Lecture 9: Computational Cognitive Modeling

Mechanistic vs. Rational models, Bounded Rationality

email address for instructors:
instructors-ccm-spring2019@nyuccl.org

course website:
https://brendenlake.github.io/CCM-site/

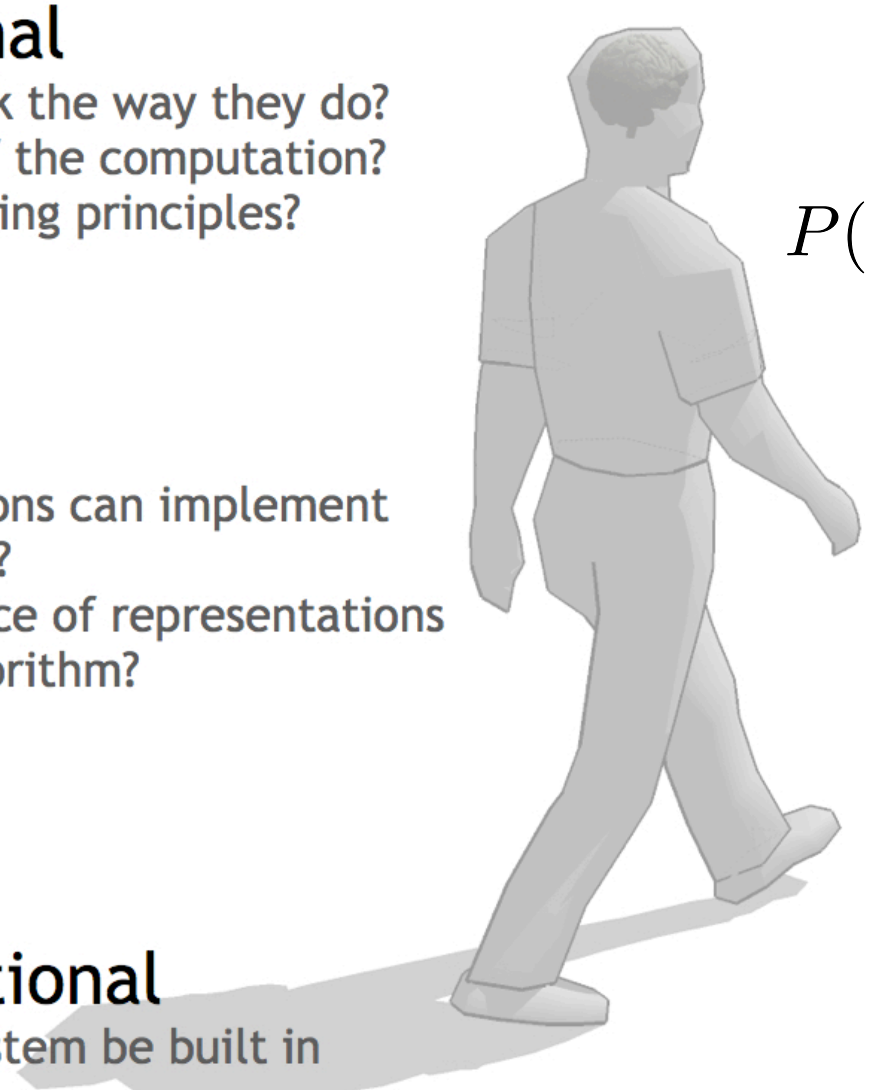
Today: A variety of topics!

- **Mechanism versus Rational modeling** - hopefully a bit of a debate!
- **Boundedly Rational Models -** sampling as a theory of cognition

Bayesian modeling is an approach for understanding inductive problems, and it typically takes a strong "top-down" strategy

Three levels of description *(David Marr, 1982)*

Computational
Why do things work the way they do?
What is the goal of the computation?
What are the unifying principles?

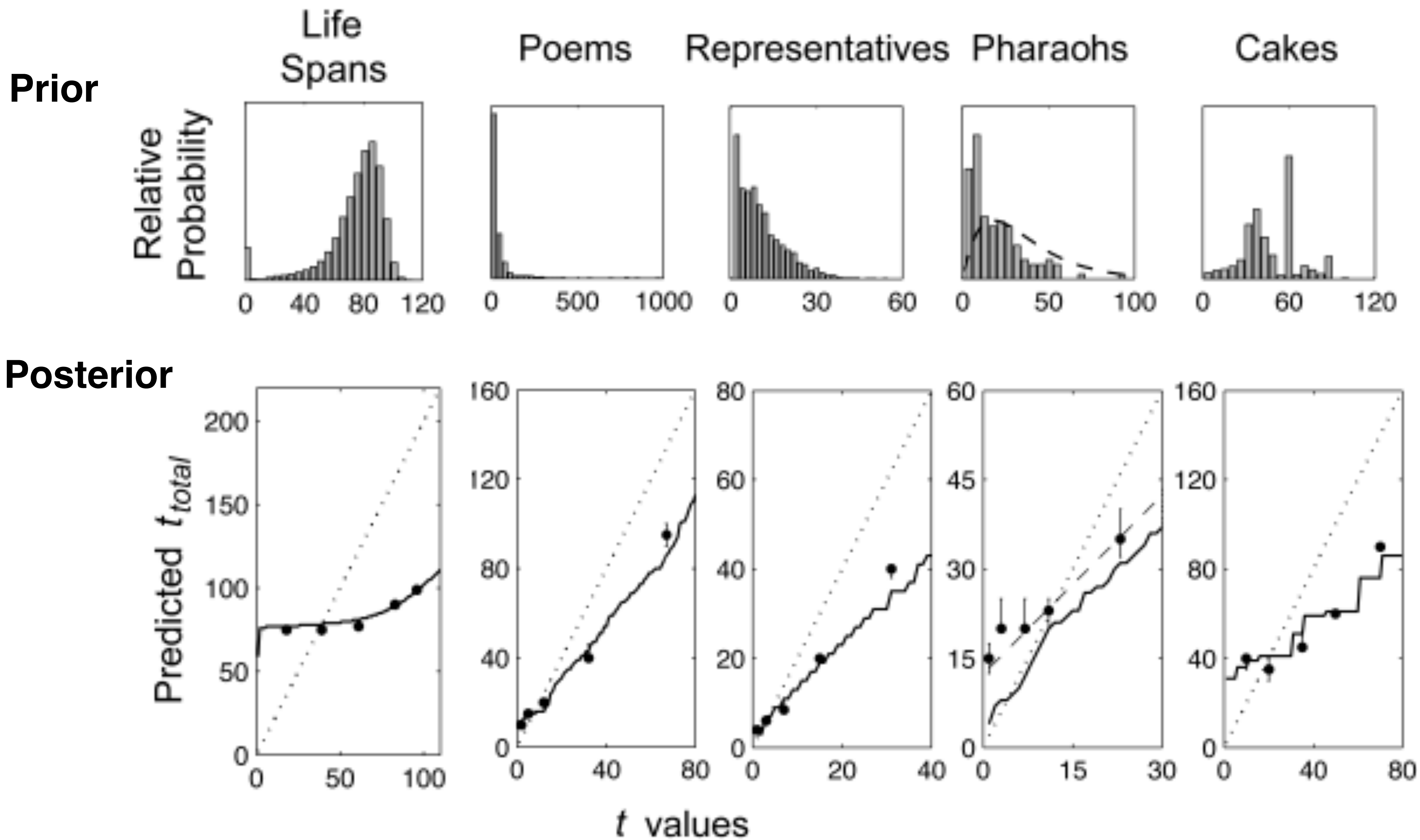


$$P(h_i|D) = \frac{P(D|h_i)P(h_i)}{\sum_j P(D|h_j)P(h_j)}$$

Algorthmic
What representations can implement such computations?
How does the choice of representations determine the algorithm?

Implementational
How can such a system be built in hardware?
How can neurons carry out the computations?

What does this mean about cognition?



Black dots are median prediction of participants
Solid lines are optimal Bayesian predictions

jones & love, 2007

Bayesian Models as a Theoretical Framework

- Brought about by mathematical/computational advances in probability and estimation theory

- Generate excitement for several reasons:

1. Offer a new interpretation of the goal of the cognitive system (behavior is seen as “rational”)

2. Assumption of Bayesian models are relatively transparent

3. Can explain very interesting and complex aspects of cognition (e.g., language acquisition or reasoning under uncertainty)

Bayesian Fundamentalism

“Fundamentalist” tenent: human behavior can be explained through rational analysis, given the correct probabilistic interpretation of the task environment. No need to look at mechanism.

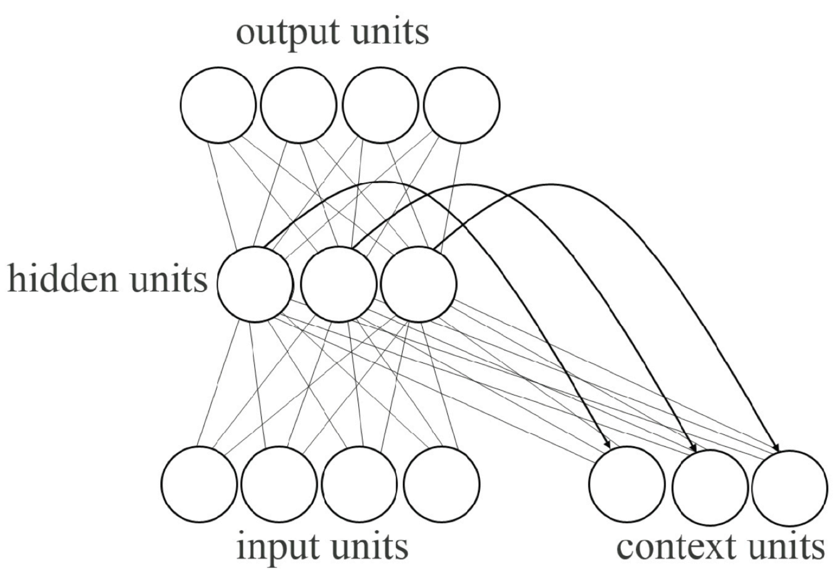
Bayesian enlightenment

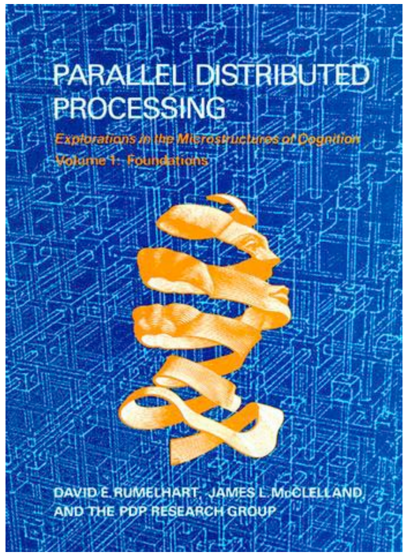
Elements of a Bayesian model treated as claims about a psychological process and representation, not mathematical conveniences.

