Data mining and modeling for behavioral sciences and beyond

Valentin Wyart professor
Lucas Benjamin teaching assistant

Lab. de Neurosciences Cognitives et Computationnelles (LNC²)
Institut National de la Santé et de la Recherche Médicale
Ecole Normale Supérieure, Université PSL

valentin.wyart@ens.psl.eu

ENS | PSL

PSL UNIVERSITÉ PARIS

Data Science Program (certifying minor at PSL)

https://psl.eu/en/programmes-gradues/programme-data

PR[AI]RIE

PaRis Artificial Intelligence Research InstitutE

Paris Artificial Intelligence Research Institute

https://prairie-institute.fr

Data mining and modeling

- We are currently facing an explosion of data across domains and disciplines.
- The ability to manipulate and understand large amounts of complex, multidimensional data has become critical in science (and for many applications outside academia).
- Can you give a few examples?

Data mining and modeling

- Data mining: manipulate data
  Afternoon classes = practical sessions
- Data modeling: understand data
  Morning classes = lectures
- Do you have eduroam working?
  a working Python environment?
  Come to Lucas during the break if not.

Data mining and modeling

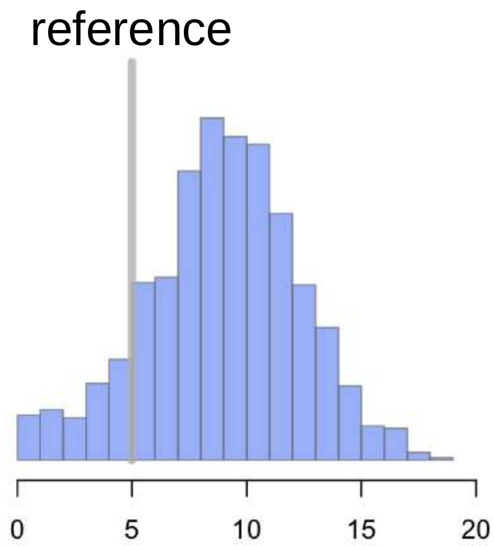
- Today's practical session: data statistics
  2.00pm, same room

- Au programme:
  - ✓ plot data
  - ✓ compute data metrics
    what metrics do you know?
    can the mean be a misleading metric? when?
    what is a probability density function?
  - ✓ identify correlations in data
    does correlation mean causation? why (not)?

Data mining and modeling

- Compare mean of data variable to reference
  Name of this test: 1-sample *t*-test

- Compare means of two data variables
  Name of this test: 2-sample *t*-test



Data mining and modeling

- Compare mean of data variable to reference
  Name of this test: 1-sample *t*-test

- Compare means of two data variables
  Name of this test: 2-sample *t*-test

- Why is it already a model of the data?

- Difference between a statistical model and
  a computational model of behavioral data

Data mining and modeling

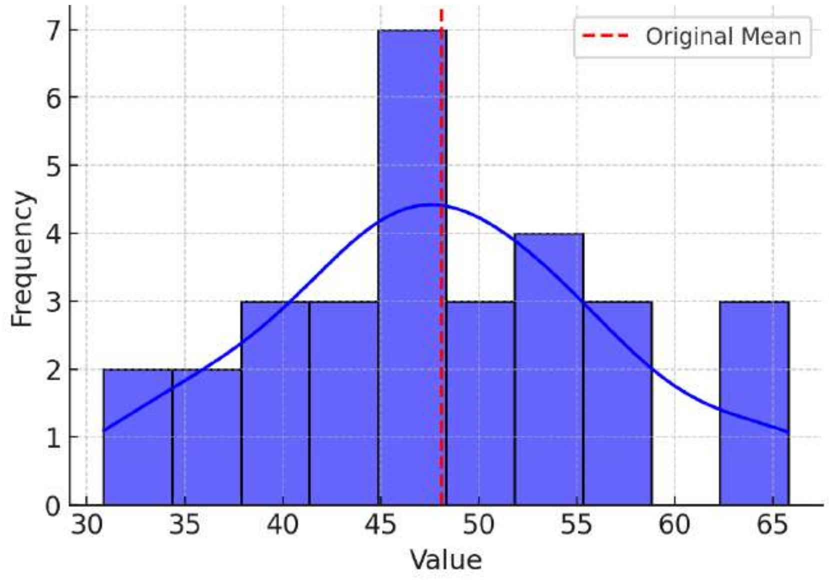
- Uncertainty about data: what does it mean?
  Name factors that influence data uncertainty.
- How do we report uncertainty about the mean?
- Difference between a point estimate and a confidence interval
- Difference between analytical and empirical measures of data uncertainty
  Example: bootstrapping of uncertainty metrics

Data mining and modeling

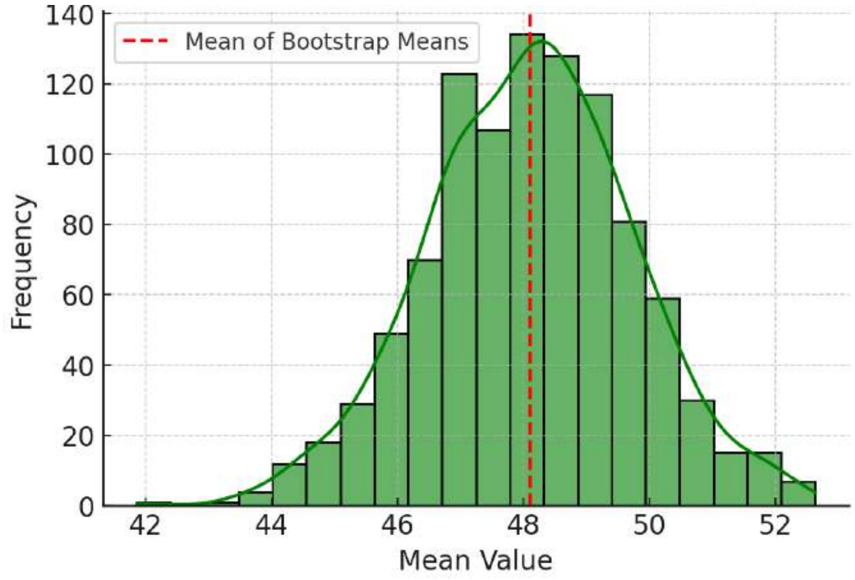
- Original sample: start with a dataset of size *n*
- Resampling: draw a bootstrap sample of size *n*, randomly selecting data points with replacement from the original dataset
- Compute statistic: calculate the statistic of interest for each bootstrap sample
- Repeat: repeat this process many times to generate a distribution of the statistic
- Estimate SE: compute the standard deviation of the distribution = bootstrap estimate of the SE of the statistic

Data mining and modeling

Original sample



Bootstrap distribution of the sample mean

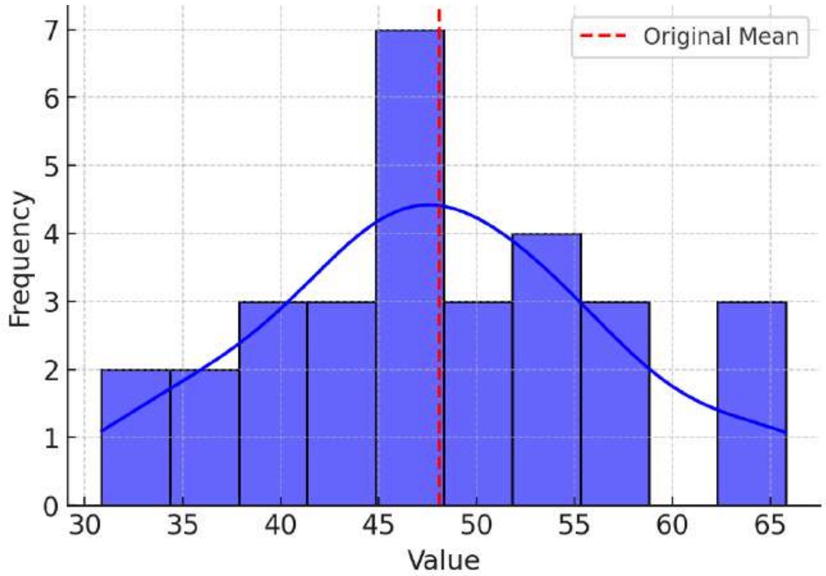


bootstrap SE = 1.63

k = 1,000 resamples

Data mining and modeling

Original sample



$$SE = \frac{s}{\sqrt{n}}$$

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

analytical SE = 1.64

bootstrap SE = 1.63

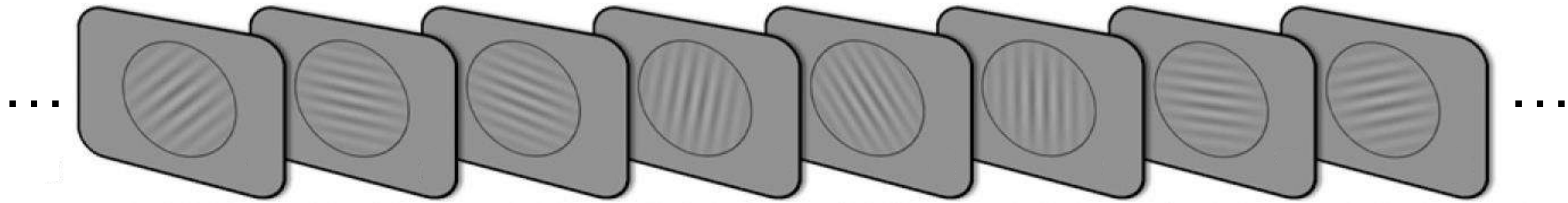
k = 1,000 resamples

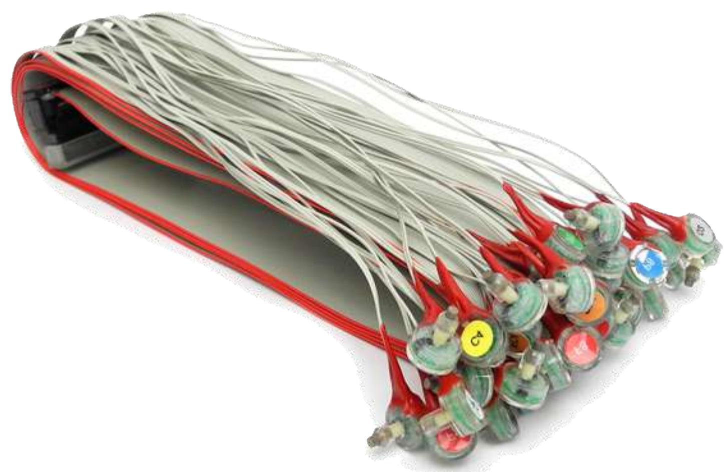
Data mining and modeling

- Data mining and modeling is not only about playing with data, but also thinking about data.
- Approaches in data mining and modeling can easily be misused, they can provide nonsensical answers, and you need to think about data to tell the difference.
- Introducing you to thinking about data is maybe the most important aim of this PSL-week.

