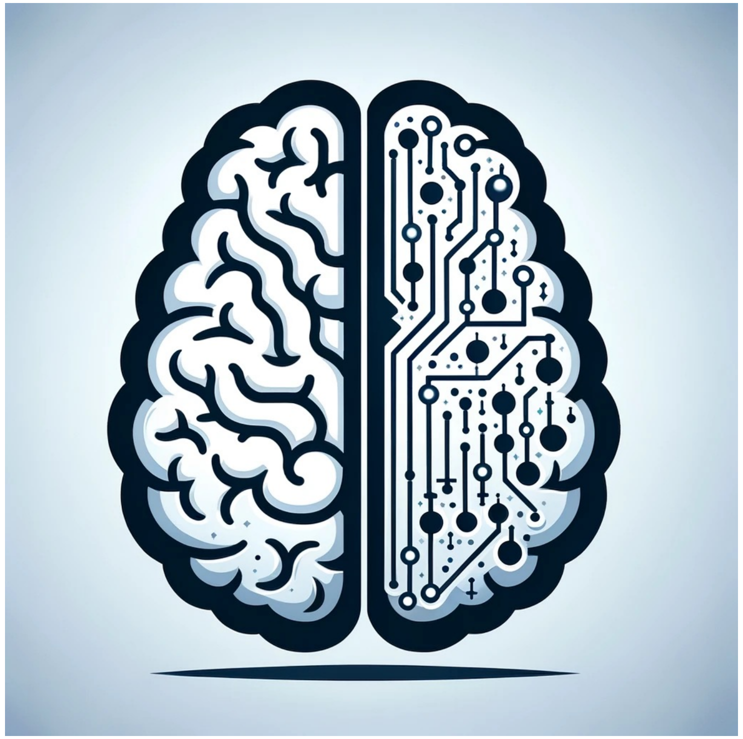
EN 601.473/601.673: Cognitive Artificial Intelligence (CogAI)



Lecture 4: Bayesian concept learning

Tianmin Shu

Bayesian concept learning

"horse"

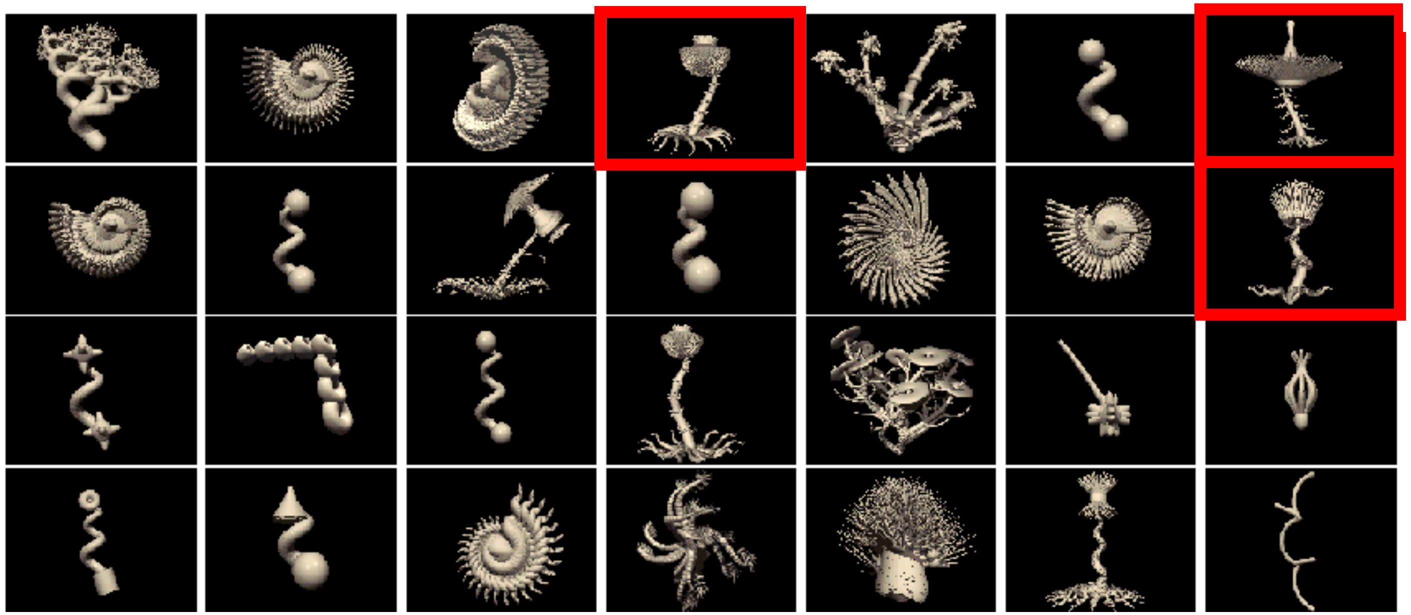
"horse"

"horse"

For Cog Sci: simplified version that gets to the core of human cognition

For AI: principled approach for solving more complex problems

"tufa"



"tufa"

"tufa"

A minimum domain for Bayesian concept learning: The number game

A Bayesian Framework for Concept Learning

by

Joshua B. Tenenbaum

Submitted to the Department of Brain and Cognitive Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1999

Author ....
Department of Brain and Cognitive Sciences
February 15, 1999

Certified by ....
Whitman A. Richards
Professor of Cognitive Science
Thesis Supervisor

Accepted by ....
Gerald E. Schneider
Chairman, Department Committee on Graduate Students

Rules and Similarity in Concept Learning

Joshua B. Tenenbaum
Department of Psychology
Stanford University, Stanford, CA 94305
jbt@psych.stanford.edu

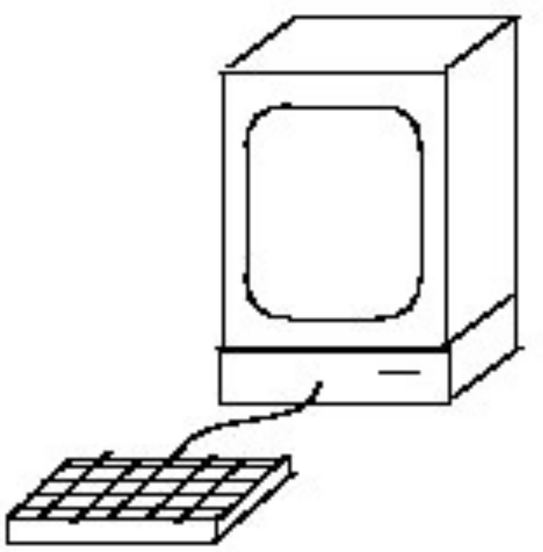
Abstract

This paper argues that two apparently distinct modes of generalizing concepts – abstracting rules and computing similarity to exemplars – should both be seen as special cases of a more general Bayesian learning framework. Bayes explains the specific workings of these two modes – which rules are abstracted, how similarity is measured – as well as why generalization should appear rule- or similarity-based in different situations. This analysis also suggests why the rules/similarity distinction, even if not computationally fundamental, may still be useful at the algorithmic level as part of a principled approximation to fully Bayesian learning.

NeurIPS 1999

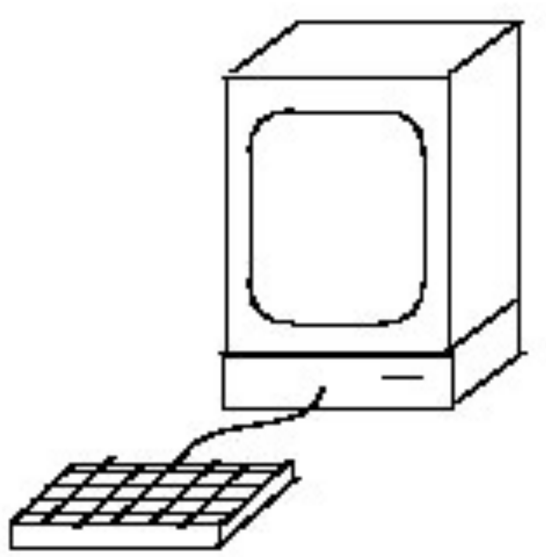
Too old? Still relevant in 2024?

The number game



- Program input: number between 1 and 100
- Program output: “yes” or “no”

The number game



- Your task:
  - Observe one or more positive (“yes”) examples.
  - Judge whether other numbers are “yes” or “no”.

The number game

One positive example: 60

What other numbers do you think are likely to be accepted?

The number game

Four positive examples: 60, 80, 10, 30

The number game

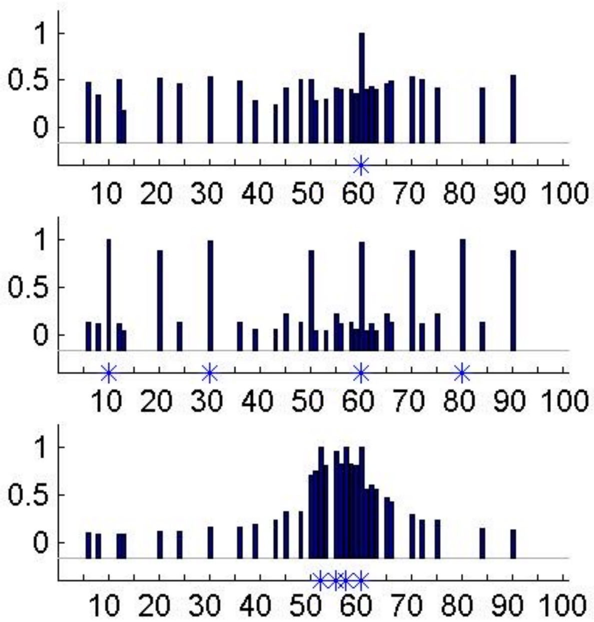
Four positive examples: 60, 52, 57, 55

The number game

Results from a human experiment

Examples of “yes” numbers	Generalization judgments (N = 20)

The number game



60 — Diffuse similarity

60 80 10 30 — Rule: “multiples of 10”

60 52 57 55 — Focused similarity: numbers near 50-60

Main phenomena to explain:

- Generalization can appear either similarity-based (graded) or rule-based (all-or-none).
- Learning from just a few positive examples.

A unifying account of (number) concept learning?

- We’re going to use this to introduce Bayesian approaches, but first consider ...
  - The “naïve programmer” approach?
  - The “modern neural network” approach?

Three levels of analysis for reverse engineering a cognitive system

- **Level 1: Computational theory**
  - What are the inputs and outputs to the computation, what is its goal, and what is the logic by which it is carried out?
- **Level 2: Representation and algorithm**
  - How is information represented and processed to achieve the computational goal?
- **Level 3: Hardware implementation**
  - How is the computation realized in physical or biological hardware?

Traditional (algorithmic level) cognitive models

- Multiple representational systems: rules and similarity
- Questions this leaves open:
  - How does each system work? How far and in ways to generalize as a function of the examples observed?
    - Which rule to choose?
      - E.g., X = {60, 80, 10, 30}: multiples of 10 vs. even numbers?
    - Which similarity metric?
      - E.g., X = {60, 53} vs. {60, 20}?
  - Why these two systems?
  - When and why does a learner switch between them?

Three levels of analysis for reverse engineering a cognitive system

- **Level 1: Computational theory**
  - What are the inputs and outputs to the computation, what is its goal, and what is the logic by which it is carried out?
- **Level 2: Representation and algorithm**
  - How is information represented and processed to achieve the computational goal?
- **Level 3: Hardware implementation**
  - How is the computation realized in physical or biological hardware?

Bayesian modeling

- H: Hypothesis space of possible concepts:
  - h_1 = {2, 4, 6, 8, 10, 12, …, 96, 98, 100} (“even numbers”)
  - h_2 = {10, 20, 30, 40, …, 90, 100} (“multiples of 10”)
  - h_3 = {2, 4, 8, 16, 32, 64} (“powers of 2”)
  - h_4 = {50, 51, 52, …, 59, 60} (“numbers between 50 and 60”)
  - . . .

Three hypothesis subspaces for number concepts

- Mathematical properties (24 hypotheses):
  - Odd, even, square, cube, prime numbers
  - Multiples of small integers
  - Powers of small integers

- Raw magnitude (5050 hypotheses):
  - All intervals of integers with endpoints between 1 and 100.

- Approximate magnitude (10 hypotheses):
  - Decades (1-10, 10-20, 20-30, …)

Bayesian modeling

- H: Hypothesis space of possible concepts:
  - Mathematical properties: even, odd, square, prime, ...
  - Approximate magnitude: {1-10}, {10-20}, {20-30}, ...
  - Raw magnitude: all intervals between 1 and 100.

- $X = \{x_1, \ldots, x_n\}$: n examples of a concept C.
- Evaluate hypotheses given data:

$$p(h \mid X) = \frac{p(X \mid h)\, p(h)}{\sum_{h' \in H} p(X \mid h')\, p(h')}$$

  - $p(h)$ [“prior”]: domain knowledge, pre-existing biases
  - $p(X|h)$ [“likelihood”]: statistical information in examples.
  - $p(h|X)$ [“posterior”]: degree of belief that h is the true extension of C.

Likelihood $p(X \mid h)$

- **Size principle**: Smaller hypotheses receive greater likelihood, and exponentially more so as *n* increases.

$$p(X \mid h) = \left[\frac{1}{\mathrm{size}(h)} \right]^n \text{ if } x_1, \ldots, x_n \in h$$

$$= 0 \text{ if any } x_i \notin h$$

- Captures the intuition of a "representative" sample, versus a "suspicious coincidence".
- A **generative** model: what kinds of numbers can be generated given the hypothesis?

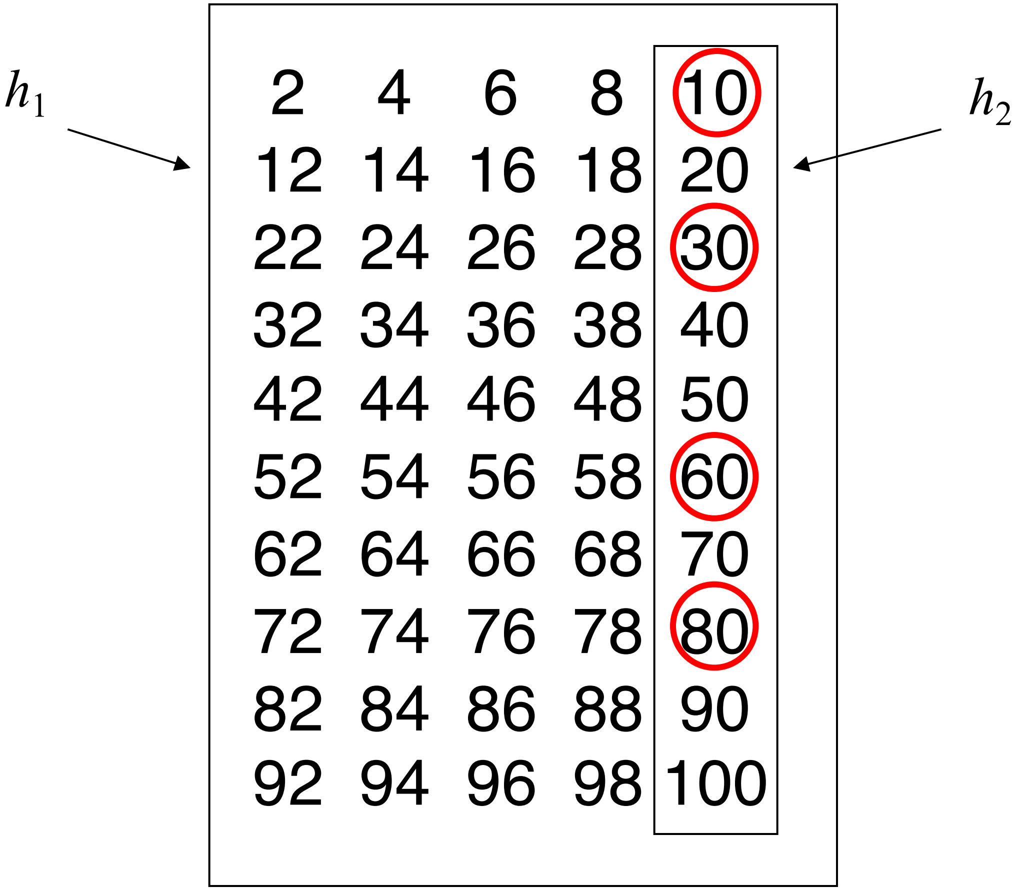
Illustrating the size principle

h_1 h_2

2	4	6	8	10
12	14	16	18	20
22	24	26	28	30
32	34	36	38	40
42	44	46	48	50
52	54	56	58	60
62	64	66	68	70
72	74	76	78	80
82	84	86	88	90
92	94	96	98	100

Data slightly more of a coincidence under h_1

Illustrating the size principle



Data *much* more of a coincidence under h_1

Prior $p(h)$

- Choice of hypothesis space embodies a strong prior:
  - Effectively, $p(h) \sim 0$ for many logically possible but conceptually unnatural hypotheses.
- Do we need this? Why not allow all logically possible hypotheses, with uniform priors, and let the data sort them out (via the likelihood)?
- Prevents overfitting by highly specific but unnatural hypotheses, e.g. “multiples of 10 except 50 and 70”.

e.g., X = {60 80 10 30}:

$$p(X \mid \text{multiples of } 10) = \left[\frac{1}{10}\right]^4 = 0.0001$$

$$p(X \mid \text{multiples of } 10 \text{ except } 50, 70) = \left[\frac{1}{8}\right]^4 = 0.00024$$

Prior $p(h)$

- $p(h)$ encodes relative weights of alternative theories:

H: Total hypothesis space

$p(H_1) = 1/5$ $\qquad$ $p(H_2) = 3/5$ $\qquad$ $p(H_3) = 1/5$

H_1: Math properties (24)	H_2: Raw magnitude (5050)	H_3: Approx. magnitude (10)
• even numbers	• 10-15	• 10-20
• powers of two	• 20-32	• 20-30
• multiples of three	• 37-54	• 30-40
….	….	….
$p(h) = p(H_1) / 24$	$p(h) = p(H_2) / 5050$	$p(h) = p(H_3) / 10$
$p(h) \sim 1/120$	$p(h) \sim 1/8500$ (on average)	$p(h) \sim 1/50$

Posterior $p(h \mid X) = \dfrac{p(X \mid h)\, p(h)}{\sum_{h' \in H} p(X \mid h')\, p(h')}$

- X = {60, 80, 10, 30}

- Why prefer "multiples of 10" over "even numbers"? $p(X|h)$.

