Probabilistic models of cognition

Lecture 9

What makes people smart?

- Memory? No.
- Deductive inference? No.

- Intuitions and inductions.

Everyday inductive leaps

How can people learn so much about the world from such limited evidence?

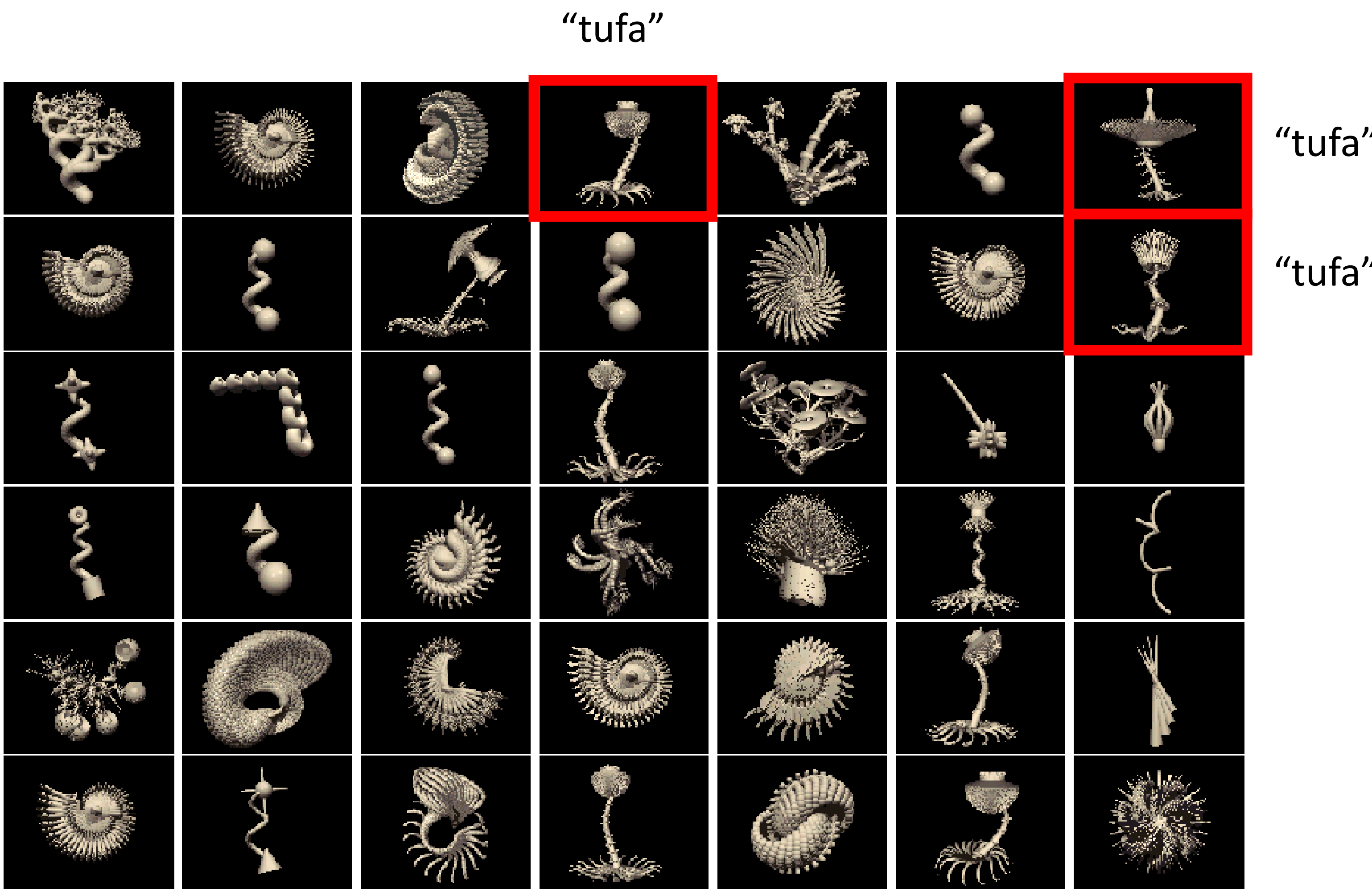
- Learning concepts from examples

"horse"

"horse"

"horse"

Learning concepts from examples



Tenenbaum

Learning concepts from examples

Tenenbaum

Learning concepts from examples

Tenenbaum

Everyday inductive leaps

How can people learn so much about the world from such limited evidence?

- Kinds of objects and their properties
- The meanings of words, phrases, and sentences
- Dynamics and durations of events
- Cause-effect relations
- The beliefs, goals and plans of other people
- Social structures, conventions, and rules

Probabilistic view

- People have prior knowledge
- Prior knowledge is often highly structured
- Learning approximates optimal statistical inference
- Ultimately this is about how to learn as much as possible from the statistics of your environment

Work by Tenenbaum, Griffiths, Kemp

1. The discovery of structural form (Kemp and Tenenbaum, 2008)
2. Optimal predictions in everyday cognition (Griffiths and Tenenbaum, 2006)
3. Markov Chain Monte Carlo with people (Sanborn and Griffiths, 2008)