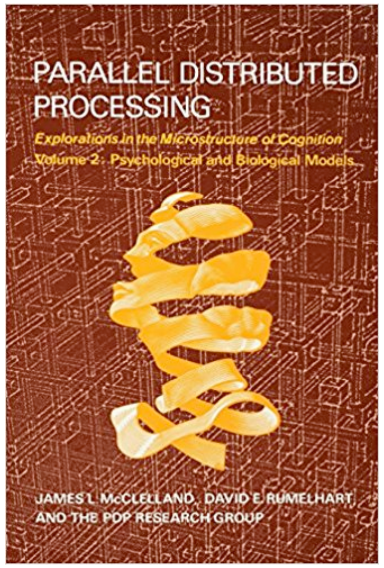
Agnostic Bayes

Bayesian inferential methods used to decide among scientific models based on empirical data.

Danger of mistaking metaphors for theories: Connectionism







- Initial excitement (once backpropogation was developed) Backpropogation allowed the training of multi-layer models which could
    - (in theory) encode any function.
    - Reaction to the computer metaphor
    - Avoided the brittleness of propositional models
    - Many studies were done applying neural nets to reproduce a wide gamut of Psychological phenomena.
- Bubble bursts
    - Incapable of capturing compositionality characteristic of language processing
    - Too opaque to offer insight
    - Biologically implausible
- Field was not able to cope with new criticisms

Summary of Bayesian inference

jones & love, 2007

- Specify the hypothesis space. A hypothesis is simply a likelihood function over patterns of observations.
- Specify the initial (prior) belief over various hypotheses
- Posterior is a sum (at least in log space)
- Hypotheses may not relate to "psychological processes”
- Most of the math is about approximating these solutions for complex distributions but not specifically about psychological processes

“Bayes’ rule”

$$\underbrace{P(h_i|D)}_{\text{posterior}} = \frac{\overbrace{P(D|h_i)}^{\text{likelihood}}\overbrace{P(h_i)}^{\text{prior}}}{\sum_j P(D|h_j)P(h_j)}$$

Hypotheses:

$h_1 = \text{John has a cold}$

$h_2 = \text{John has emphysema}$

$h_3 = \text{John has a stomach flu}$

An opposing view What do we mean by “mechanistic”? An example

- Compare Bayesian model of the “optimal predictions” to an algorithm like Q-learning
- Q-learning specifies when and what type of update should happen on each trial
- Each trial is also divided into a phase where there is learning (q-value updates) and decision making (softmax or epsilon greedy)
- Model is lower level and makes commitments to the order of operations, specific parameters that explain or capture regularity in behavior

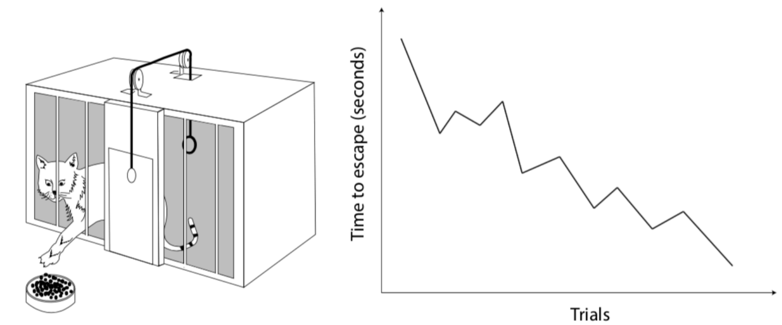


Figure 1: *Left*: An illustration of Thorndike's puzzle box experiments. *Right*: The time recorded to escape the box is reduced over repeated trials as the cat becomes more efficient at selecting the actions which lead to escape.

$$\Delta Q_t(s_t, a_t) = \alpha[r_{t+1} + \gamma Q_t(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t)]$$

Things that simple computational-level analyses (often) can’t tell you about:

- How will we move our eyes? Eyes might fixate options prior to choosing which give addition measures about value
- What is the sources of our errors? Softmax? Learning rate? Initial values?
- How long will it take someone to decide? More difficult choices might have q-values that are more similar

Consequences of the denial of mechanism

- Problematic to toss out everything below computational level
  - Mechanistic explanations often make more useful predictions than ad hoc explanations.
  - Optimal decision making is not always possible: how does the brain deal with this problem? (stay tuned!)
  - E.g., mechanistic (connectionist) accounts of semantic memory have inspired useful models of brain damage and even treatments.
  - Manipulations unrelated to rational form, such as making information part of a cue or an outcome within a learning trial can yield surprising results. Or manipulating the order of items in learning strongly influences what is learned by Bayes is “exchangable”
- Also, non-Bayesian applications of cognitive models tend to rely on mechanistic descriptions: e.g., neurology or psychopharmacology

Bayes and Evolutionary Theory

jones & love, 2007

- Animal behavior is adapted by natural selection to increase fitness
  - Thus an animal's behavior should reflect whatever was adaptive in its ancestral habitat.
  - Focus on environment can sometimes come at the cost of mechanism.
  - Not any function is optimized; only those relevant to fitness (more constrained than Bayesian fundamentalism).
  - Proliferation of modules; e.g., snake avoidance module.
- Like the modules in evolutionary Biology, Bayesian analyses are often ad hoc solutions for a given behavior.
  - E.g., children's ability to give the number of items requested.
  - This comes at the expense of identifying general mechanisms and architectural characteristics (e.g., working memory) that are applicable across tasks.
- Bayesian models are in a sense less constrained than evolutionary theories, which are restricted to functions that are relevant to fitness.

Remember the dutch book from last week!

Agent 1		Agent 2		Outcome for Agent 1			
Proposition	Belief	Bet	Stakes	$a \wedge b$	$a \wedge \neg b$	$\neg a \wedge b$	$\neg a \wedge \neg b$
a	0.4	a	4 to 6	–6	–6	4	4
b	0.3	b	3 to 7	–7	3	–7	3
$a \vee b$	0.8	$\neg(a \vee b)$	2 to 8	2	2	2	–8
				–11	–1	–1	–1

Figure 13.2 Because Agent 1 has inconsistent beliefs, Agent 2 is able to devise a set of bets that guarantees a loss for Agent 1, no matter what the outcome of a and b.

Bayes and Evolutionary Theory

- “Just so” stories; Bayesian fundamentalism labels a behavior as “rational” simply because a probabilistic model can fit it.
    - Both kinds of explanations can predict any outcome post hoc by changing the formulation of the problem.
- Both approaches deemphasize mechanisms to their detriment. constraints on their assumptions.
    - *“Completely sidestepping mechanistic considerations when considering optimality leads to absurd conclusions. To illustrate, it may not be optimal or evolutionarily advantageous to ever age, become infertile, and die, but these outcomes are universal and follow from biological constraints.”*
    - *“Evolution is survival of the best current design, not survival of the globally optimal design.”* (Marcus, 2008)

An example - Cognitive Development

jones & love, 2007

- A key question of development: what develops?
  - **In rational models, nothing develops, only more information is accumulated.**
  - Some aspects of development clearly reflect maturational changes in mechanism.
  - Some measures of child development are indexed to prefrontal development development are indexed to prefrontal development.
- A fundamentalist Bayesian approach:
  - Posit rational theories that are collections of discrepant causal models (hypothesis spaces)
  - Each model is intended to correspond to a different stage This does not explain how or why the transitions occur
- A satisfying approach needs to consider how selection mechanisms change throughout development, or how the mechanisms of cognition shift.

Bayesian Enlightenment!

jones & love, 2007

- In a Psychological model, the prior and hypothesis space should be beliefs of the subject.
  - Information in brain != ground truth.
  - Accuracy of information encoded can be separated from whether the participant acts on that information rationally.
  - Constructing the generative model can be the most substantial part of learning.
- Priors could be another area of inquiry; conjugate priors might be easier for the brain to work with.
- **Approximation algorithms could also be taken seriously as mechanistic models**.
- An appeal of Bayesian models is the transparent view of their operation (as opposed to, say, connectionism); but design assumptions made available are rarely viewed critically in practice.