- Why prefer "multiples of 10" over "multiples of 10 except 50 and 70"? $p(h)$.

- Why does a good generalization need both high prior and high likelihood? $p(h|X) \sim p(X|h)\, p(h)$

Generalizing to new objects

- **From hypotheses to predictions:**

  How do we compute the probability that C applies to some new object y, given the posterior $p(h|X)$?

Hypothesis averaging

In general, we have the law of total probability:

$$p(A=a)=\sum_z p(A=a \mid Z=z)\ p(Z=z)$$

$$p(A=a \mid B=b)=\sum_z p(A=a \mid Z=z, B=b)\ p(Z=z \mid B=b)$$

…especially useful if *A* and *B* are independent conditioned on *Z*:

$$p(A=a \mid B=b)=\sum_z p(A=a \mid Z=z)\ p(Z=z \mid B=b)$$

Can you think of examples of this reasoning pattern?

Tuesday, November 6, 2018

The New York Times

Today's Paper

Your Tuesday Briefing
Here's what you need to know to start your day.

Listen to 'The Daily'
Four themes and four races to watch in today's elections.

In the 'Smarter Living' Newsletter
The cognitive and social benefits of being in a bad mood.

46°F 60° 49°
Woburn, MA

With Control of Congress at Stake, Americans Head to Polls

With Control of Congress at Stake, Americans Head to Polls

- In the House, Democrats need to pick up 23 seats to win c… appear to have the a… Senate map is tricki… because some incum… defend their seats i… Trump won.
- Here, we have mapp… when their polls clos… earliest point by which full results could be known.

1m ago

Two Visions of Patriotism Clash in the Midterm Elections

Mr. Trump's critics and supporters express sharply differing views on policies and issues. But increasingly …ng more …neans to be

… where to

9m ago

Bad Weather, Known to Lower Turnout, Will Greet Many Voters

Rain can decrease voter numbers, which studies show tends to help Republicans. "I hope it rains hard tomorrow," one Republican candidate said.

10h ago

6 Types of Misinformation to Watch For Today, and What to Do if You Spot Any

Be careful of rumors and hoaxes about the voting and polling places. Here are some tips for spotting and avoiding false information.

Bad Weather, Known to Lower Turnout, Will Greet Many Voters

Rain can decrease voter numbers, which studies show tends to help Republicans. "I hope it rains hard tomorrow," one Republican candidate said.

10h ago

Hypothesis averaging

In general, we have the law of total probability:

$$p(A=a)=\sum_z p(A=a \mid Z=z)\ p(Z=z)$$

$$p(A=a \mid B=b)=\sum_z p(A=a \mid Z=z, B=b)\ p(Z=z \mid B=b)$$

…especially useful if *A* and *B* are independent conditioned on *Z*:

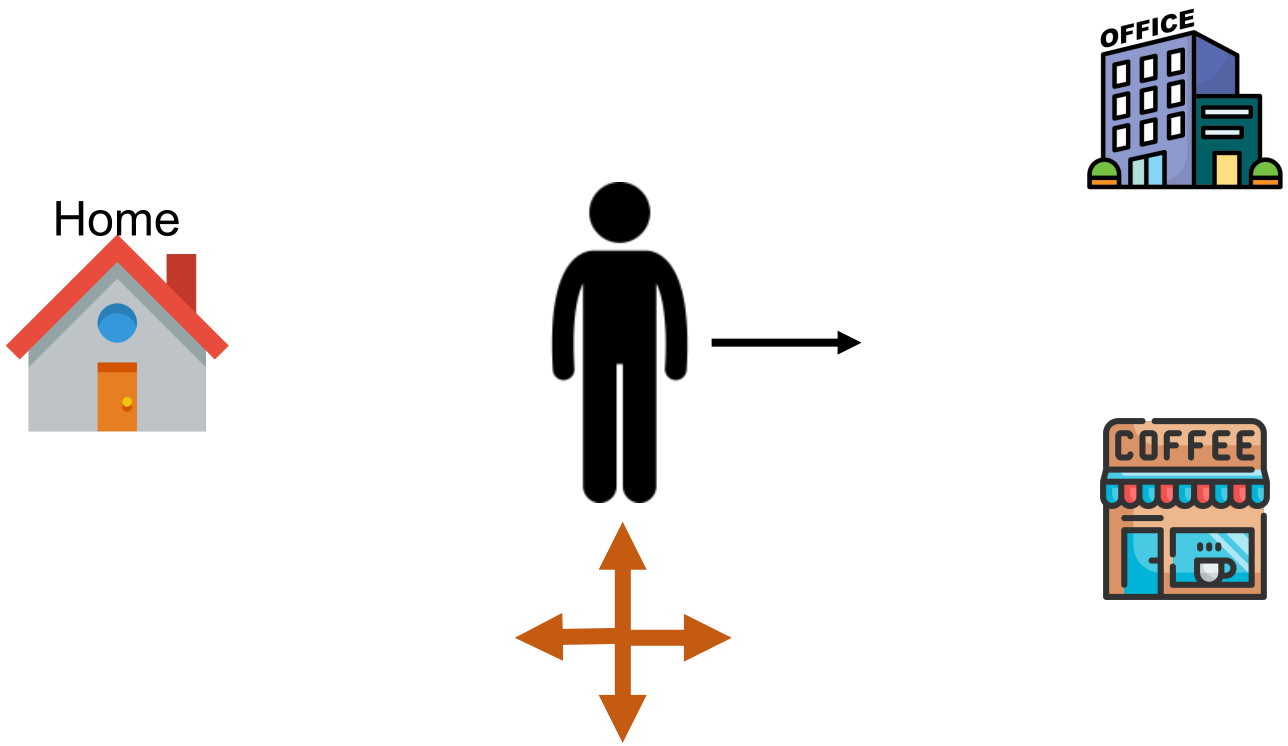
$$p(A=a \mid B=b)=\sum_z p(A=a \mid Z=z)\ p(Z=z \mid B=b)$$

What is the probability that the republican will win the election, given that the weather man predicts rain?

$$p(\text{Republican win} \mid \text{Weather report: “Rain storm”}) =$$

$$\sum_{\substack{w \in weather \\ conditions}} p(\text{Republican win} \mid w)\ p(\ w \mid \text{Weather report: “Rain storm”})$$

Model-based action prediction



Predicting the future action of the person based on their past behavior

$$p(\text{action} \mid \text{past behavior}) = \sum_{goal} p(\text{action} \mid goal)\, p(goal \mid \text{past behavior})$$

Generalizing to new objects

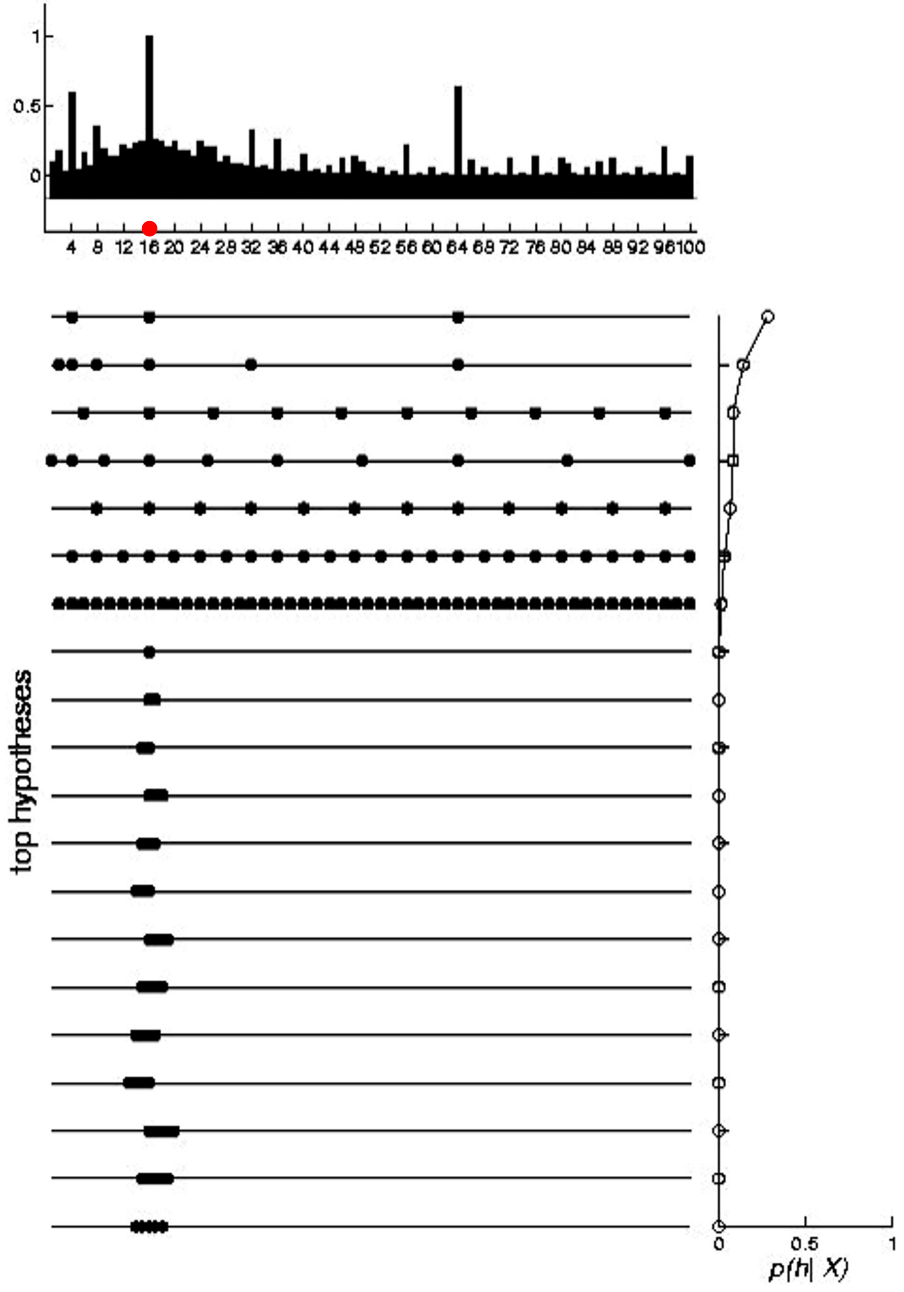
Hypothesis averaging:

Compute the probability that *C* applies to some new object *y* by averaging the predictions of all hypotheses *h*, weighted by *p*(*h*|*X*):

$$p(y \in C \mid X) = \sum_{h \in H} \underbrace{p(y \in C \mid h)}_{= \begin{cases} 1 \text{ if } y \in h \\ 0 \text{ if } y \notin h \end{cases}} \; p(h \mid X)$$