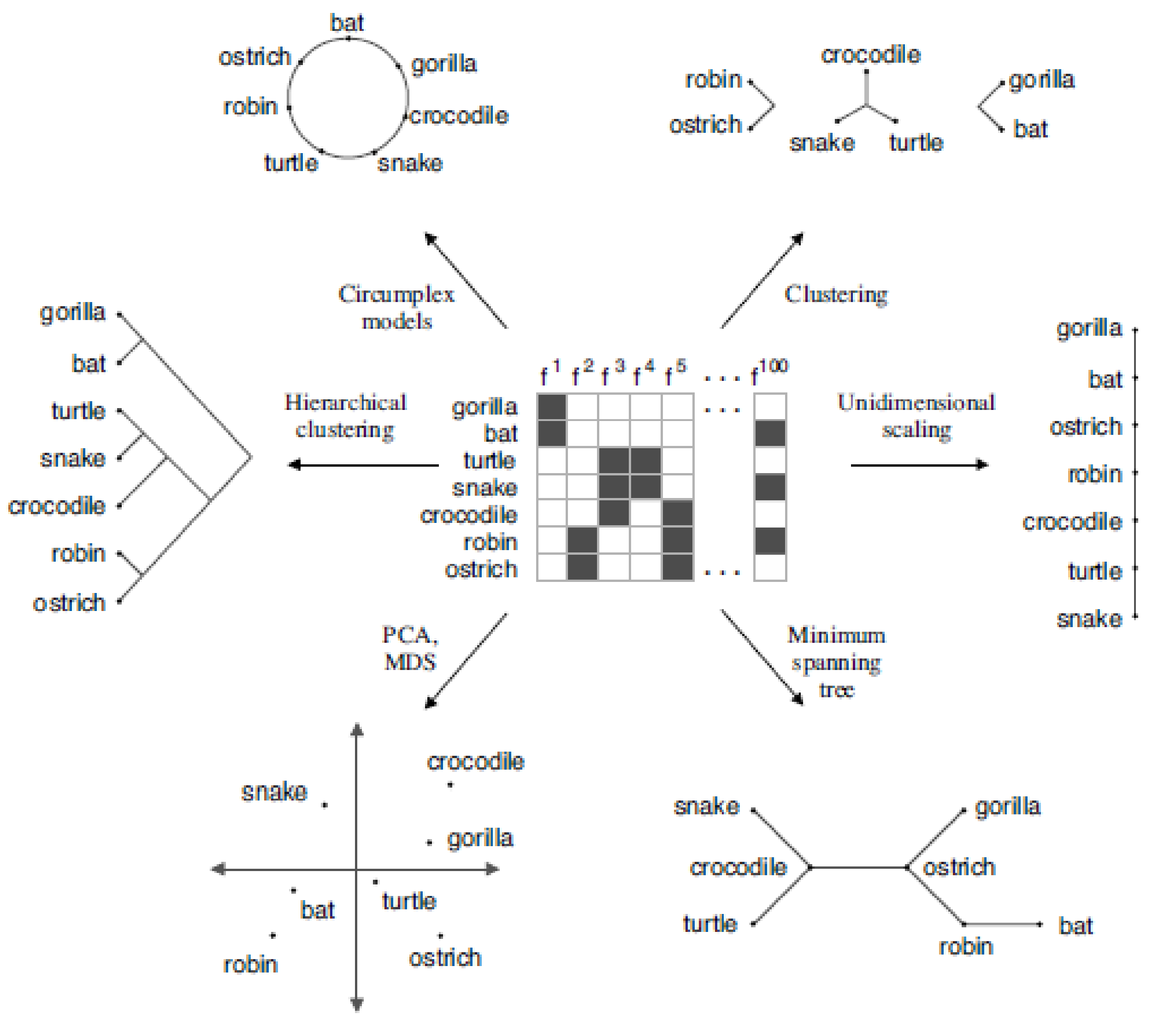
Learning structures

- Scientists discover structure in their data:
  - Linnaeus: classifying species in tree
  - Mendeleev: periodic system of elements
- Children discover structure in the world
  - Poodle can be dog and animal, but not dog and cat
  - Friendship networks are cliques
  - Days of week, seasons are cyclical

Two questions in structure learning

1. What form does the structure have? Tree, ring, linear order, clusters, etc.
2. Given a particular form, what instance of this form explains the data?

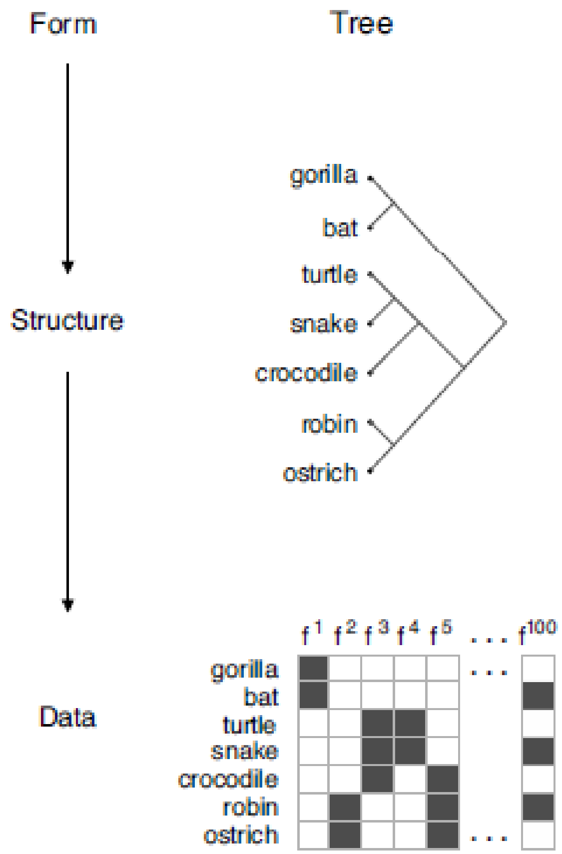
- Structure-learning algorithms only concerned with second problem.



bat
ostrich
gorilla
robin
crocodile
turtle
snake
crocodile
robin
ostrich
snake
turtle
gorilla
bat
Circumplex models
Clustering
gorilla
bat
turtle
snake
crocodile
robin
ostrich
Hierarchical clustering
f1 f2 f3 f4 f5 . . . f100
gorilla
bat
turtle
snake
crocodile
robin
ostrich
Unidimensional scaling
gorilla
bat
ostrich
robin
crocodile
turtle
snake
PCA, MDS
Minimum spanning tree
snake
crocodile
gorilla
bat
turtle
robin
ostrich
snake
gorilla
crocodile
ostrich
turtle
bat
robin

Kemp

Hierarchical generative model



Kemp

Hierarchical Bayesian inference

- Data D
- Structure S
- Form F

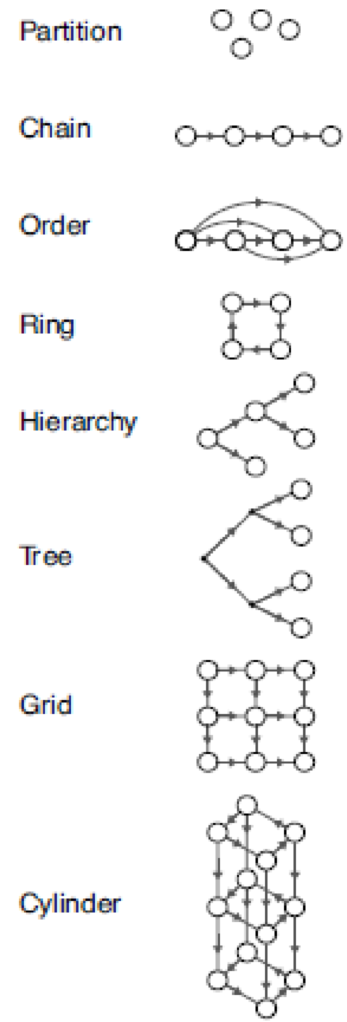
uniform prior

similarity metric

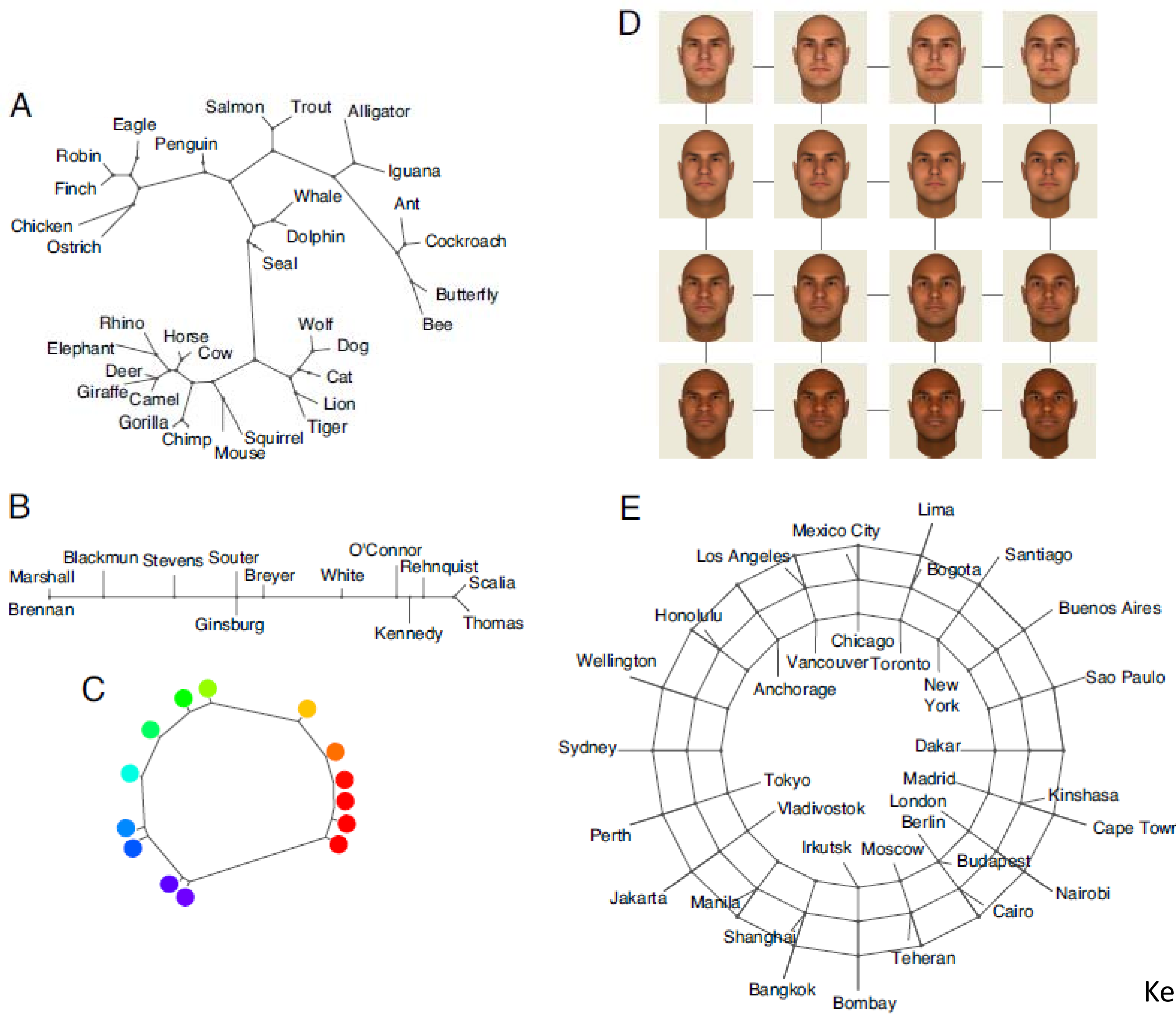
$$p(S,F\,|\,D) \propto p(F)\,p(S\,|\,F)\,p(D\,|\,S)$$