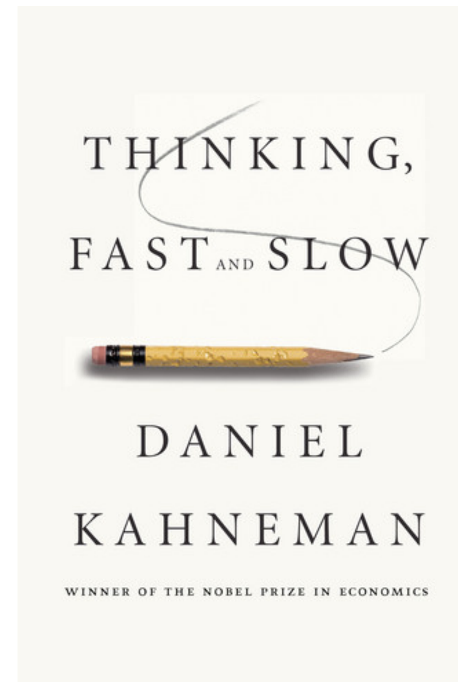
Bayesian inference is hard

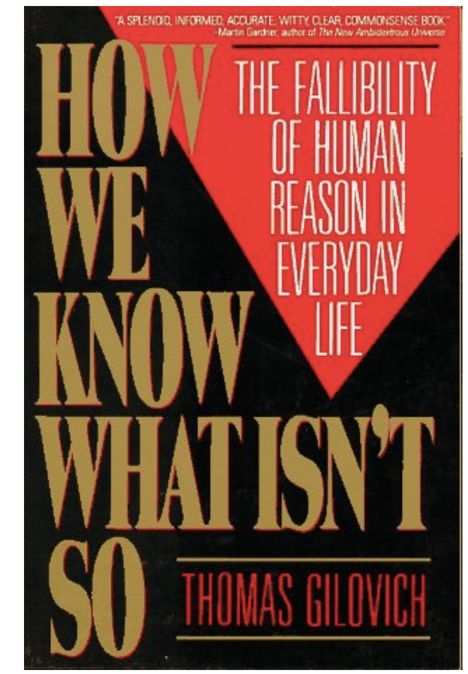
The goal: A *predictive* account of behavior that can *emulate* both the general successes and the occasional failures of human performance, and predict which will happen when.

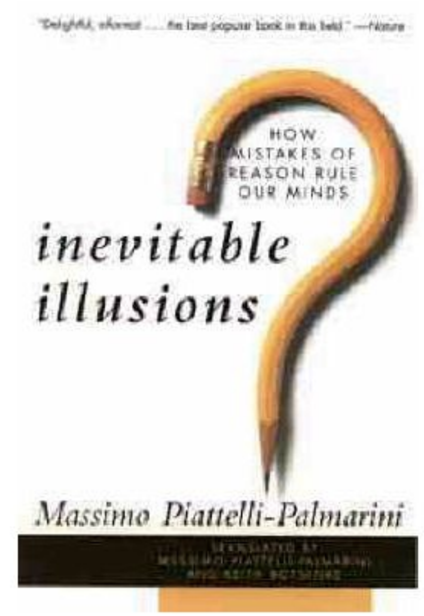
Why *predict?* Why *emulate?* Why emulate *failures?*

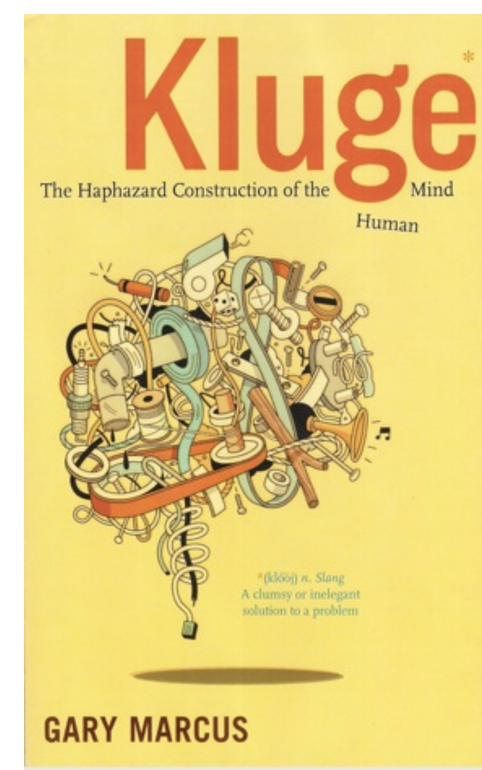
- In realistic problems, the number of possible hypotheses can be huge
  - e.g., more than 100,000 clusterings of 10 objects
- In the worst case, the time required to perform exact Bayesian inference increases linearly in the number of hypotheses

Rationality vs. Heuristics



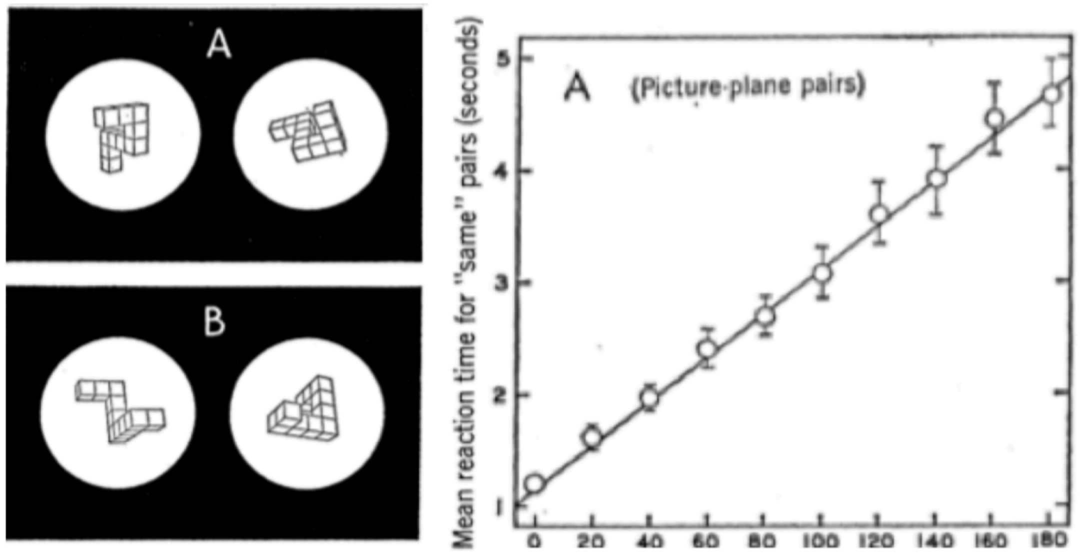






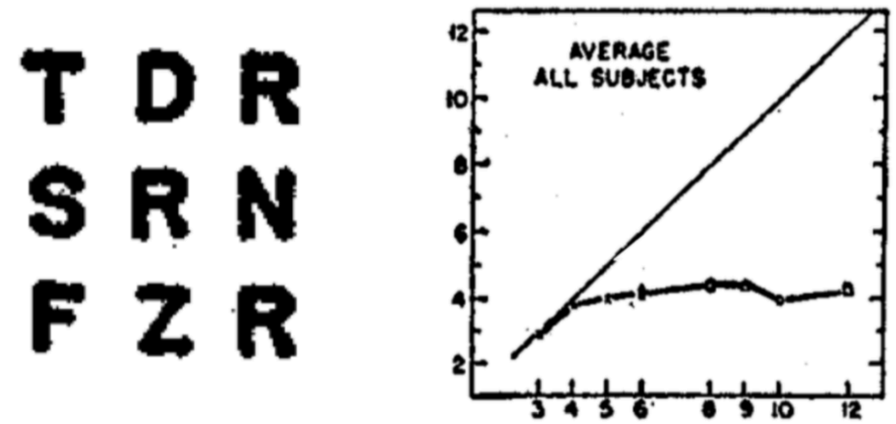
Resource constraints on cognition

Thinking takes time



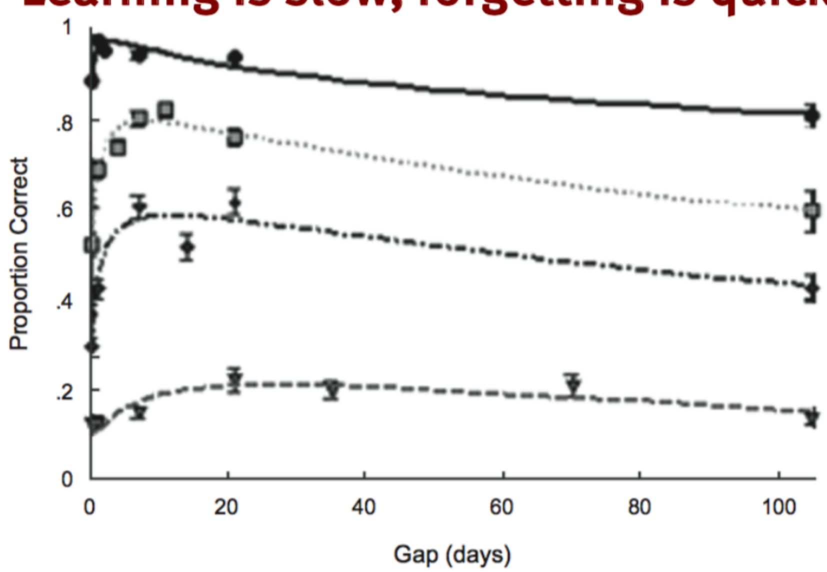
Shepard & Metzler (1971)

Memory stores are limited



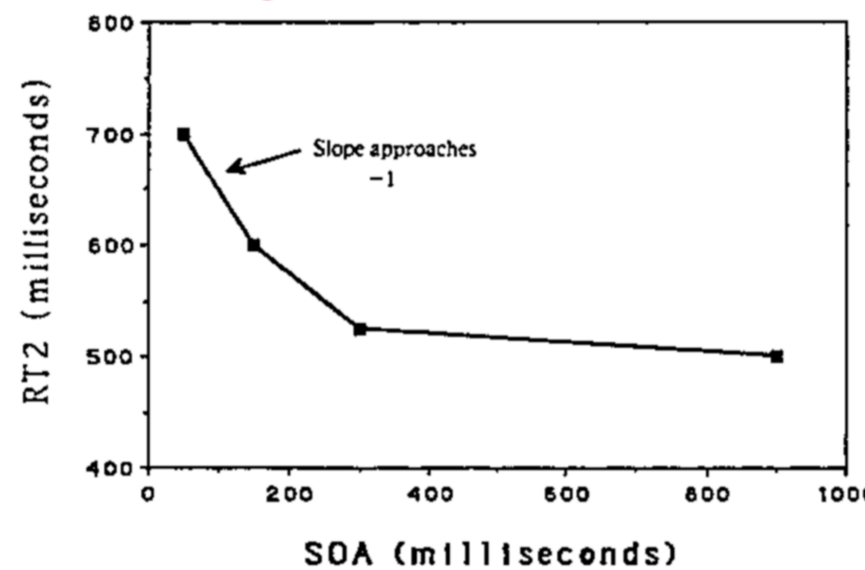
Sperling (1960)