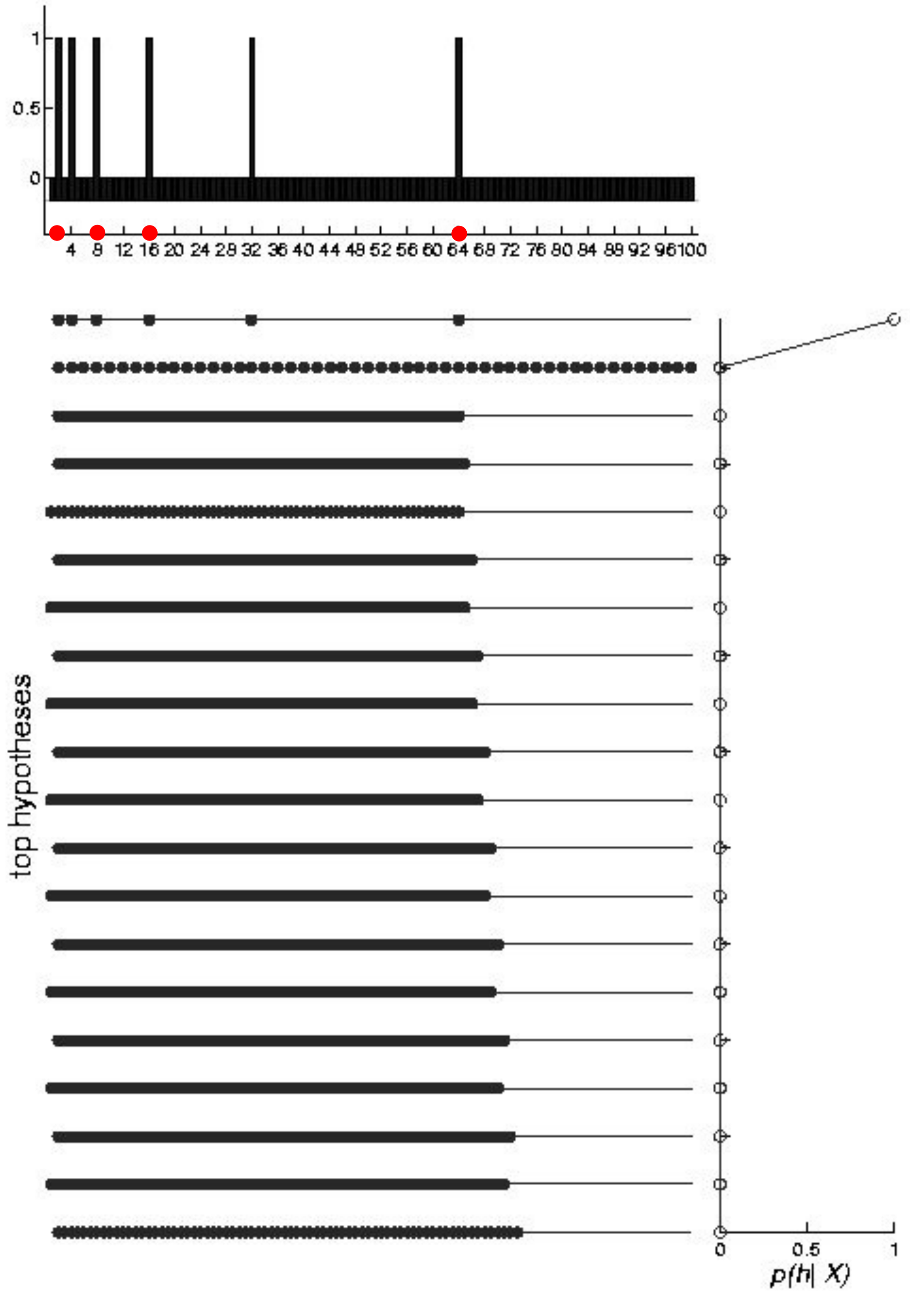
Concept inference

Examples:
16



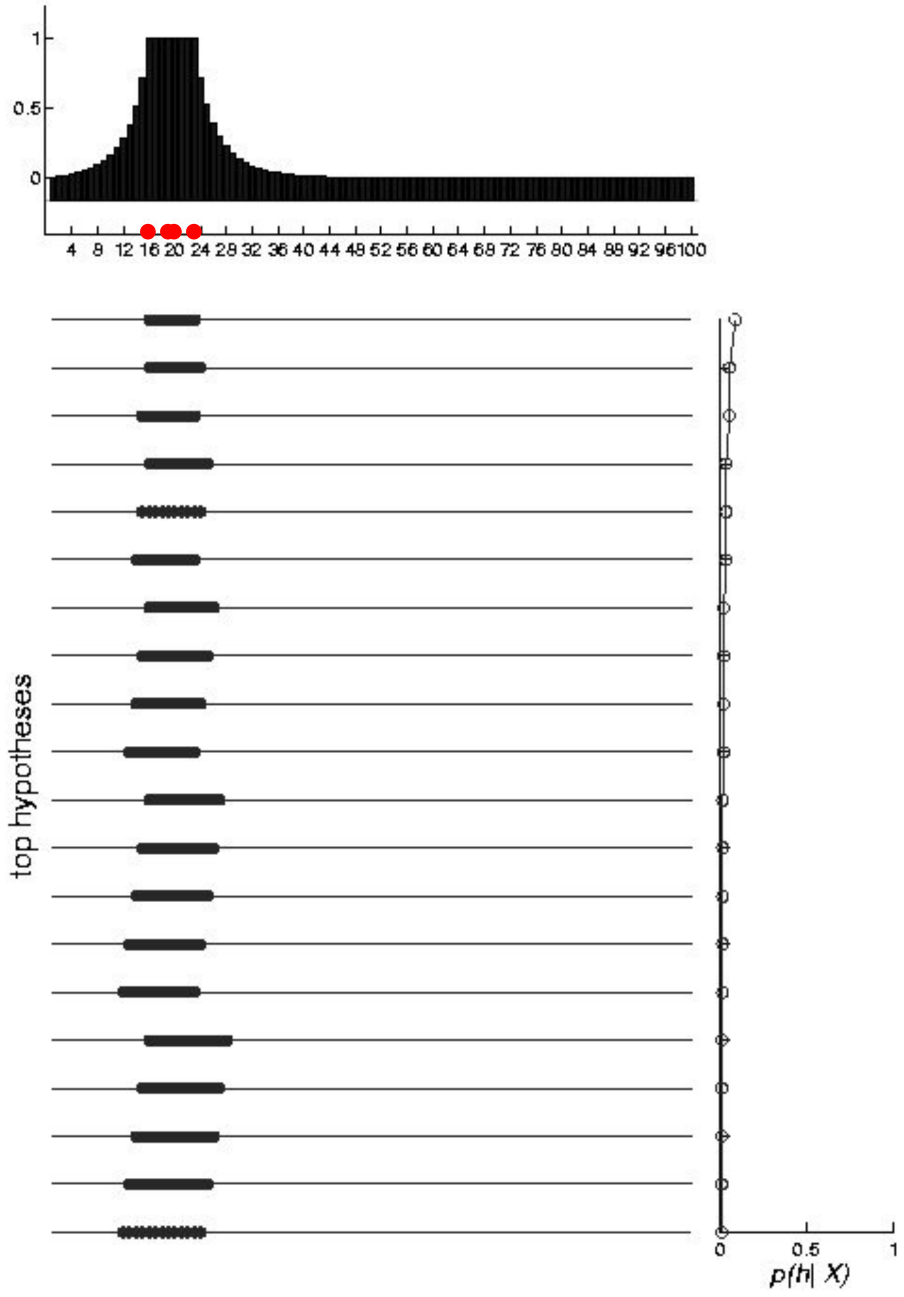
Concept inference

Examples:
16
8
2
64



Concept inference

Examples:
16
23
19
20



+ Examples | Human generalization | Bayesian Model

60

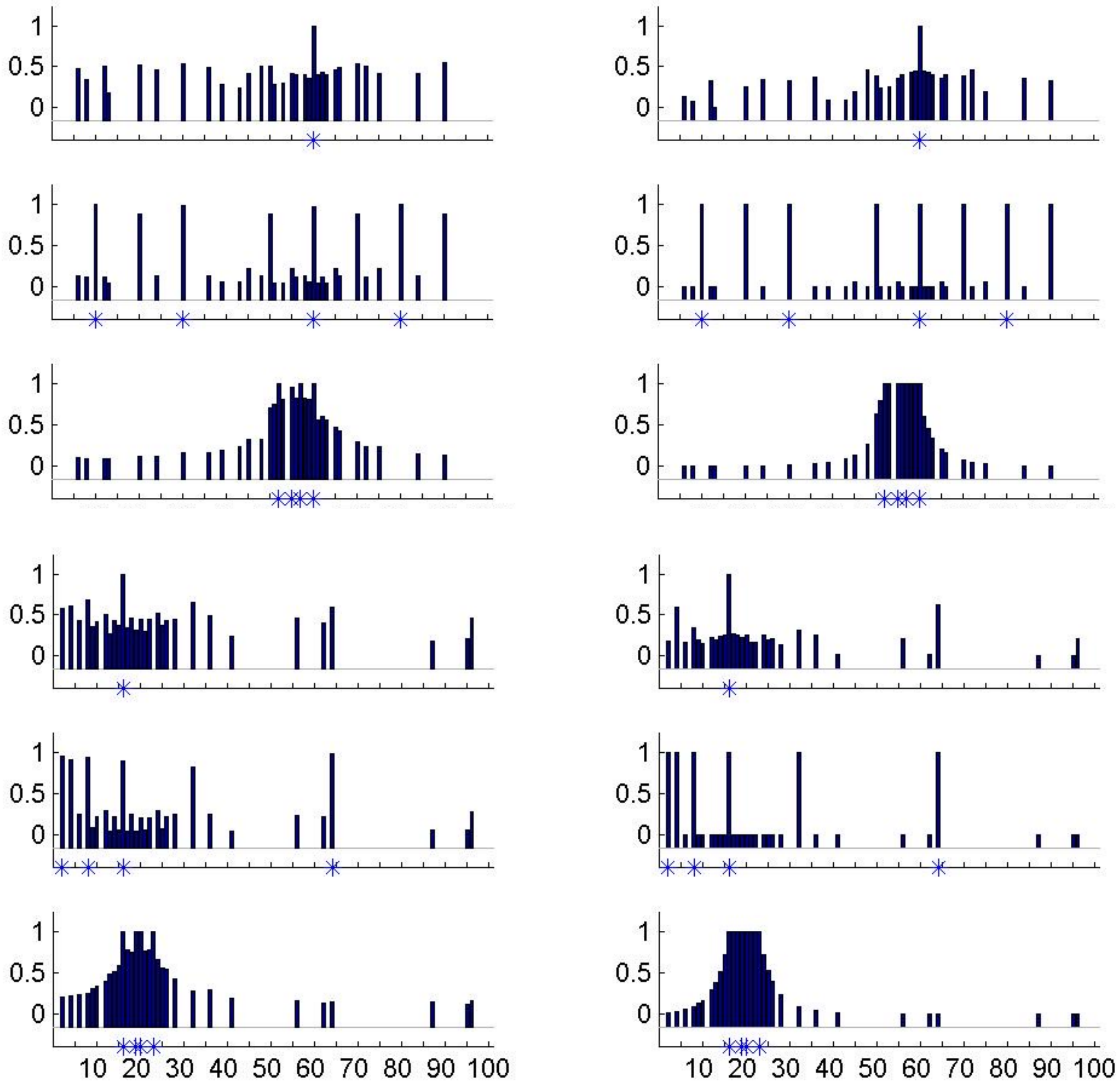
60 80 10 30

60 52 57 55

16

16 8 2 64

16 23 19 20



Summary of the Bayesian Model

- How do the statistics of the examples interact with prior knowledge to guide generalization?

$$\textbf{posterior} \propto \textbf{likelihood} \times \textbf{prior}$$

- Why does generalization appear rule-based or similarity-based?

$$\textbf{hypothesis averaging} + \textbf{size principle}$$

$$\Downarrow$$

broad $p(h\|X)$: similarity gradient
narrow $p(h\|X)$: all-or-none rule

$$p(y \in C \mid X) = \sum_{h \in H} \underbrace{p(y \in C \mid h)}_{= \begin{cases} 1 \text{ if } y \in h \\ 0 \text{ if } y \notin h \end{cases}} \; p(h \mid X)$$

Summary of the Bayesian Model

- How do the statistics of the examples interact with prior knowledge to guide generalization?

$$\textbf{posterior} \propto \textbf{likelihood} \times \textbf{prior}$$

- Why does generalization appear rule-based or similarity-based?

$$\textbf{hypothesis averaging} + \textbf{size principle}$$

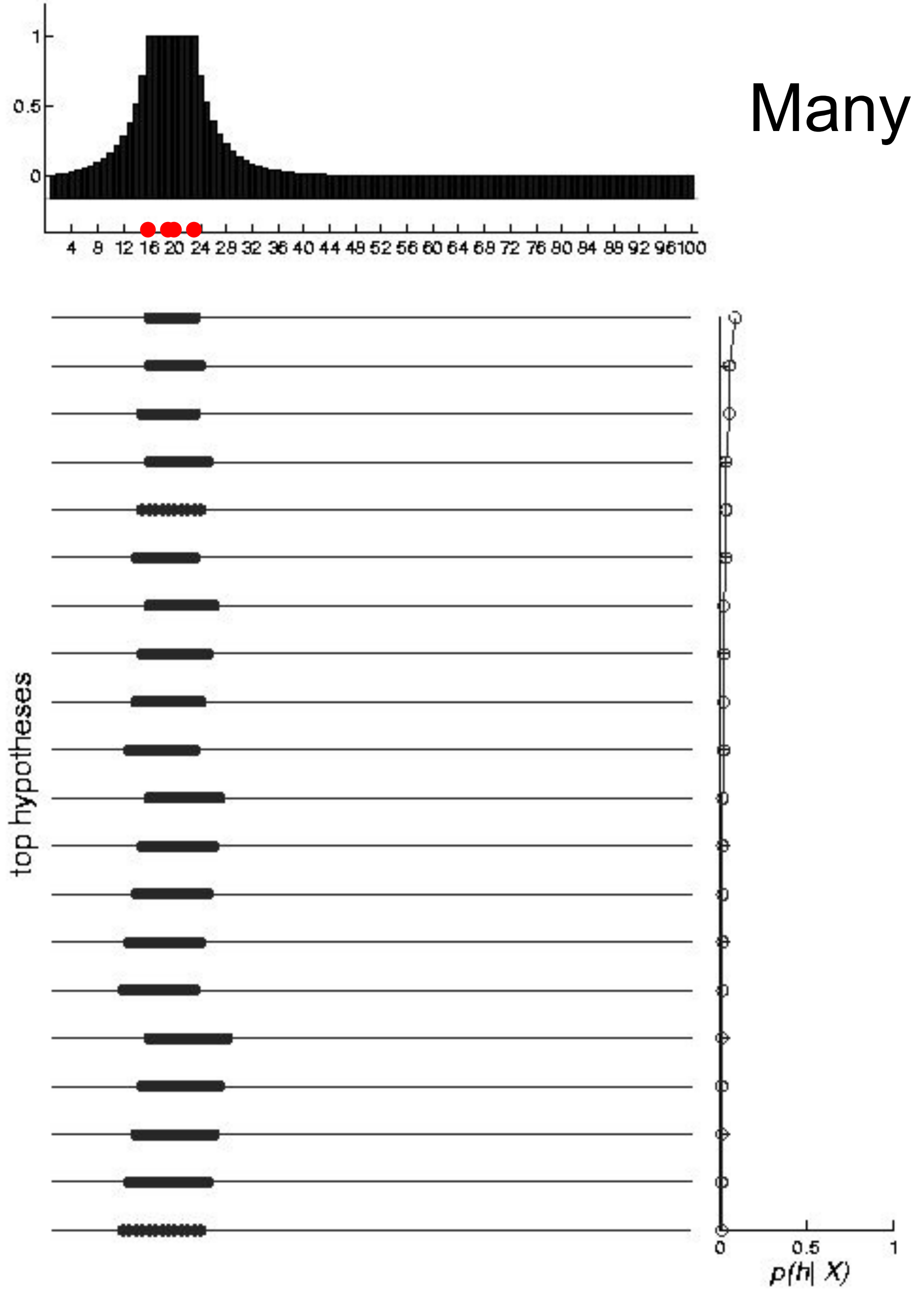
$$\Downarrow$$

broad $p(h \mid X)$: Many h of similar size, or very few examples (i.e. 1)
narrow $p(h \mid X)$: One h is much smaller

