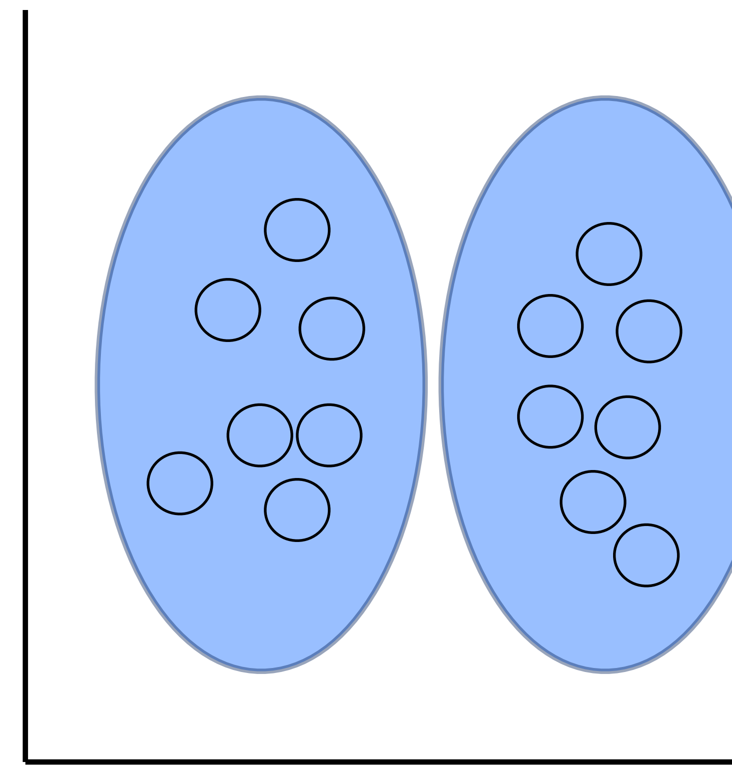


Low probability

And a set of entities that need names

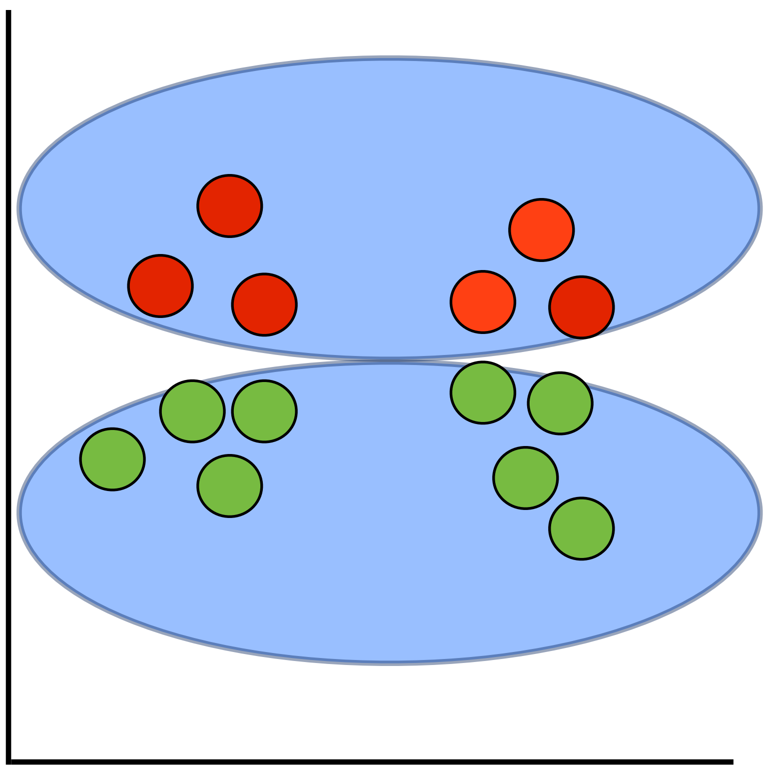


High probability

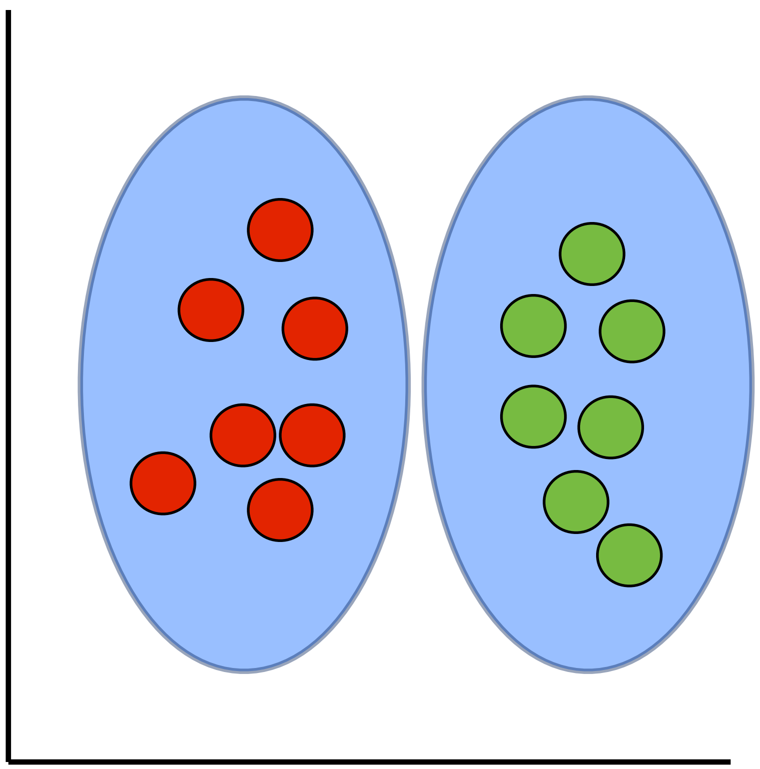


Low probability

If the standard theory were correct…



Your language <u>should</u> do this (high prior)



Your language should <u>not</u> do this (low prior)

The devil is in the details…

$$\ell = P(y|x)$$

A language provides labels y for entities x

It says <u>nothing</u> about which entities x will be observed

Griffiths & Kalish (2005, 2007)

"laser"

My language has a word for laser… does that really tell me <u>nothing at all</u> about my chances of encountering one?

Let's see what happens if you believe languages supply other biases…

$$\ell = P(y|x)$$

A language provides labels y for entities x

It says nothing about which entities x will be observed

Griffiths & Kalish (2005, 2007)

$$\ell = P(y, x)$$

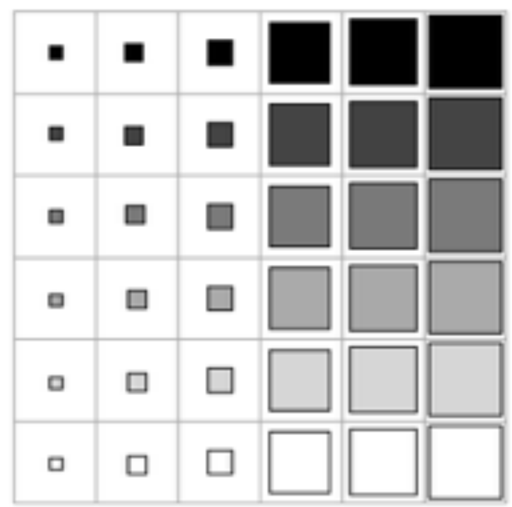
A language provides labels y for entities x, but it also makes assumptions about which entities x are likely to appear

Perfors & Navarro (2011, 2014)

Huh.

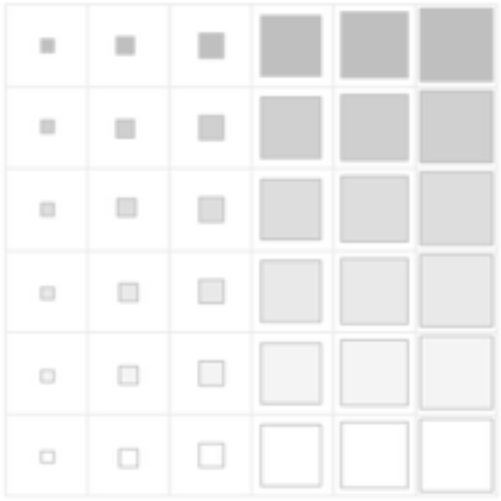
$\ell = P(y\|x)$	$\ell = P(y, x)$
A language provides labels y for entities x It says nothing about which entities x will be observed	A language provides labels y for entities x, but it also makes assumptions about which entities x are likely to appear
Labels converge to the prior distribution	Labels converge to the expected posterior given the entities (sort of)