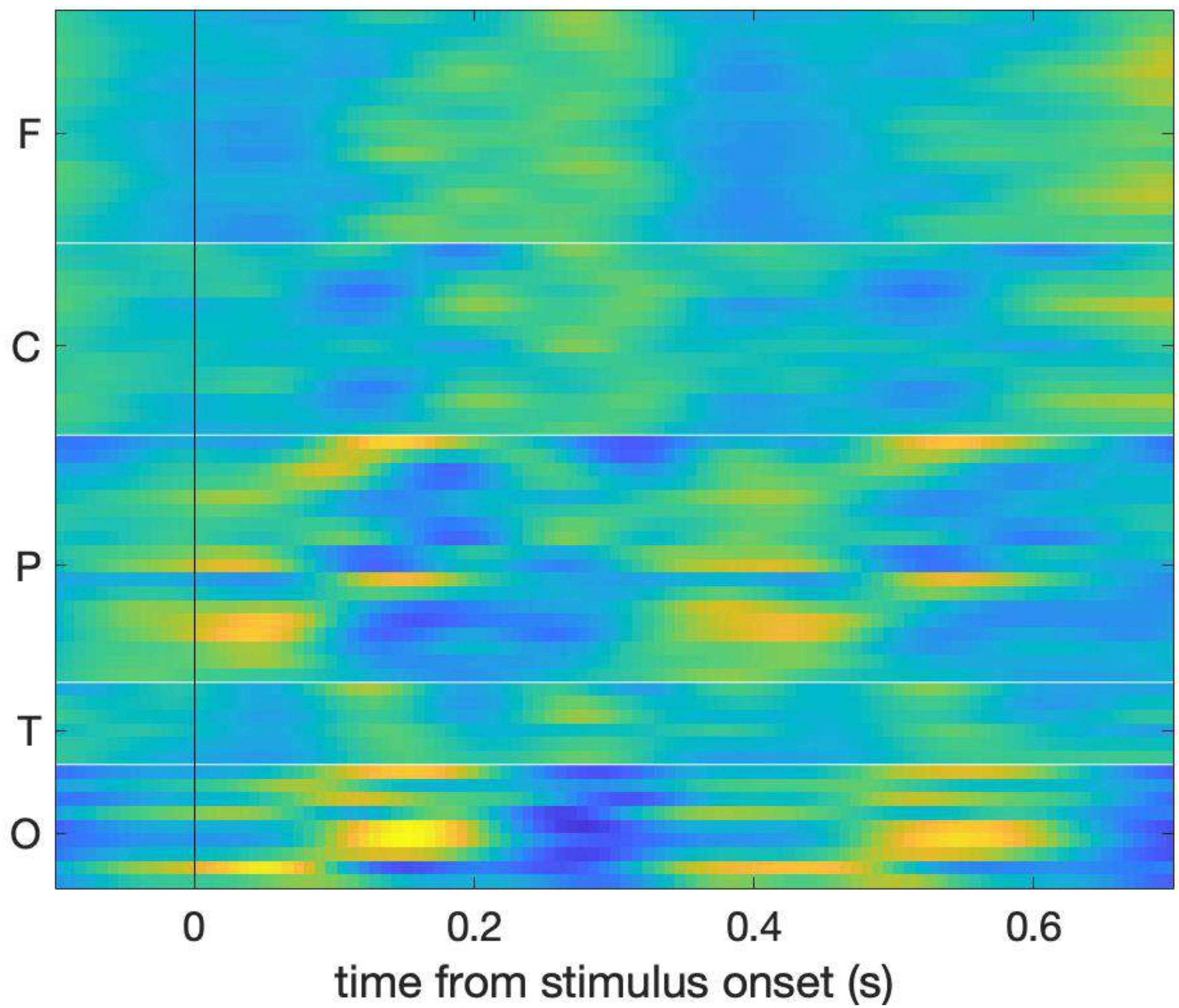
Data mining and modeling

Stream of oriented visual patterns

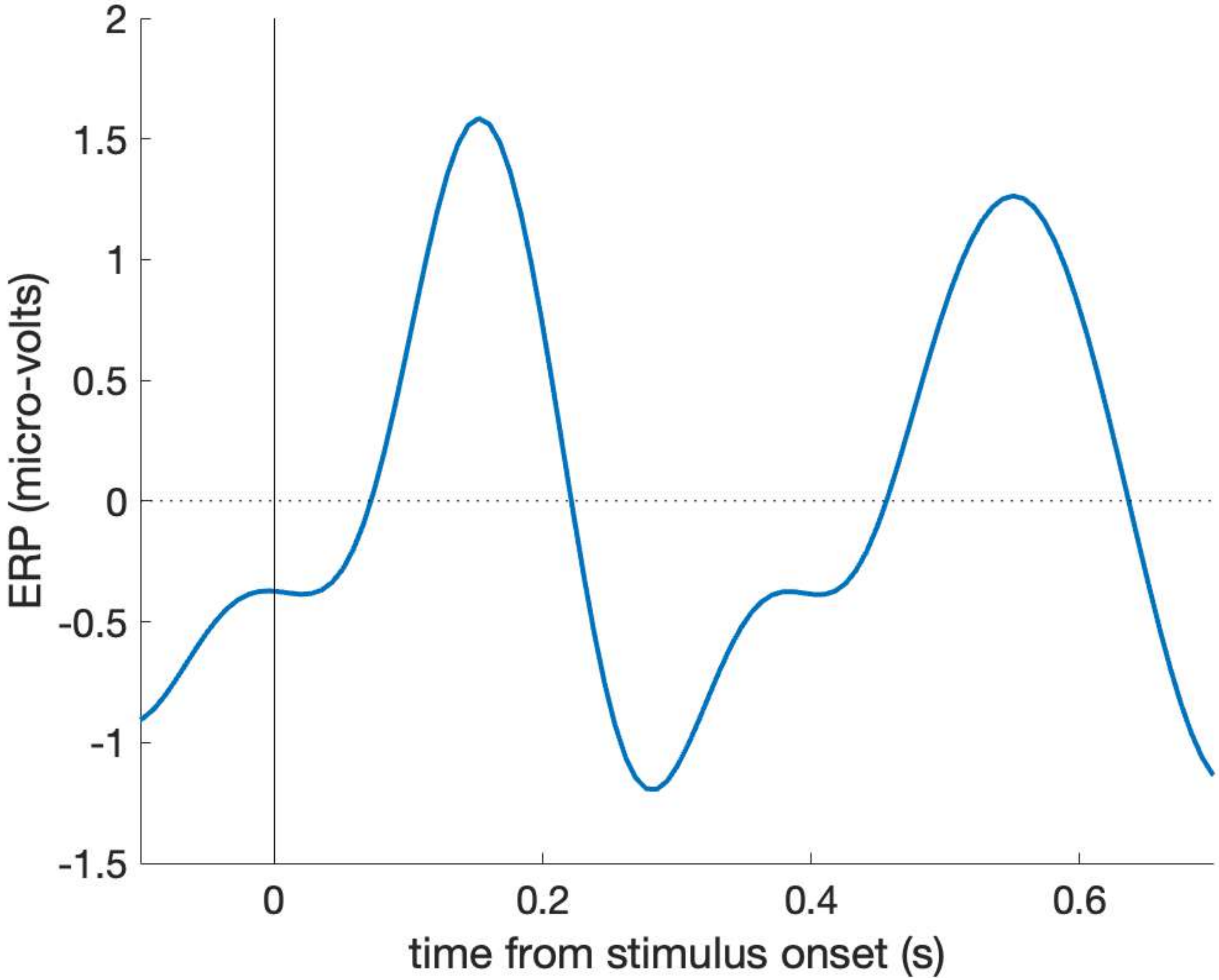




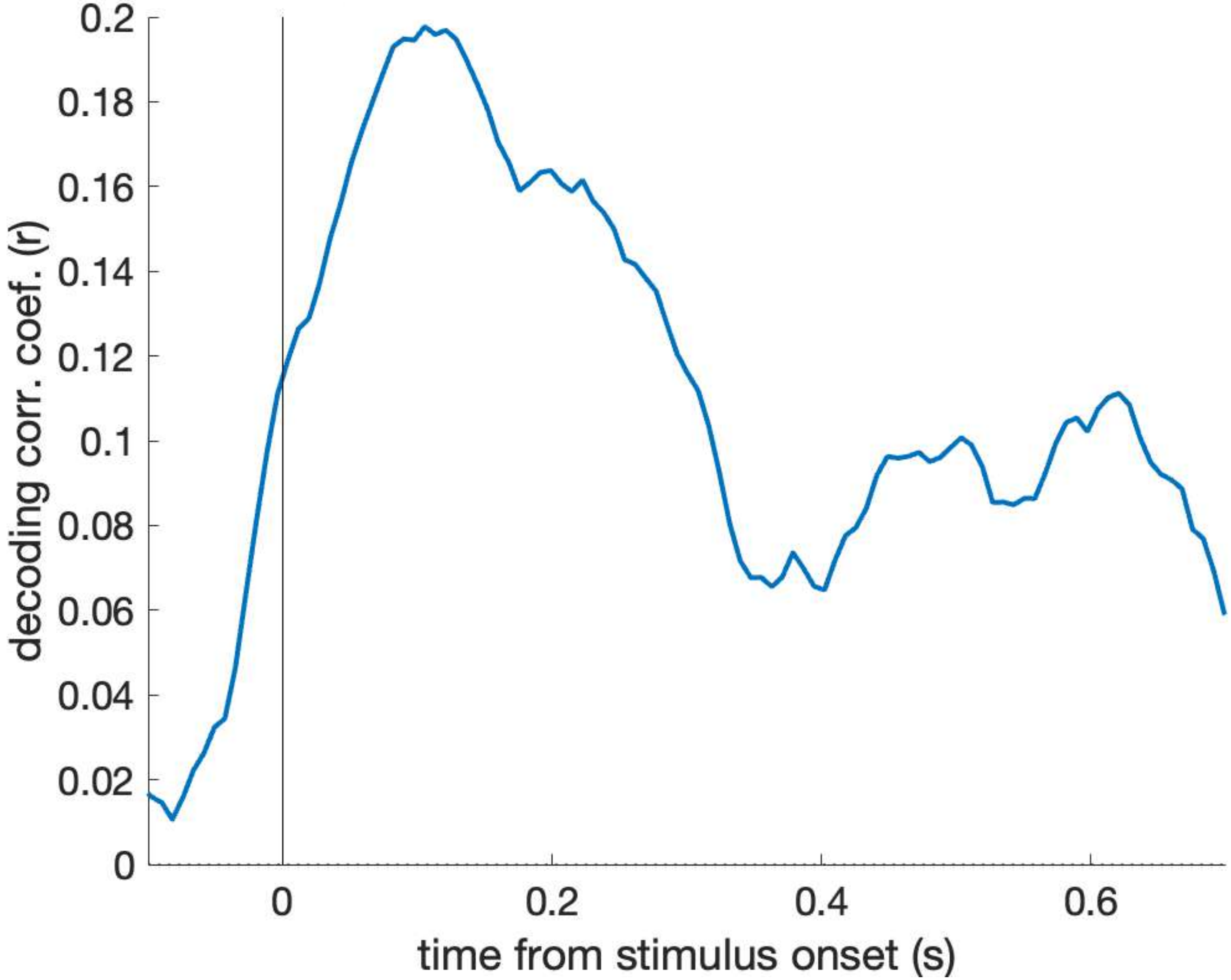
Data mining and modeling



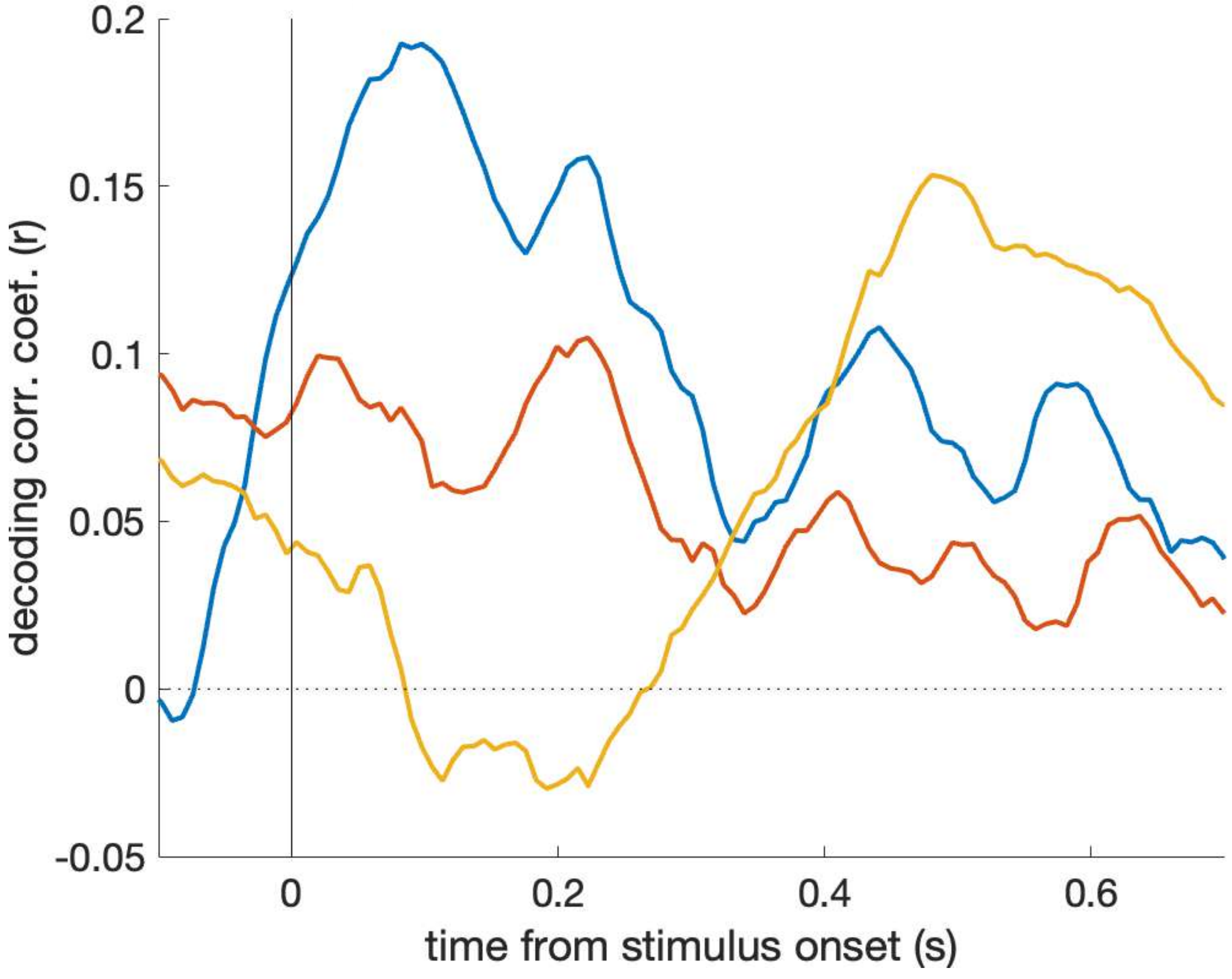
Data mining and modeling



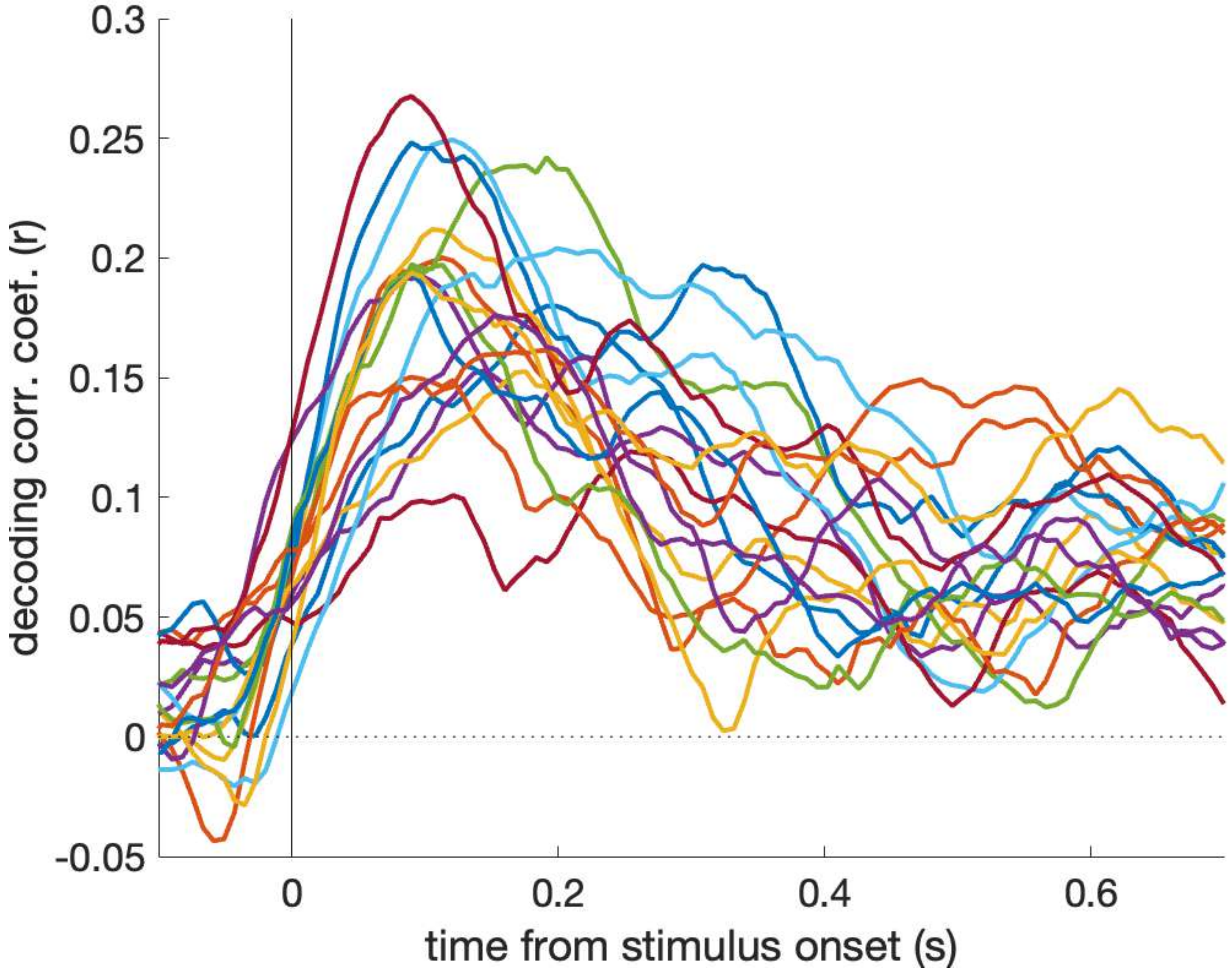
Data mining and modeling



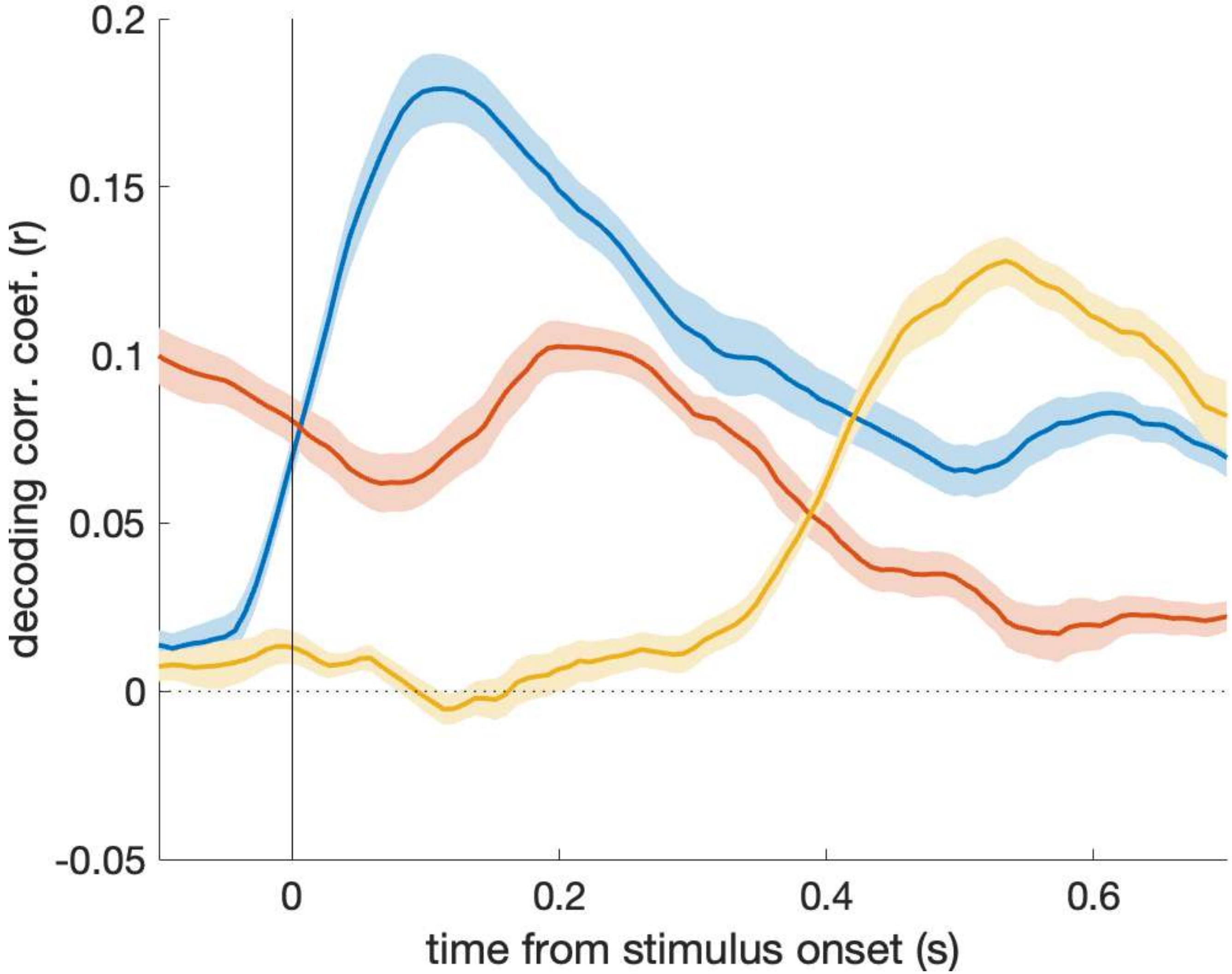
Data mining and modeling