joint posterior over structure and form

higher when S has fewer clusters



Kemp



A
Salmon
Trout
Alligator
Eagle
Penguin
Robin
Iguana
Finch
Whale
Ant
Chicken
Dolphin
Cockroach
Ostrich
Seal
Butterfly
Bee
Rhino
Wolf
Horse
Dog
Elephant
Cow
Deer
Cat
Giraffe
Camel
Lion
Gorilla
Tiger
Chimp
Squirrel
Mouse
D
B
O'Connor
Blackmun
Stevens
Souter
Rehnquist
Marshall
Breyer
White
Scalia
Brennan
Ginsburg
Kennedy
Thomas
C
E
Lima
Mexico City
Los Angeles
Santiago
Bogota
Honolulu
Buenos Aires
Chicago
Wellington
Vancouver
Toronto
Anchorage
New
York
Sao Paulo
Sydney
Dakar
Tokyo
Madrid
Kinshasa
Vladivostok
London
Berlin
Cape Town
Perth
Irkutsk
Moscow
Budapest
Jakarta
Manila
Nairobi
Cairo
Shanghai
Teheran
Bangkok
Bombay

Kemp

Work by Tenenbaum, Griffiths, Kemp

1. The discovery of structural form (Kemp and Tenenbaum, 2008)
2. Optimal predictions in everyday cognition (Griffiths and Tenenbaum, 2006)
3. Markov Chain Monte Carlo with people (Sanborn and Griffiths, 2006)

Perception vs cognition

- “Perception may be optimal, but cognition is sloppy.”
- “Cognitive judgments under uncertainty are often characterized as the result of error-prone heuristics that are insensitive to prior probabilities.”

Griffiths

Movie grosses: Imagine you hear about a movie that has taken in 10 million dollars at the box office, but don't know how long it has been running. What would you predict for the total amount of box office intake for that movie?

Poem lengths: If your friend read you her favorite line of poetry, and told you it was line 5 of a poem, what would you predict for the total length of the poem?

Life spans: Insurance agencies employ actuaries to make predictions about people's life spans—the age at which they will die—based upon demographic information. If you were assessing an insurance case for an 18-year-old man, what would you predict for his life span?

Reigns of pharaohs: If you opened a book about the history of ancient Egypt to a page listing the reigns of the pharaohs, and noticed that at 4000 BC a particular pharaoh had been ruling for 11 years, what would you predict for the total duration of his reign?

Lengths of marriages: A friend is telling you about an acquaintance whom you do not know. In passing, he happens to mention that this person has been married for 23 years. How long do you think this person's marriage will last?

Movie run times: If you made a surprise visit to a friend, and found that they had been watching a movie for 30 minutes, what would you predict for the length of the movie?

Terms of U.S. representatives: If you heard a member of the House of Representatives had served for 15 years, what would you predict his total term in the House would be?

Baking times for cakes: Imagine you are in somebody's kitchen and notice that a cake is in the oven. The timer shows that it has been baking for 35 minutes. What would you predict for the total amount of time the cake needs to bake?

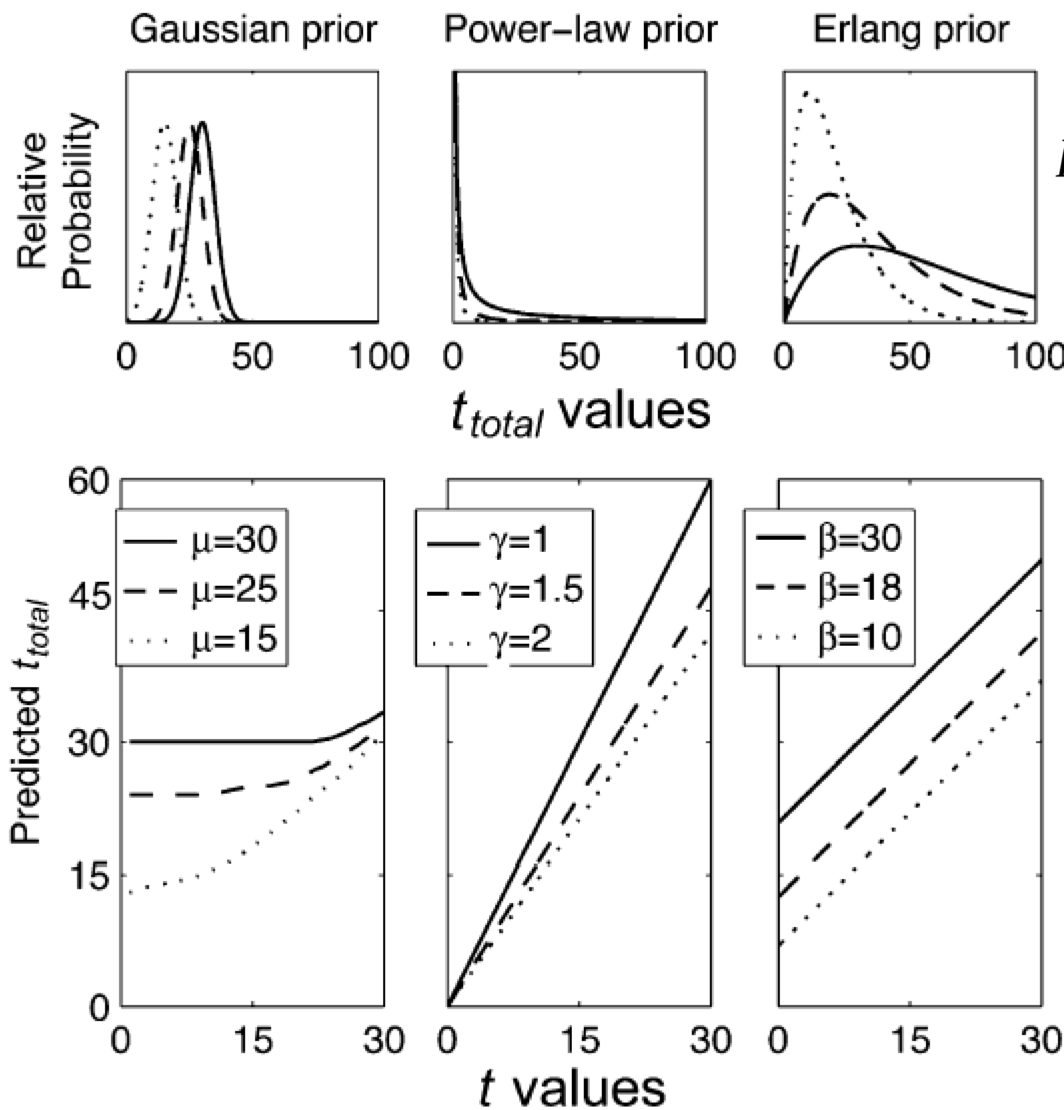
Waiting times: If you were calling a telephone box office to book tickets and had been on hold for 3 minutes, what would you predict for the total time you would be on hold?