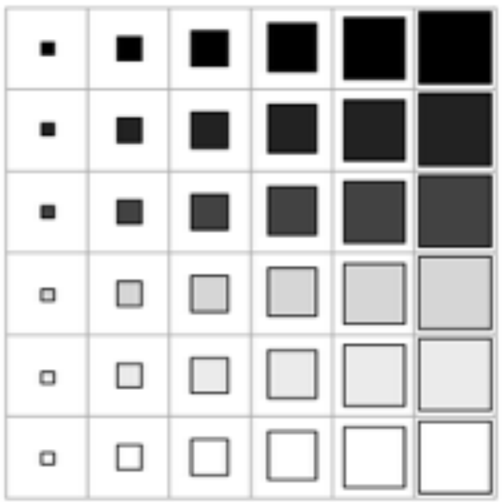
Three worlds…



Name objects based
on their size

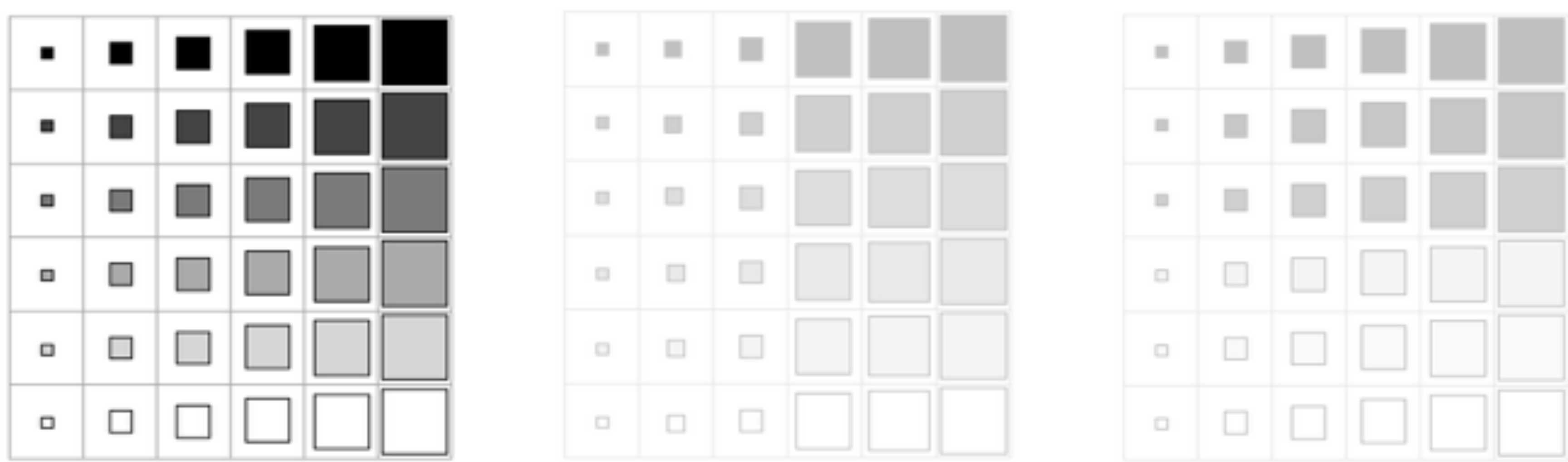
Three worlds…



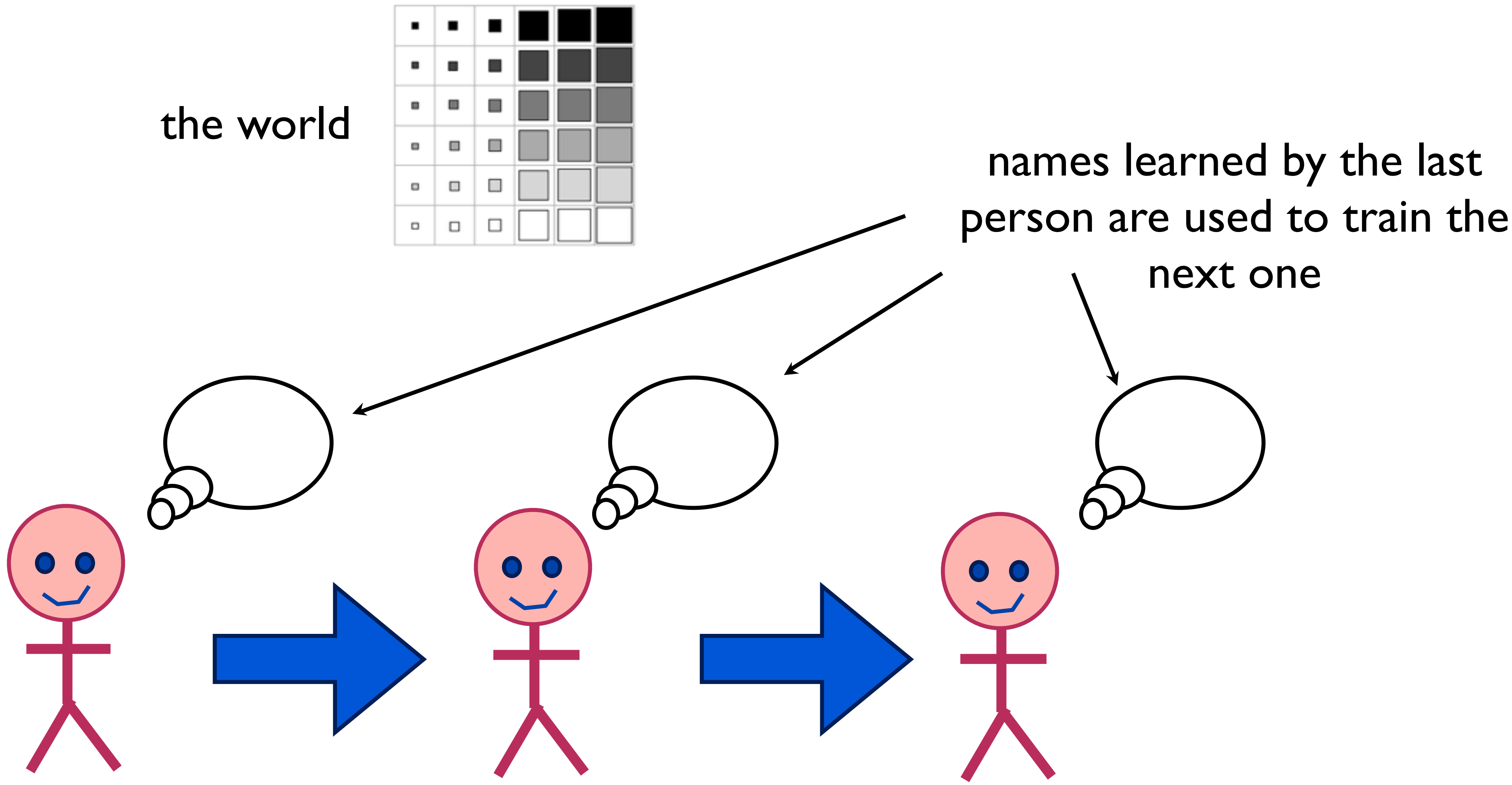


Name objects based
on their colour

Three worlds…

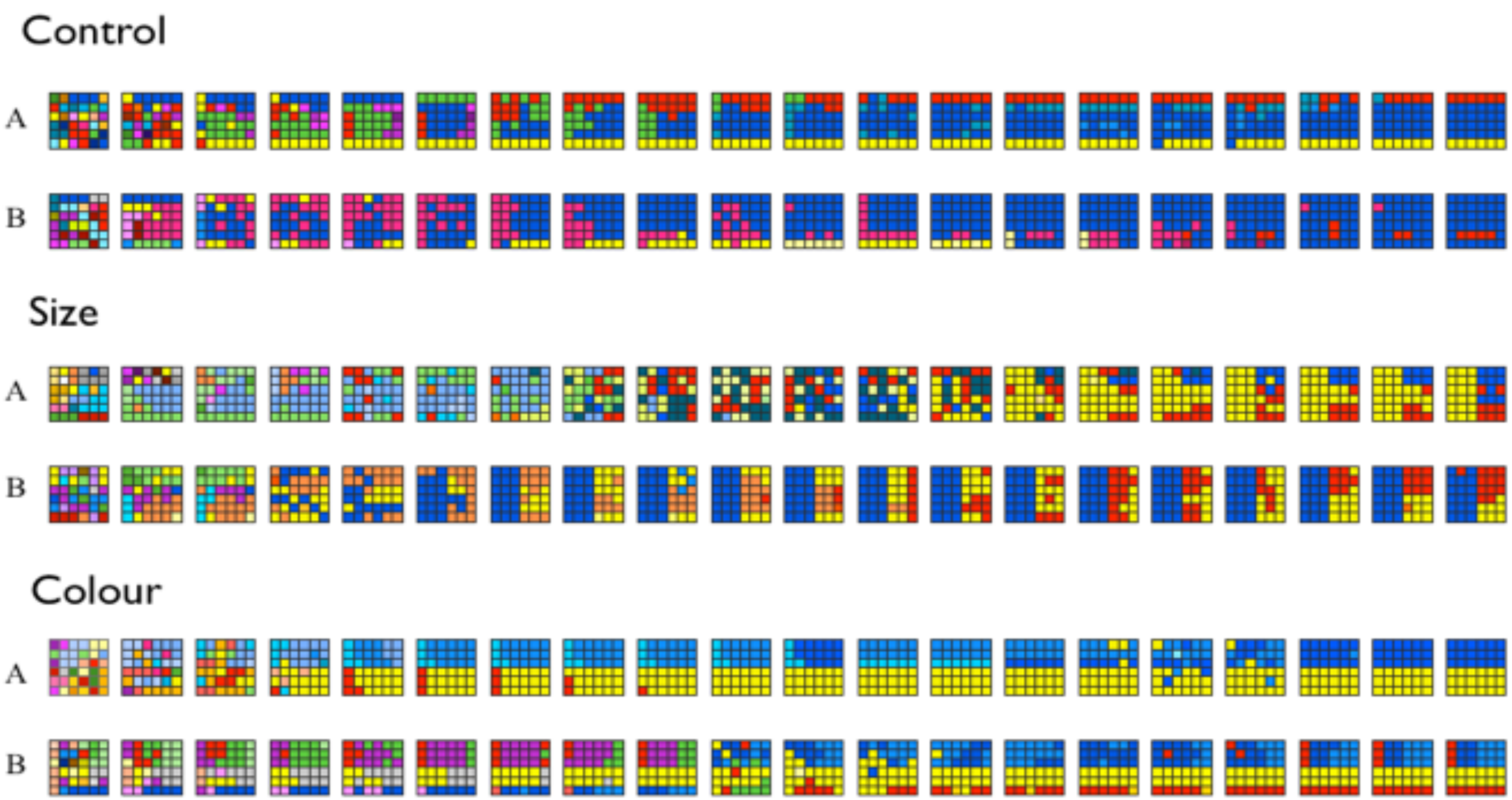


Doesn't really matter



Experiment!

A sequence of human participants each trying to learn the word names, using the previous person's responses as the training data



6 empirically observed Markov chains

	Expected Size	Expected Colour
Control	-0.0204	0.0618
Size	0.704	0.079
Colour	0.065	0.696

The naming systems adapt to match the environment that the speakers are exposed to

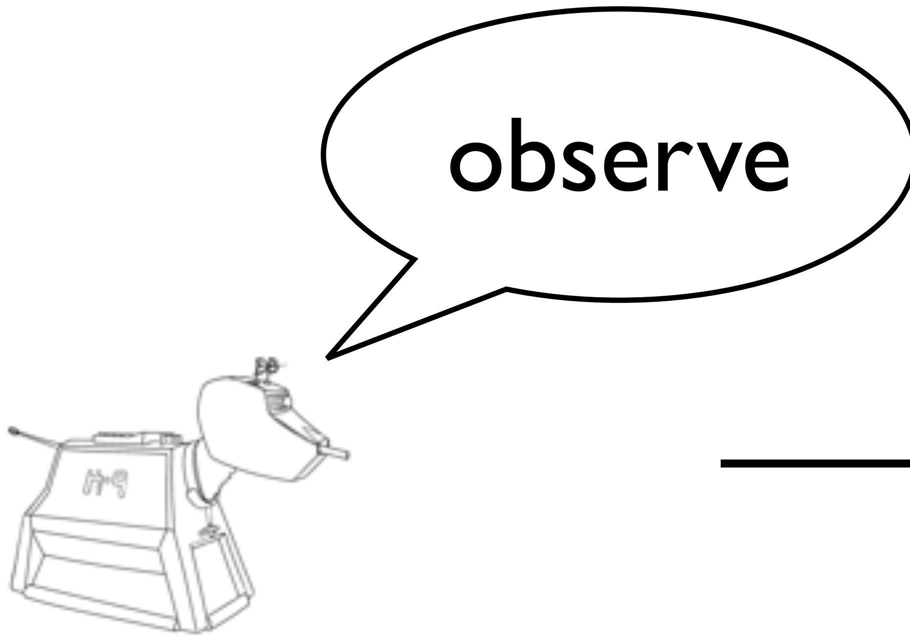
Yes, human communicative codes adapt to suit the cognitive biases of the learner <u>and</u> the operating environment

(Seems like this shouldn't need to be said, but unfortunately there's been some very silly overreach caused by people not reading the G&K proof in sufficient detail)

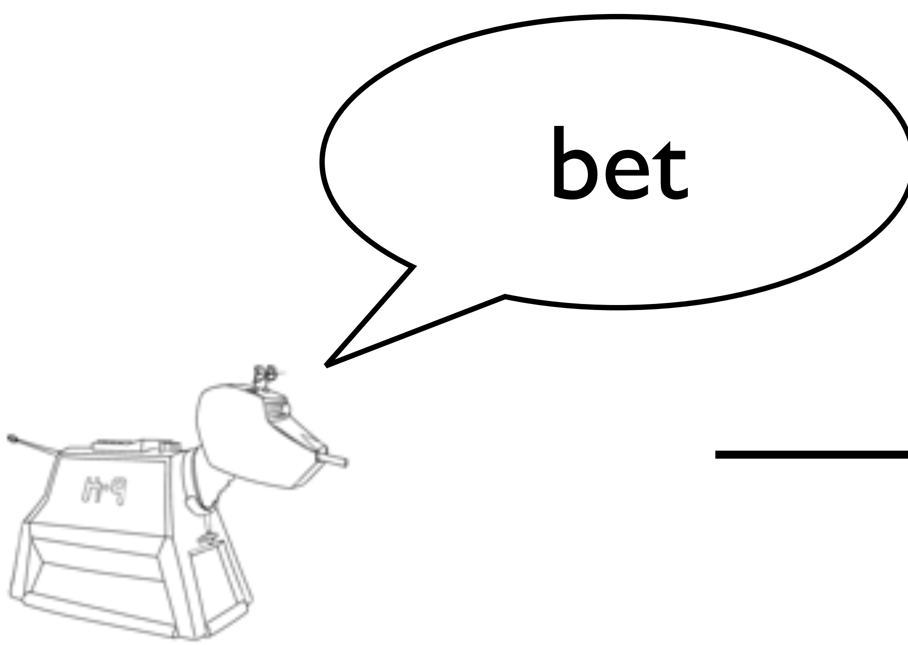
Decision making as stochastic planning: Making good choices in a changing world

Navarro, Newell & Schulze (under review)

The “observe or bet” task



information
without reward

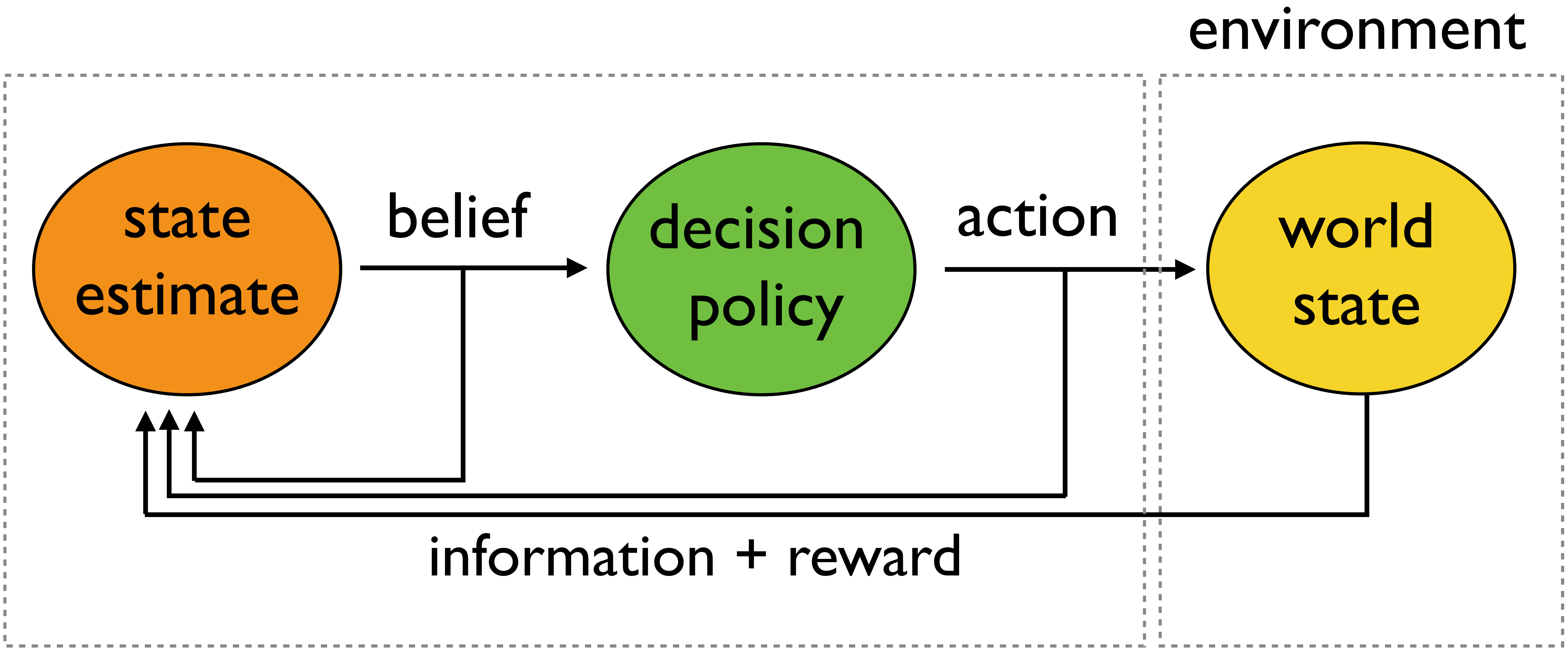


(hidden) reward
without information

(real world analogs)

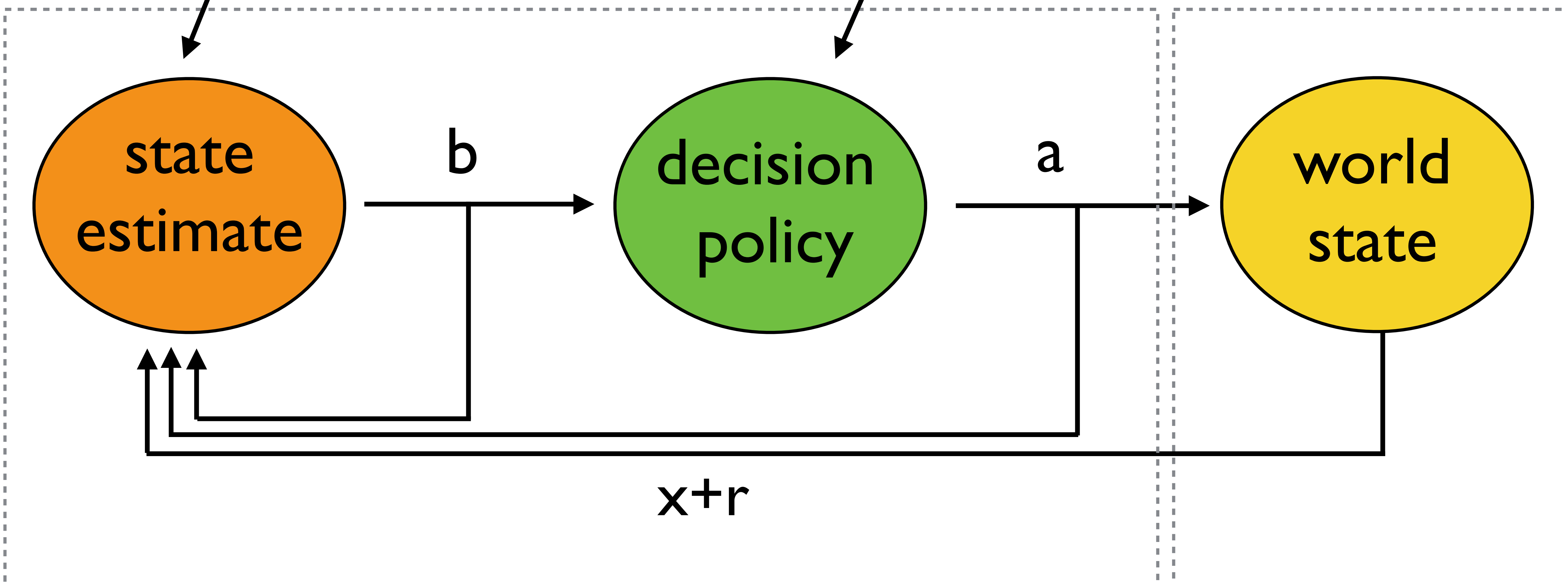
collecting evidence, doing background research, etc → information without reward

"delayed reward" situations where the results of your actions aren't obvious until much later → (hidden) reward without information


environment
state estimate
belief
decision policy
action
world state
information + reward

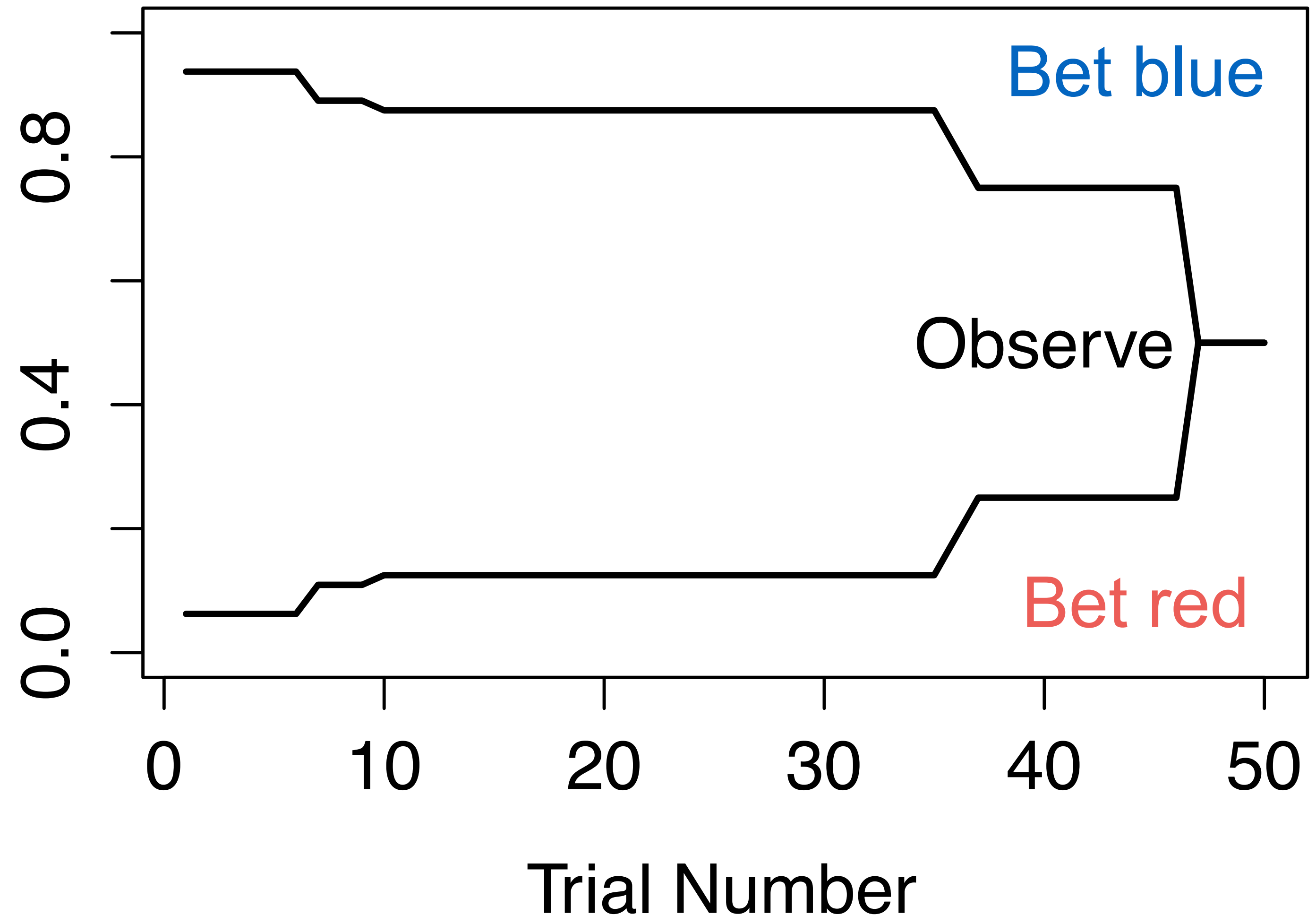
$$\mathbf{b} := P(\mathbf{w}|\mathbf{x})$$

$$U(\mathbf{b}) = R(\mathbf{b}) + \gamma \max_a \sum_{\mathbf{b}'} P(\mathbf{b}'|a, \mathbf{b})U(\mathbf{b}')$$