Learning is slow, forgetting is quick



Cepeda, Vul, Rohrer, Wixted, Pahler, 2008

Thinking has a central bottleneck



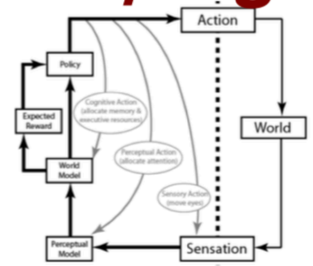
Pashler (1994)

Computation-algorithm intersection

Computation | **Process / Algorithm** | **Implementation**

$$P(h \mid D) = \frac{P(D \mid h)P(h)}{\int_{h'} P(D \mid h')P(h')}$$

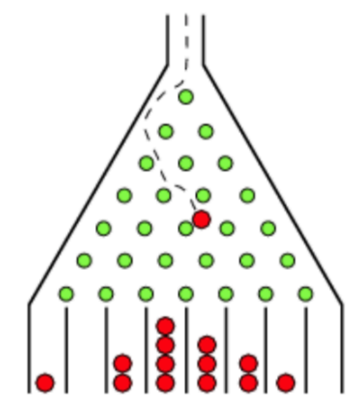
$$\arg\max_{A} = \int_{h} U(A;h)P(h \mid D)$$



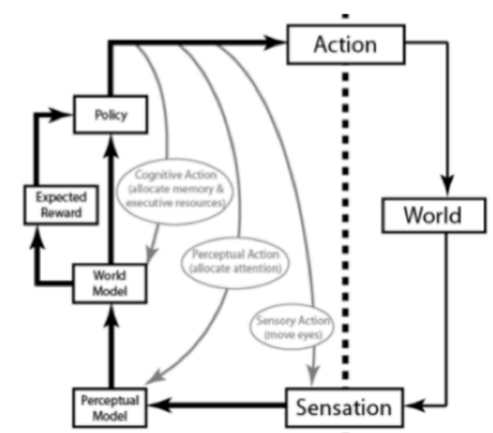
Marr, 1982

Challenges at the intersection of probabilistic reasoning and cognitive resource constraints:

- How does the mind represent uncertainty?



- How do we use limited cognitive resources?

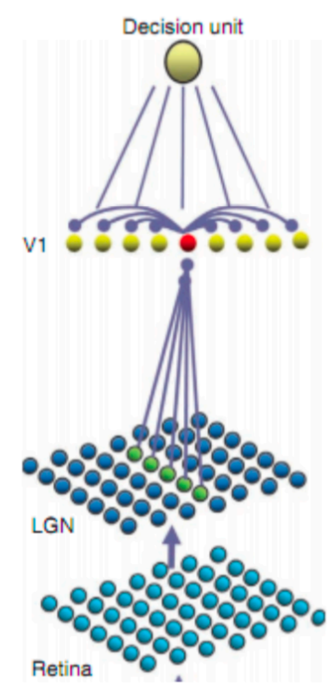


Computation-algorithm intersection

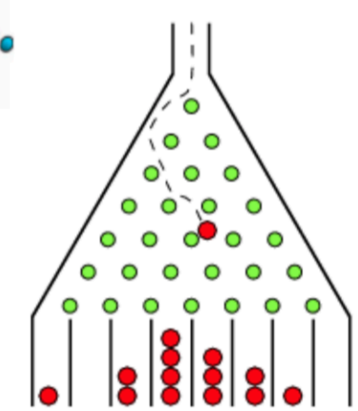
- **Analytical**
  i.e., equations, including variational methods (Friston)

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$$

- **Tabular / grids**
  e.g., probabilistic population codes (Pouget/Ma/Beck)

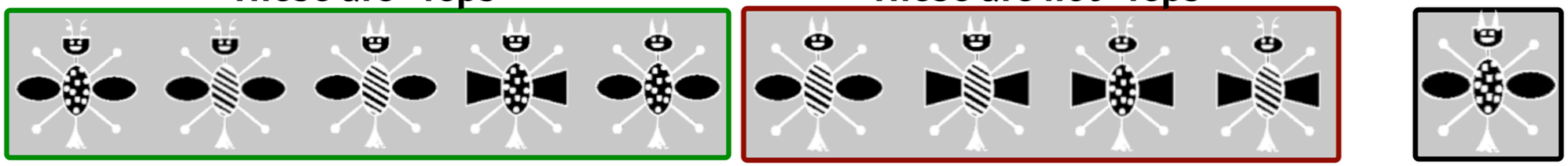


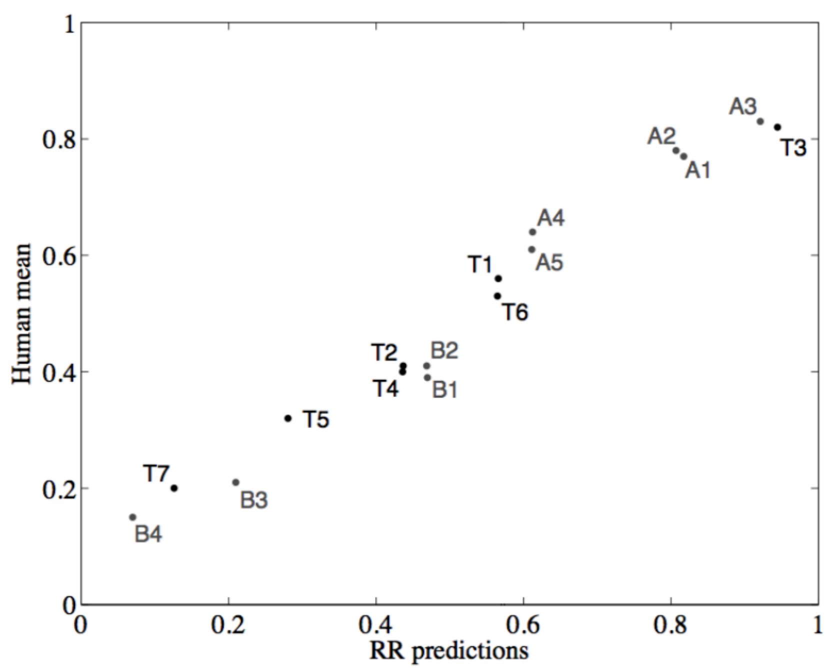
- **Sampling**
  (Goodman/Griffiths/Sanborn/Tenenbaum/Vul, etc.)

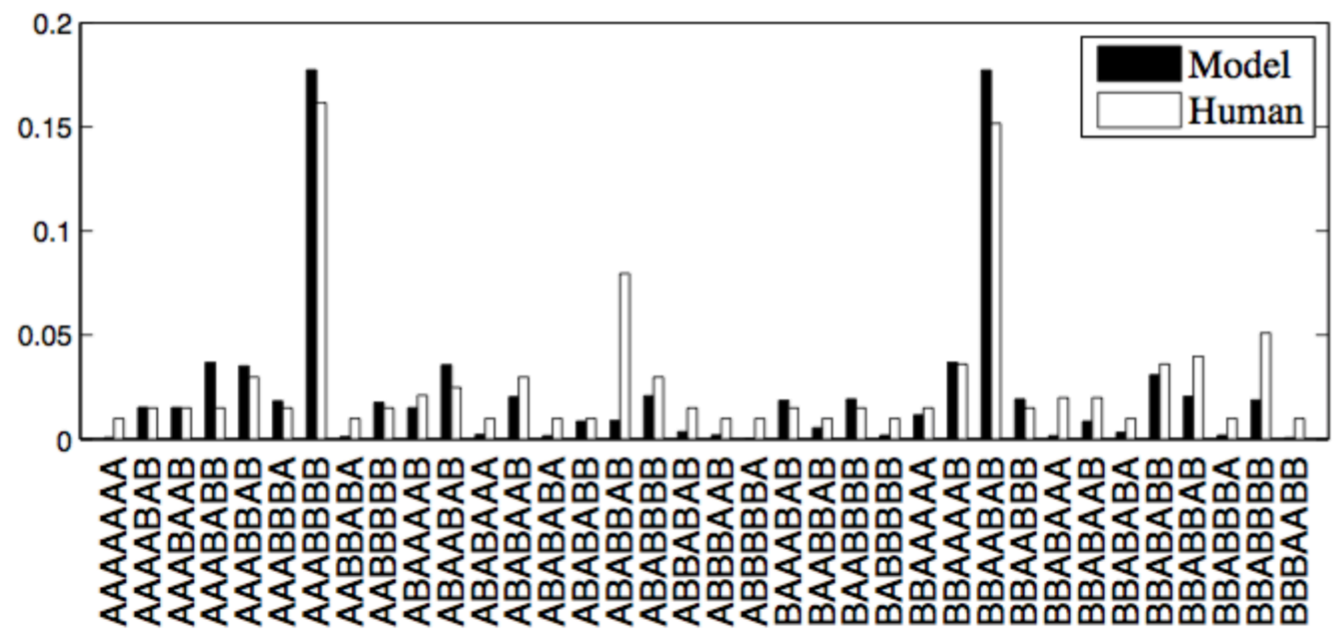


Sampling categorization rules

These are "feps" These are *not* "feps"