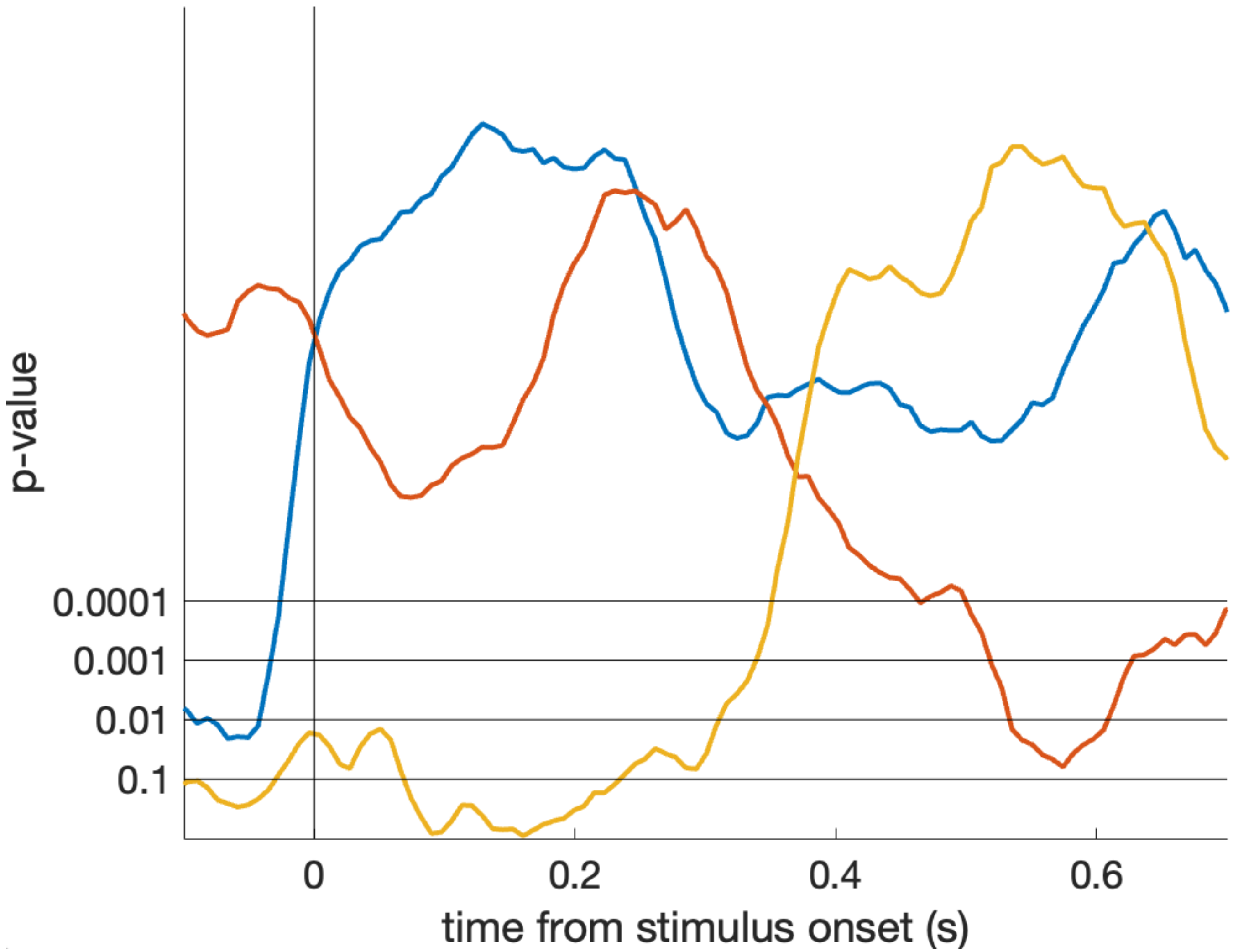
Data mining and modeling



Data mining and modeling



Data mining and modeling



What is a model of cognition?

- Statistical models are used to summarize and describe behavioral data.
- Computational models are used to understand how behavioral data have been generated.
- This morning, we will discuss computational models of cognition = cognitive modeling.