Optimal policy

Learner's
confidence that blue
is more likely

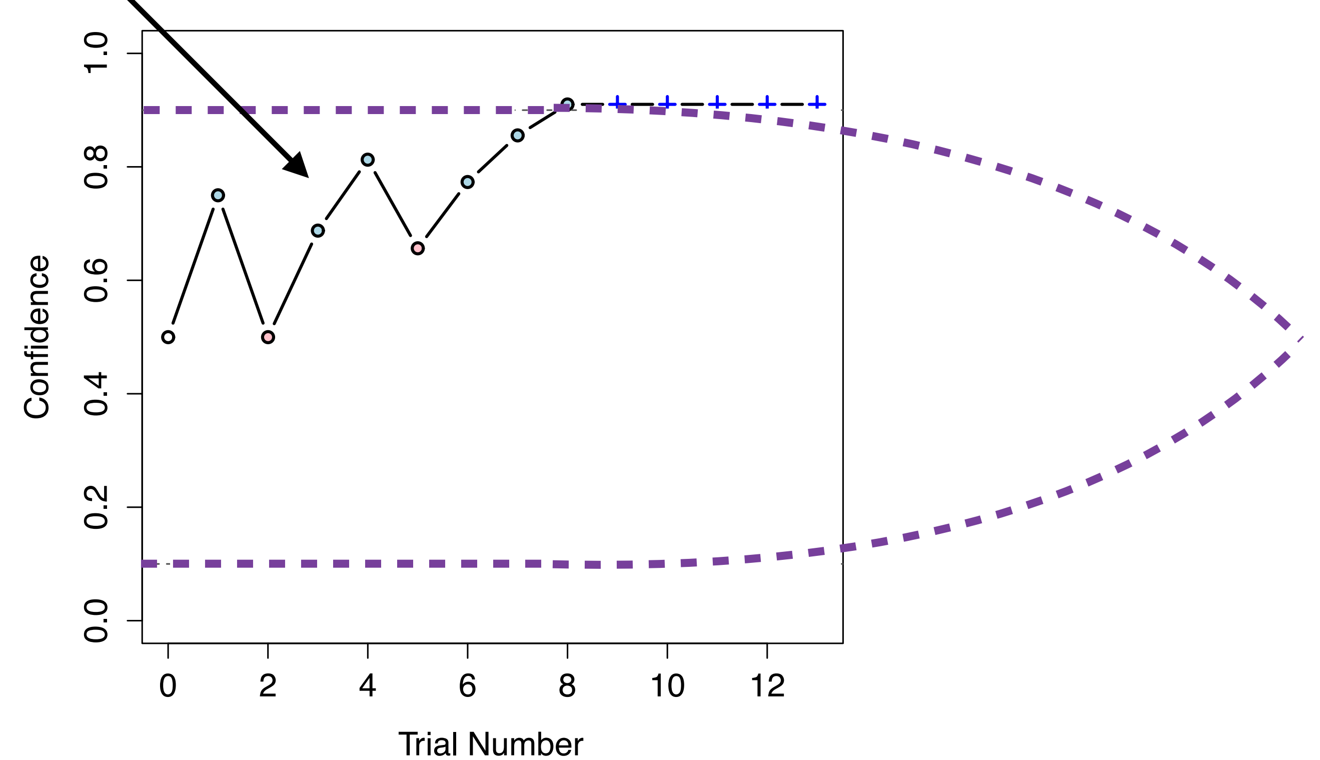


Behaviour of a naive Bayesian agent following the optimal decision policy

Behaviour of a naive Bayesian agent following the optimal decision policy

O O O O O O O B B B B B B B B B B B B B B B B B B

Keep making observations until you figure out the right betting strategy

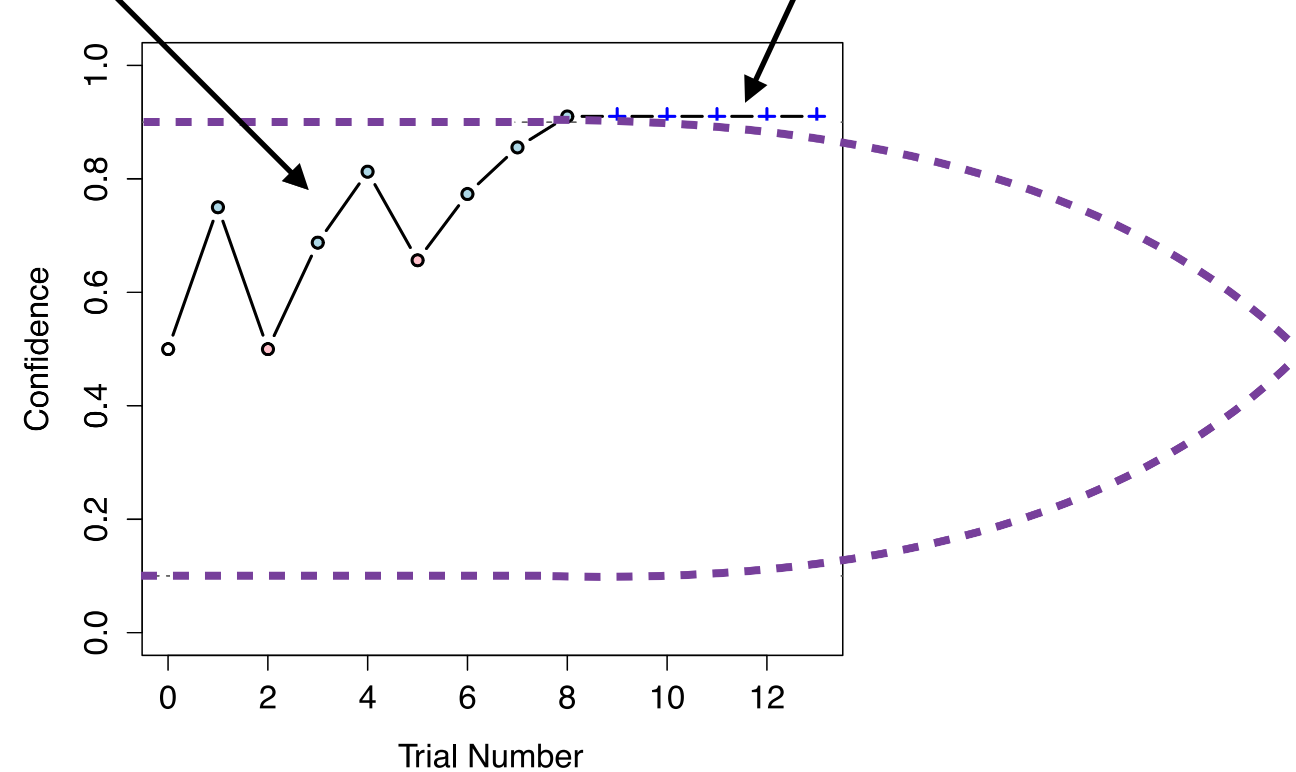


Behaviour of a naive Bayesian agent following the optimal decision policy

O O O O O O O B B B B B B B B B B B B B B B B B B

Keep making observations until you figure out the right betting strategy

Then trust blindly in your betting strategy, never revisiting



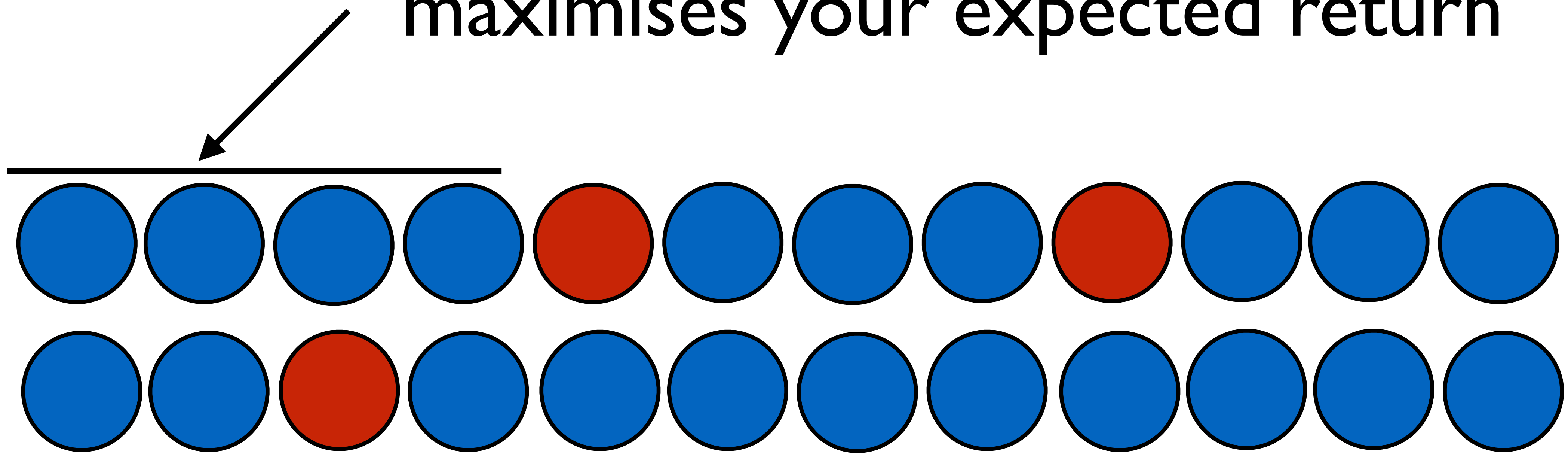
Humans don't do this

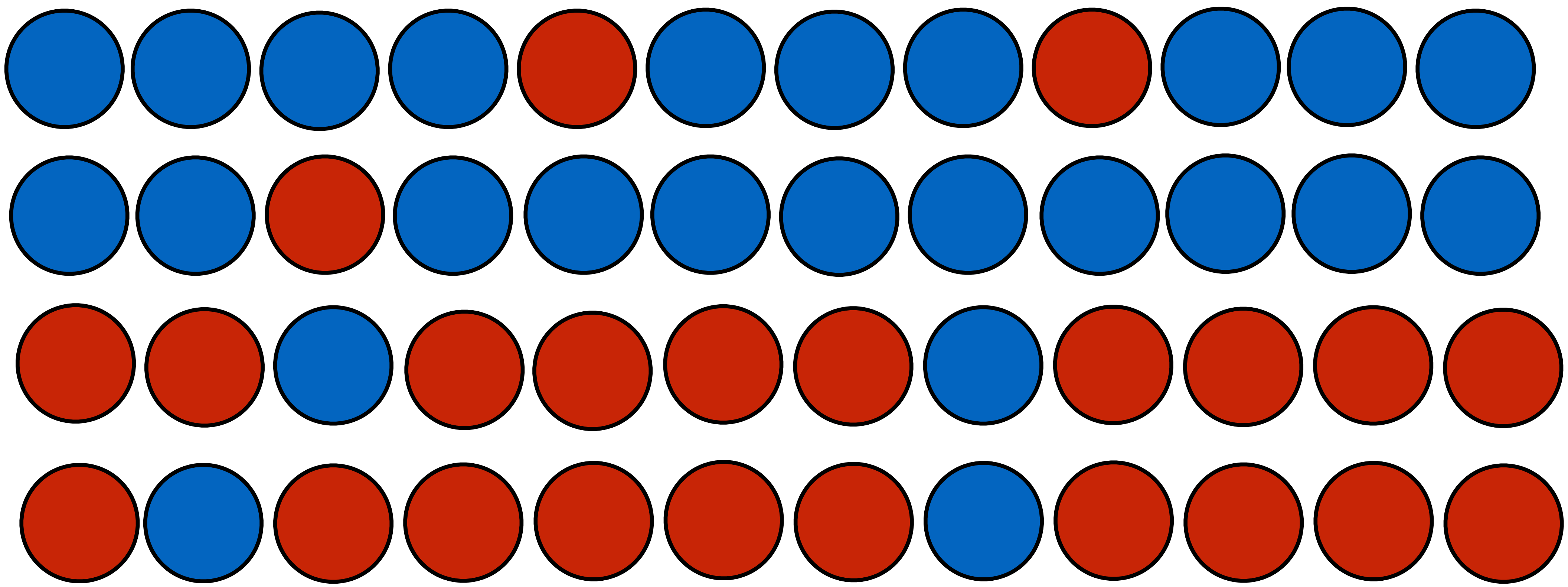
O O O O O O O B B B B B B B B B B B B B B B B B B

O O O O O O O B B B B B B O O B B B B B B O B B B

Humans don't seem to trust their betting strategy: they constantly "check" to see if it is still working.