Concept inference

Many *h* of similar size

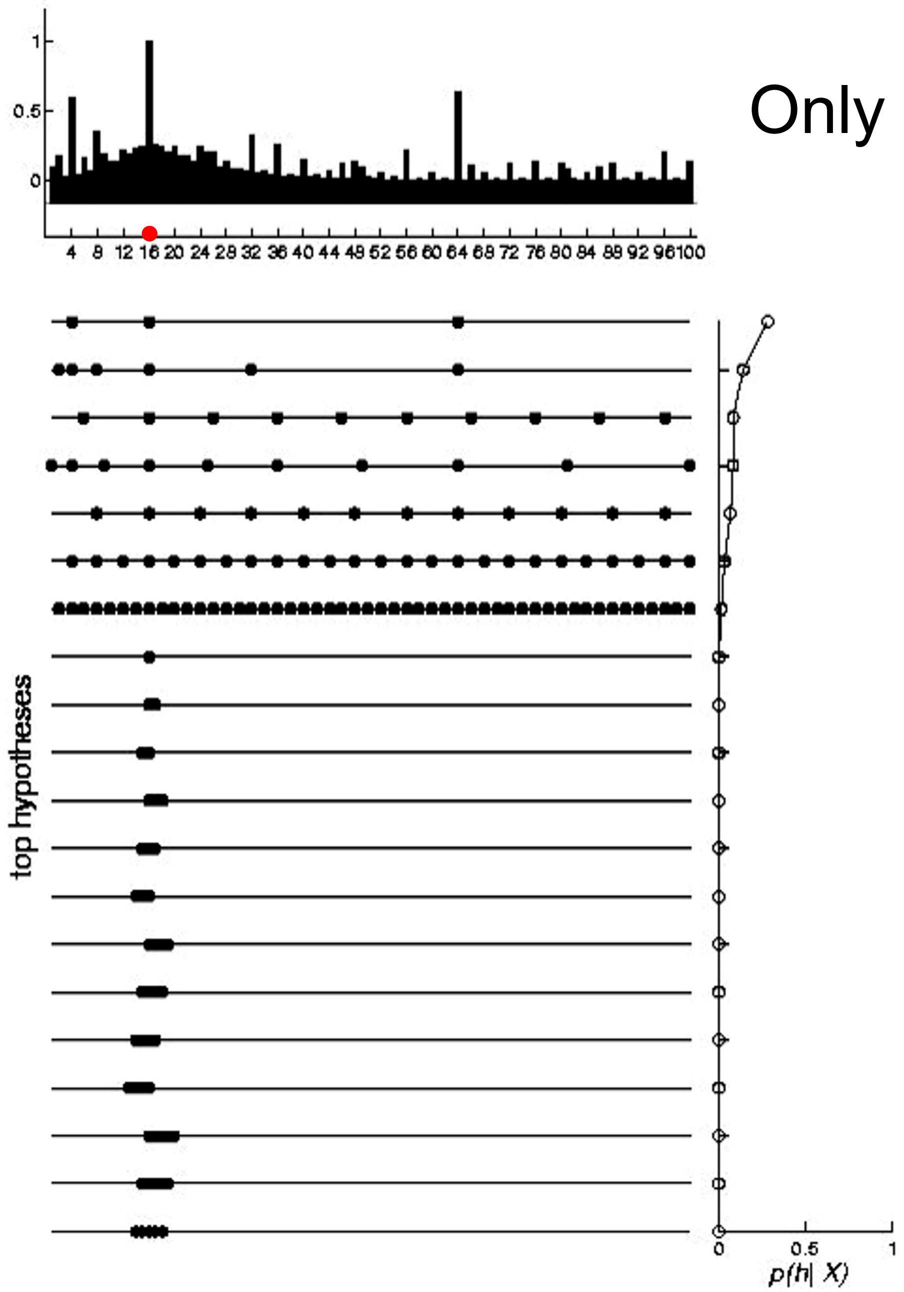
Examples:
16
23
19
20



Concept inference

Only 1 example

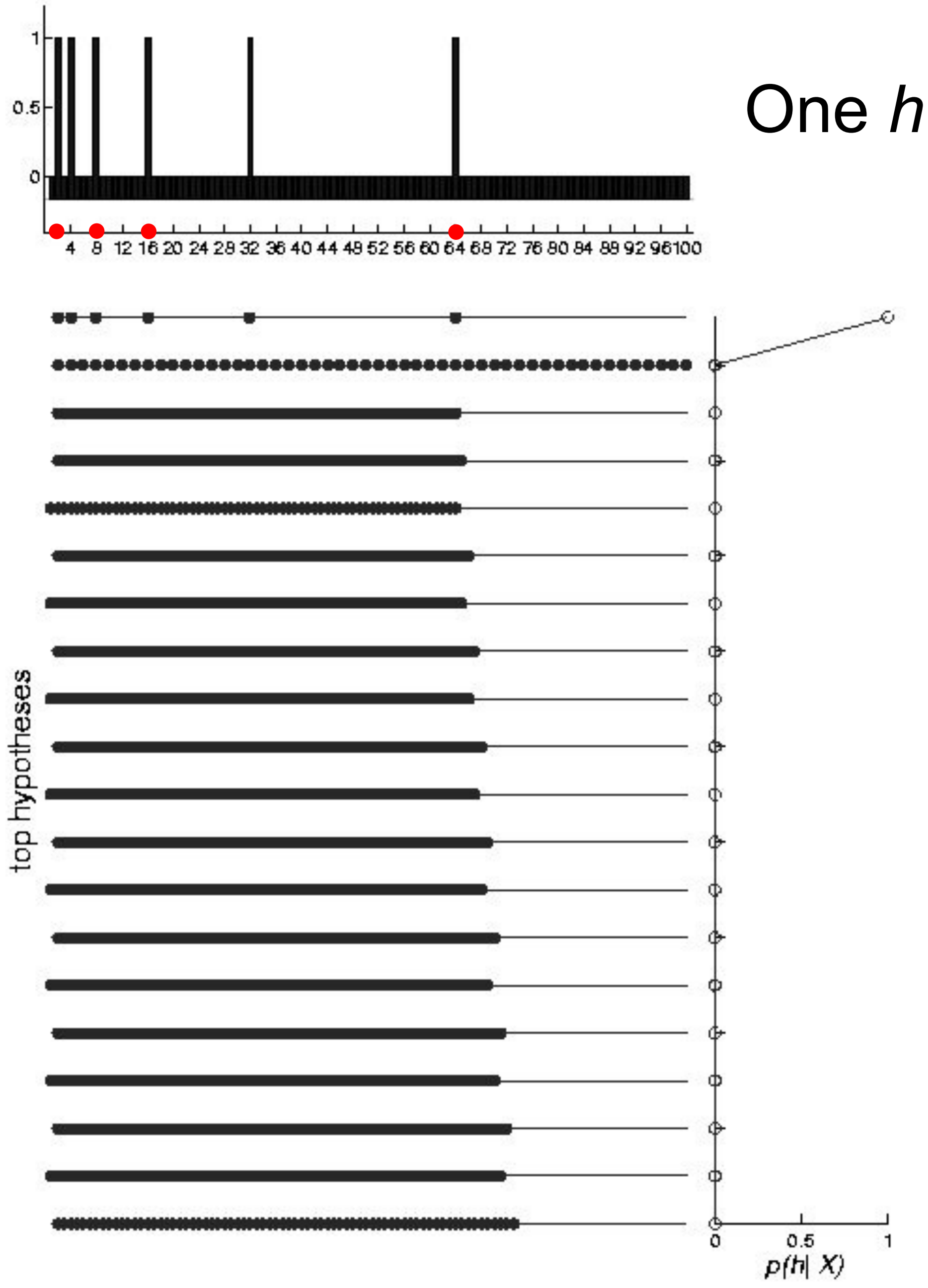
Examples:
16



Concept inference

One *h* is much smaller

Examples:
16
8
2
64



Model variants

1. Bayes with weak sampling

$$\textbf{posterior} \propto \textbf{likelihood} \times \textbf{prior}$$

$$\textbf{hypothesis averaging} + \cancel{\textbf{size principle}}$$

$$p(X \mid h) \propto 1 \text{ if } x_1, \ldots, x_n \in h$$
$$= 0 \text{ if any } x_i \notin h$$

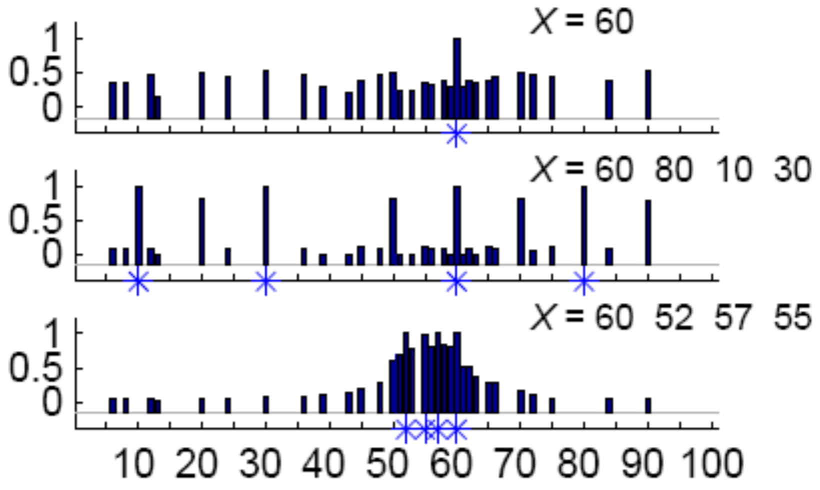
2. Maximum a posteriori (MAP)
   Maximum likelihood /subset principle

$$\textbf{posterior} \propto \textbf{likelihood} \times \textbf{prior}$$

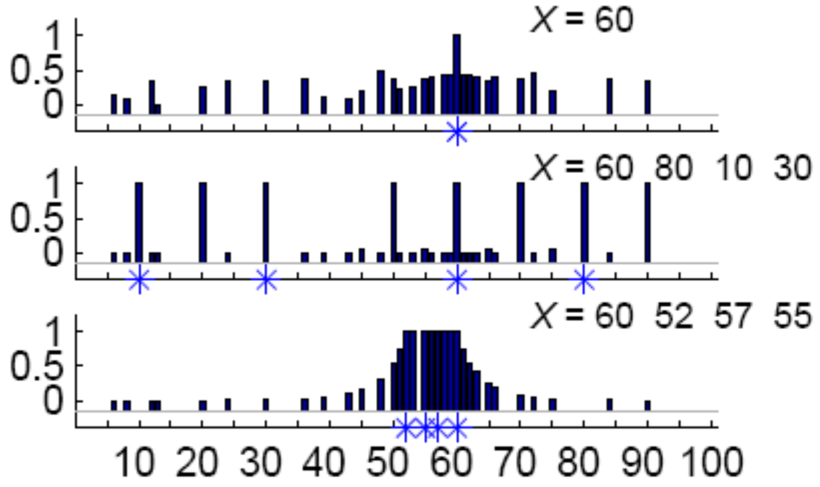
$$\cancel{\textbf{hypothesis averaging}} + \textbf{size principle}$$

$$p(y \in C \mid X) = 1 \text{ if } y \in h^*; \; h^* = \arg\max_{h \in H} p(h \mid X)$$
$$= 0 \text{ if } y \notin h^*$$

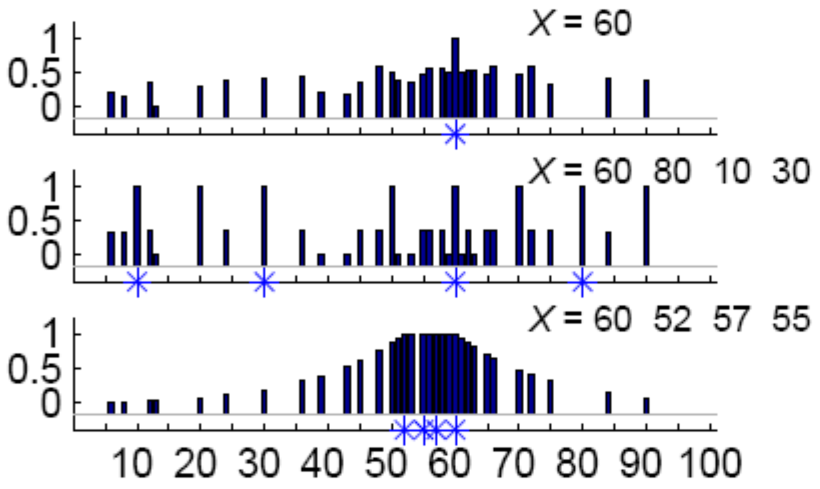
Human generalization



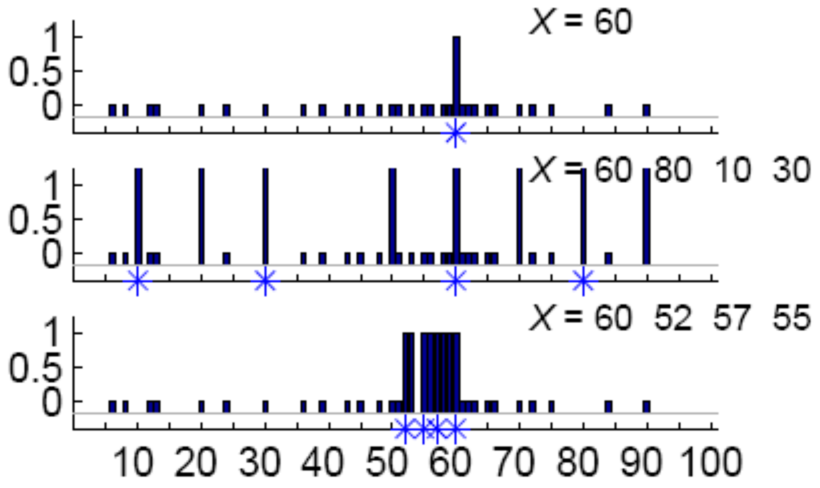
Full Bayesian model



Bayes with weak sampling
(no size principle)



Maximum a posteriori (MAP) / subset
principle (no hypothesis averaging)



Taking stock

- A model of high-level, knowledge-driven inductive reasoning that makes strong quantitative predictions.
  - High correlation between human judgments and model judgments.

- Explains qualitatively different patterns of generalization (rules, similarity) as the output of a single general-purpose rational inference engine.
  - Marr level 1 (Computational theory) explanation of phenomena that have traditionally been treated only at Marr level 2 (Representation and algorithm).

Problem set 1

- Write Python code to implement the Bayesian model for the number game

- If successful, your model’s predictions should be close to human responses, i.e., correlation between model predictions and human responses should be higher than 0.95 in each condition (given a specific set of example numbers)

Human responses

Given example 60

6	0.46905
8	0.34429
12	0.50786
13	0.18429
20	0.52571
24	0.46571
30	0.53524
36	0.49238
39	0.27762
43	0.23595
45	0.4169
48	0.50071
50	0.51143
51	0.27595
53	0.29857
55	0.41643
56	0.40143
58	0.39976
59	0.35262
60	1
61	0.39548
62	0.43595
63	0.39976
65	0.4669
66	0.4869
70	0.53524
72	0.51024
75	0.41643
84	0.4169
90	0.54952

Problem set 1

- Implement and run the code in jupyter notebook and export it as a pdf file. Only upload the pdf file.

- Make sure that the correlation between the model predictions and human responses is higher than 0.95

- Due by Feb 18, 11:59 pm

Learning more natural concepts

“horse”

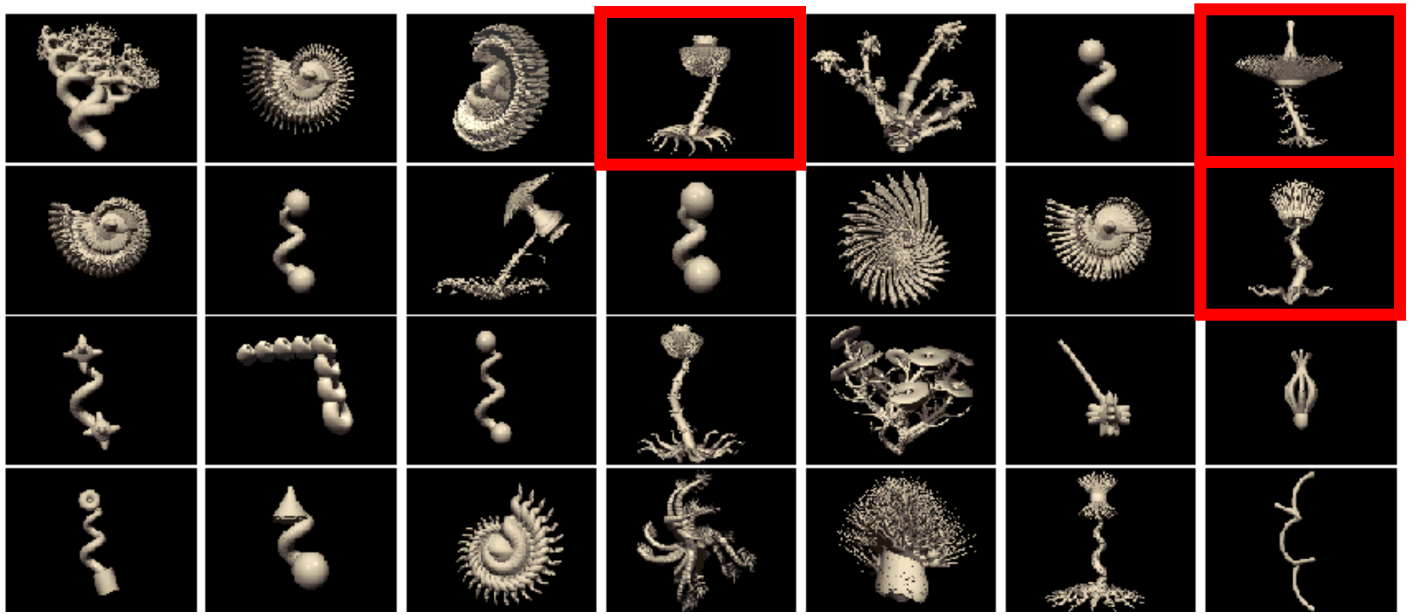
“horse”

“horse”

For Cog Sci: simplified version that gets to the core of human cognition

For AI: principled approach for solving more complex problems

“tufa”