What is a model of cognition?

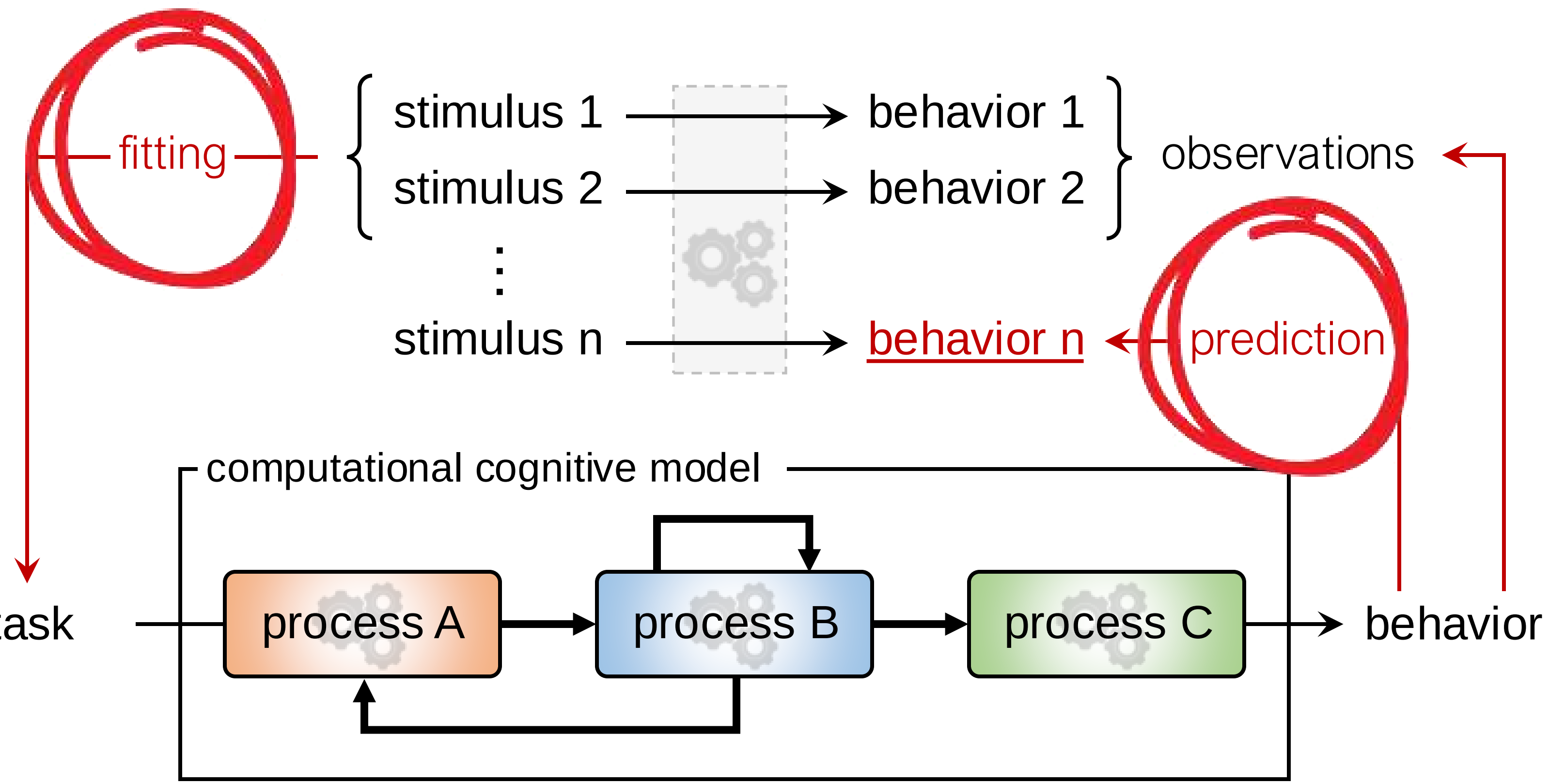
- Cognitive modeling aims at understanding behavior in terms of its underlying cognitive processes.
- Cognitive modeling proceeds by building mathematical descriptions and computer algorithms of these processes that are able to reproduce the studied behavior.

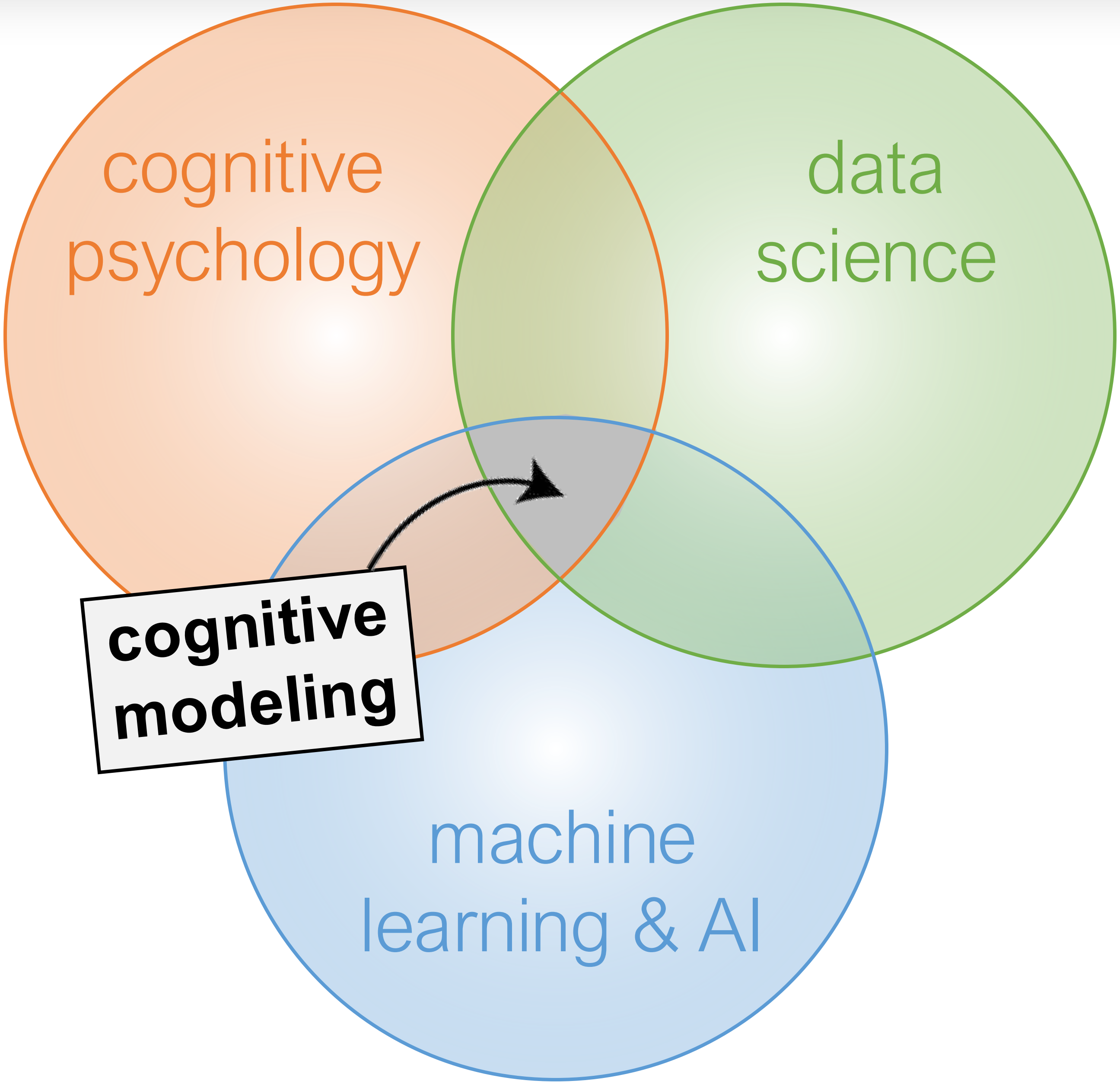
What is a model of cognition?