“front loading” all observations maximises your expected return





unless, of course, the rules can change…

Learning when changes can't happen

Static world: today's posterior is tomorrow's prior

$$P(\theta|\mathbf{x}_t) \propto P(x_t|\theta)P(\theta|\mathbf{x}_{t-1})$$

Learning when changes can happen

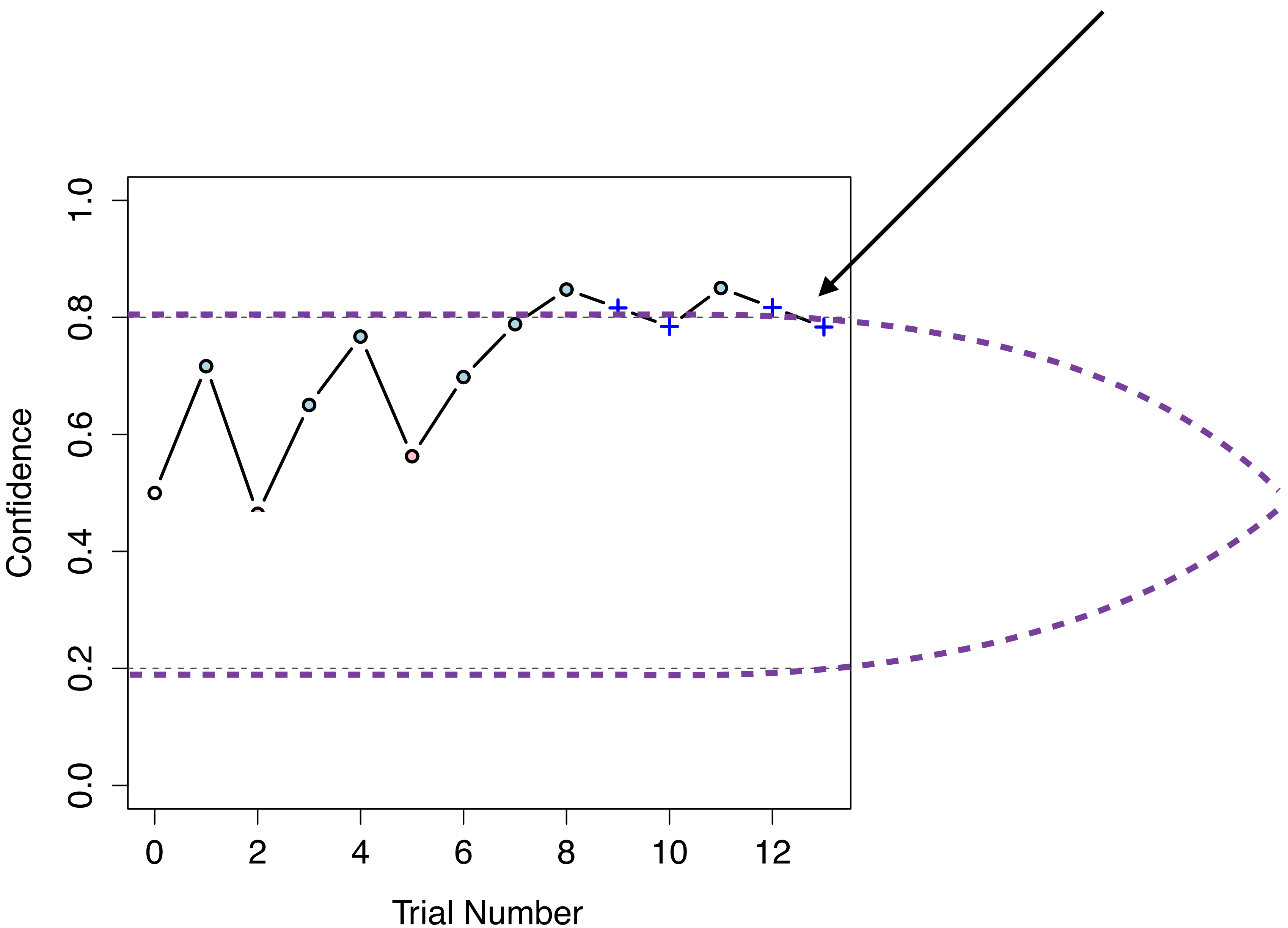
Static world: today's posterior is tomorrow's prior

$$P(\theta|\mathbf{x}_t) \propto P(x_t|\theta)P(\theta|\mathbf{x}_{t-1})$$

Dynamic world: today's posterior shapes tomorrow's prior, but needs to track changes that happen in the interim…

$$P(\theta_t|\mathbf{x}_t) \propto P(x_t|\theta_t) \int_0^1 P(\theta_t|\theta_{t-1})P(\theta_{t-1}|\mathbf{x}_{t-1})\ d\theta_{t-1}$$

Rational agents operating in dynamic environments produce human-like strategy <u>switching</u>

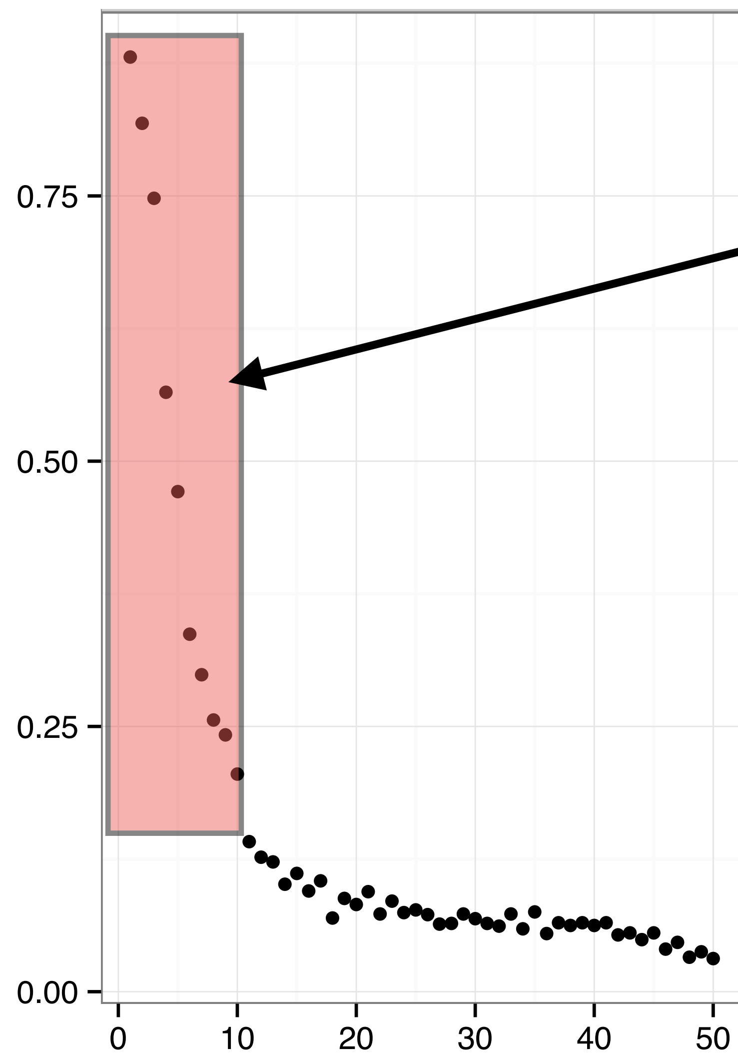


Back to the lab!

What happens when humans do the task in a stationary vs dynamic environment?