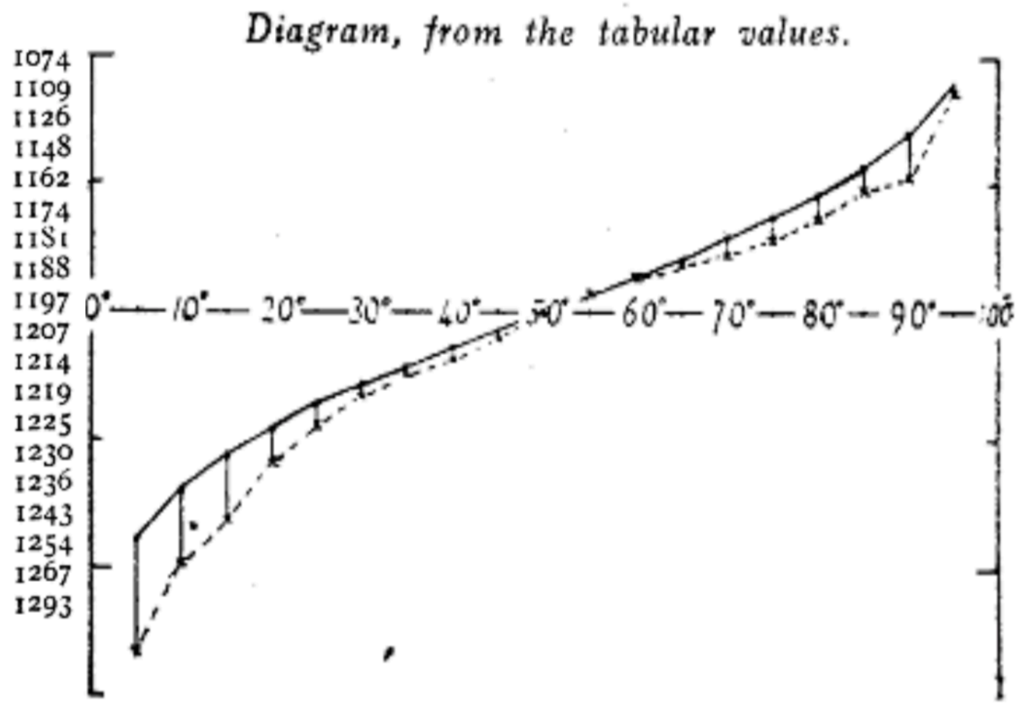


Generalization patterns suggest that each individual was not using a probability distribution over rules, but just one sampled rule.

Goodman, Tenenbaum, Feldman, & Griffiths, 2008

Wisdom of Crowds

Galton, 1907: Vox Populi
How much does an ox weigh?



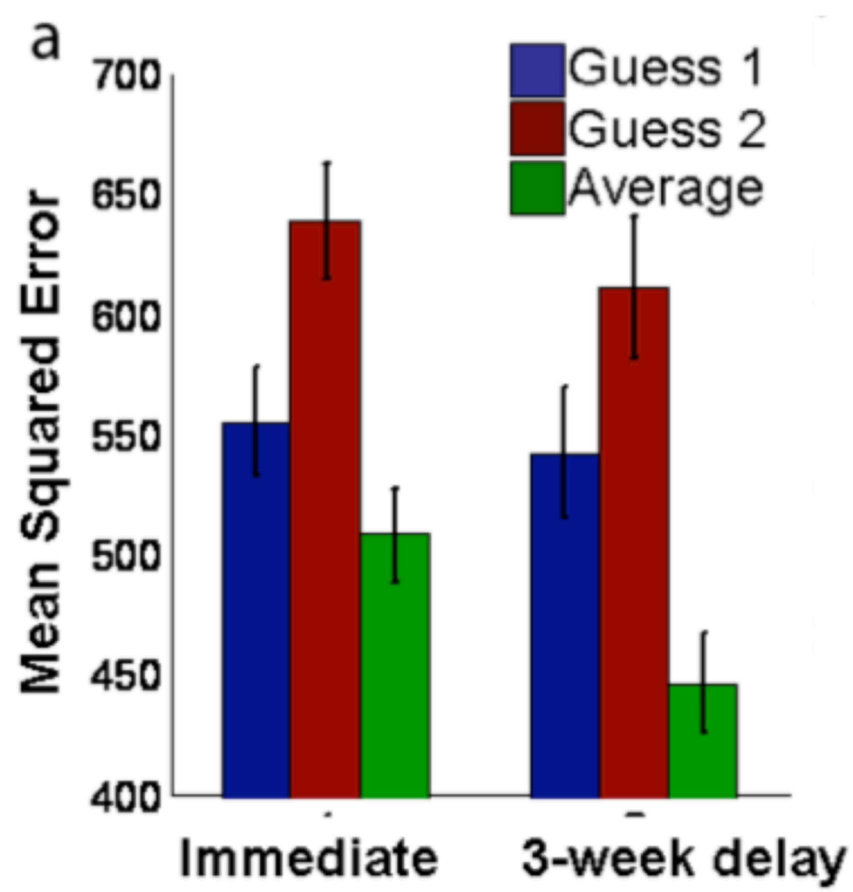
Mean was (1206) was closer to the correct answer (1207) than any one guess.

Benefit of averaging multiple guesses holds so long as errors are independent samples.

Do we get the same effect *within* individuals?

Wisdom of Crowd Within

- What percent of the world's airports are in the United States?
- Saudi Arabia consumes what percentage of the oil it produces?
- What percentage of the world's countries have a higher life expectancy than the United States?



Benefit of averaging multiple guesses from a single individual:
Estimation errors do not arise only from individual biases,
but reflect *sampling* under uncertainty.

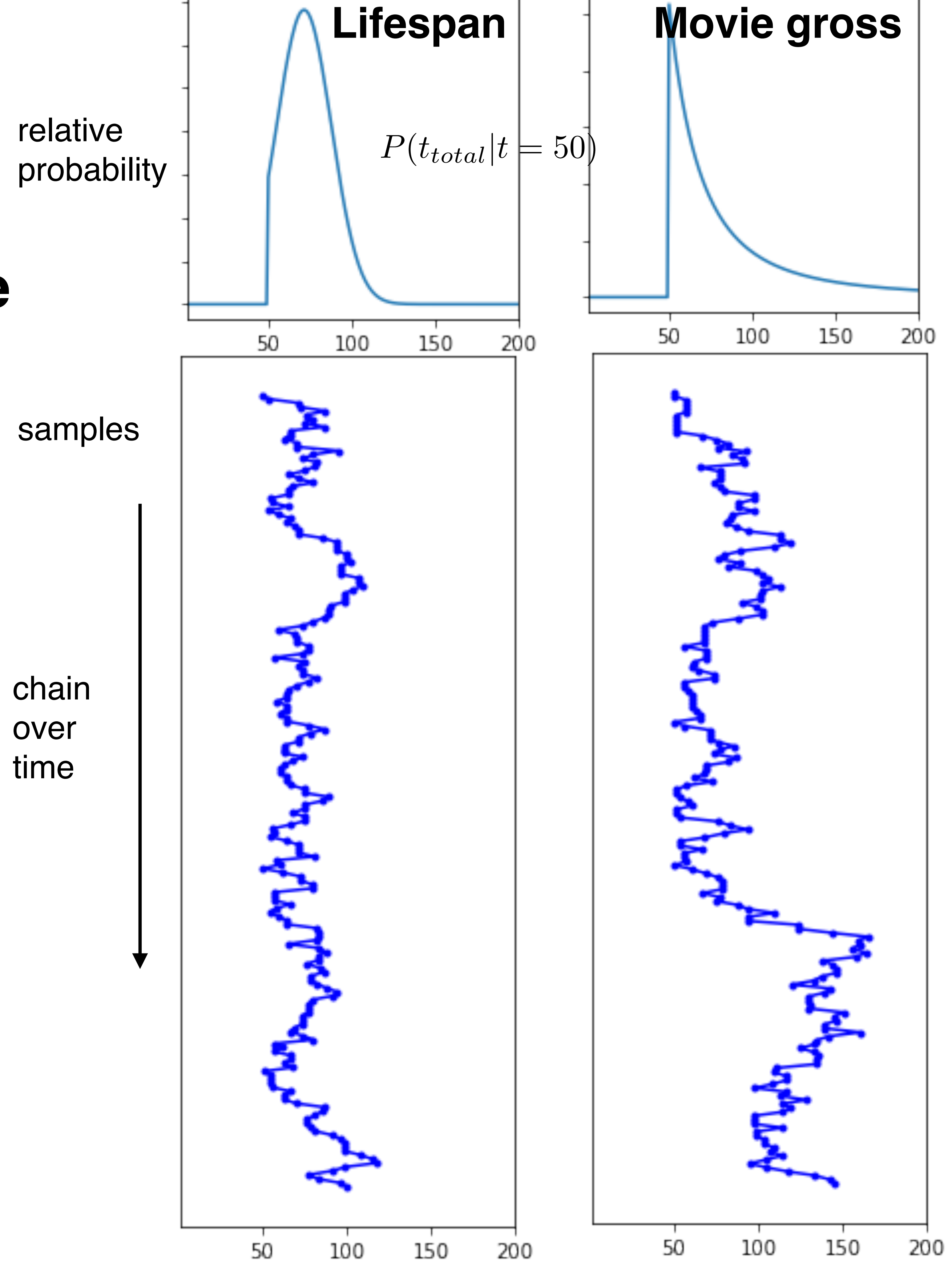
Vul & Pashler (2008)