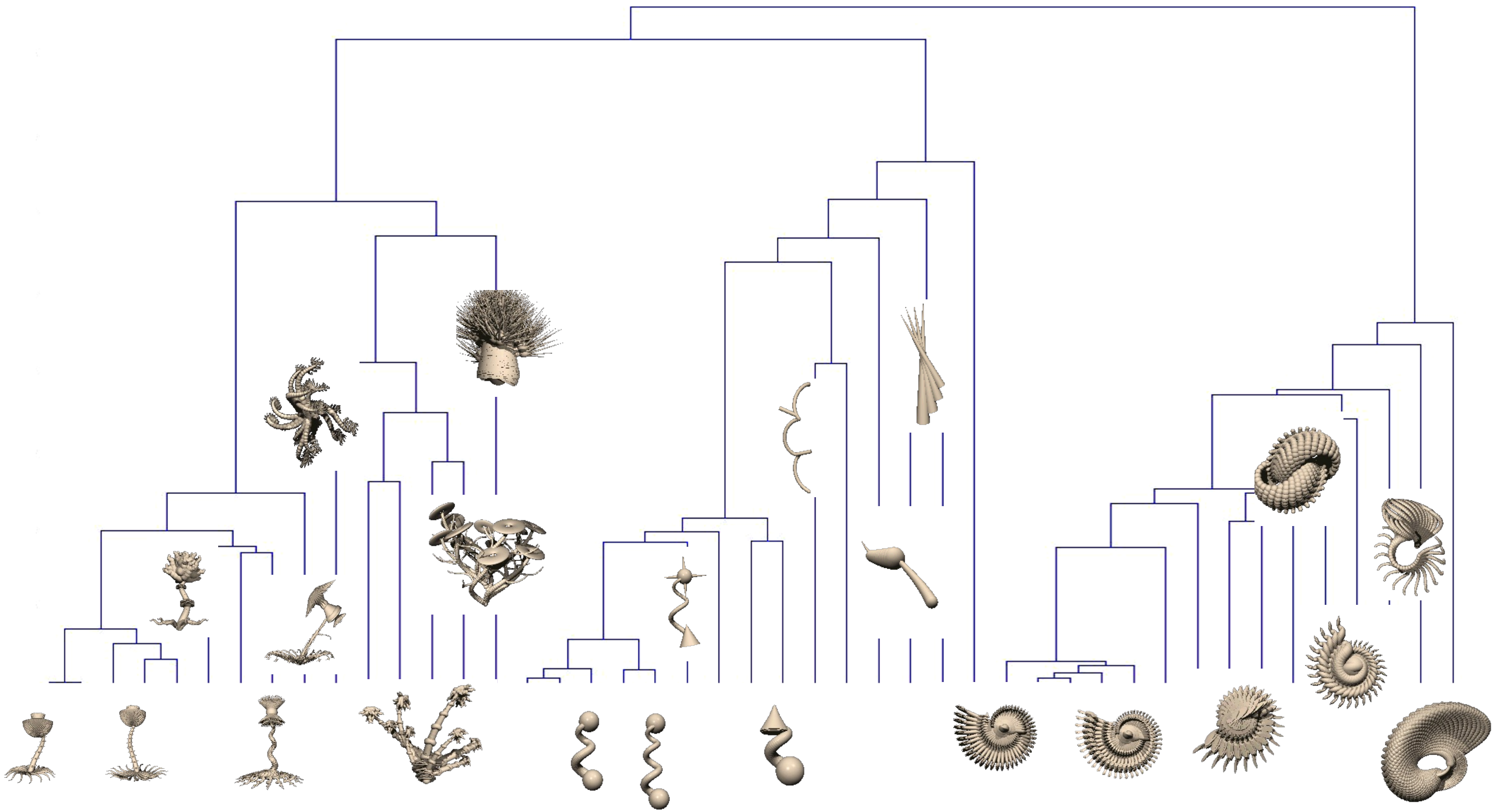


“tufa”

“tufa”

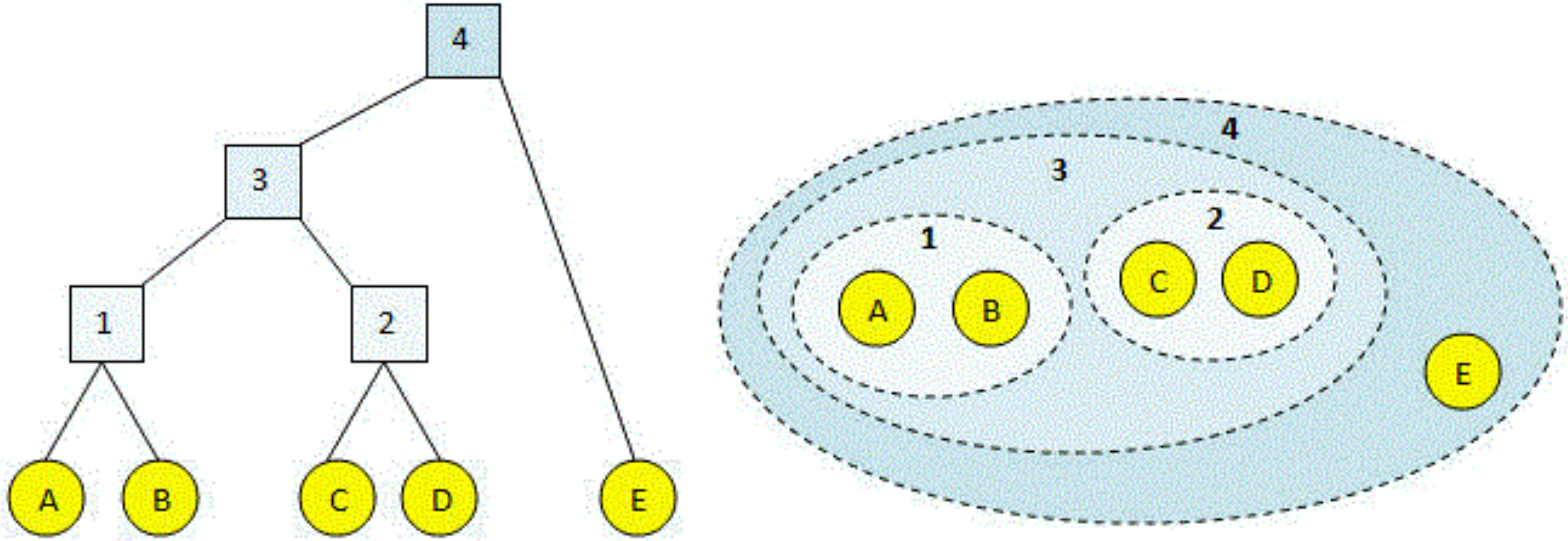
Modeling word learning (Xu & Tenenbaum, 2007)

Heterarchical clustering
based on similarity between two examples (human judgement)

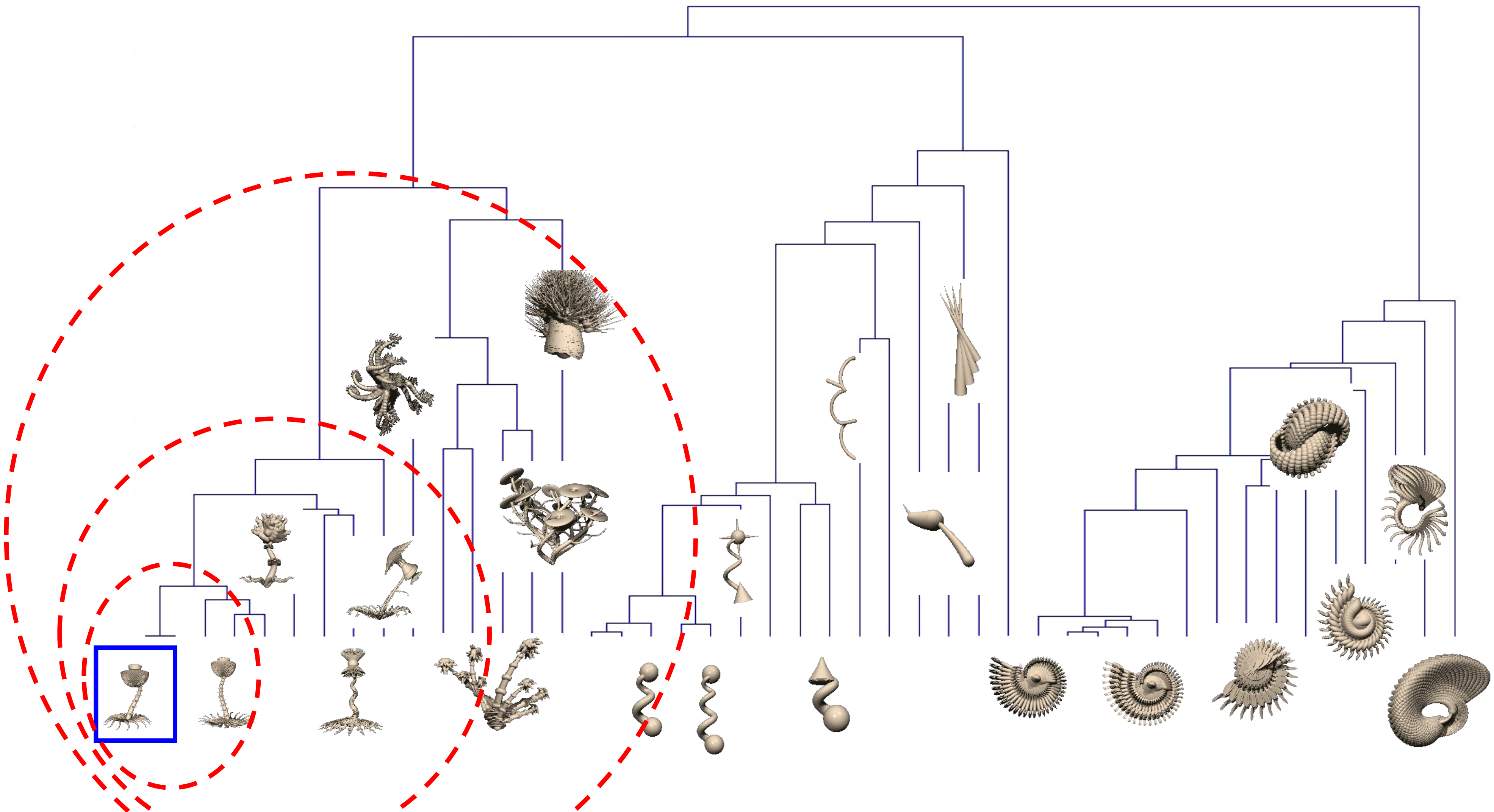


Modeling word learning (Xu & Tenenbaum, 2007)

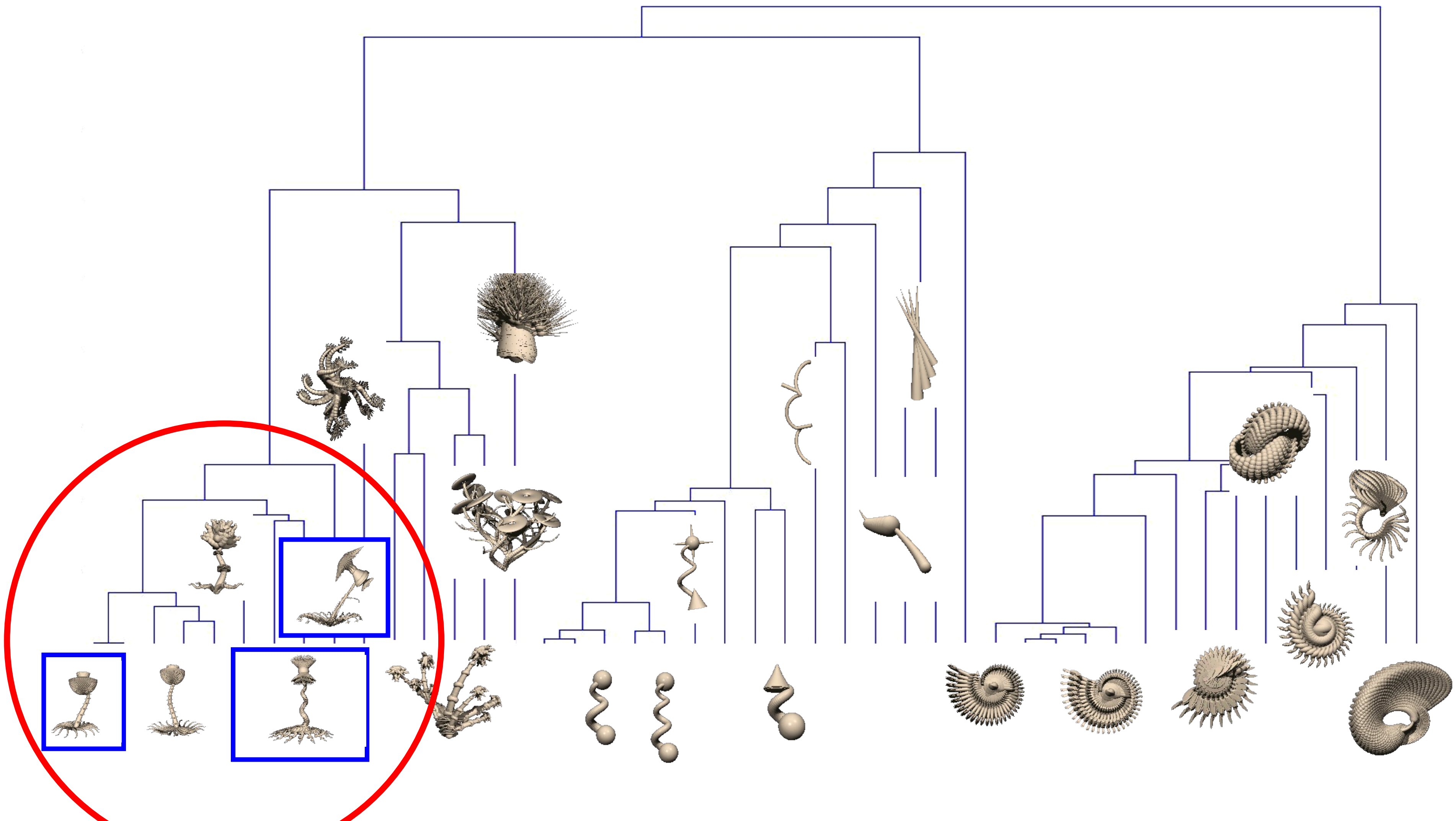
Hierarchical clustering



Modeling word learning (Xu & Tenenbaum, 2007)

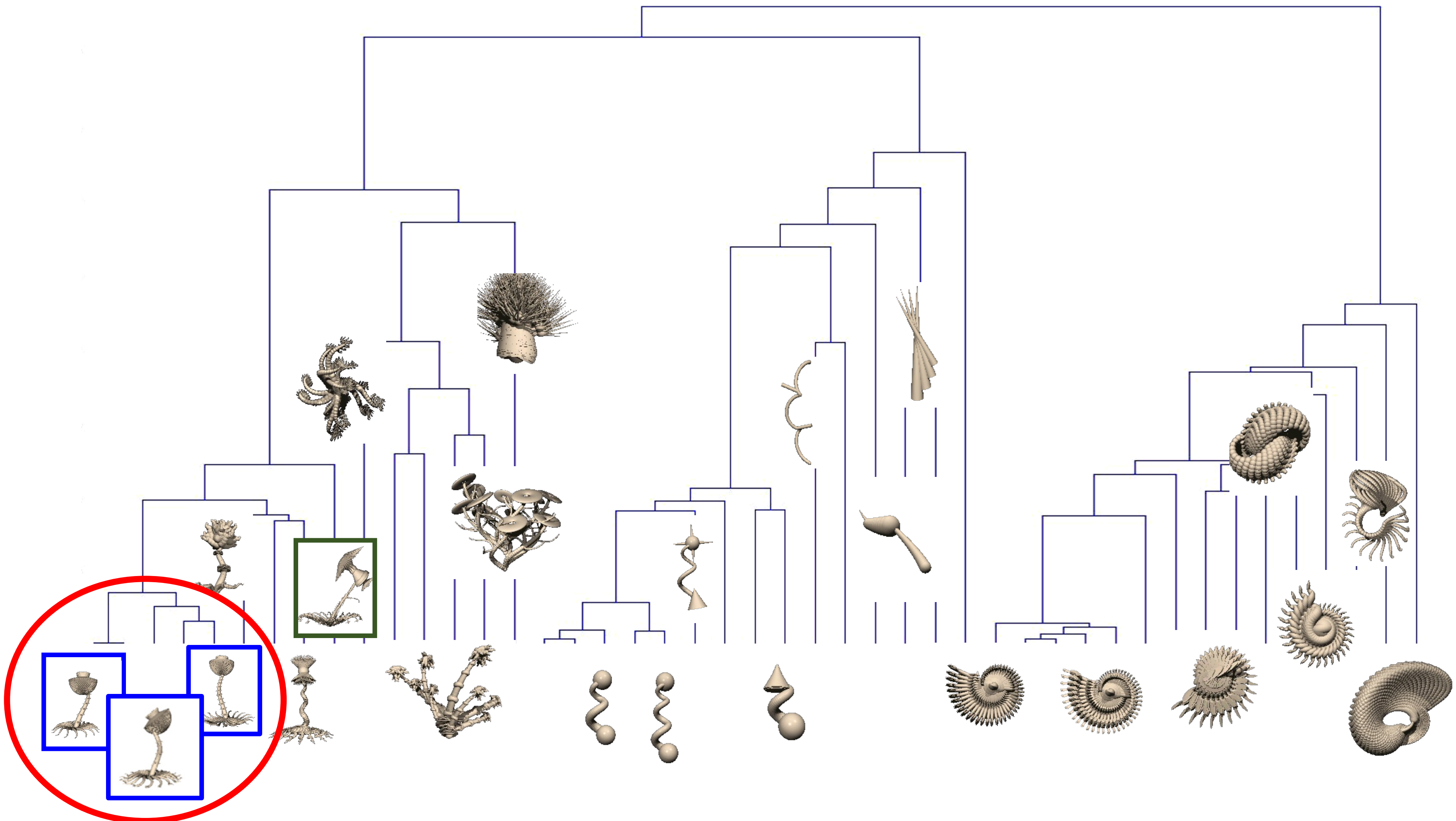


Modeling word learning (Xu & Tenenbaum, 2007)