- <u>Not</u> a statistical model of effect size: a *t*-test of response time differences between experimental conditions is <u>not</u> a cognitive model.
- Cognitive modeling aims at understanding not whether choices differ between conditions, but why choices differ between <u>these</u> conditions.
- Cognitive models are computational instances of theories of cognition and behavior.

What is a model of cognition?

- Fitting vs predicting behavior




cognitive
psychology
data
science
machine
learning & AI
cognitive
modeling

What is a model of cognition?

- Connections with cognitive psychology
  - ✓ <u>shared goal:</u>
    understand the human mind
  - ✓ <u>shared techniques:</u>
    design controlled experiments that target <u>specific</u> cognitive processes

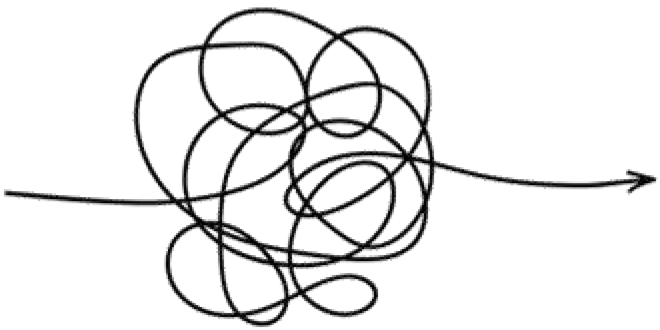
What is a model of cognition?

- Cognitive psychology uses rhetorical statements to explain differences between conditions.
- Such qualitative statements can be flawed by internal inconsistencies, logical contradictions, and theoretical weaknesses.
- A cognitive model is used as a quantitative proof of the internal coherence and completeness of the theory it is based upon.

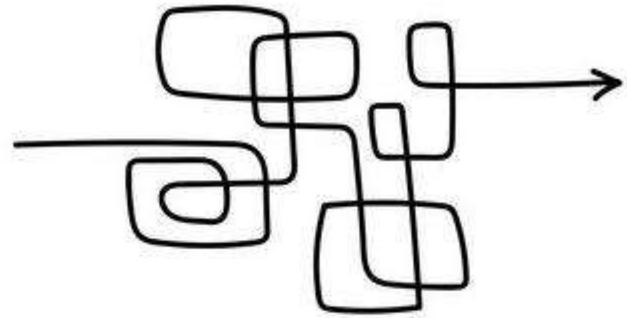
What is a model of cognition?

- Connections with data science
  - ✓ shared goal:
    build computer algorithms to explain/predict behavioral data
  - ✓ shared techniques:
    formulate, simulate, fit, compare, validate computer algorithms against behavioral data
- Data science aims (first) at predicting data, cognitive models aim at understanding behavior.