Static condition

Proportion of participants observing

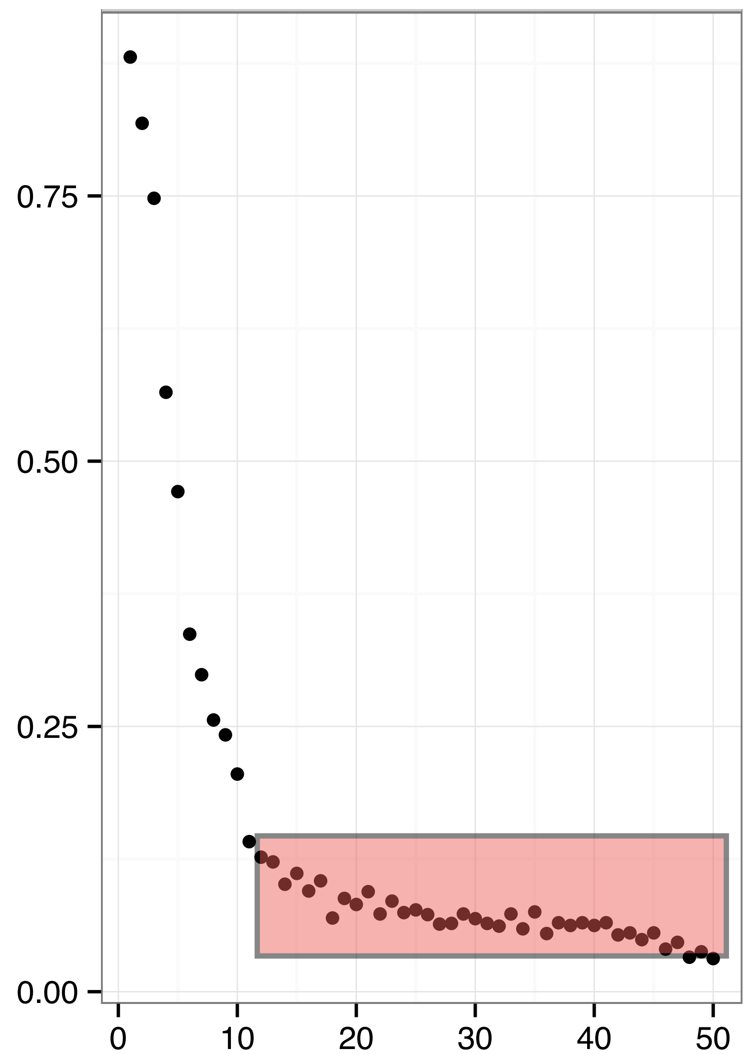


Trial Number

As expected, initial evidence collection to reach threshold…

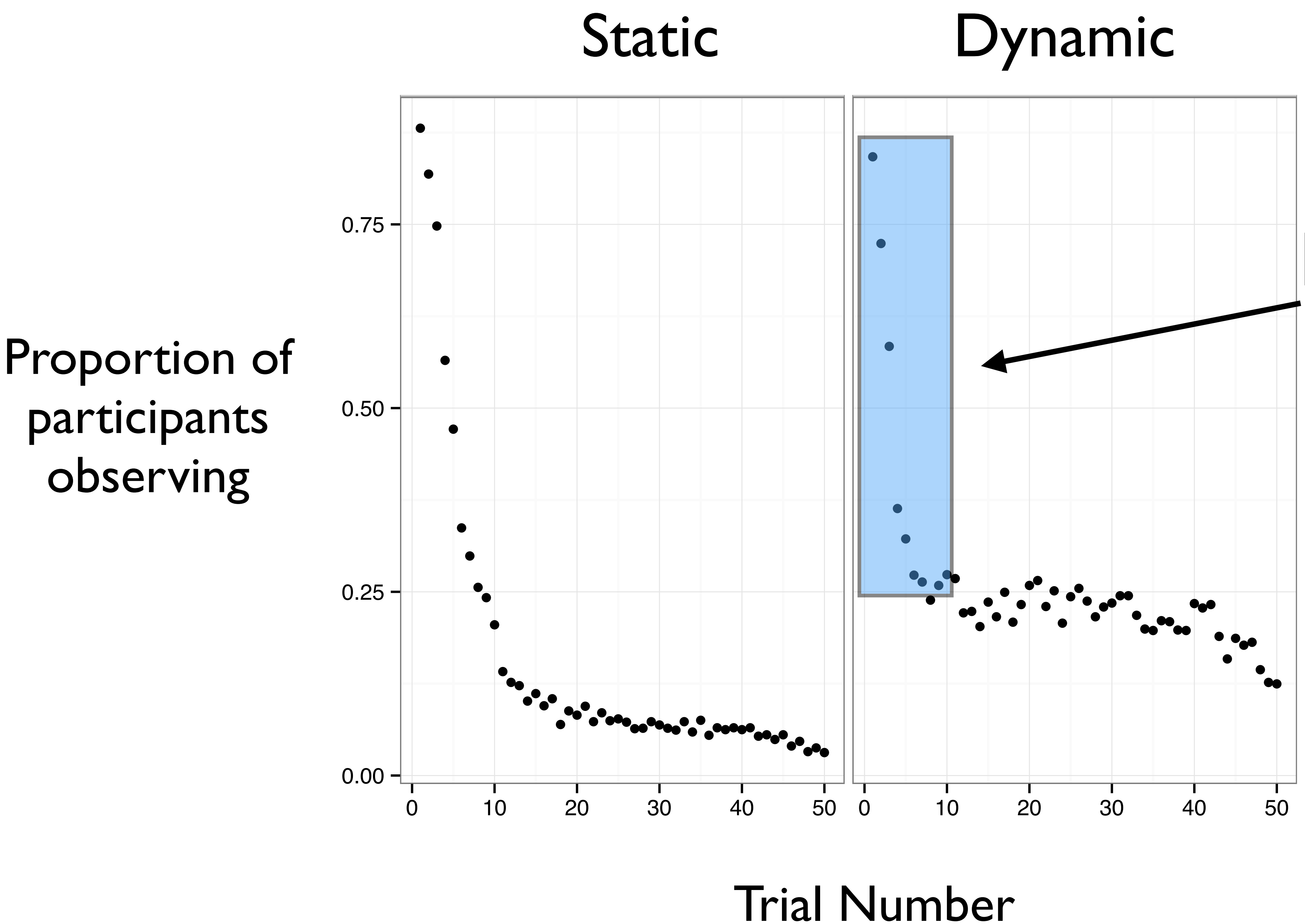
Static condition

Proportion of participants observing

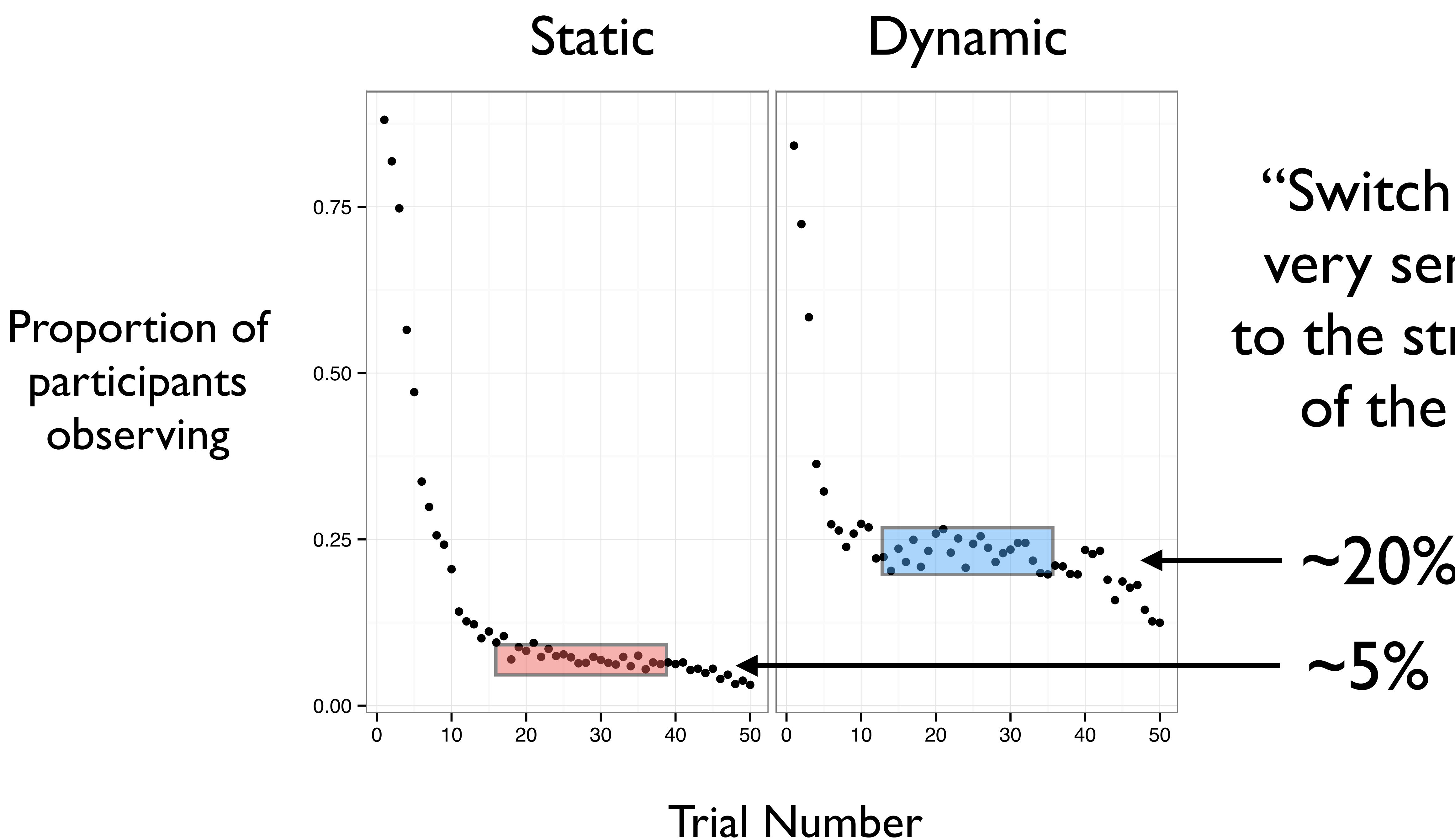


Trial Number

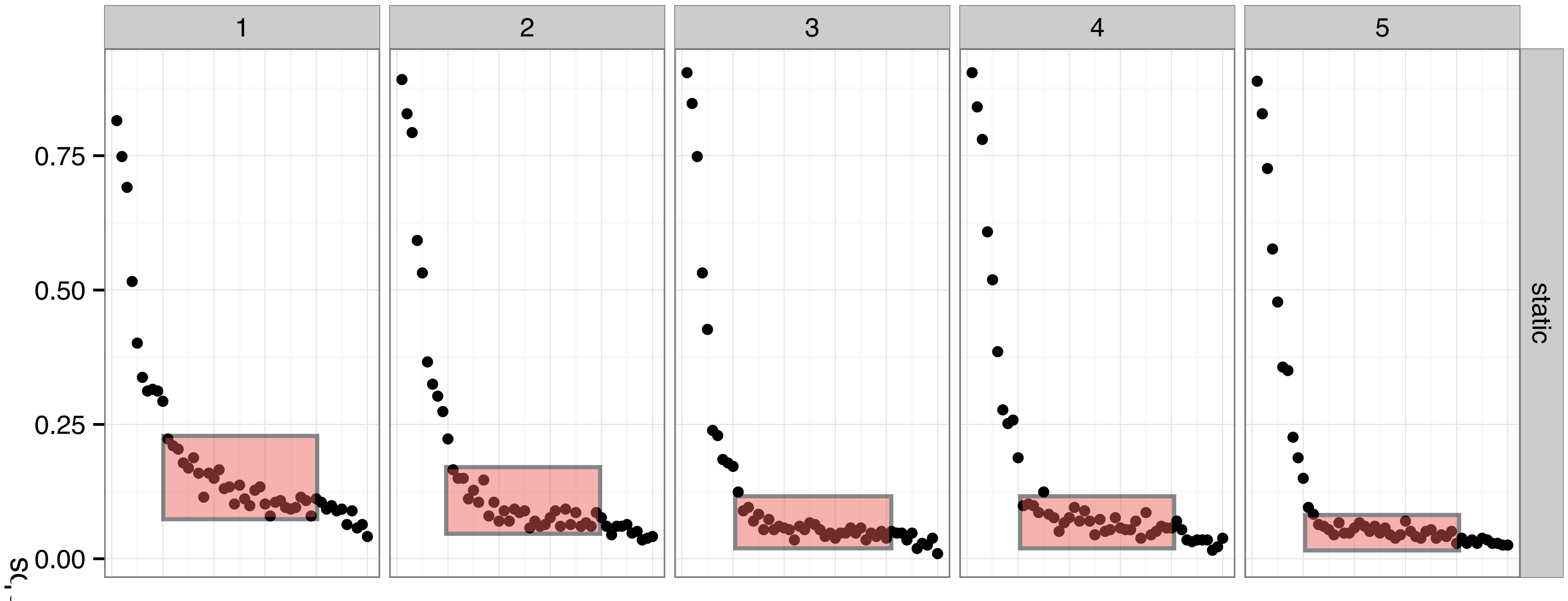
Followed by mostly bets



Similar <u>initial</u> pattern for the dynamic condition…



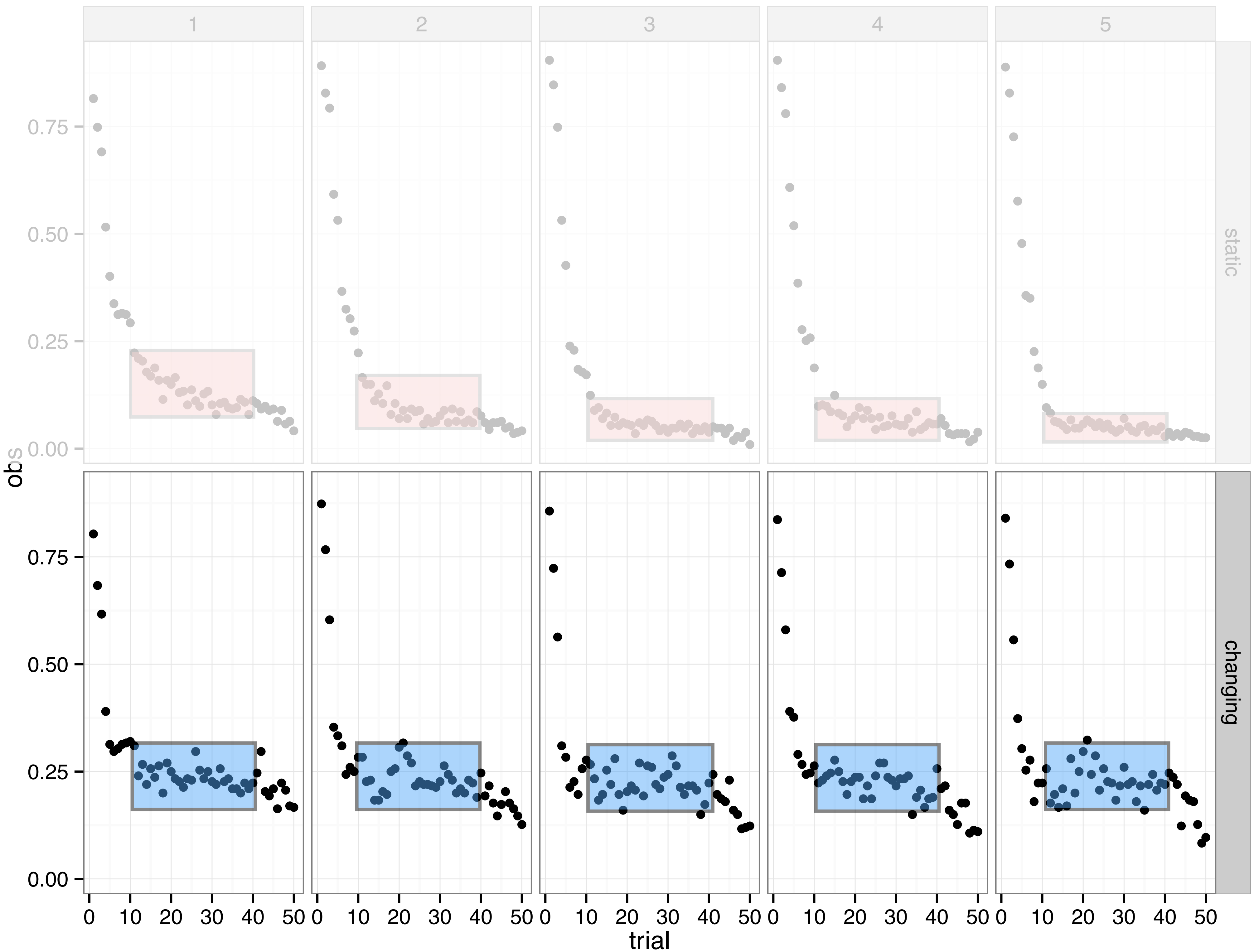
"Switching" is very sensitive to the structure of the task



static condition

On game 1, people are not front loading all the observations

By game 5, the decision policy is close to optimal

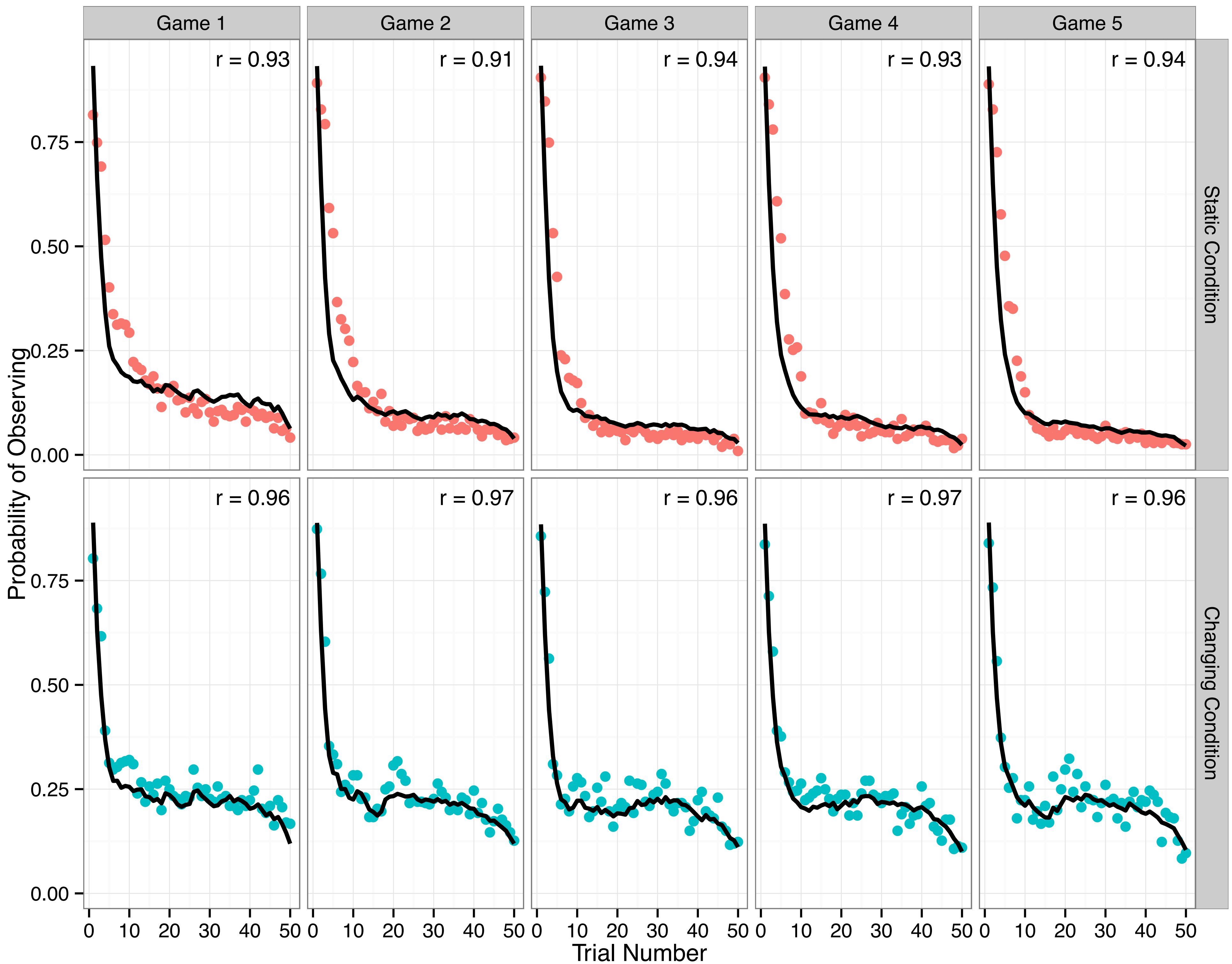


static condition

dynamic condition

No learning across games in the dynamic condition because people were already pretty well-calibrated?

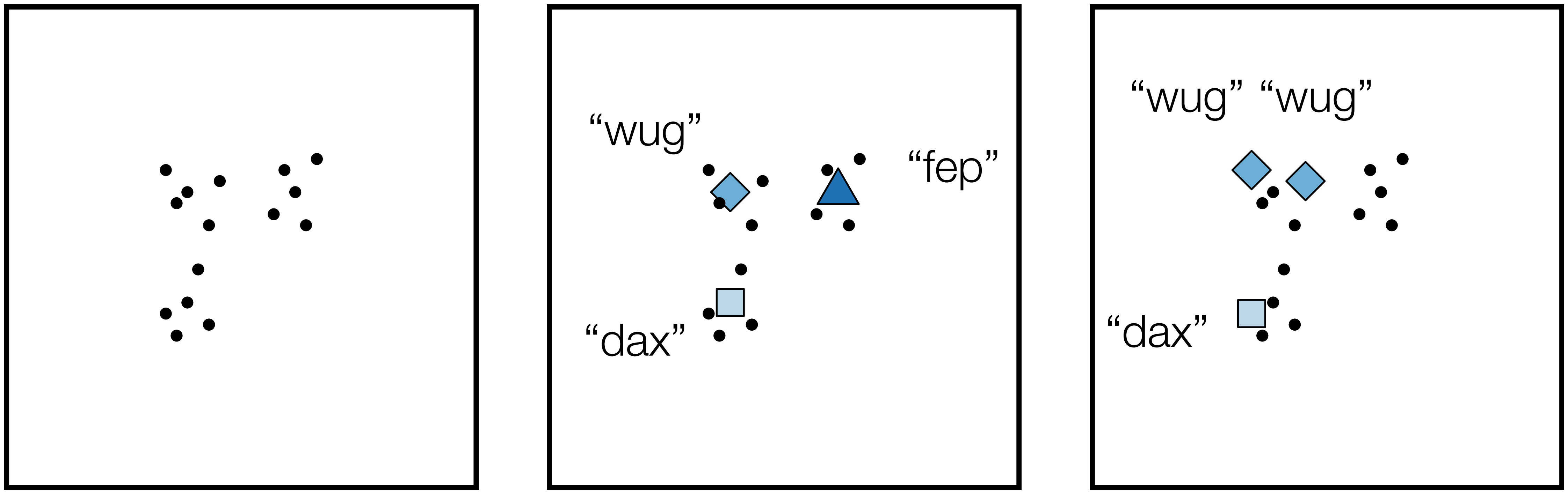
Humans closely mirror approximate versions of the POMDP model



The “irrational” strategy that people use actually is optimal in a changing environment.

(we also ran some experiments showing that people <u>know</u> that their default strategy is suboptimal in a stationary world, but that you need direct experience to <u>learn</u> not to expect change)

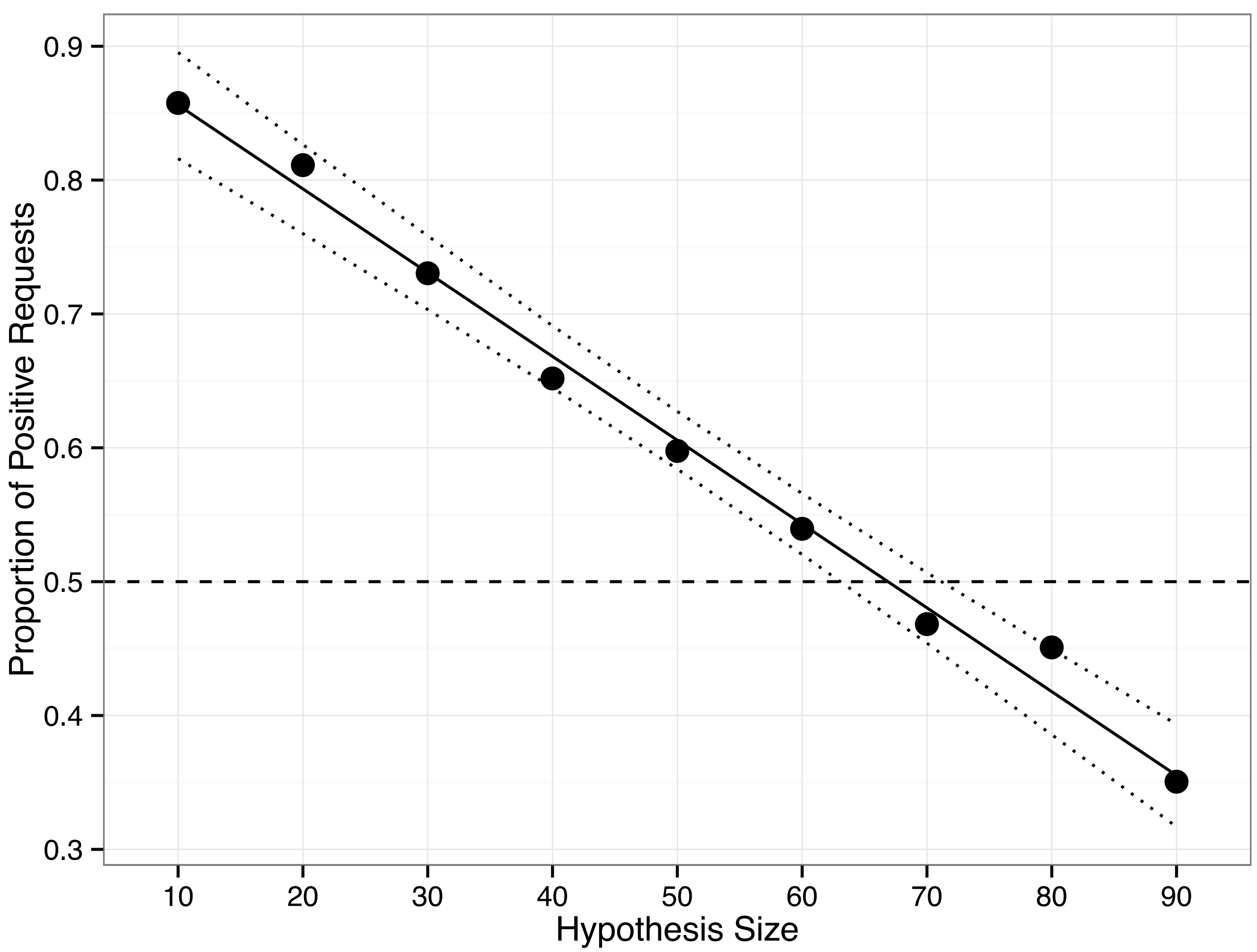
What else?