Griffiths

$$p\left(t_{\text{total}} \mid t\right) \propto p\left(t \mid t_{\text{total}}\right) p\left(t_{\text{total}}\right)$$

$$\frac{1}{t_{\text{total}}}$$

Median read-out (what cost function?)



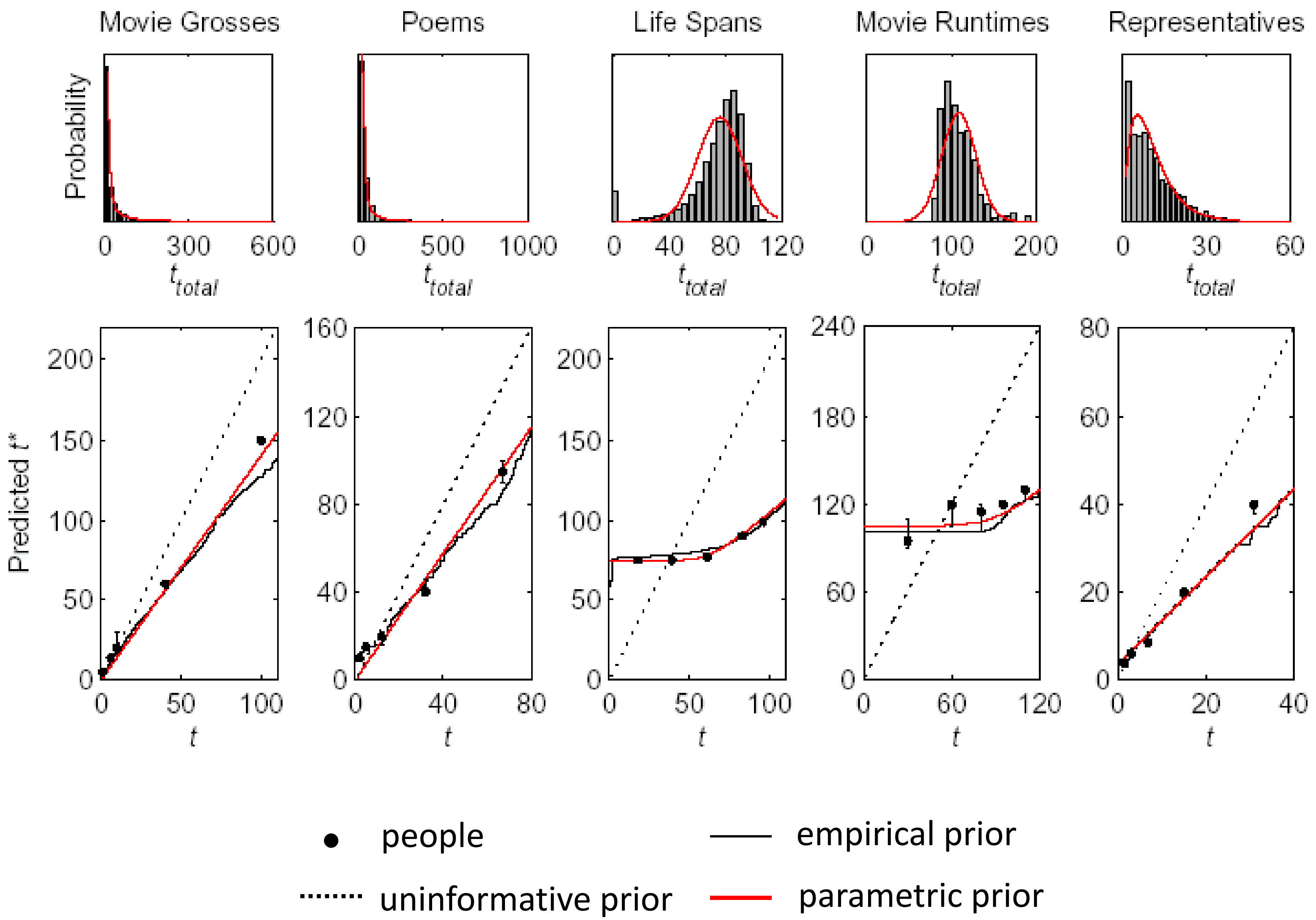
$$p\left(t_{\text{total}}\right) \propto t_{\text{total}}^{k-1} e^{-\frac{t_{\text{total}}}{\beta}}$$

Griffiths

Evaluating human predictions

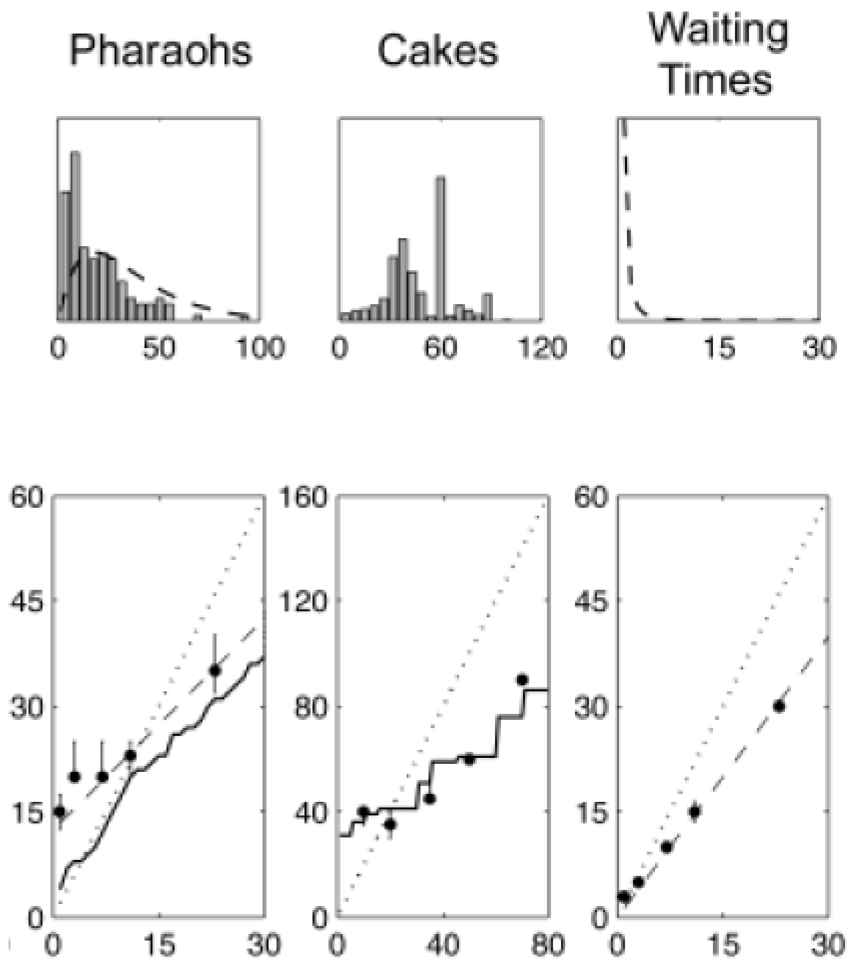
- Different domains with different priors:
  - a movie has made $60 million [power-law]
  - your friend quotes from line 17 of a poem [power-law]
  - you meet a 78 year old man [Gaussian]
  - a movie has been running for 55 minutes [Gaussian]
  - a U.S. congressman has served for 11 years [Erlang]
- Prior distributions derived from actual data
- Use 5 values of t for each
- People predict t_{total}