model A

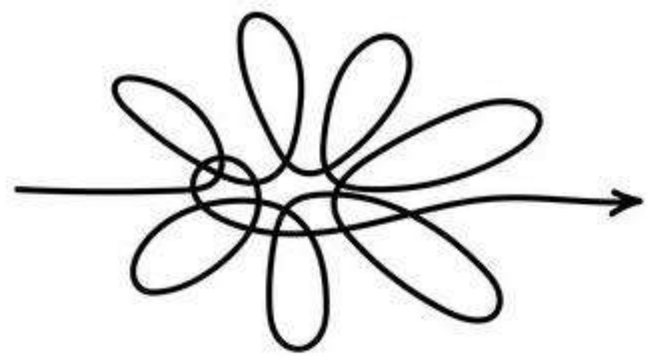


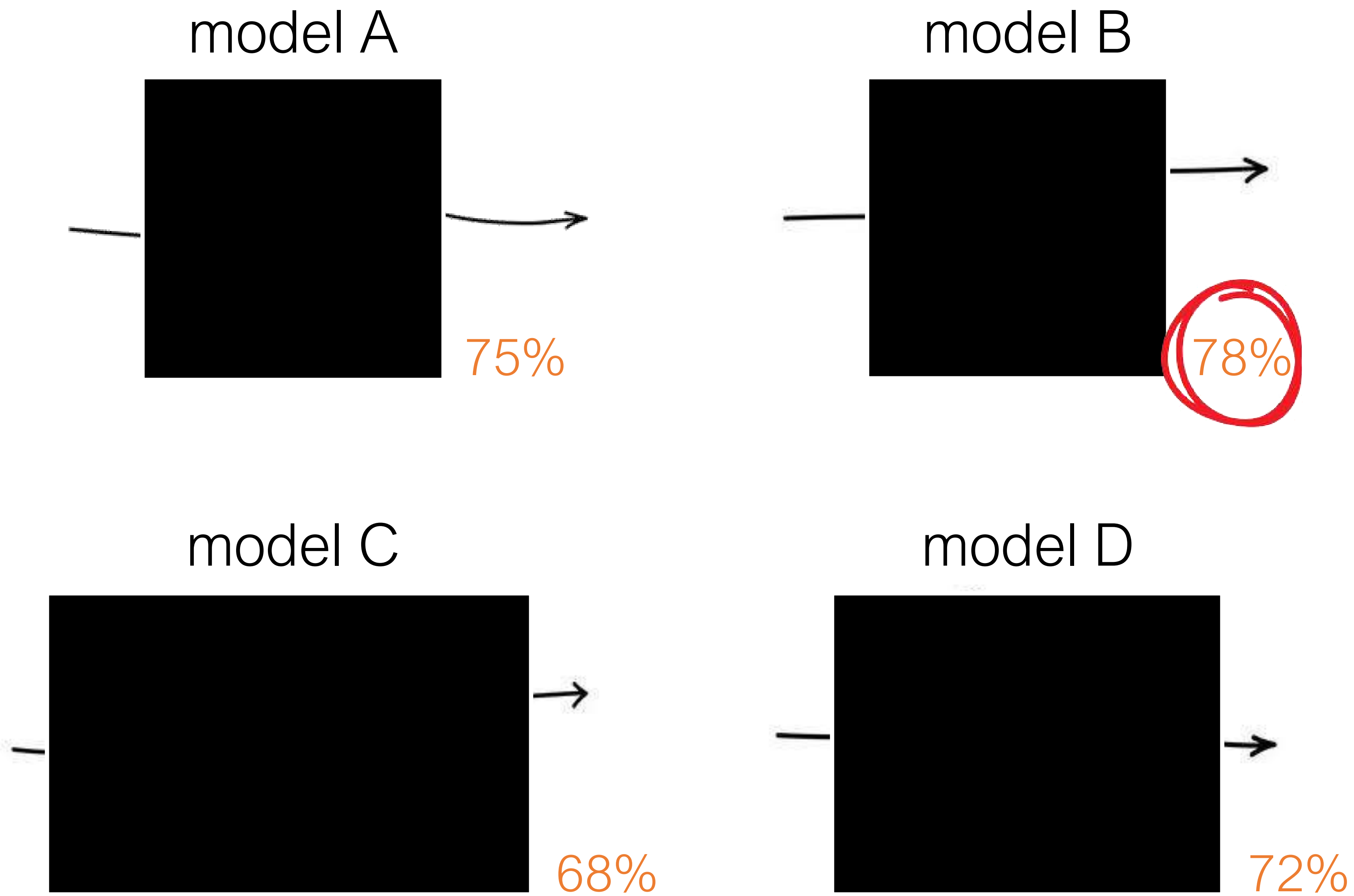
model B



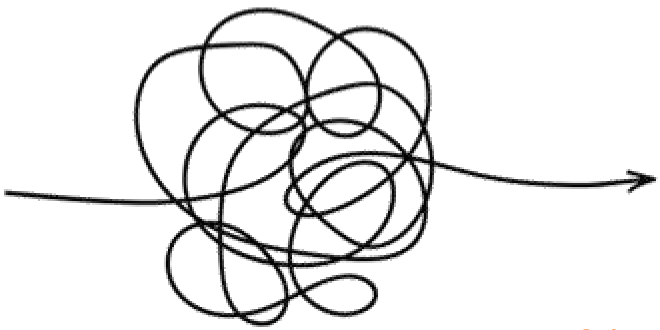
model C

model D




model A
75%
model B
78%
model C
68%
model D
72%

model A

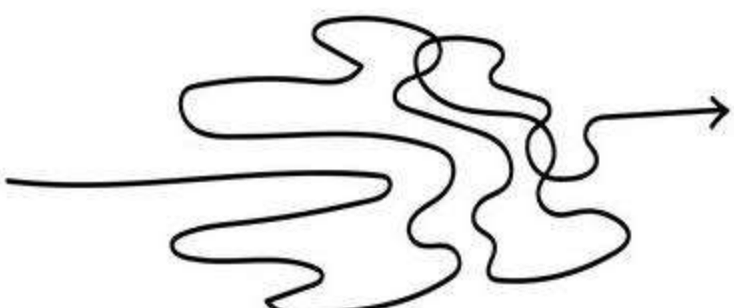


75%

model B

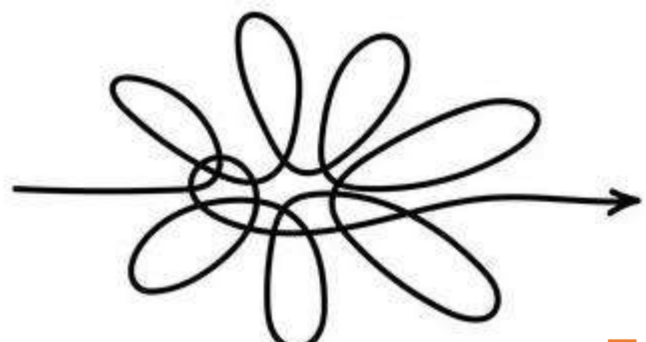
78%

model C



68%

model D

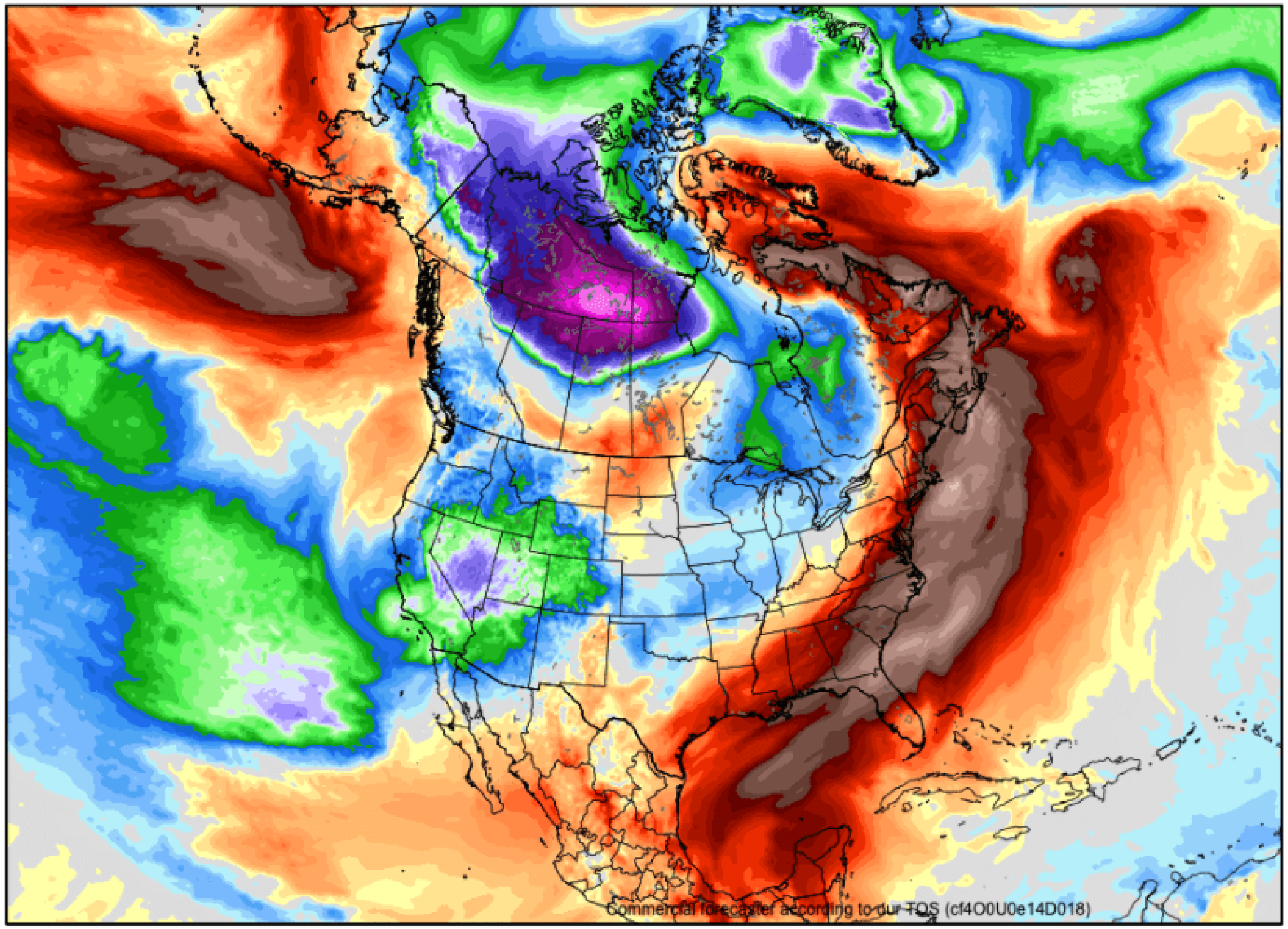


72%

What is a model of cognition?

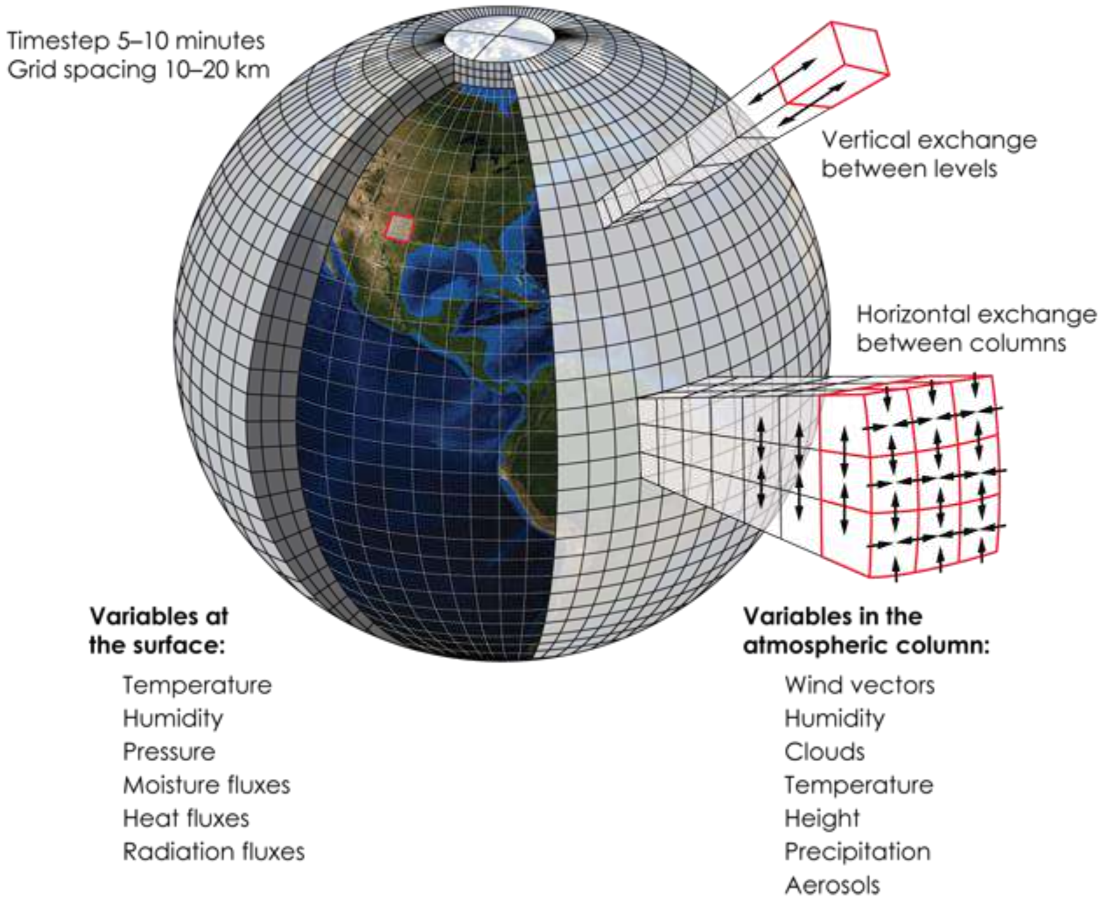
- Data science cares about predictive accuracy, not about the data generation process.
- Cognitive modeling cares about the underlying processes that have generated the data.

What is a model of cognition?



What is a model of cognition?

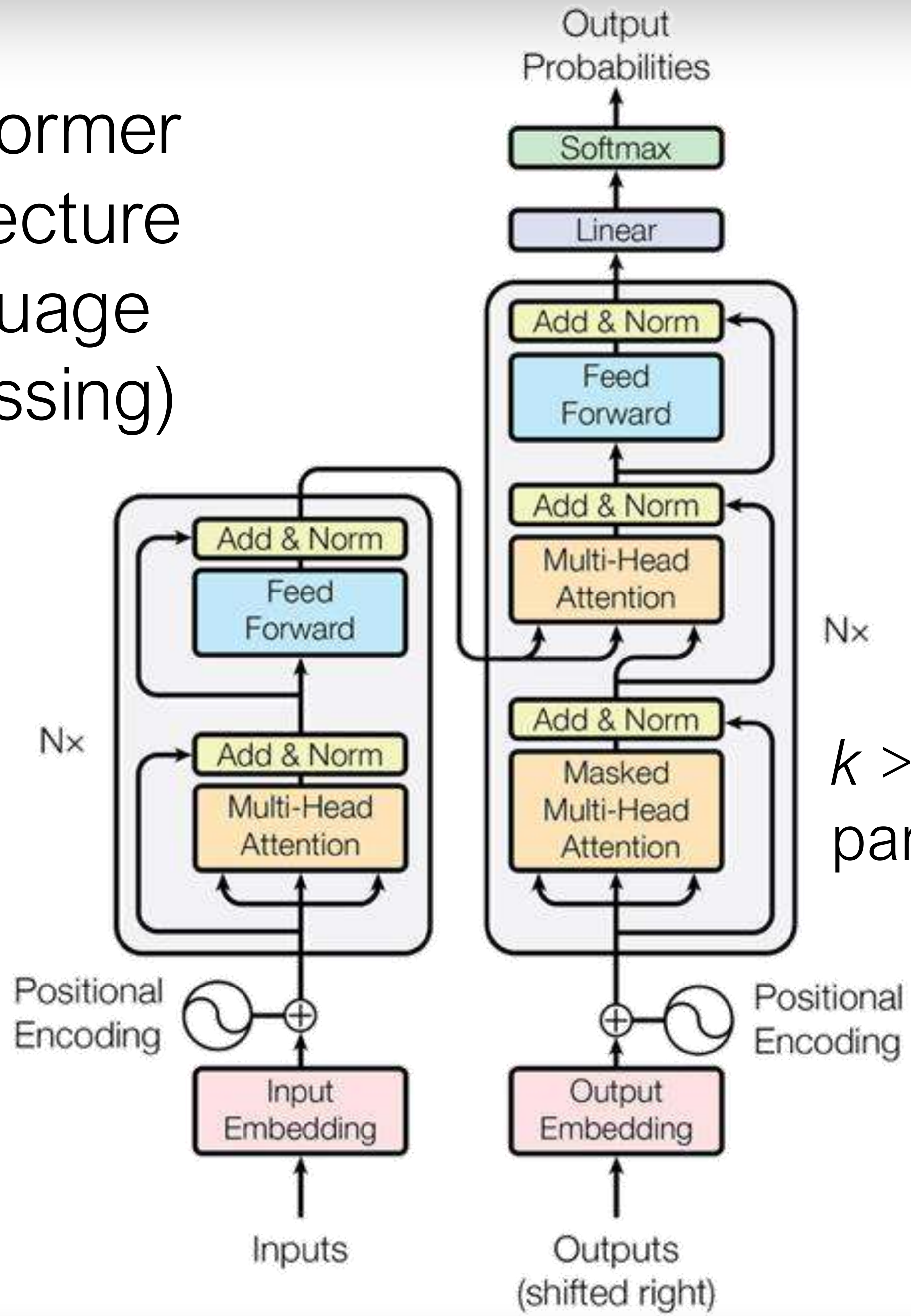
weather modeling



What is a model of cognition?

- Connections with machine learning & AI
  - ✓ shared goal:
    build models of the mind
  - ✓ shared techniques:
    reinforcement learning, particle filtering, pattern classification, neural networks…
- AI aims at maximizing performance, cognitive modeling aims at understanding the human mind

Transformer architecture (language processing)

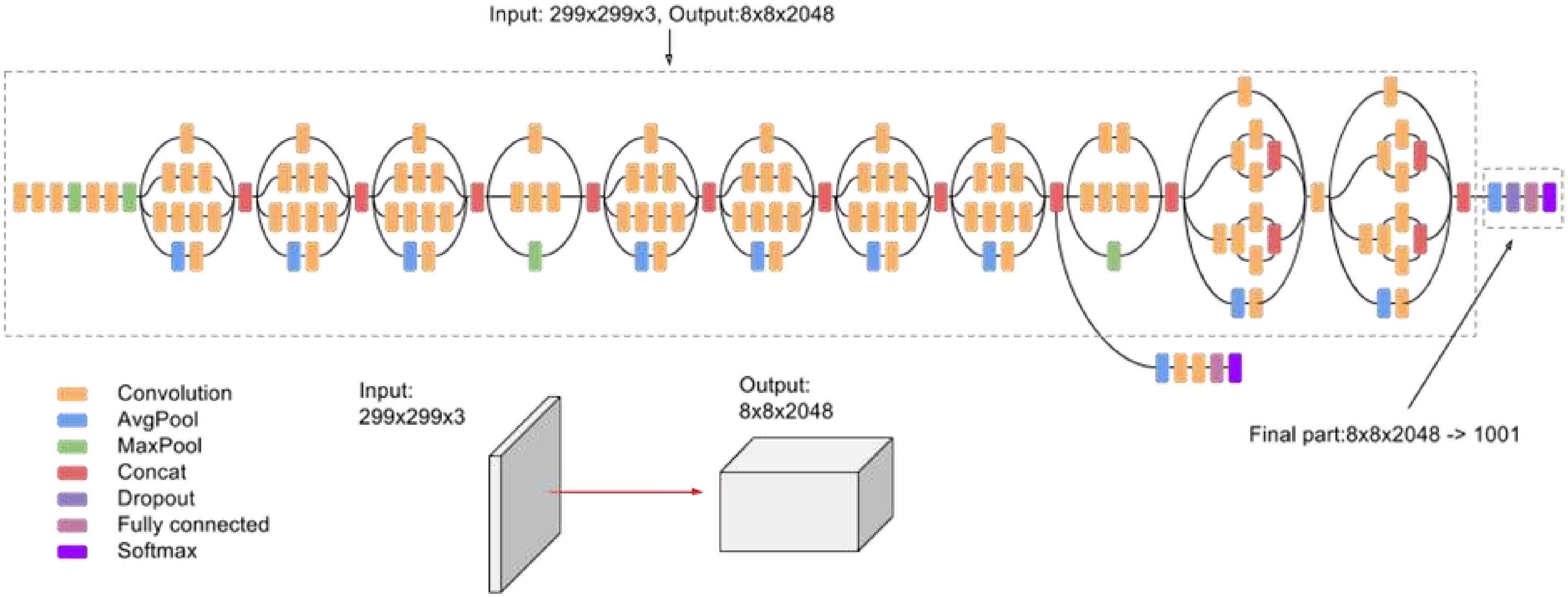


trained to predict next word/token

$k > 10^9$ free parameters

Inception v3 architecture (image recognition)