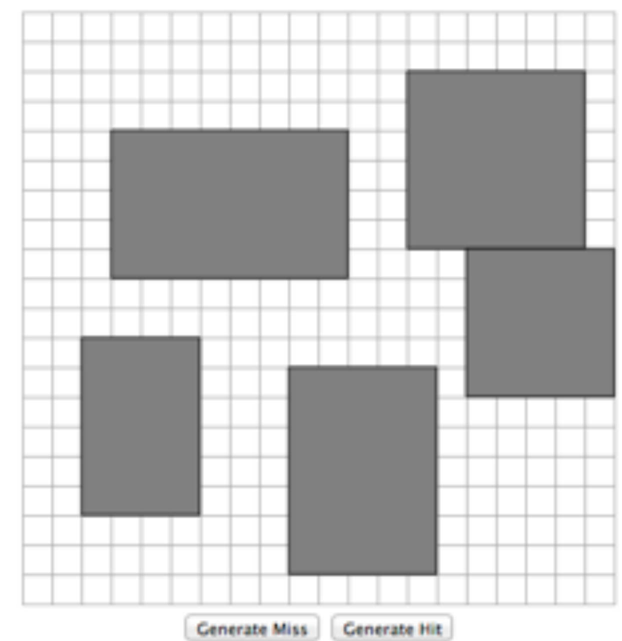


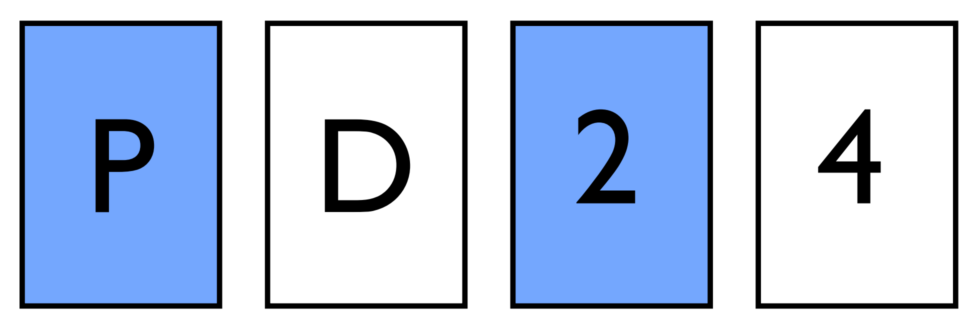
You can understand how people learn from sparsely labelled data

Vong, Perfors & Navarro (under review, 2014)

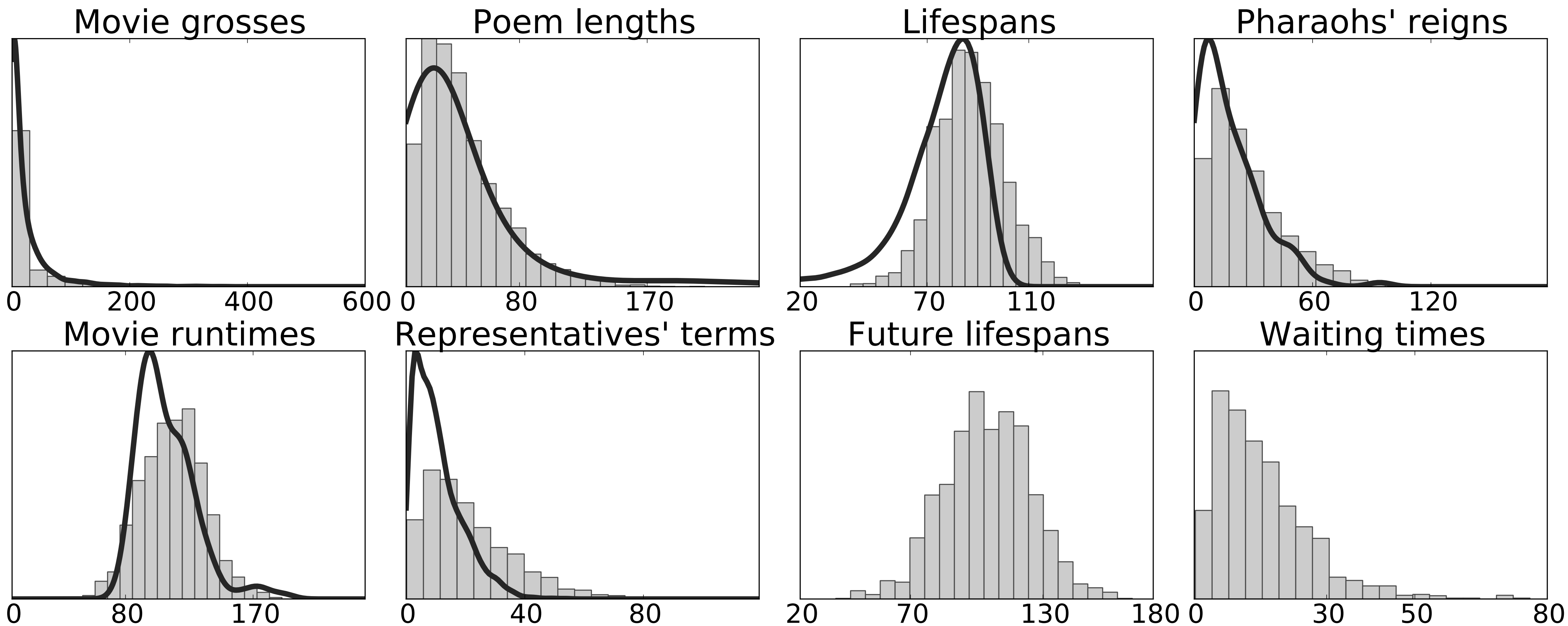
You can show that intuitive hypothesis testing maximises expected information gain







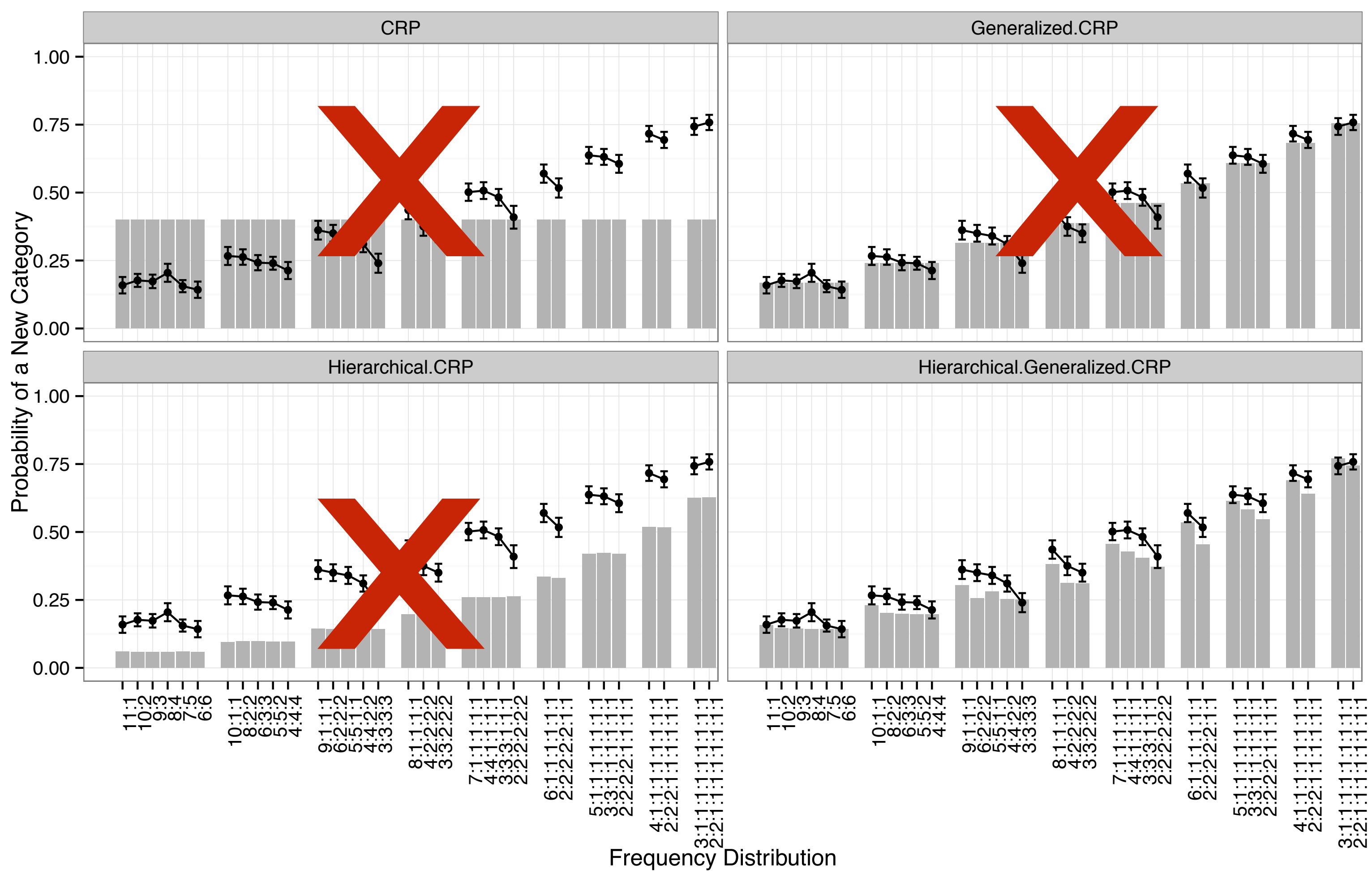
Hendrickson, Perfors & Navarro (under review, 2014)

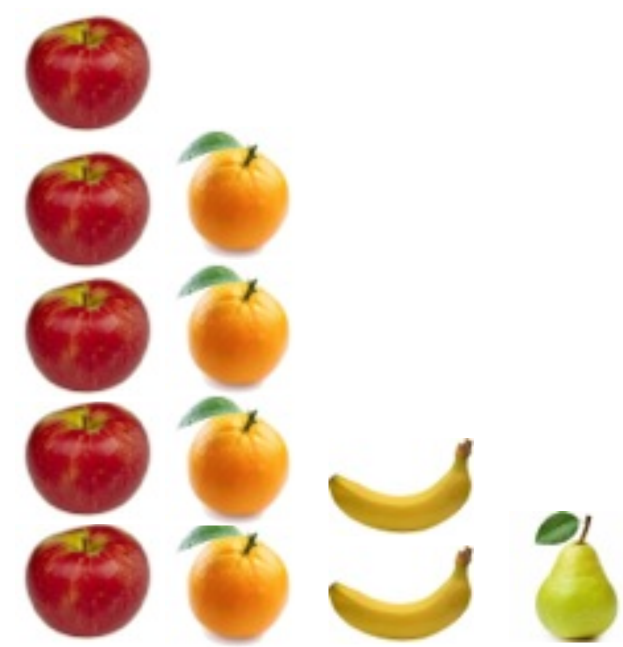


You can infer people's priors from judgments and compare them to veridical distributions

Tauber, Navarro, Perfors & Steyvers (in preparation)

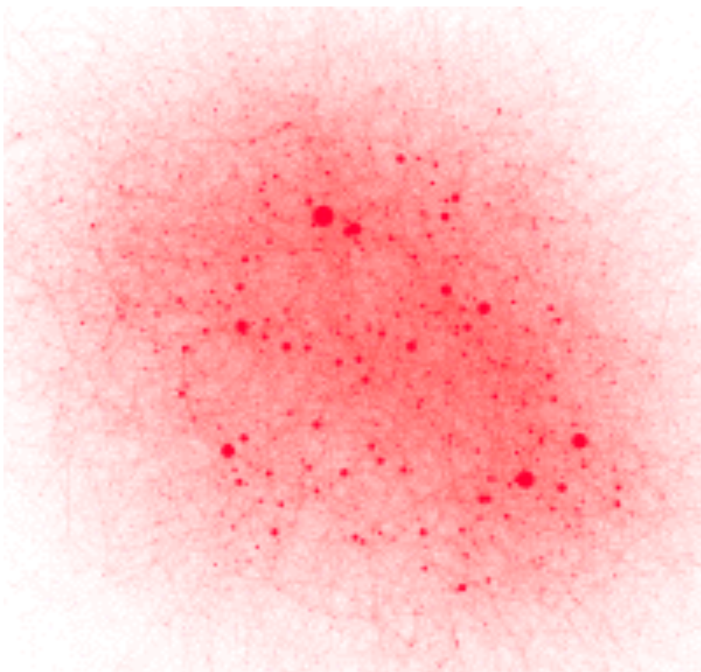
You can describe how people use <u>frequencies</u> to infer <u>distributions</u>





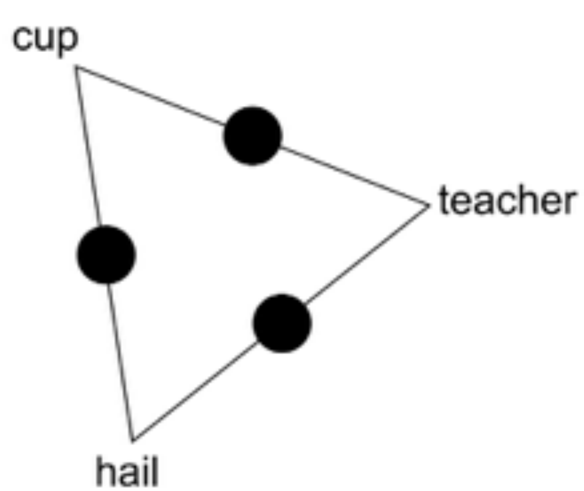
$$P(\mathrm{new}|\boldsymbol{n}) = \int_0^1 \int_{-\alpha}^{\infty} \frac{\theta + K\alpha}{\theta + N} P(\theta, \alpha|\boldsymbol{n})\, d\theta\, d\alpha$$

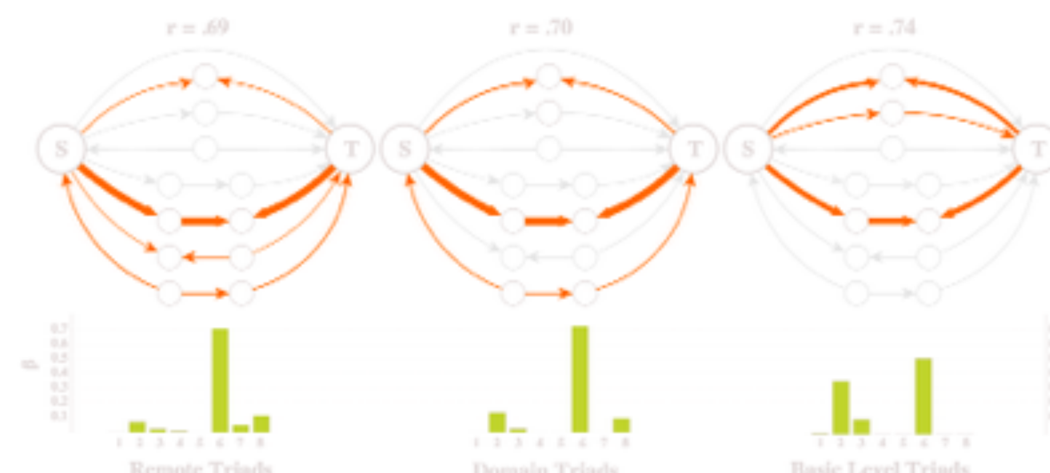
Navarro & Kemp (in preparation)



You can create lexical semantic networks (~12k nodes, ~100k edges, ~3m responses) …