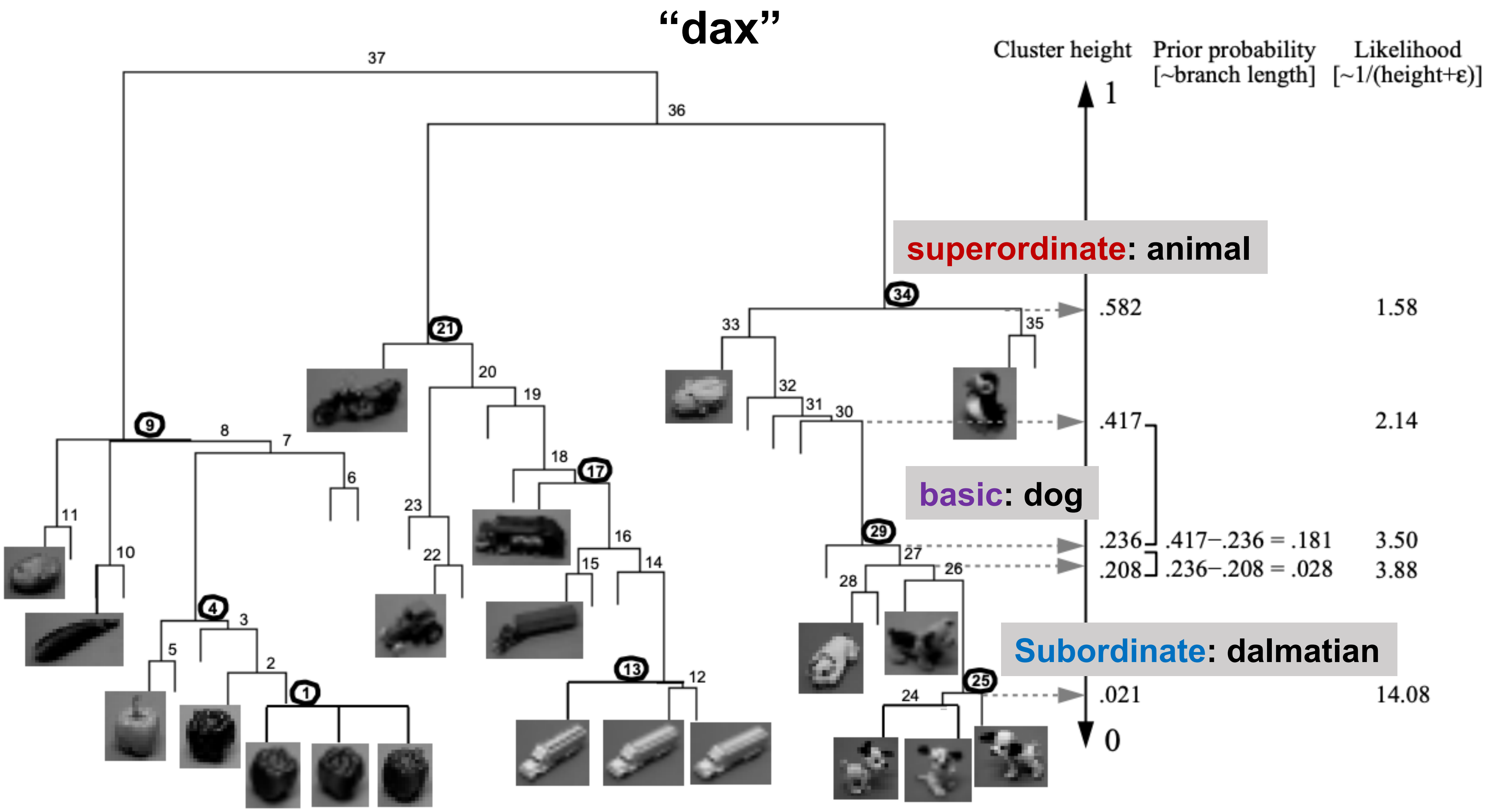


Modeling word learning (Xu & Tenenbaum, 2007)

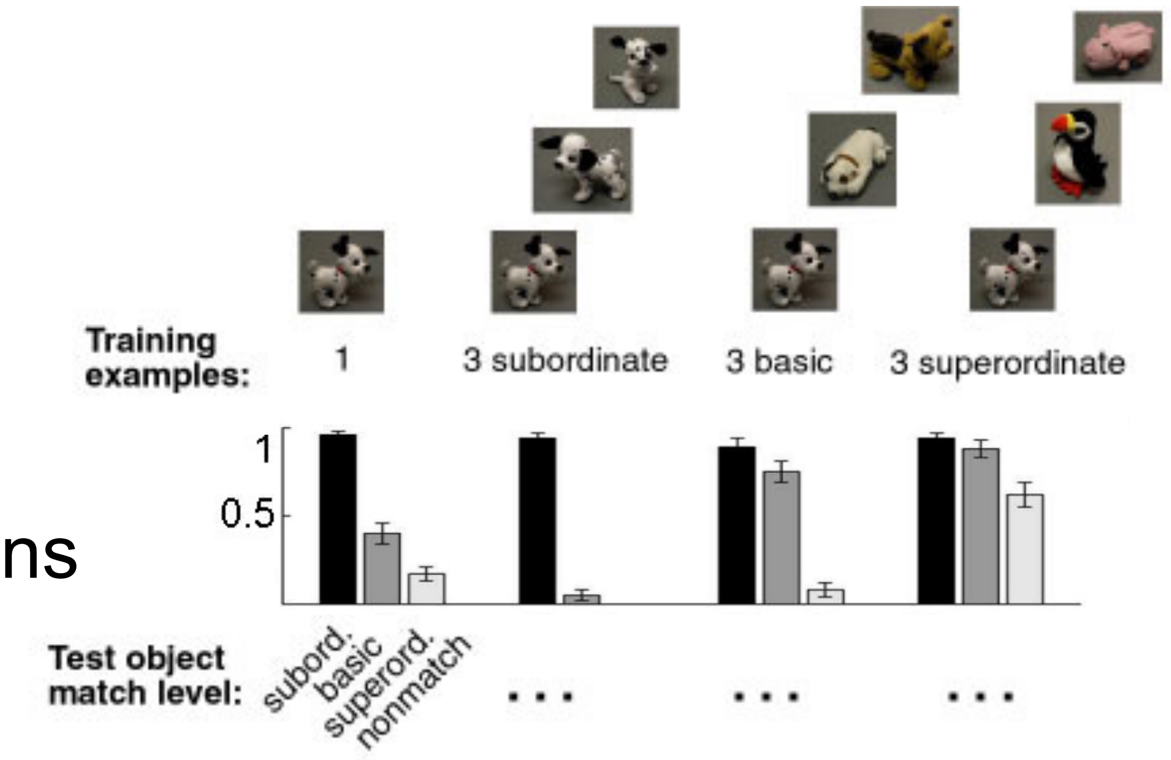
Suspicious coincidence



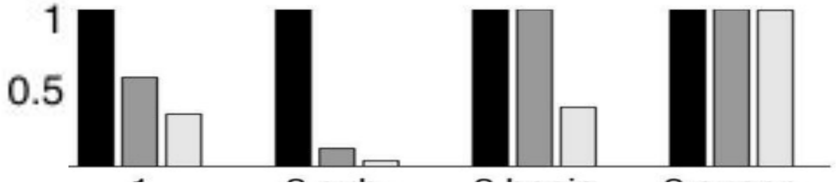
Modeling word learning (Xu & Tenenbaum, 2007)



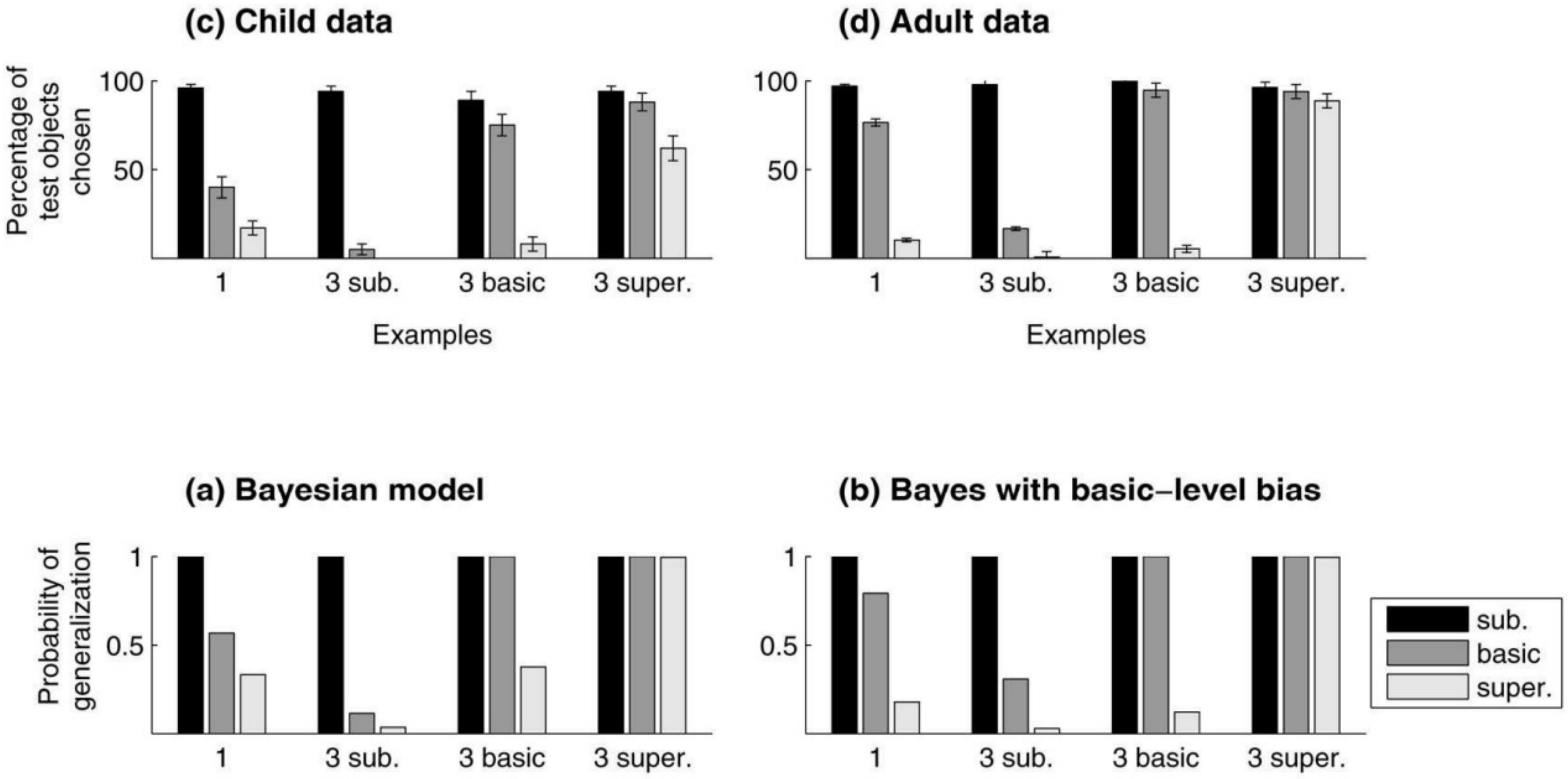
Children’s generalizations



Bayesian
concept learning with
tree-structured
hypothesis space



An additional inductive bias in adults: the basic level



Looking forward (beyond basic Bayes)

- Can we see these ideas at work in more natural cognitive function, not just toy problems and games?
  - How might differently structured hypothesis spaces, different likelihood functions or priors, be needed?
- Can we move from ‘weak rational analysis’ to ‘strong rational analysis’ in the priors, as with the likelihood?
  - “Weak”: behavior consistent with some reasonable prior.
  - “Strong”: behavior consistent with the “correct” prior given the structure of the world.
- Can we work with more flexible priors, not just restricted to a small subset of all logically possible concepts?
  - Would like to be able to learn any concept, even very complex ones, given enough data
  - E.g., Language-like hypothesis space?
- Can we describe formally how these hypothesis spaces and priors are generated by abstract knowledge or theories?
- Can we explain how people learn these rich priors?

Looking forward (beyond basic Bayes)

- Can we see these ideas at work in more natural cognitive function, not just toy problems and games?
  - How might differently structured hypothesis spaces, different likelihood functions or priors, be needed?
- Can we move from 'weak rational analysis' to 'strong rational analysis' in the priors, as with the likelihood?
  - "Weak": behavior consistent with some reasonable prior.
  - "Strong": behavior consistent with the "correct" prior given the structure of the world.
- Can we work with more flexible priors, not just restricted to a small subset of all logically possible concepts?
  - Would like to be able to learn any concept, even very complex ones, given enough data
  - E.g., Language-like hypothesis space?
- Can we describe formally how these hypothesis spaces and priors are generated by abstract knowledge or theories?
- Can we explain how people learn these rich priors?

Learning more natural concepts

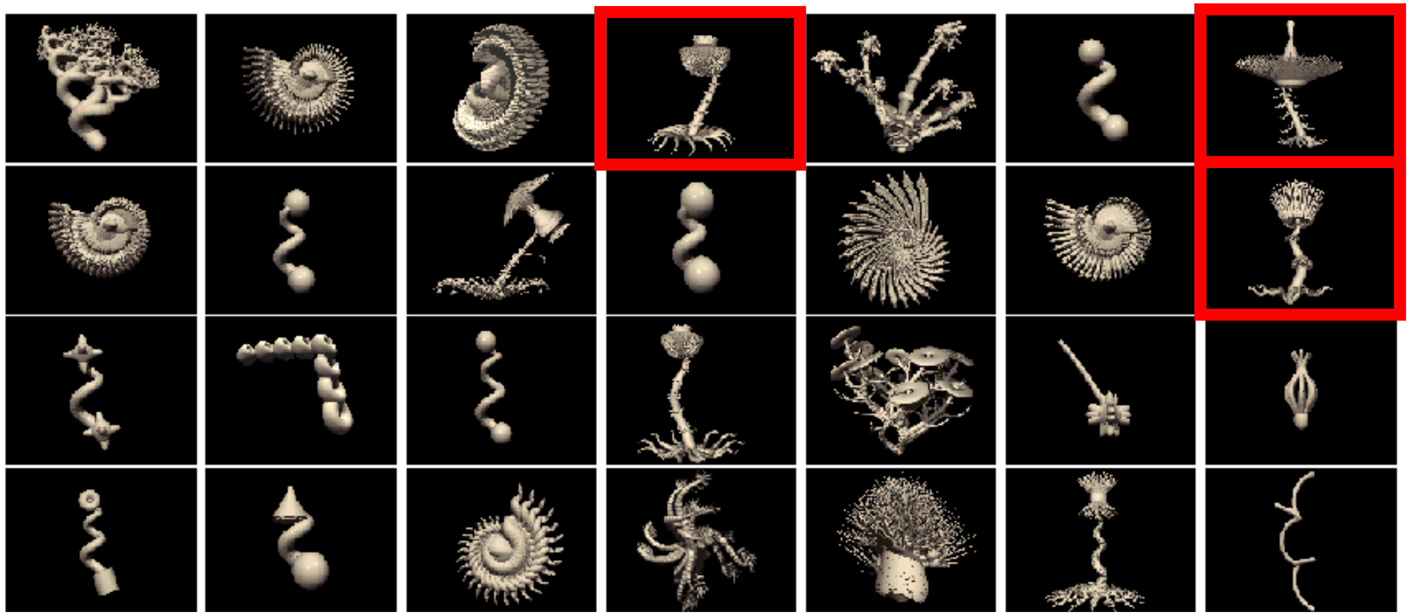
"horse"

"horse"

"horse"

For Cog Sci: simplified version that gets to the core of human cognition

"tufa"



For AI: principled approach for solving more complex problems

"tufa"

"tufa"

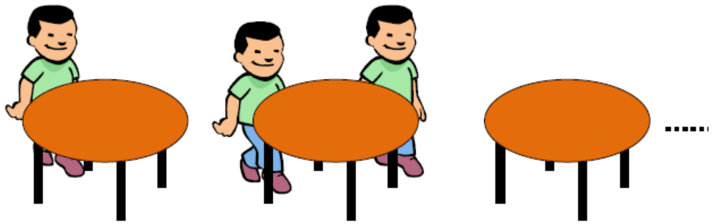
More sophisticated Bayesian models
+
Computer vision techniques (feature extraction)