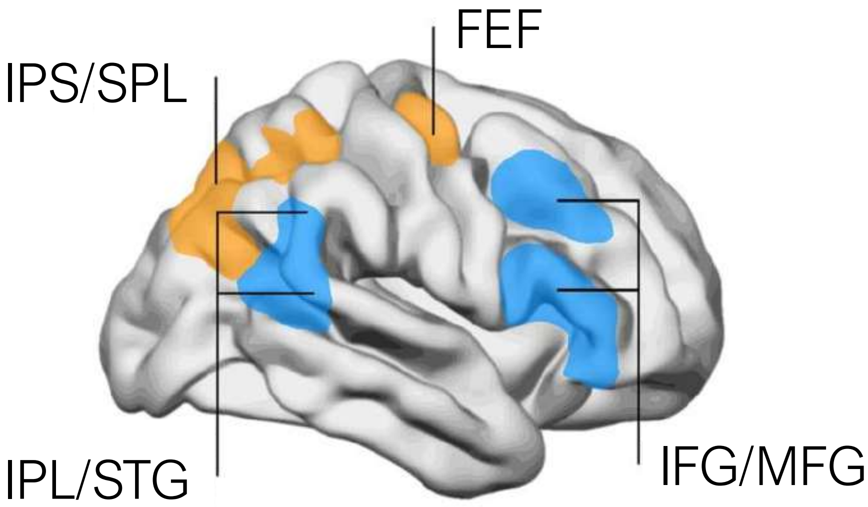
$k > 10^9$ free parameters



trained to maximize recognition accuracy

Why modeling cognition? We can look into brains!

- Does mind = brain?



Organization for Human Brain Mapping
Advancing Understanding of the Human Brain

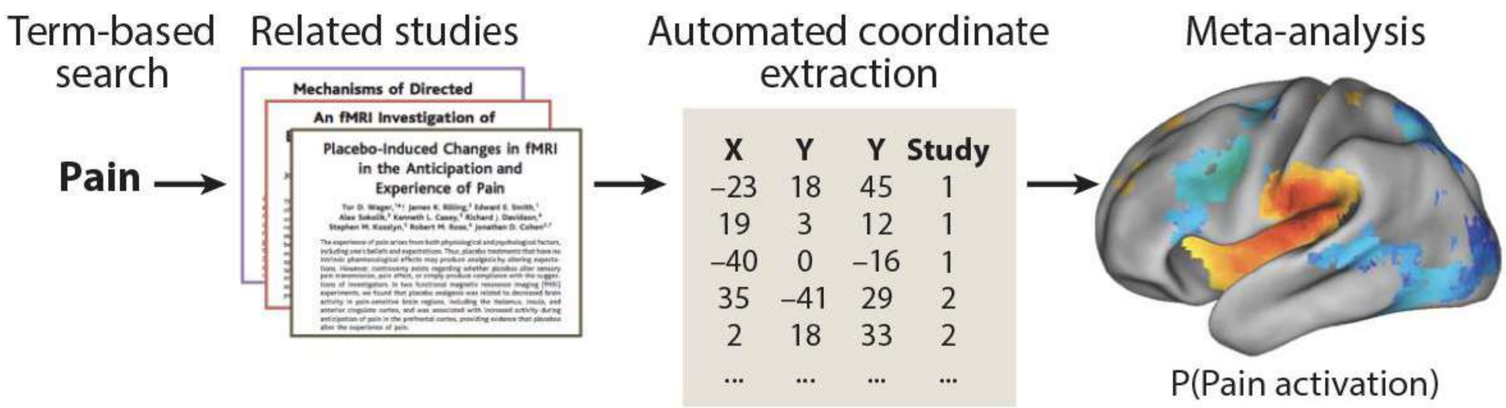
Why modeling cognition?
We can look into brains!

- Does mind = brain?

- <u>Definition of brain mapping:</u>
  “the creation of a visual representation of the brain in which different cognitive functions are assigned to different brain regions.”
  (American Psychological Association)

Why modeling cognition?
We can look into brains!

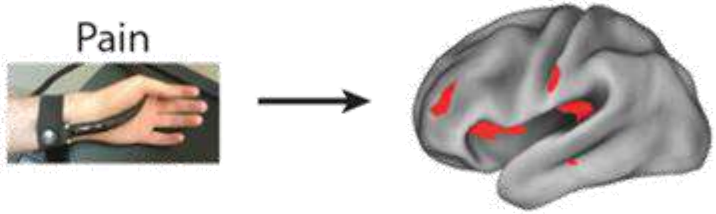
- Mapping psychological constructs on the brain is notoriously tricky



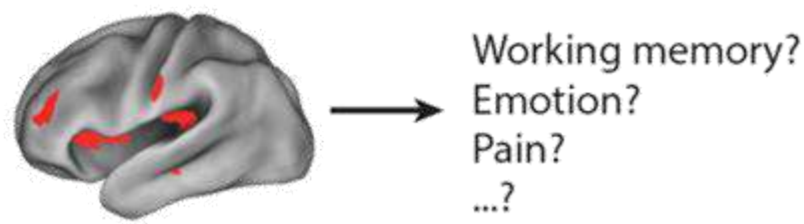
Yarkoni and Poldrack (2016) From brain maps to cognitive ontologies. *Annu. Rev. Psychol.*

Why modeling cognition?
We can look into brains!

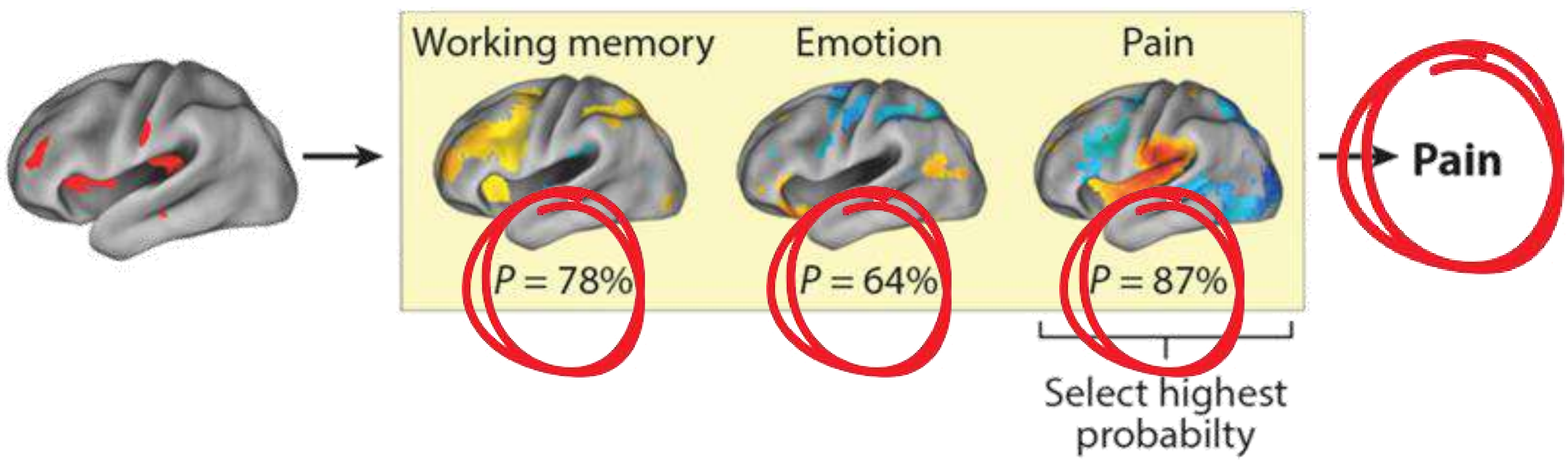
forward inference



<u>reverse</u> inference



classification (machine learning)



Why modeling cognition?
~~We can look into brains!~~

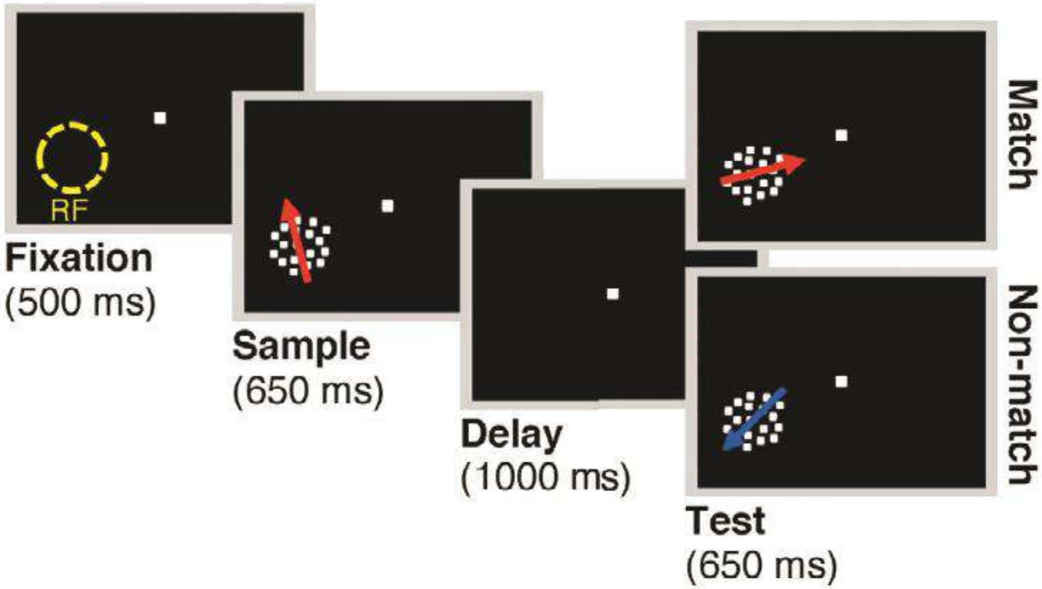
- Shimon Edelman: “the mind (as cognitive system) is best defined not in terms of its physical substrate, but in terms of the relations that states of the system have to one another, and to the outside world.”