… and predict the similarity between apparently arbitrary concepts

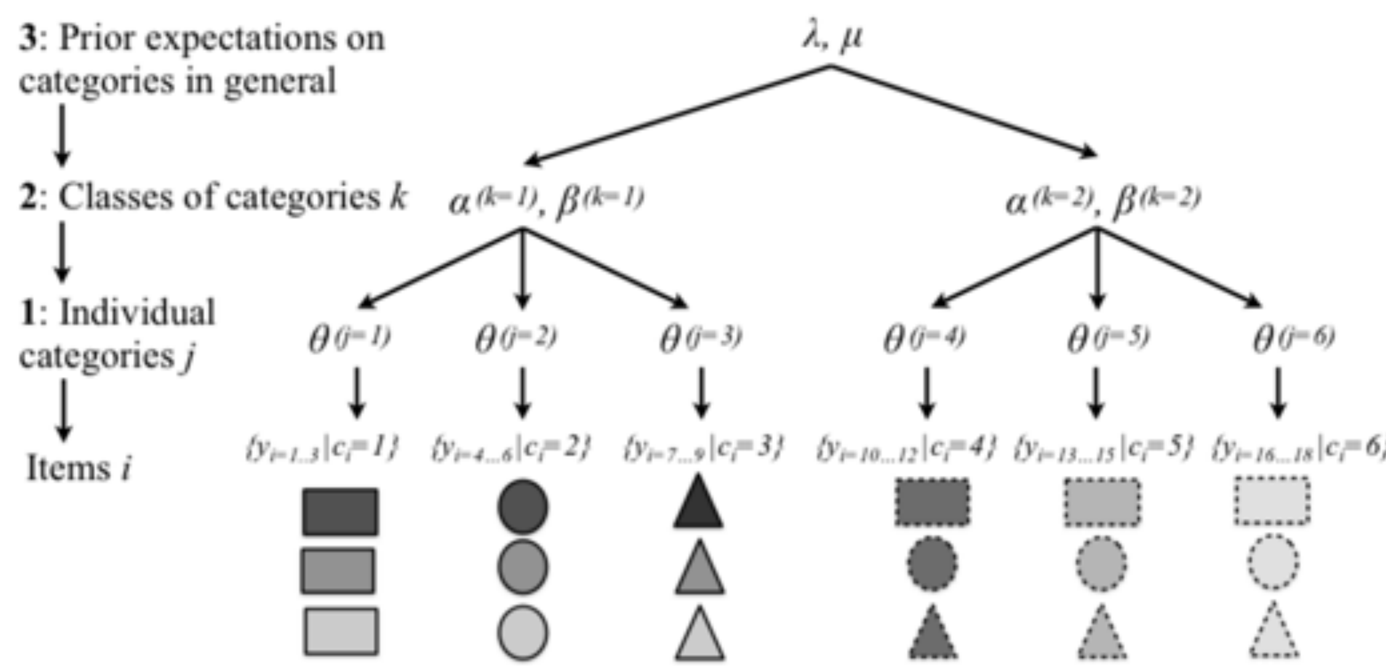




De Deyne, Navarro, Perfors & Storms (2012)
De Deyne, Navarro & Storms (2013)

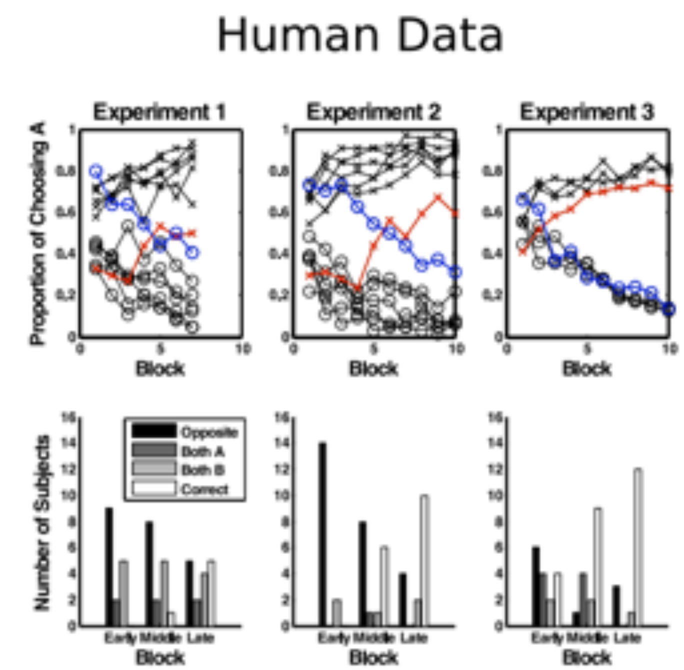
You can explore how people acquire different kinds of biases for different ontological kinds

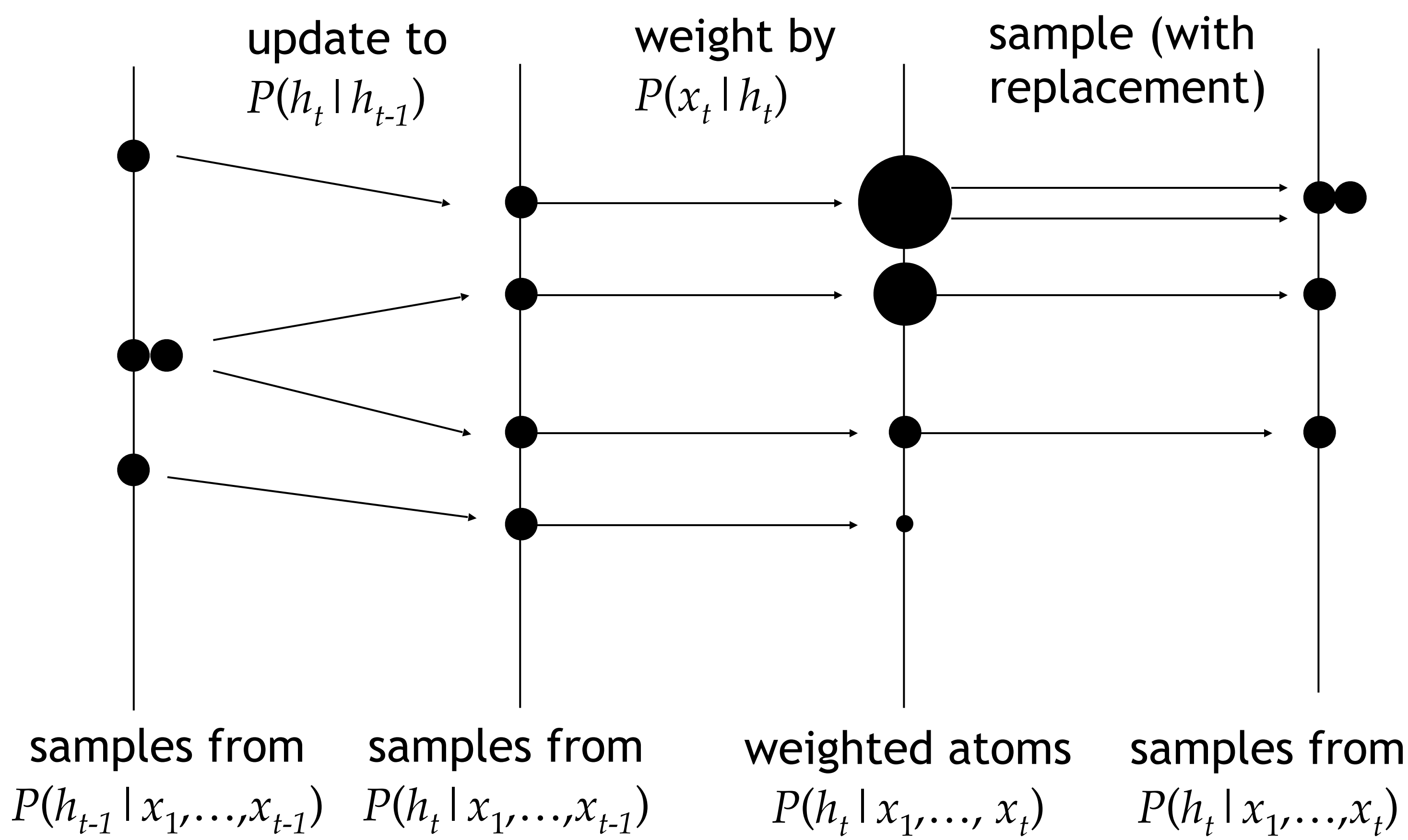
Class	Category	Instance	Shape	Color
Object	Cat	1	*Quadruped*	Tabby
		2	*Quadruped*	White
		3	*Quadruped*	Grey
	Ball	1	*Sphere*	Blue
		2	*Sphere*	Red
		3	*Sphere*	Green
Substance	Mud	1	Pile	*Brown*
		2	Puddle	*Brown*
		3	Spherical	*Brown*
	Flour	1	Pile	*White*
		2	Cloud	*White*
		3	Scatter	*White*

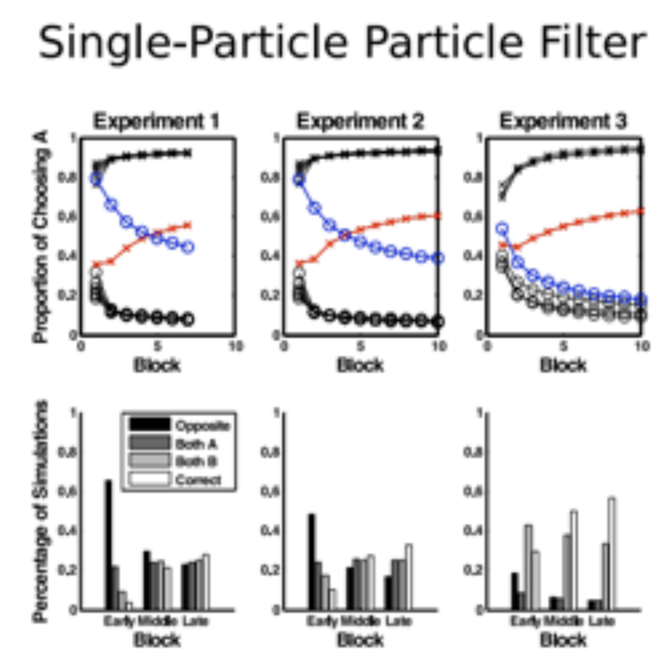


Perfors, Navarro & Tenenbaum (under review)

You can look at how people approximate the solutions to computationally intractable problems

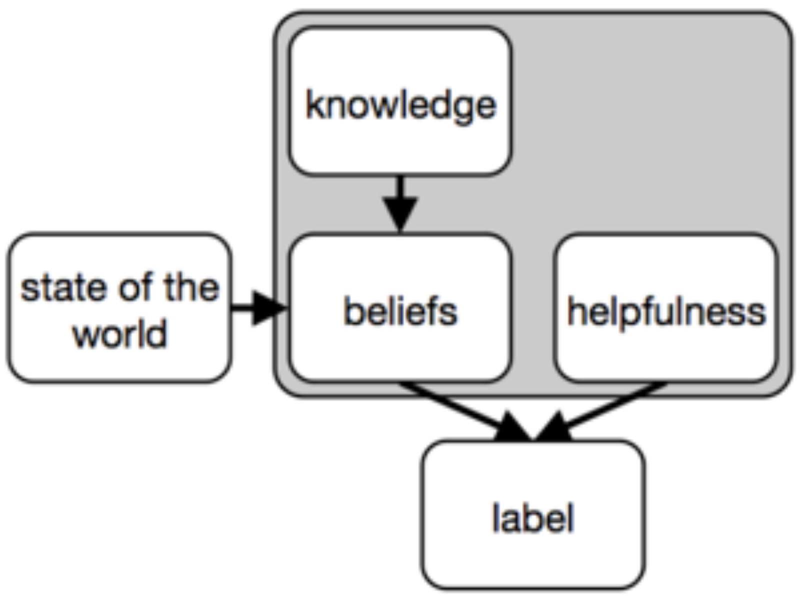


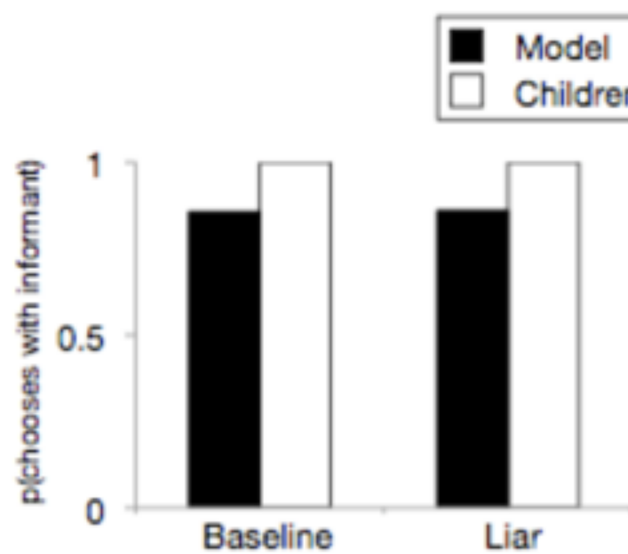




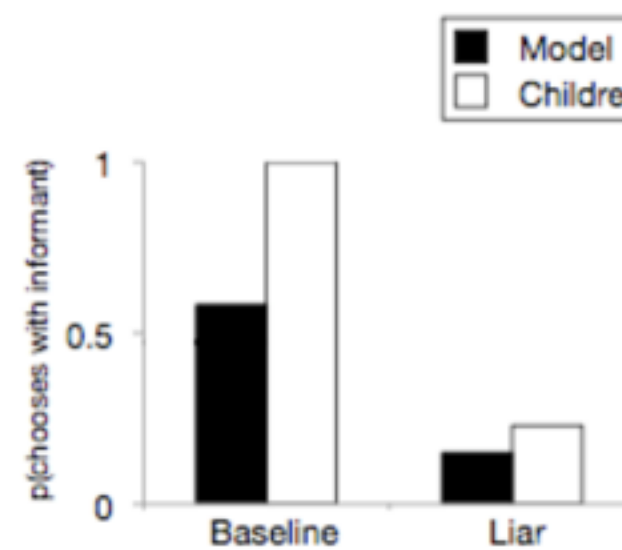
Sanborn, Griffiths & Navarro (2010)

You can even work out the statistical explanation for why 4 year olds catch liars better than 3 year olds





3 yrs



4 yrs

Shafto, Eaves, Navarro & Perfors (2012)

So, what <u>does</u> statistics tell us about human cognition?

So, what <u>does</u> statistics tell us about human cognition?

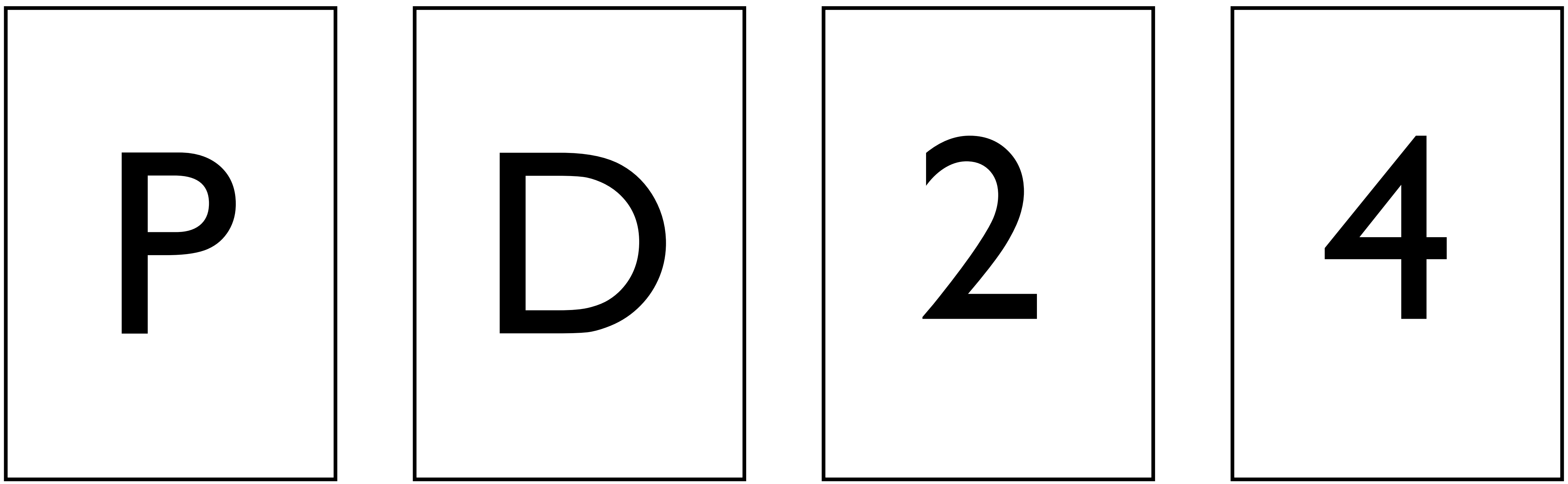
A lot.