One-Shot Learning with a Hierarchical Nonparametric Bayesian Model

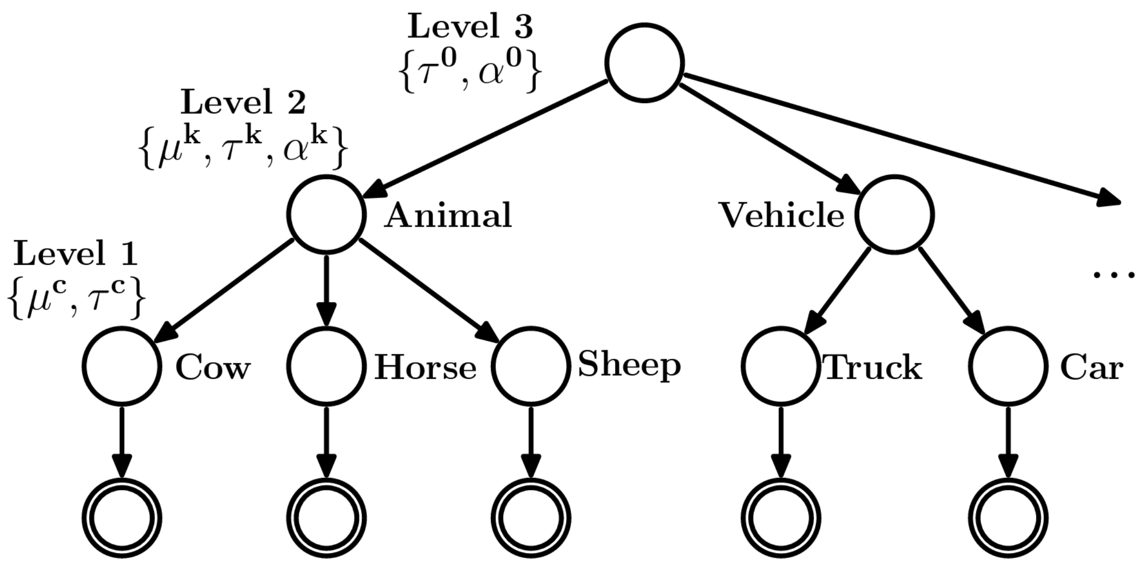
Ruslan Salakhutdinov RSALAKHU@MIT.EDU
Department of Statistics, University of Toronto
Toronto, Ontario, Canada

Josh Tenenbaum JBT@MIT.EDU
Department of Brain and Cognitive Sciences, MIT
Cambridge, MA, USA

Antonio Torralba
CSAIL, MIT
Cambridge, MA, USA



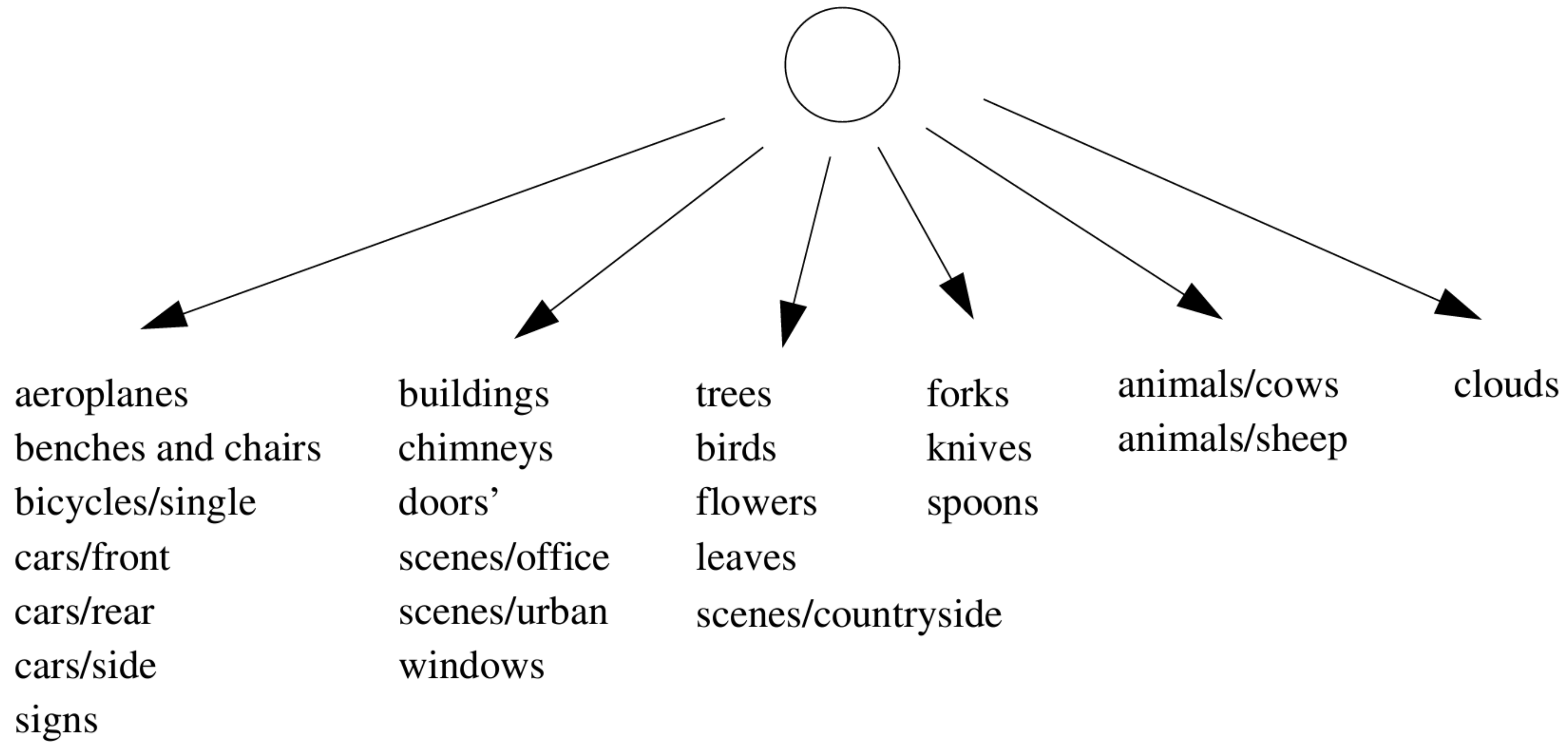
nested Chinese Restaurant Process (nCRP)



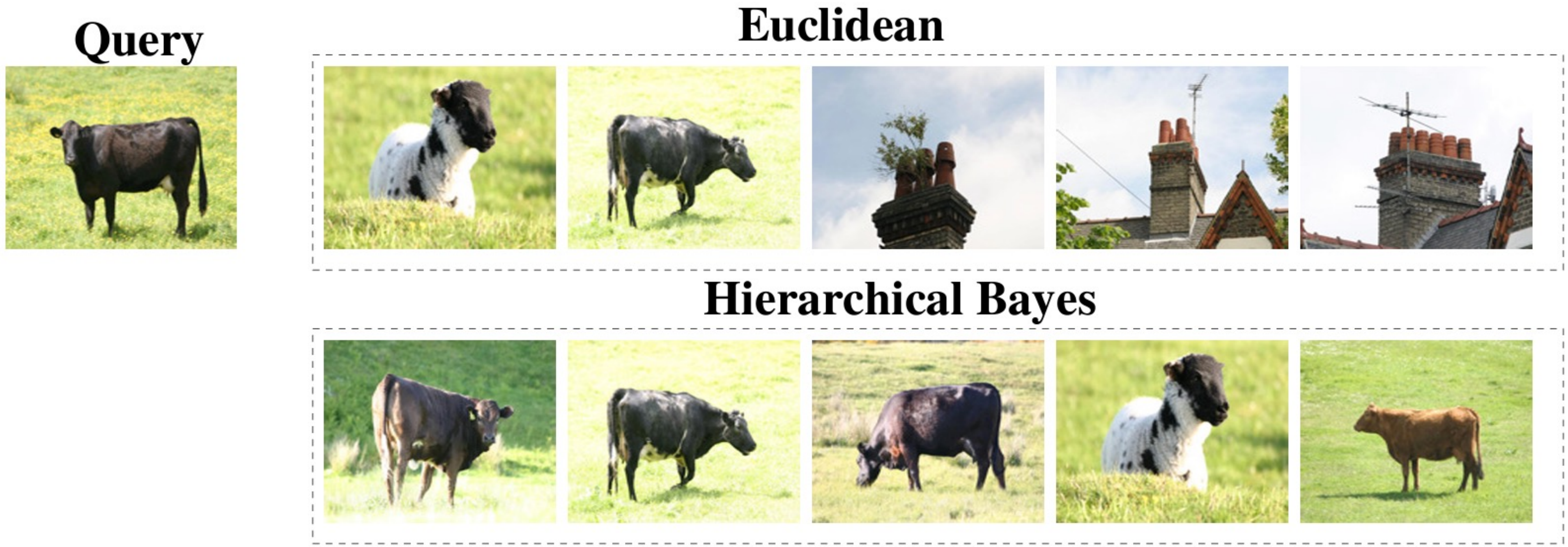
- For each super-category $k = 1, .., \infty$:
  draw θ^2 using Eq. 4.
- For each basic category $c^k = 1, .., \infty$, placed under each super-category k:
  draw θ^1 using Eq. 2.
- For each observation $n = 1, ..., N$
  draw $\mathbf{z}_n \sim \text{nCRP}(\gamma)$
  draw $\mathbf{x}^n \sim \mathcal{N}(\mathbf{x}^n|\theta^1, \mathbf{z}_n)$ using Eq. 1

Results on naturalistic images

- 24 categories in MSR Object Recognition Dataset



Results on naturalistic images



Looking forward (beyond basic Bayes)

- Can we see these ideas at work in more natural cognitive function, not just toy problems and games?
  - How might differently structured hypothesis spaces, different likelihood functions or priors, be needed?
- Can we move from ‘weak rational analysis’ to ‘strong rational analysis’ in the priors, as with the likelihood?
  - “Weak”: behavior consistent with some reasonable prior.
  - “Strong”: behavior consistent with the “correct” prior given the structure of the world.
- Can we work with more flexible priors, not just restricted to a small subset of all logically possible concepts?
  - Would like to be able to learn any concept, even very complex ones, given enough data
  - E.g., Language-like hypothesis space?
- Can we describe formally how these hypothesis spaces and priors are generated by abstract knowledge or theories?
- Can we explain how people learn these rich priors?

Is cognition rational?

- Weak version of Bayesian cognition:
  - Is cognition well described by a rational (Bayesian) model of inference, using some sensible priors and likelihoods?

E.g., Coin flipping, number game, tufa

- Strong version of Bayesian cognition:
  - Is cognition well described by an optimal (Bayesian) model of inference, using the “right” priors and likelihoods?
  - Is this possible?
  - At least in some cases, perhaps yes.

Everyday prediction problems

PSYCHOLOGICAL SCIENCE

Research Article

Optimal Predictions in Everyday Cognition

Thomas L. Griffiths[1] and Joshua B. Tenenbaum[2]

[1]Department of Cognitive and Linguistic Sciences, Brown University, and [2]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Everyday prediction problems

- You stopped by a friend's apartment, and she has been watching a movie for 15 minutes. What would you predict for the length of the movie in total?

- You stopped by a friend's apartment, and she has been watching a movie for 75 minutes. What would you predict for the length of the movie in total?

- A movie has grossed 15 million dollars at the box office, but you don't know how long it's been running. How much will it gross in total?

- You encounter a phenomenon or event with an unknown extent or duration, t_{total}, at a random time or value of $t_{past} < t_{total}$. What is the total extent or duration t_{total}?

Bayesian modeling

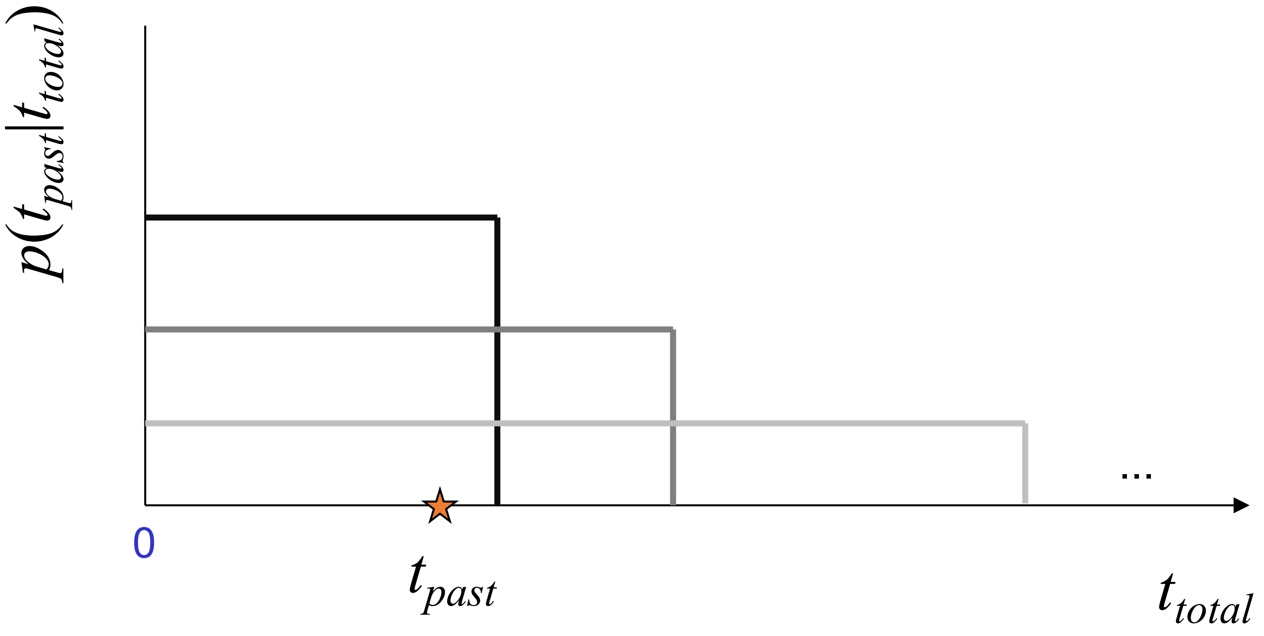
$$p(t_{total}|t) \propto p(t|t_{total})p(t_{total})$$

$$p(t|t_{total}) = \frac{1}{t_{total}}$$

Likelihood

$$p(t|t_{total}) = \frac{1}{t_{total}}$$

Assume uniformly sample *t* from the range (0, t_{total})



Bayesian modeling

$$p(t_{total}|t) \propto p(t|t_{total})p(t_{total})$$

$$p(t|t_{total}) = \frac{1}{t_{total}}$$

$$p(t_{total}|t) \propto \frac{p(t_{total})}{t_{total}}$$

Form of $p(t_{total})$?