Griffiths



Movie Grosses
Poems
Life Spans
Movie Runtimes
Representatives
Probability
t_{total}
Predicted t^*
t
people
empirical prior
uninformative prior
parametric prior

Griffiths


Pharaohs
Cakes
Waiting Times
0
50
100
0
60
120
0
15
30
60
45
30
15
0
0
15
30
160
120
80
40
0
0
40
80
60
45
30
15
0
0
15
30

Griffiths

- People can make accurate predictions even in contexts where priors lack a simple form.
- People’s predictions can be used to reconstruct the priors they were using.
- People make inaccurate predictions when they know the appropriate form of the prior, but not the details of its parameters.
  - Pharao reigns
  - Maybe people might pick a prior by analogy when confronted with an unfamiliar prediction task

Work by Tenenbaum, Griffiths, Kemp

1. The discovery of structural form (Kemp and Tenenbaum, 2008)
2. Optimal predictions in everyday cognition (Griffiths and Tenenbaum, 2006)
3. Markov Chain Monte Carlo with people (Sanborn and Griffiths, 2006)

Learning by sampling

- People learn by modifying their beliefs about hypotheses.
- How do people learn probability distributions?
- Machines: sampling, e.g. Monte Carlo
- Markov Chain Monte Carlo: Markov chain that has the target distribution as stationary distribution
- Initialize with any state, guaranteed to converge after many iterations.

Metropolis-Hastings algorithm

(Metropolis et al., 1953; Hastings, 1970)

Step 1: propose a state (we assume symmetrically)

$$Q(x^{(t+1)} | x^{(t)}) = Q(x^{(t)}) | x^{(t+1)})$$

Step 2: decide whether to accept, with probability

$$A(x^{(t+1)}, x^{(t)}) = \min\left(1, \frac{p(x^{(t+1)})}{p(x^{(t)})}\right)$$

Metropolis acceptance function

$$A(x^{(t+1)}, x^{(t)}) = \frac{p(x^{(t+1)})}{p(x^{(t+1)}) + p(x^{(t)})}$$

Barker acceptance function

Griffiths

A task

Ask subjects which of two alternatives comes from a target category

Which animal is a frog?

Griffiths

A Bayesian analysis of the task

$$h_1 : x_1 \text{ is from } p(x|c); x_2 \text{ is from } g(x)$$

$$h_2 : x_2 \text{ is from } p(x|c); x_1 \text{ is from } g(x)$$

$$p(h_1|x_1, x_2) = \frac{p(x_1|c)g(x_2)p(h_1)}{p(x_1|c)g(x_2)p(h_1) + p(x_2|c)g(x_1)p(h_2)}$$

Assume:

$$p(h_1) = p(h_2)$$

$$g(x_1) = g(x_2)$$

Griffiths

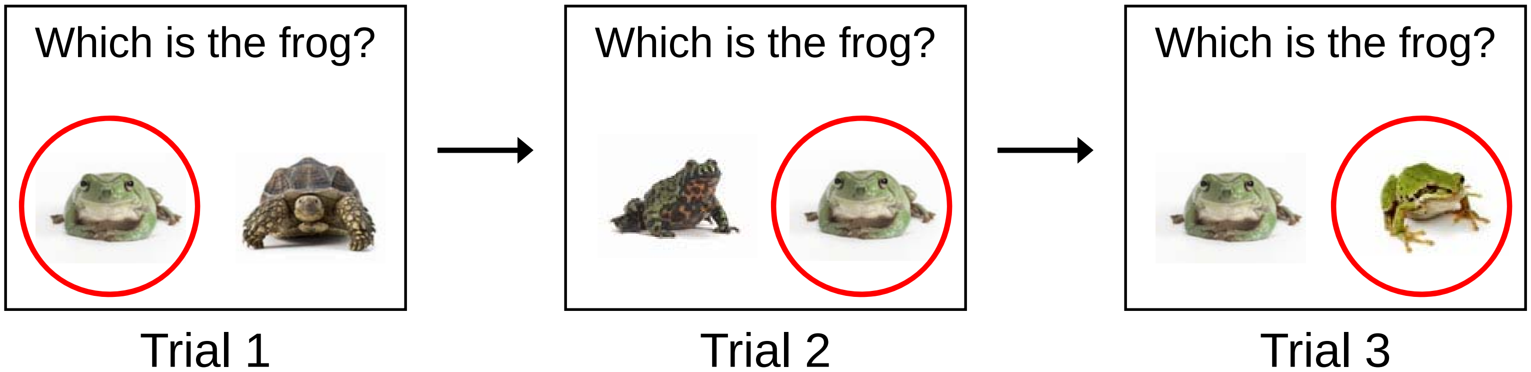
Response probabilities

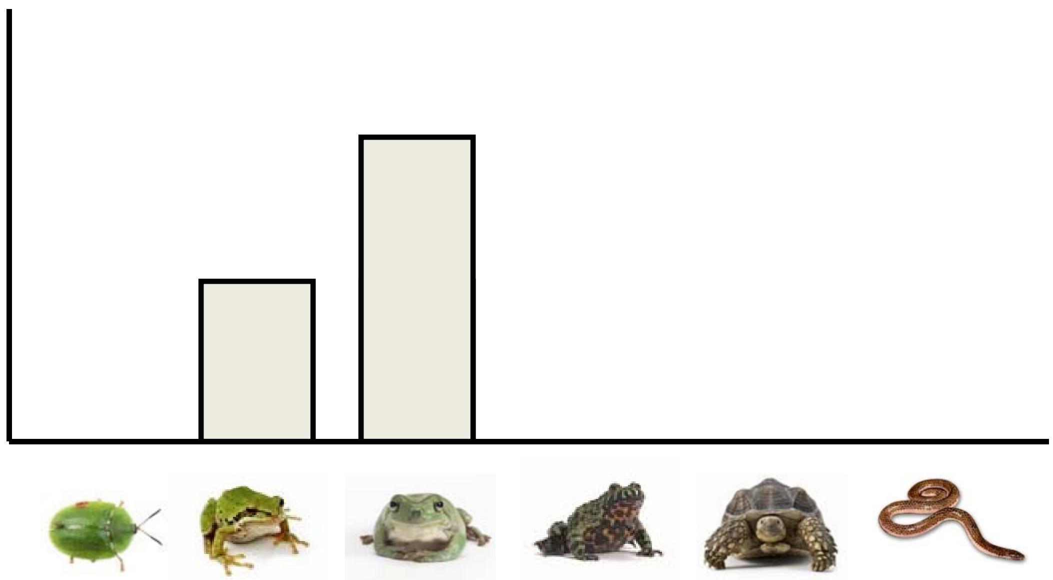
$$\frac{p(x_1|c)}{p(x_1|c) + p(x_2|c)}$$

If people probability match to the posterior, response probability is equivalent to the Barker acceptance function for target distribution $p(x|c)$