The Monte Carlo principle

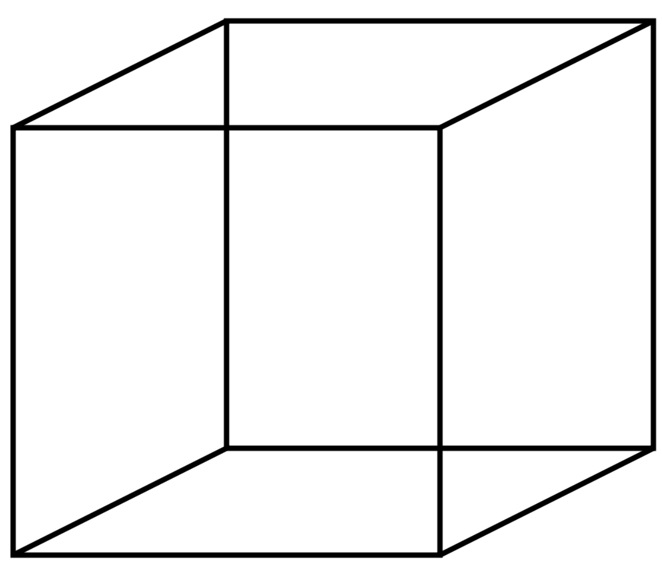
The expectation of f with respect to P can be approximated by

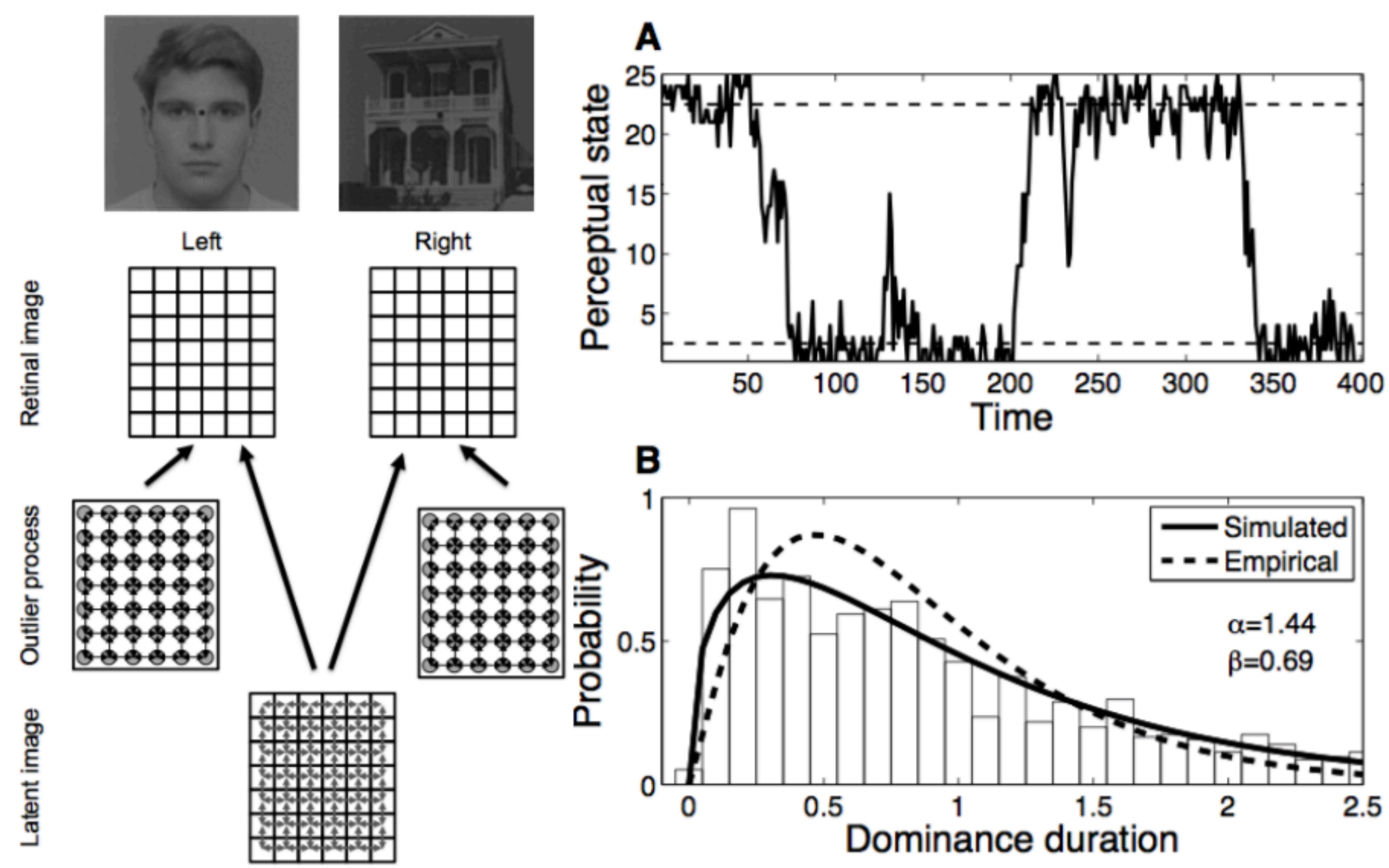
$$E_{P(x)}\left[f(x)\right] \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i)$$

where the x_i are sampled from $P(x)$



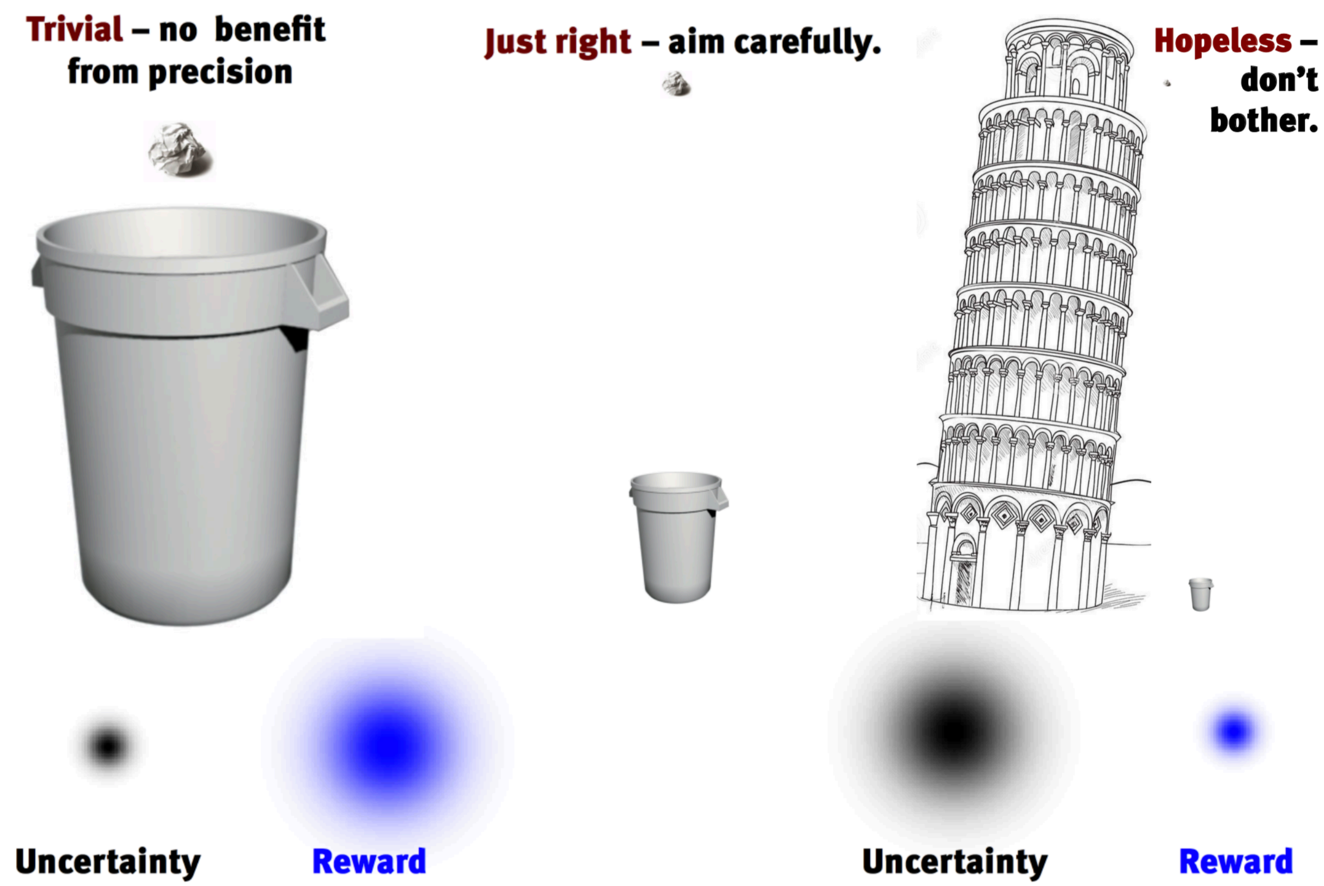
The Monte Carlo principle





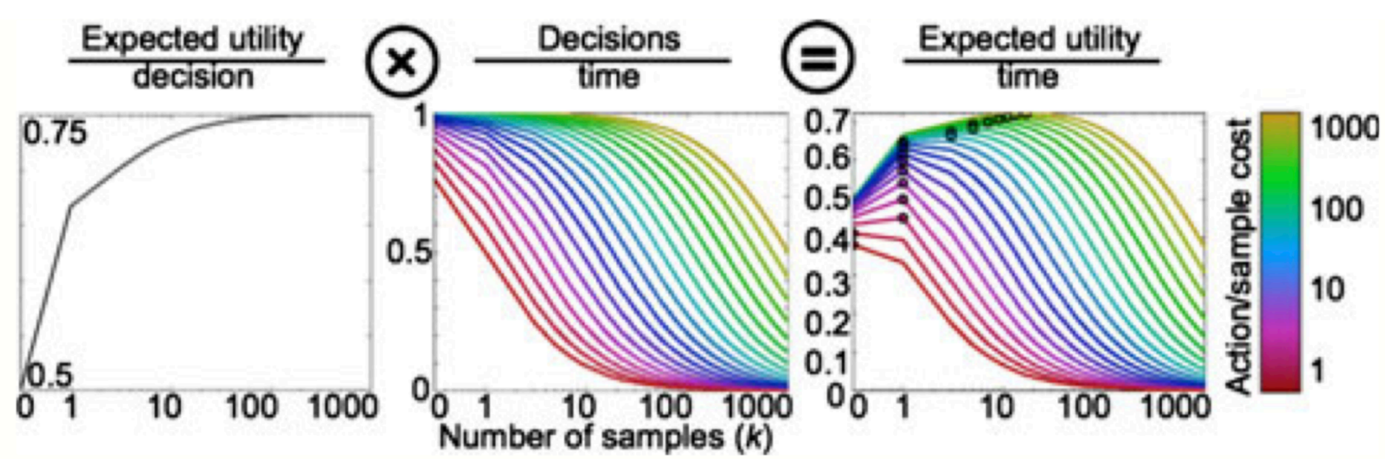
Gershman, Vul, & Tenenbaum; 2011

How many samples?



$$A^* = \arg\max_A \sum_S U(A;S)P(S|D).$$

What does adaptive behavior look like if there is a cost (at least of time) to drawing new samples (computation)?