Uninformative prior

- Not matter what domains (movie gross, movie runtimes, etc.), a single uninformative prior:

$$p(t_{total}) \propto 1/t_{total}$$

nature

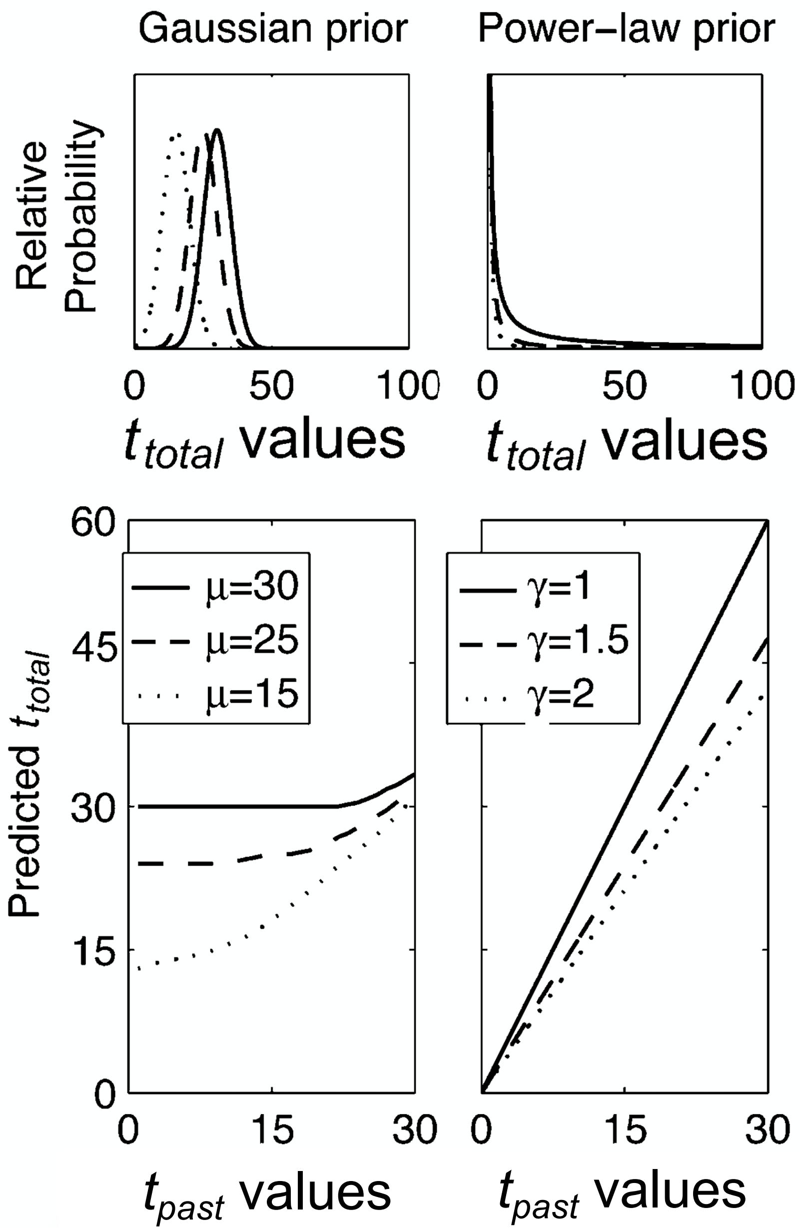
Implications of the Copernican principle for our future prospects

J. Richard Gott III

Nature **363**, 315–319 (1993) | Cite this article

Making only the assumption that you are a random intelligent observer, limits for the total longevity of our species of 0.2 million to 8 million years can be derived at the 95% confidence level. Further consideration indicates that we are unlikely to colonize the Galaxy, and that we are likely to have a higher population than the median for intelligent species.

Different priors have qualitatively different predictions



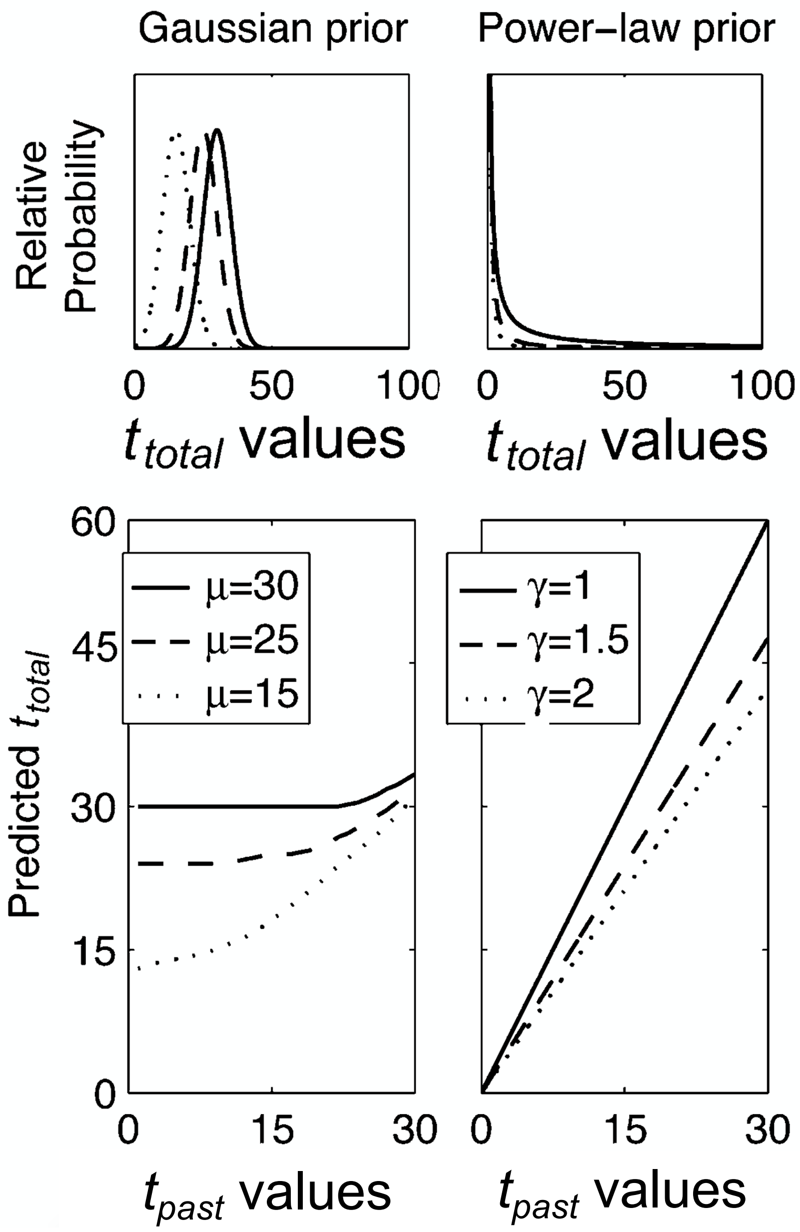
Gaussian prior:

$$P(t_{total}) \propto \exp(-\frac{1}{2\sigma^2}(t_{total}-\mu)^2)$$

Power-law prior:

$$P(t_{total}) \propto t_{total}^{-\gamma}$$

Different priors are appropriate in different domains

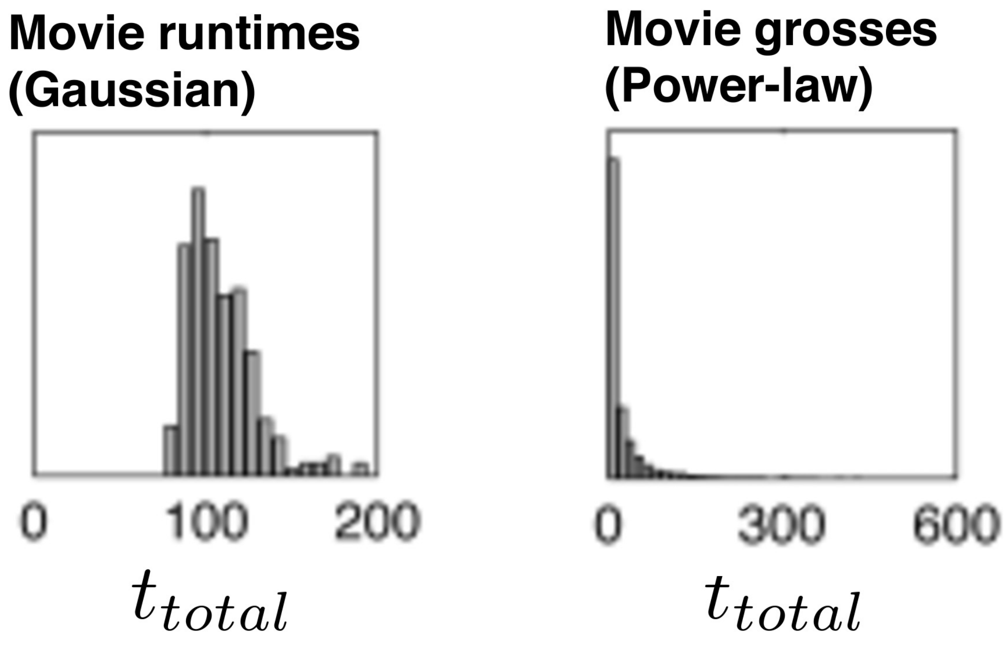


Gaussian prior:

$$P(t_{total}) \propto \exp(-\frac{1}{2\sigma^2}(t_{total} - \mu)^2)$$

Power-law prior:

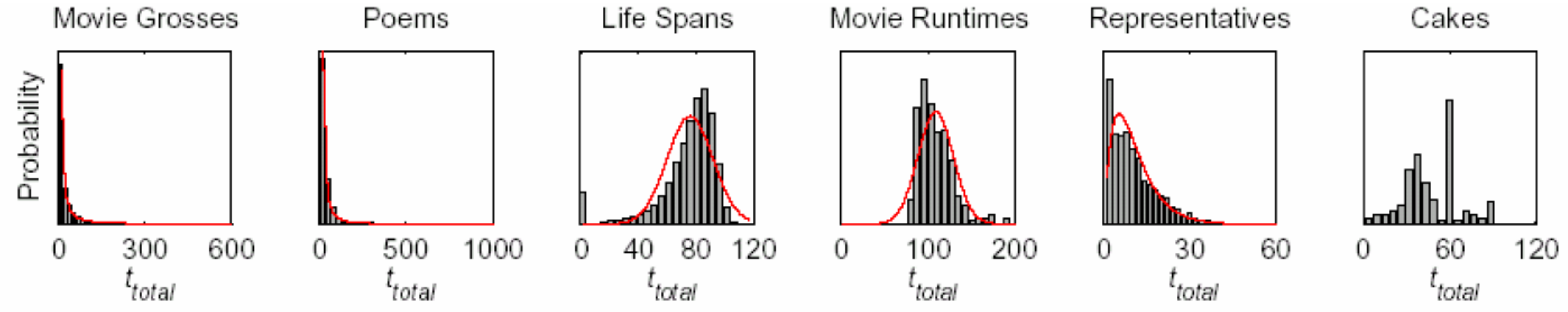
$$P(t_{total}) \propto t_{total}^{-\gamma}$$



Evaluating human predictions in different domains

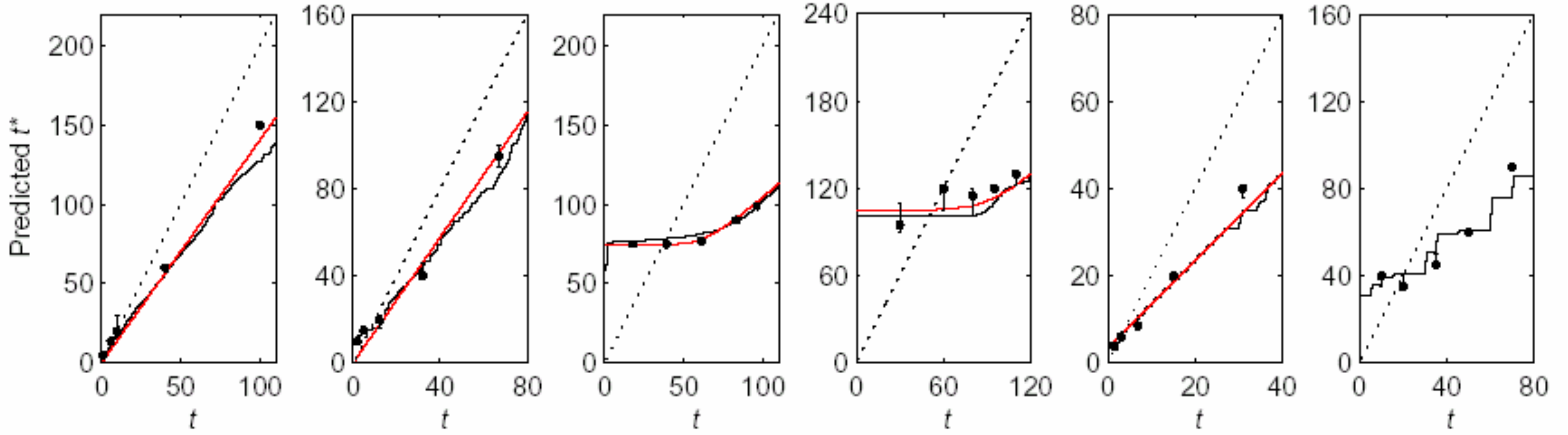
- **Movie gross:** You read about a movie that has made $60 million to date. How much money will it make in total?
- **Poem lengths:** Your friend quotes to you from line 17 of his favorite poem. How long is the poem?
- **Life spans:** You meet someone who is 78 years old. How long will they live?
- **Movie runtimes**: You stopped by a friend's apartment, and she has been watching a movie for 55 minutes. What would you predict for the length of the movie in total?
- **Terms of U.S. representatives:** You meet a US congressman who has served for 11 years. How long will he serve in total?
- **Baking times for cakes:** You see that something has been baking in the oven for 34 minutes. How long until it's ready?
- Use values of t_{past} for each domain, creating 5 events in each domain
- People predict t_{total} for each event

Priors $p(t_{total})$ based on empirically measured durations or magnitudes for many real-world events in each class:



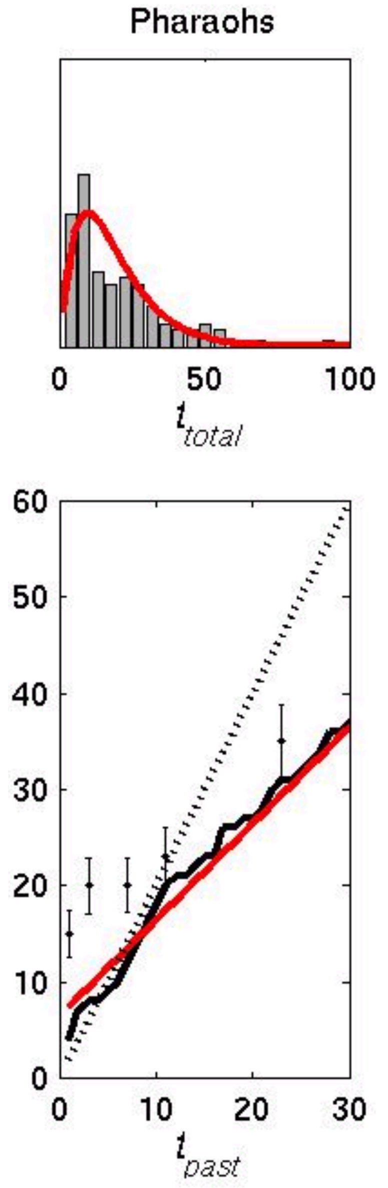
Median human judgments of the total duration or magnitude t_{total} of events in each class, given one random observation at a duration or magnitude t, versus Bayesian predictions (median of $p(t_{total}|t)$).

Dot: human
Solid line: emp. priors
Dashed line: uninform. prior



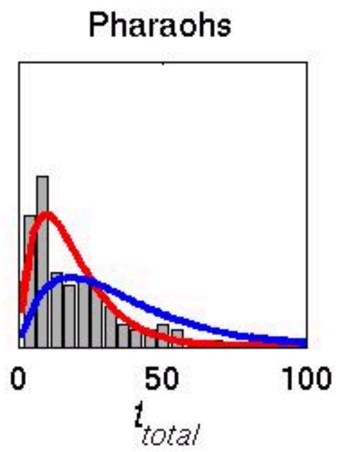
A special case

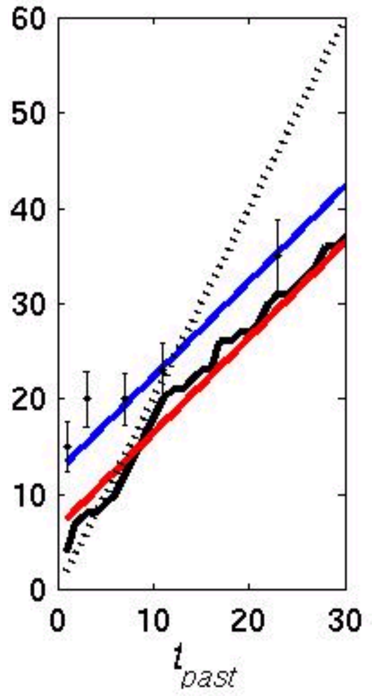
- Reigns of pharaohs: You learned that in ancient Egypt, there was a great flood in the 11th year of a pharaoh's reign. How long did he reign?



A special case

- Reigns of pharaohs: You learned that in ancient Egypt, there was a great flood in the 11th year of a pharaoh's reign. How long did he reign?





- Ask 35 undergraduate students: How long did the typical pharaoh reign in ancient Egypt?
- Recalibrate prior based on their responses

Is cognition rational?

- Strong version of Bayesian cognition:
  - Is cognition well described by an optimal (Bayesian) model of inference, using the “right” priors and likelihoods?
  - Is this possible?
  - At least in some cases, perhaps yes.
    - Priors vary with the qualitative form (power-law vs. exponential) + correct parameter values
    - People seem to accurately absorb the statistics of their environment for everyday quantitates
      - In ML/AI: learning priors from prior experience
    - In the absence of concrete experience (e.g., Pharaoh), priors may be generated by qualitative background knowledge
      - In ML/AI: query a knowledge base / language model