Griffiths

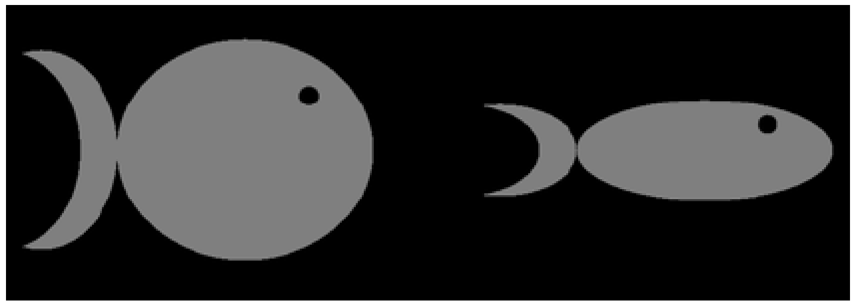
Collecting the samples





Griffiths

Experiment 1: Recovering trained distributions using MCMC



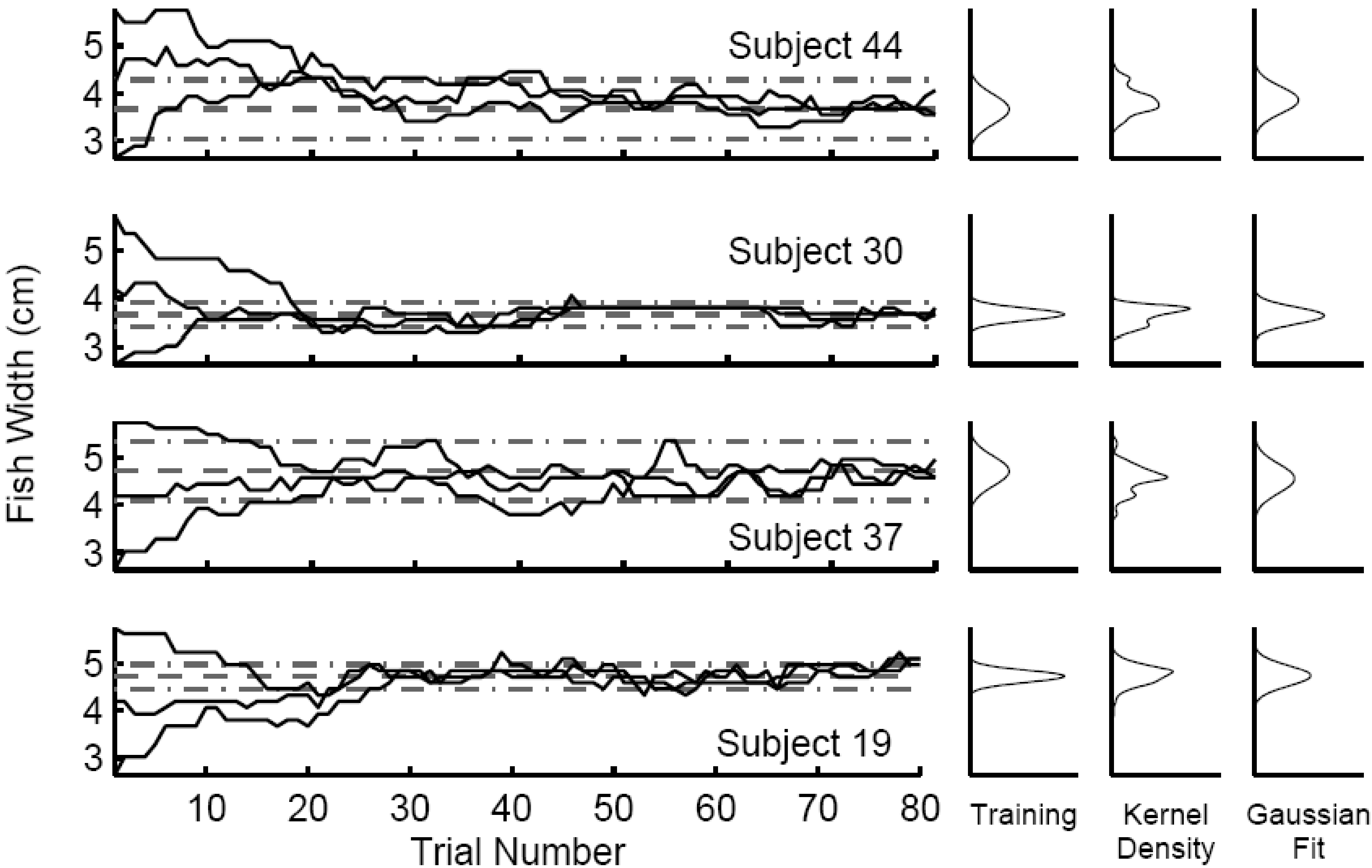
- Fish height is the only variable
- Ocean fish have uniformly distributed height
- Farm fish have normally distributed height
- First train subjects on both categories

Griffiths

After training

- Present two fish on each trial
- Task: which one was farm fish?
- One presented fish was state of the Markov chain
- Other presented fish was proposal (drawn from pseudo-Gaussian distribution)
- Training and MCMC trials in alternating blocks
- Interleave 3 MCMC chains with different initial values
- Does MCMC distribution converge to training distribution?

Markov chains

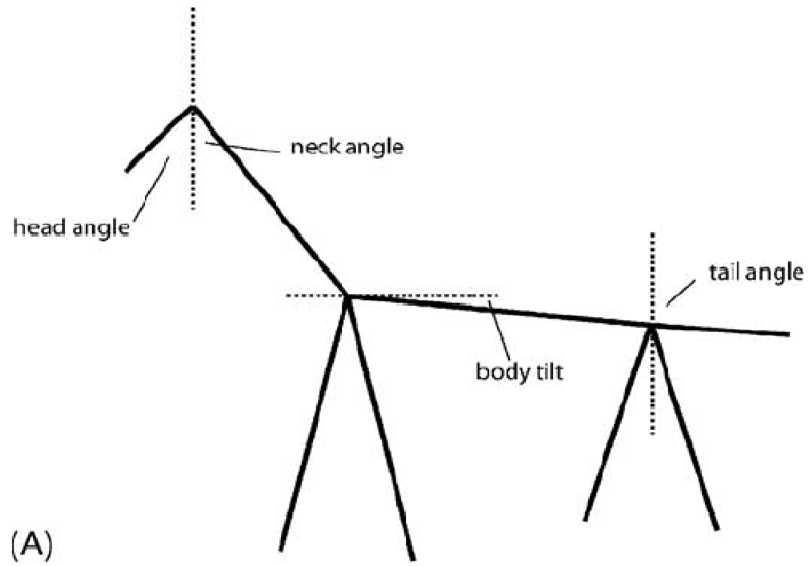


Griffiths

- This is a method for sampling from subjective probability distributions.
- So far only validation. What about unknown distributions?

Experiment 2: Sampling from natural categories

Examined distributions for four natural categories: giraffes, horses, cats, and dogs