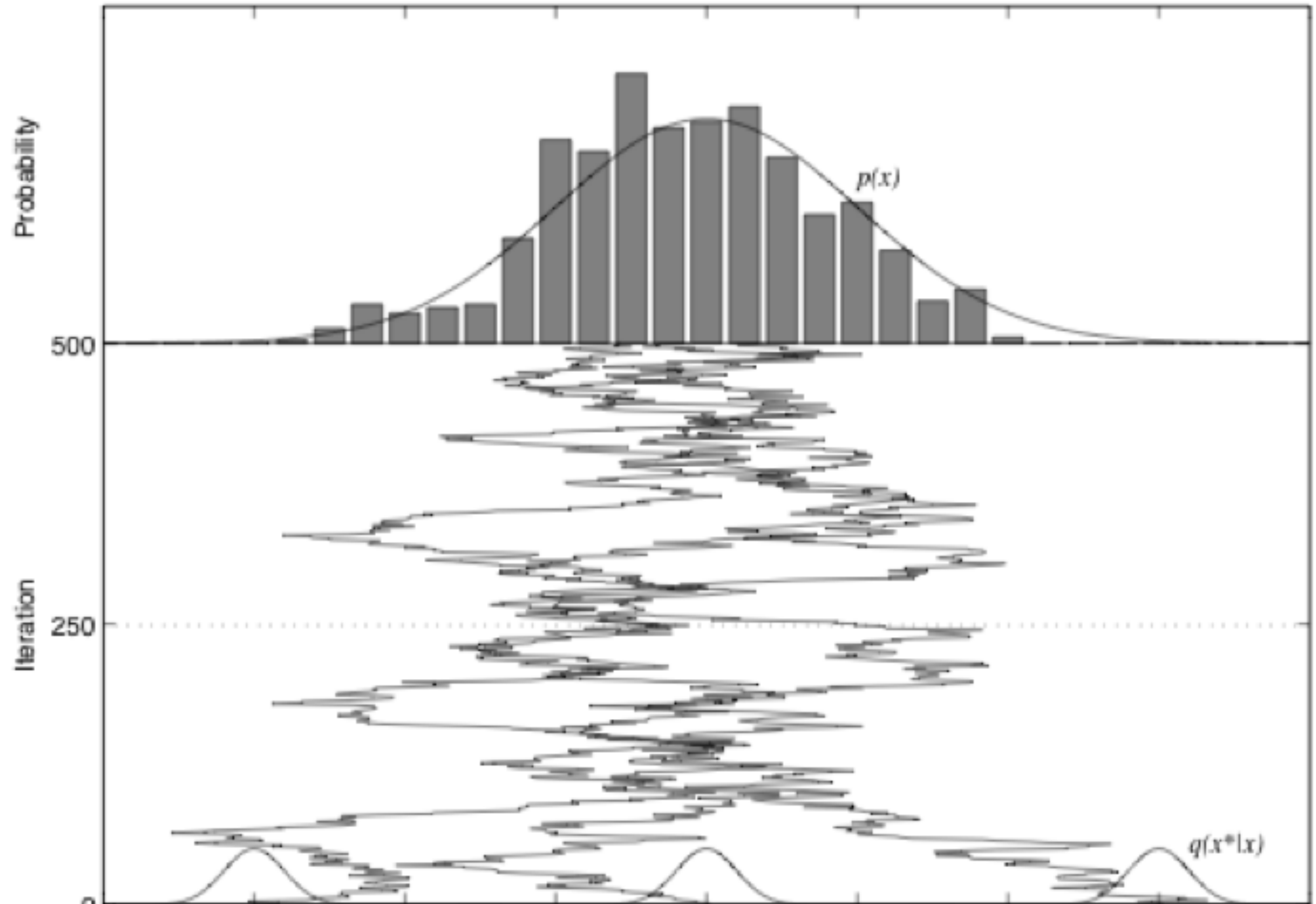
cost of computation
= cost of thinking

Vul, Goodman, Griffiths, & Tenenbaum (2014)

Anchoring and adjustment

- Is the population of Chicago greater or less than 200,000 people?
- Now guess the population of Chicago
- People give lower estimates when given a lower anchor (200,000) than a higher anchor (5 million) (Jacowitz & Kahneman, 2005)

Metropolis-Hastings



Predicting and Detecting Change

S.D. Brown, M. Steyvers / Cognitive Psychology 58 (2009) 49–67

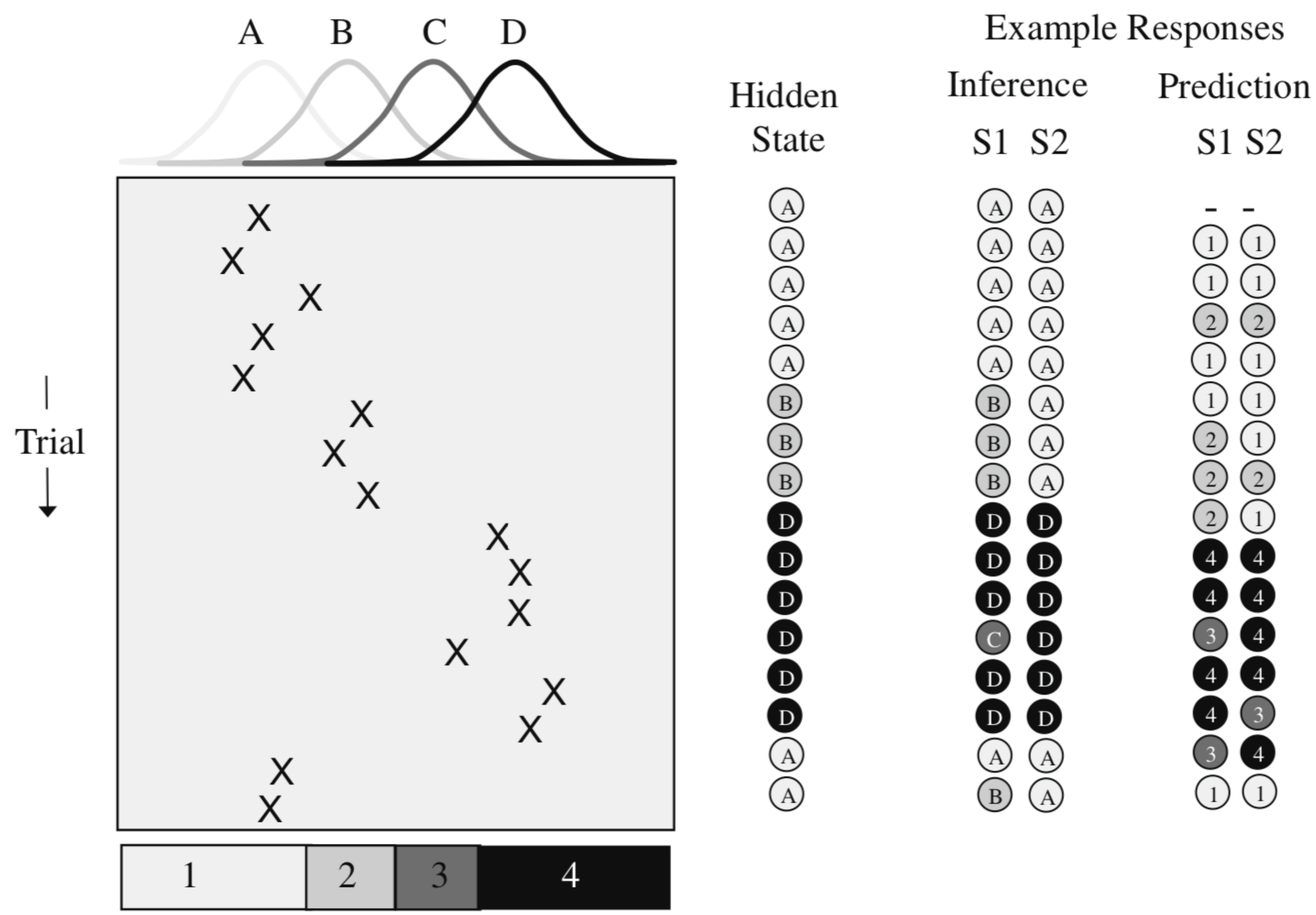


Fig. 1. An illustration of the change detection paradigm. Each stimulus is sampled from one of four normal distributions, but the selection of the hidden state (generating distribution) is not observed by the participant. At each timestep, there is a small probability that the hidden state is switched to a new distribution. Example data for two participants are shown on the right for two different tasks; an inference task where the goal is to identify the hidden state for the last stimulus, and a prediction task where the goal is to predict the most likely region where the next stimulus will appear. See text for additional details.