Done.

Explaining intuitive hypothesis testing using information theory

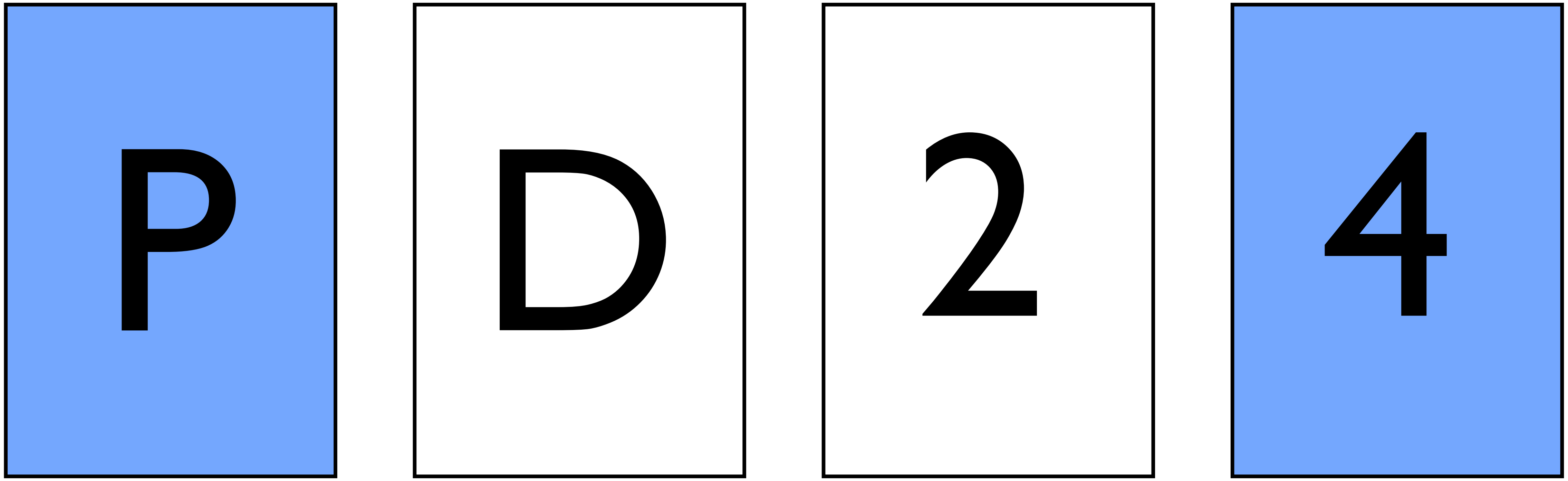
Hendrickson, Perfors & Navarro (under review)
Navarro & Perfors (2011)

Hypothesis testing



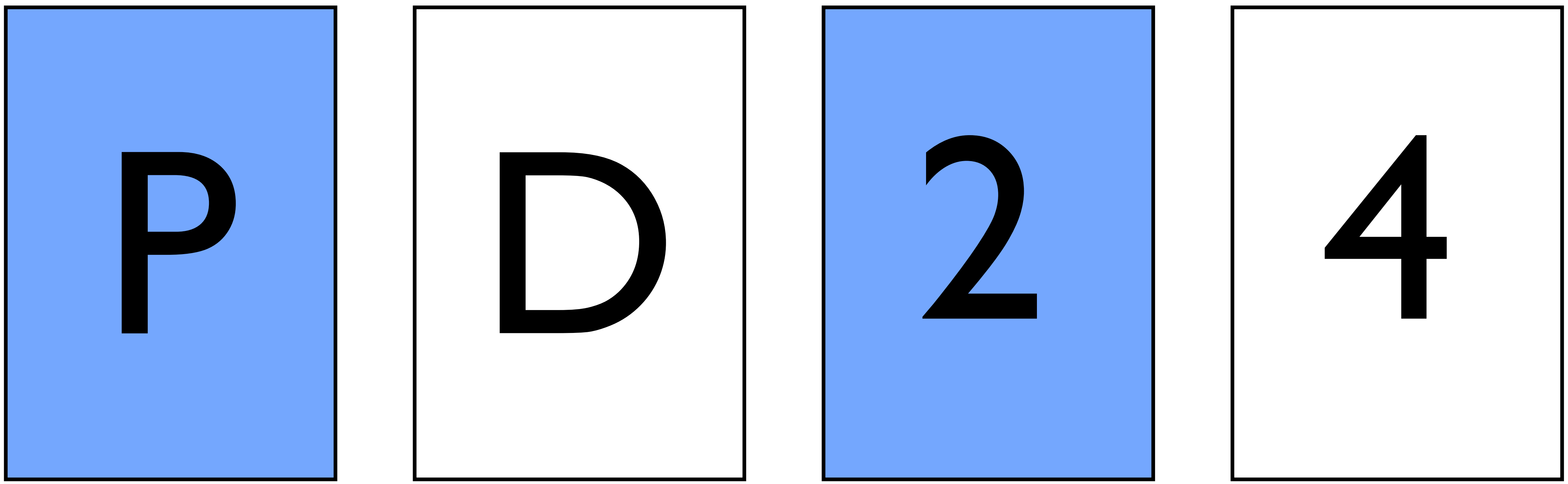
- Cards have a letter and a number
- Test the hypothesis that "if P then 2"

Falsificationist answer



Popper (1937)

The “positive test strategy”



Wason (1968)

Positive tests maximize expected information gain…

- Klayman & Ha (1987): Positive tests more likely to produce belief change in the rule learning game
- Oaksford & Chater (1994): Positive tests in the four-card selection tasks yield maximum information gain about a hypothesis
- Austerweil & Griffiths (2007): Positive tests in deterministic rule learning tasks are optimal in the Bayesian sense
- Navarro & Perfors (2011): Positive tests yield faster convergence to the true hypothesis under realistic assumptions about limited working memory capacity

… but only when the expected "size" of the true category is small

- Small categories contain a minority of entities:
  - Few animals are PETS
  - Few numbers are DIVISIBLE BY 10

- Large categories are the opposite:
  - Most animals are MOTILE
  - Most numbers are COMPOSITE

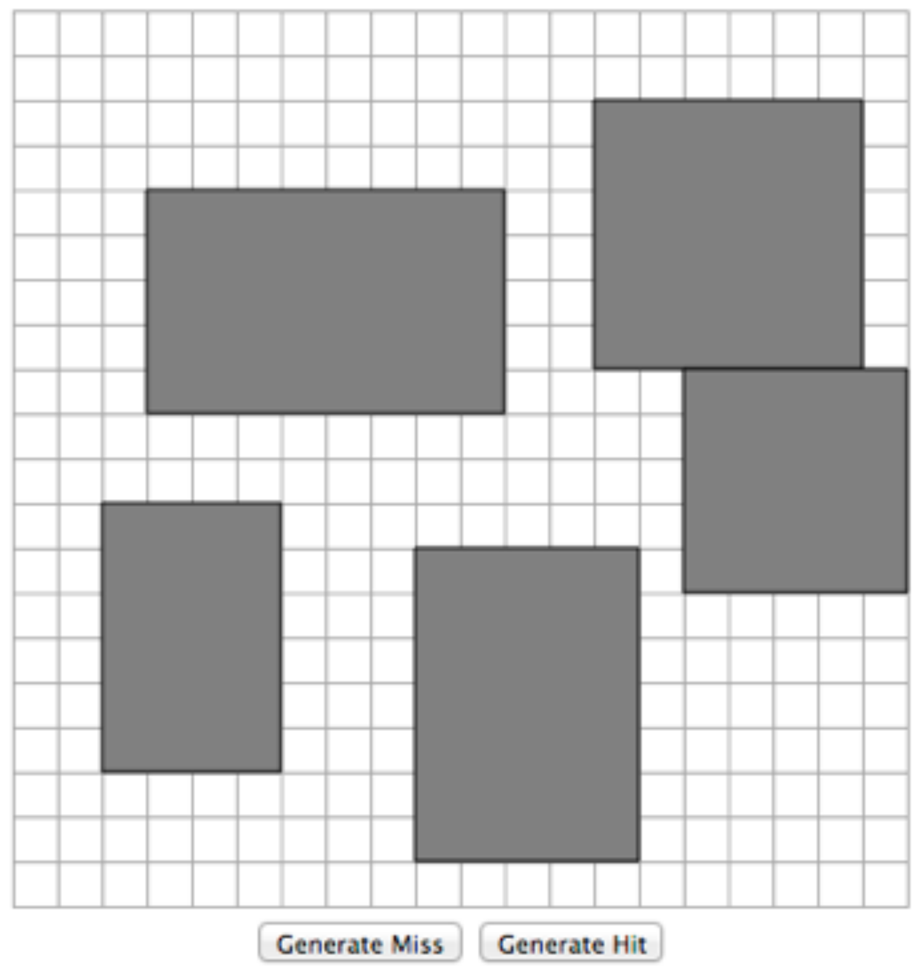
I demand another empirical test!

What happens when you systematically manipulate category size?

Hendrickson, Perfors & Navarro (under review)

The battleships task!

Click Generate Miss and Generate Hit to get information about where the hidden ships are located. Move the grey ships to show where you think the hidden ships might be located.

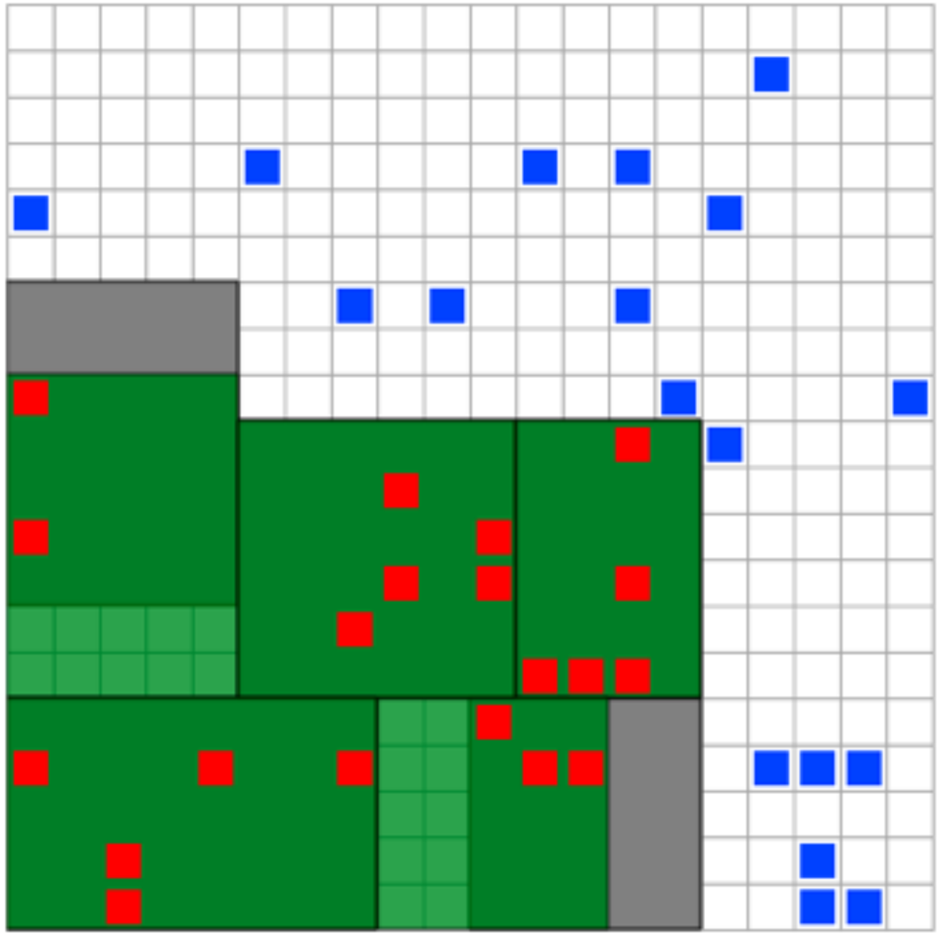


When you think you know where the hidden ships are positioned, move the grey ships to those positions and click Done. Your score will be calculated based on how close you are to correct and how many queries you made.

Done

The battleships task!

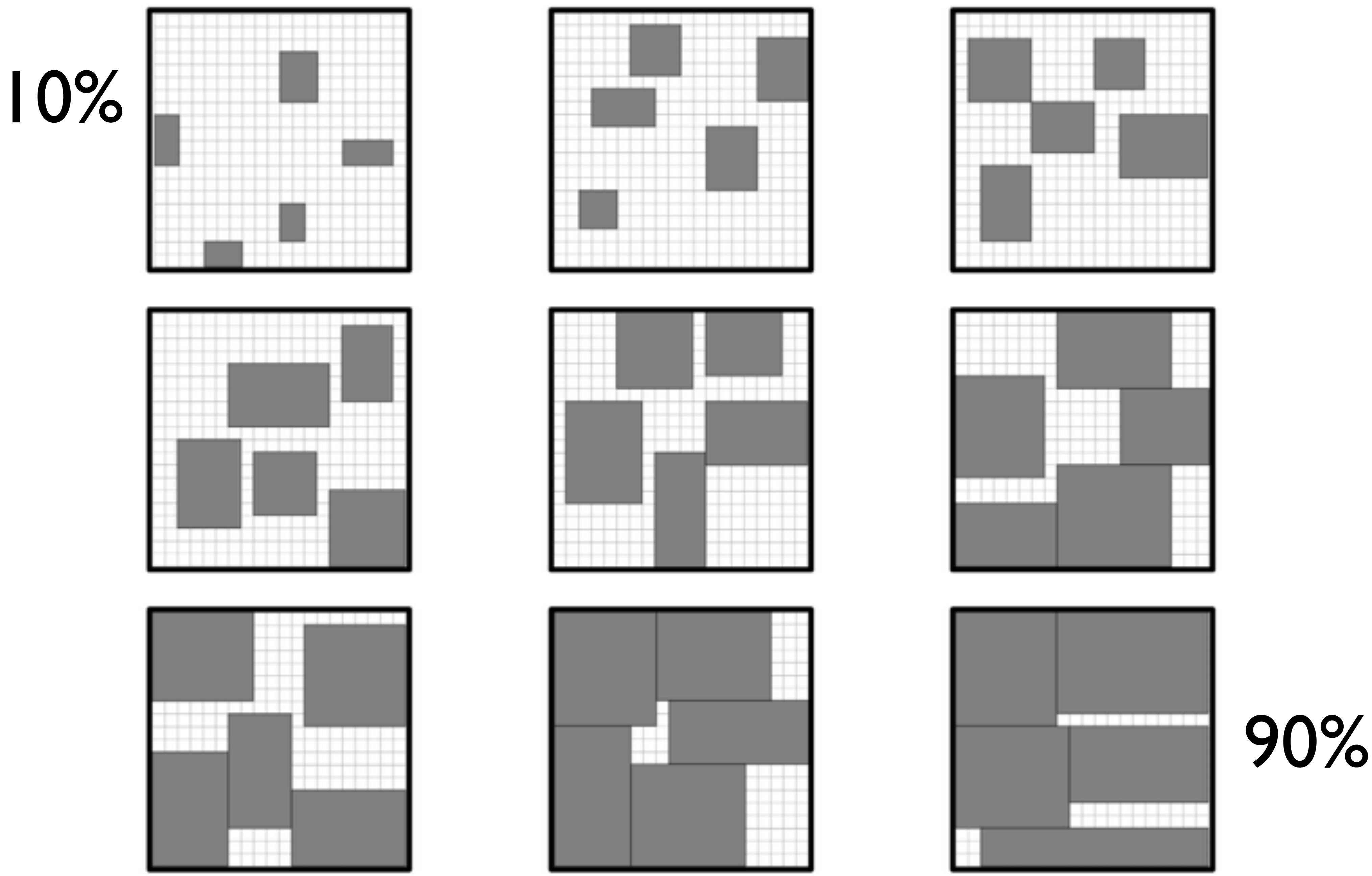
Shown in green below are where the hidden ships were.



Your score on this round is 1111 points out of 10,000 possible points. You made 38 position requests. Your score per request was 29. Please press Next to continue when you are ready.

Next

Manipulate the size of the ships…



People are very sensitive to hypothesis size