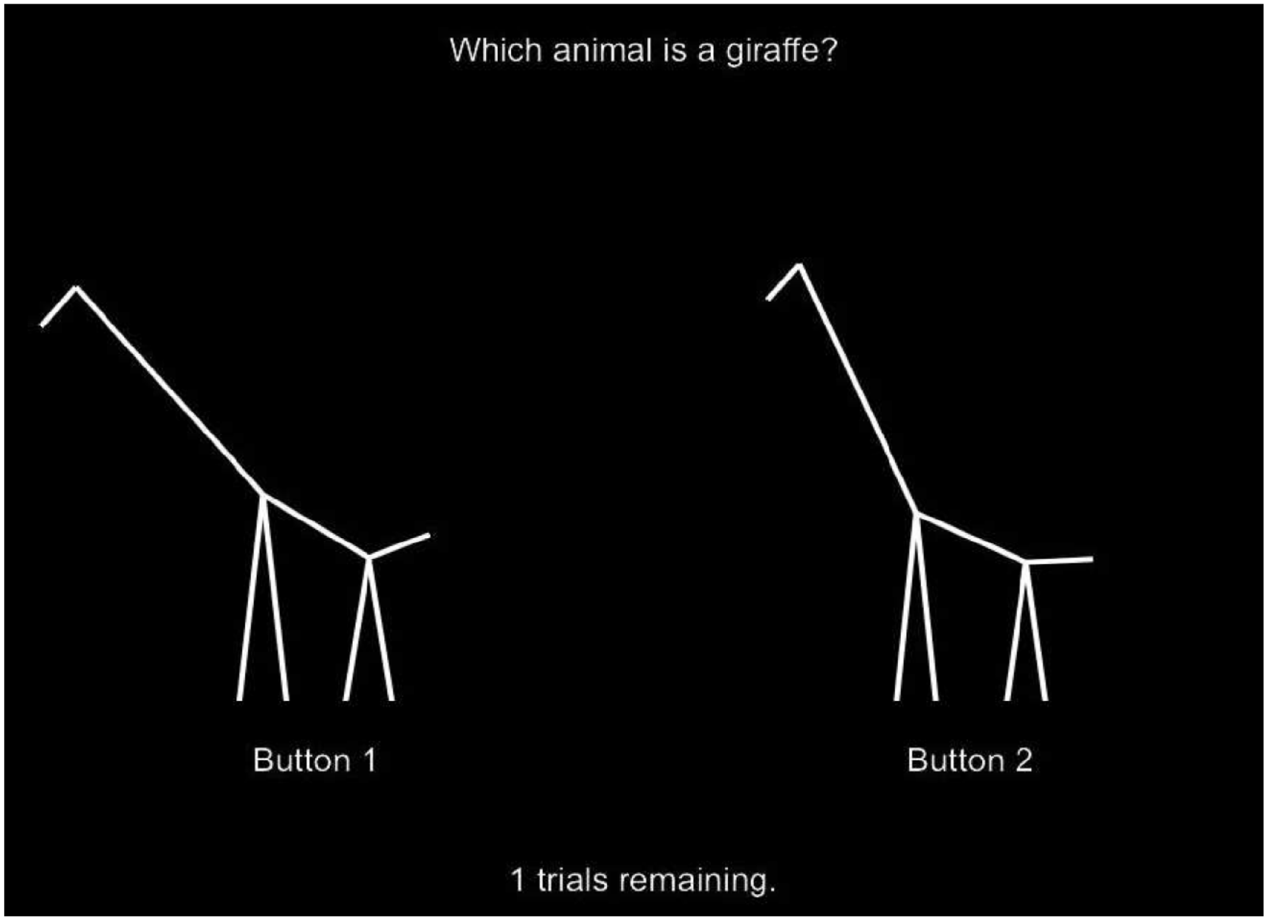


Presented stimuli with nine-parameter stick figures

(Olman & Kersten, 2004)