Predicting and Detecting Change

S.D. Brown, M. Steyvers / Cognitive Psychology 58 (2009) 49–67

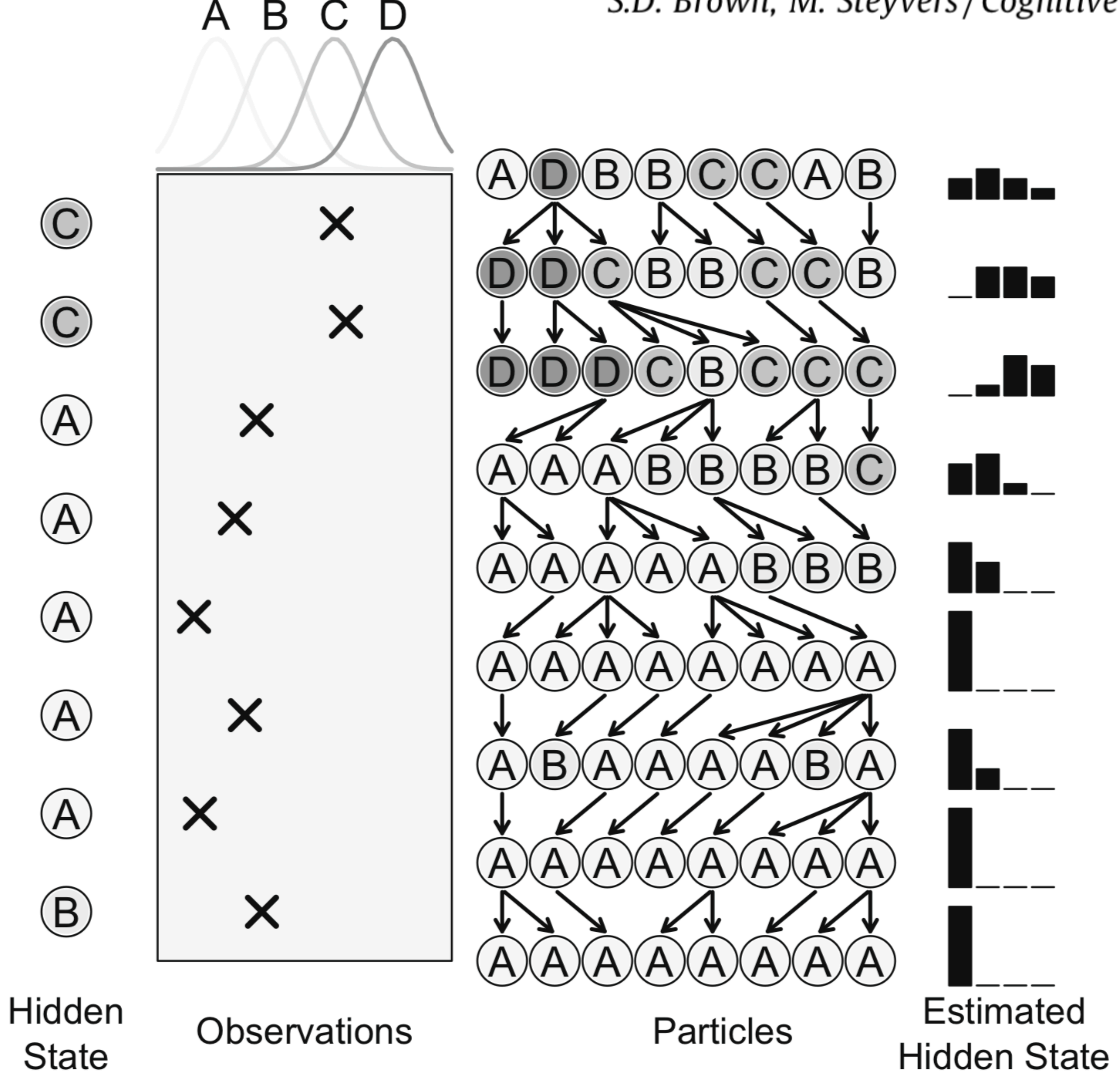


Fig. 2. Illustrative example of the particle filter algorithm.

Predicting and Detecting Change

S.D. Brown, M. Steyvers / Cognitive Psychology 58 (2009) 49–67

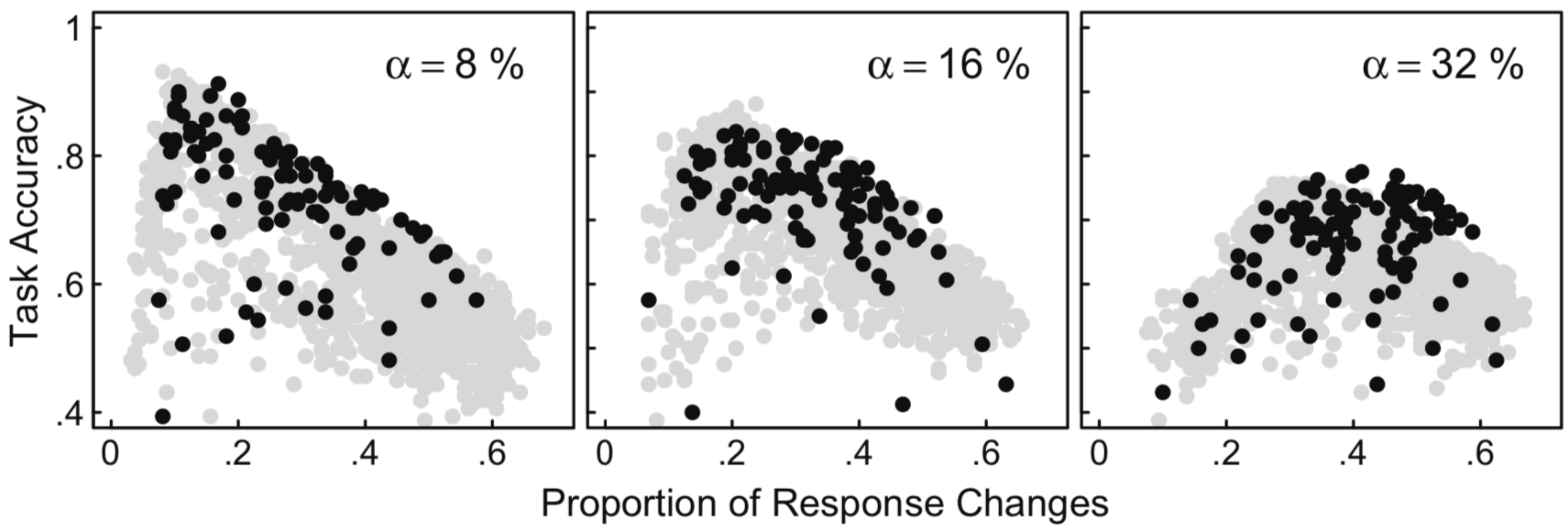


Fig. 4. Accuracy vs. response variability (the proportion of trials on which the participant changed their response). From left to right, the panels show data from the three conditions of Experiment 1: low, medium and high probabilities of a change (α) in the hidden state. Black dots show data, the gray dots shows predictions from the particle filter model, under all possible parameter values.

- Model explains variability in performance by committing to process assumptions

Sampling as mechanism

- Each sampling solution yields different samples:
  Variability across decisions;
  Variability across trials;
  Variability across participants

- Solutions outside of the "convergent" regime:
  Systematic deviations from optimal decisions that will reflect the biases of the sampling algorithm.

Biases

- Probability matching
  Vul et al. 2009
- Anchoring and adjustment
  Lieder et al. 2013
- Sequential effects
  Sanborn et al. 2010
- Garden-pathing
  Levy et al. 2009
- Dynamics of belief change
  Gershman et al. 2012
- Memory reproduction
  Shi et al. 2010

We now face tradeoffs between the effort of computation, variability, and bias of answers. These tradeoffs yield tricky meta-cognitive decisions.

Three levels of description *(David Marr, 1982)*

Computational

Why do things work the way they do?
What is the goal of the computation?
What are the unifying principles?

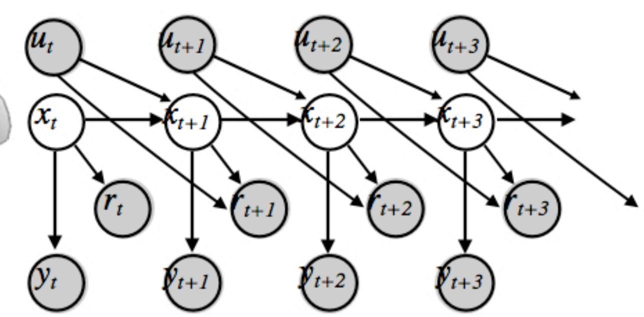
maximize:

$$R_t = r_{t+1} + r_{t+2} + \cdots + r_T$$

Bellman

Algorthmic

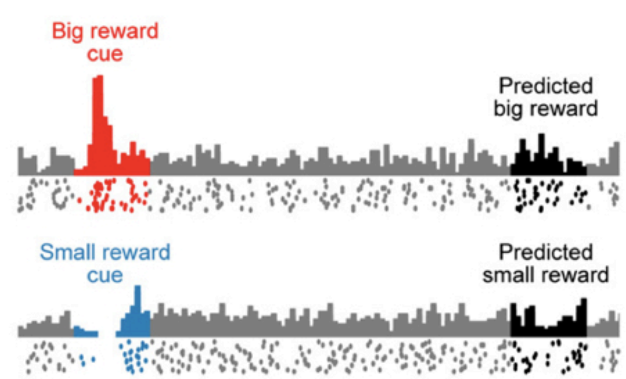
What representations can implement such computations?
How does the choice of representations determine the algorithm?



Dynamic programming, TD methods, Monte Carlo

Implementational

How can such a system be built in hardware?
How can neurons carry out the computations?



Neural firing patterns, prediction errors, system level neuroscience

Slide Credits

Tom Griffiths (bounded rationality)
Ed Vul (bounded rationality)
John McDonnell (Love/Jones paper)