Griffiths

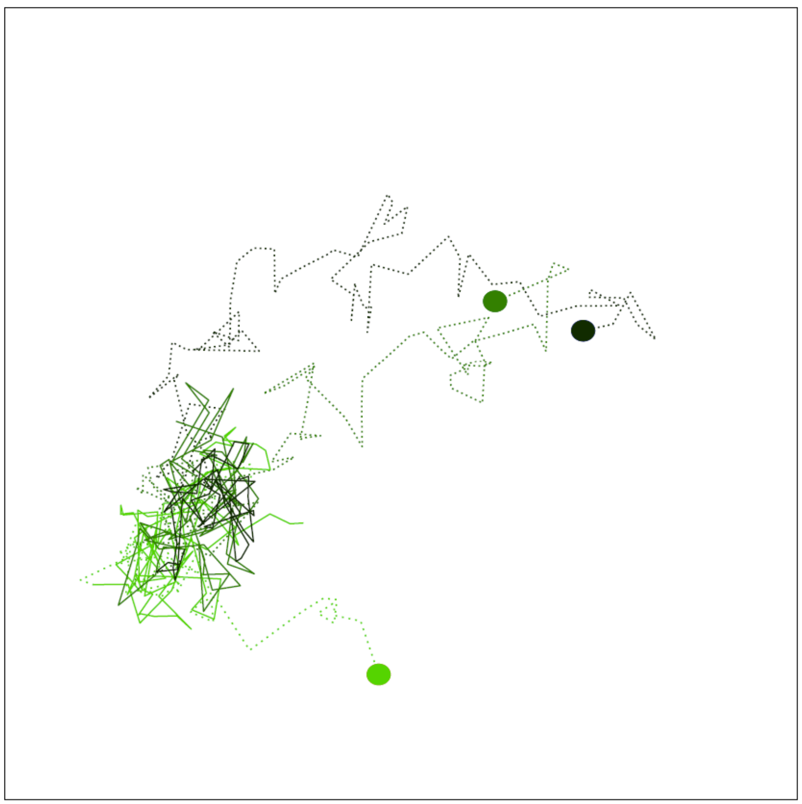
Choice task

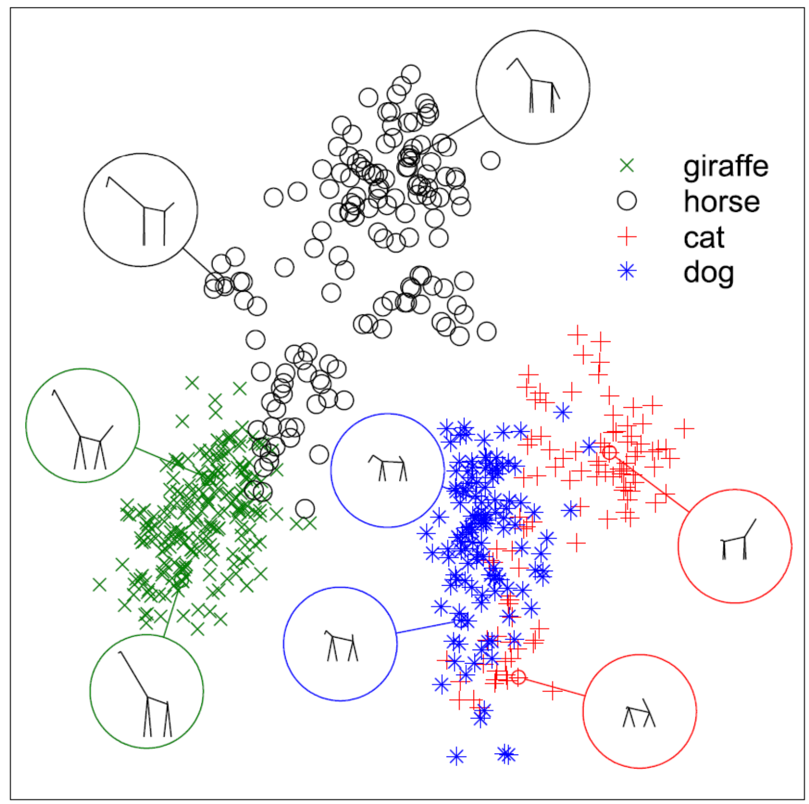


Griffiths

Samples from Subject 3

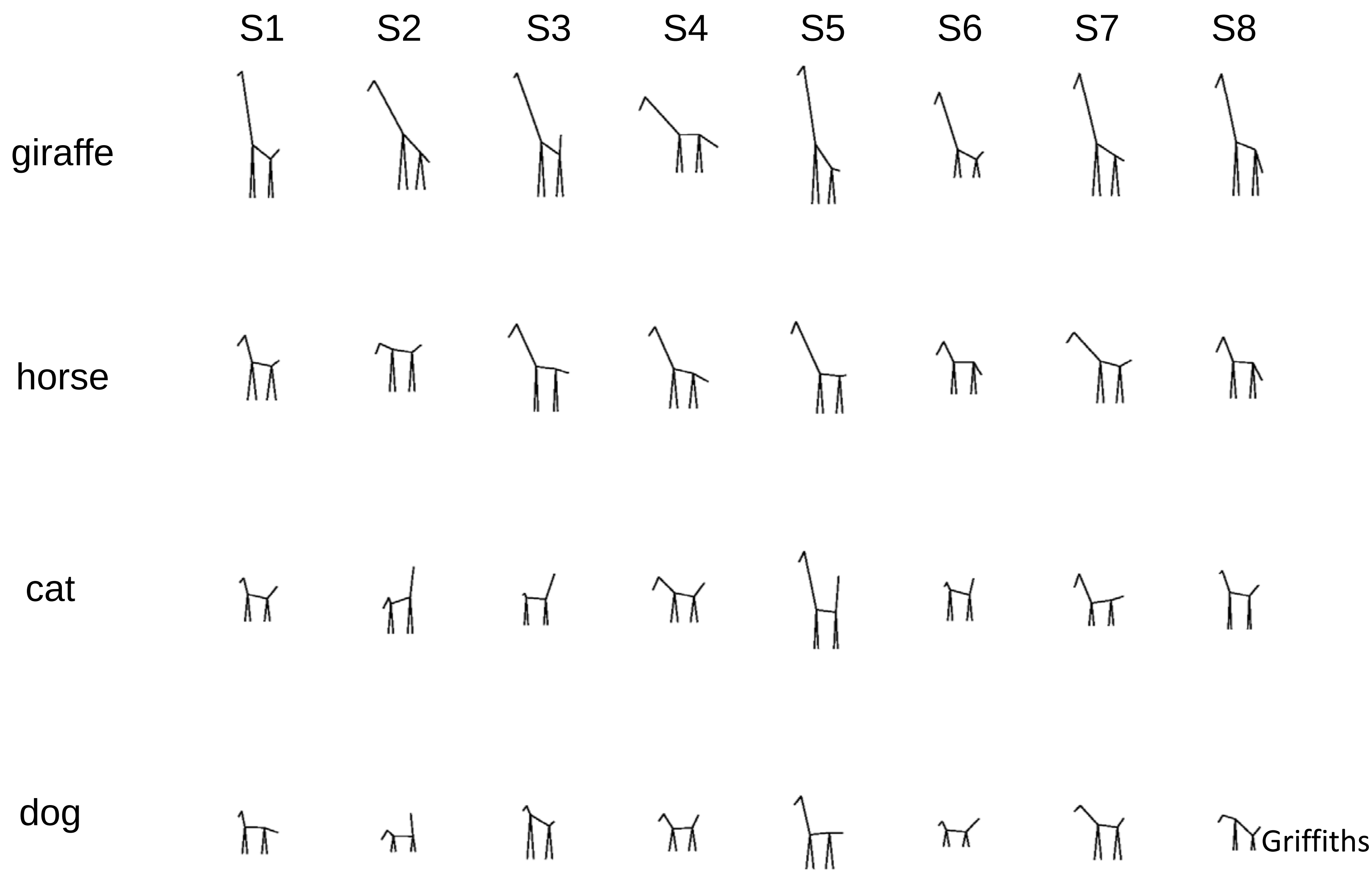
(projected onto a plane)



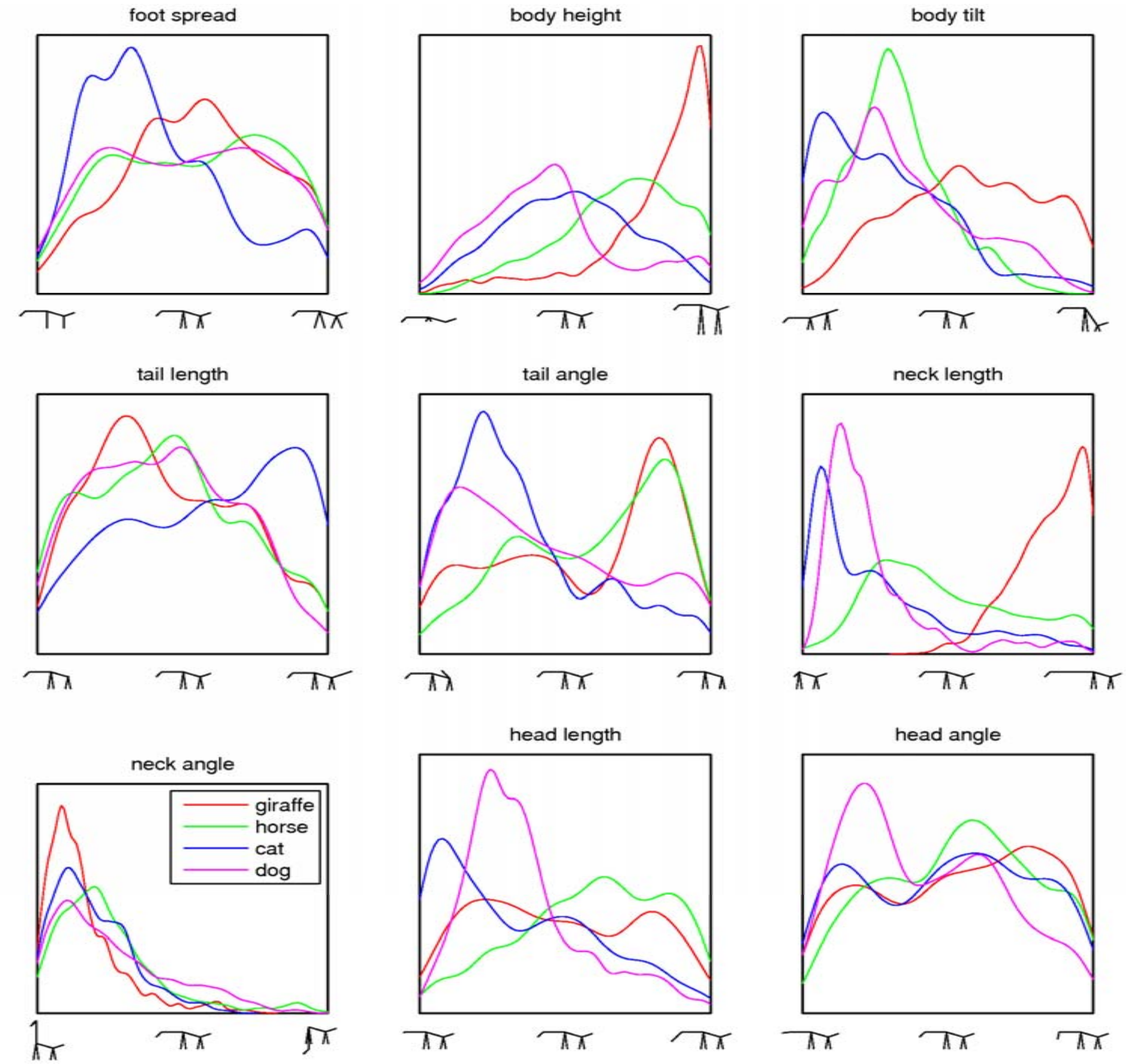


Griffiths

Mean animals by subject



Marginal densities
(aggregated across subjects)



Giraffes are distinguished by neck length, body height and body tilt

Horses are like giraffes, but with shorter bodies and nearly uniform necks

Cats have longer tails than dogs

Griffiths

Markov chain Monte Carlo with people

- Probabilistic models can guide the design of experiments to measure psychological variables
- Markov chain Monte Carlo can be used to sample from subjective probability distributions
  - category distributions (Metropolis-Hastings)
  - prior distributions (Gibbs sampling)
- Effective for exploring large stimulus spaces, with distributions on a small part of the space

Griffiths

Conclusion

- Probabilistic models give us a way to explore the knowledge that guides people's inferences
- Basic problem for both cognition and perception: identifying subjective probability distributions
- Two strategies:
  - natural statistics
  - Markov chain Monte Carlo with people

Griffiths

Further reading

Tenenbaum, Griffiths, Kemp (2006). *Theory-based Bayesian models of inductive learning and reasoning*, TICS