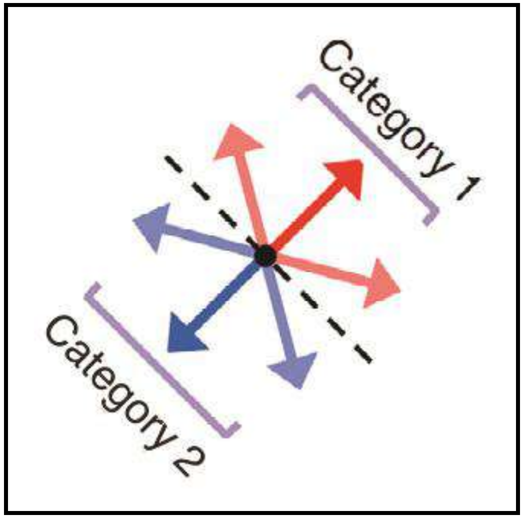
- Correspondences can be identified between physically dissimilar cognitive systems in terms of shared computations.

Why modeling cognition?

~~We can look into brains!~~

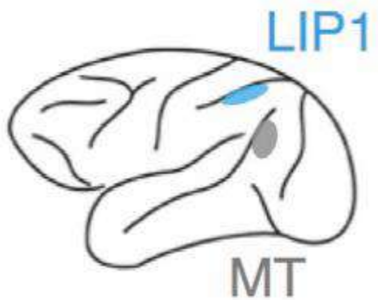
Delayed match-to-sample task



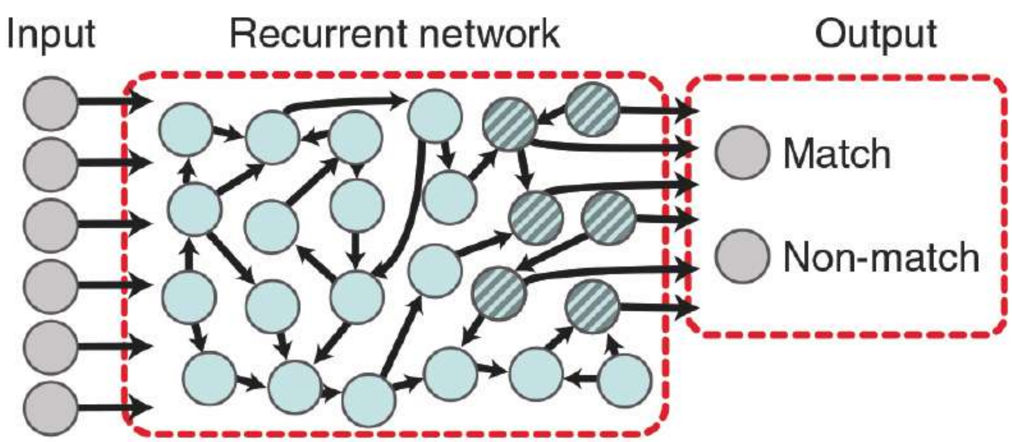


Why modeling cognition?
~~We can look into brains!~~

Experiment 1

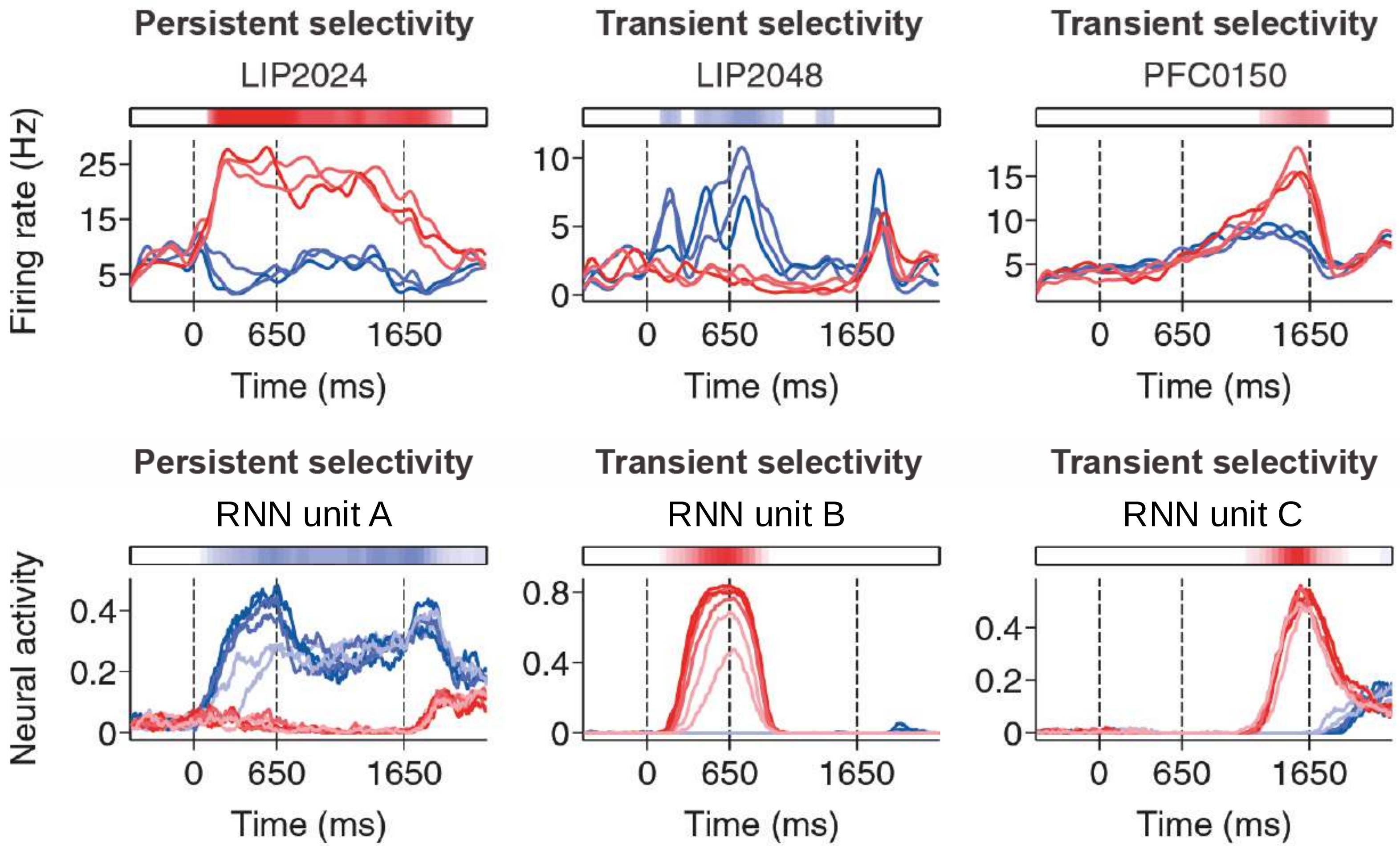


Experiment 2



Why modeling cognition?

~~We can look into brains!~~

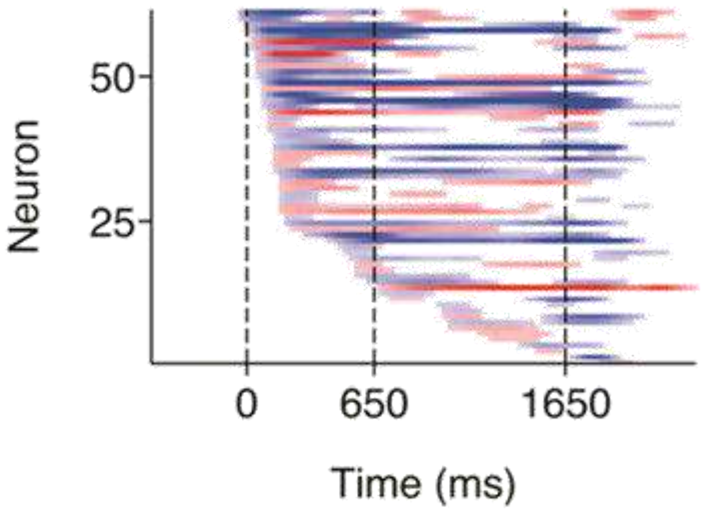


Why modeling cognition?

~~We can look into brains!~~

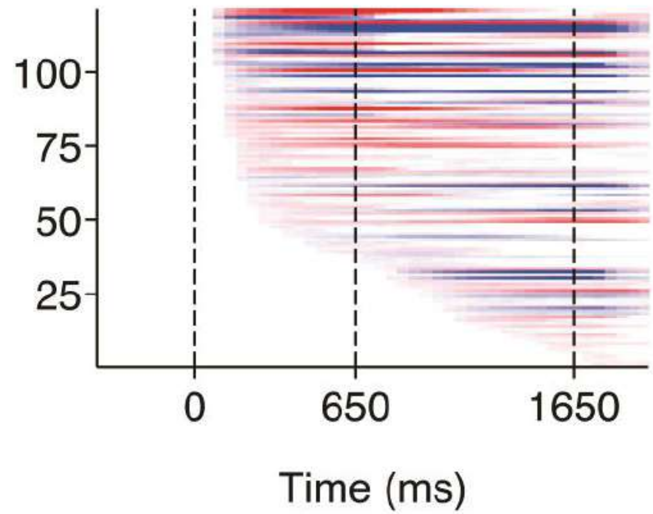
Category selectivity
of individual units

LIP1



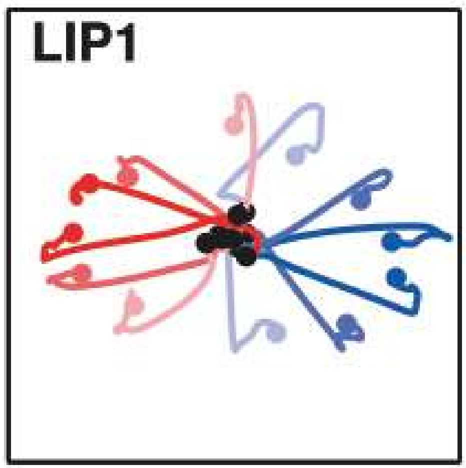
≈

MODEL

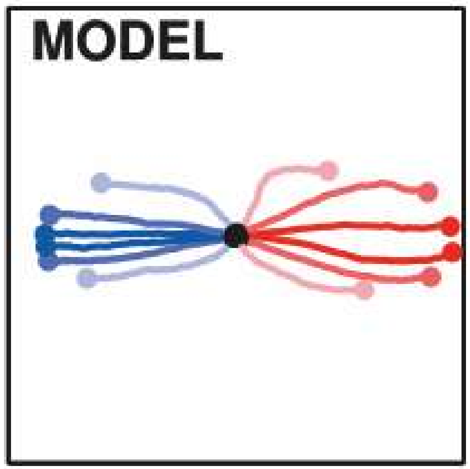


Why modeling cognition?
~~We can look into brains!~~

Low-dimensional projection
of population activity



≈



How to model cognition?

- Three main ways of modeling cognition:
- 1/ Theorizing cognition
- 2/ Experimenting cognition in the laboratory
- 3/ Studying cognition in real-world conditions
- Do you see the benefits and limits of each?
- Let's see an example in the case of human exploration strategies under uncertainty.

How to model cognition?

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Current Opinion in
Neurobiology

The algorithmic architecture of exploration in the human brain

Eric Schulz and Samuel J Gershman

Check for updates

Balancing exploration and exploitation is one of the central problems in reinforcement learning. We review recent studies that have identified multiple algorithmic strategies underlying exploration. In particular, humans use a combination of random and uncertainty-directed exploration strategies, which rely on different brain systems, have different developmental trajectories, and are sensitive to different task manipulations. Humans are also able to exploit sophisticated structural knowledge to aid their exploration, such as information about correlations between options. New computational models, drawing inspiration from machine learning, have begun to formalize these ideas and offer new ways to understand the neural basis of reinforcement learning.

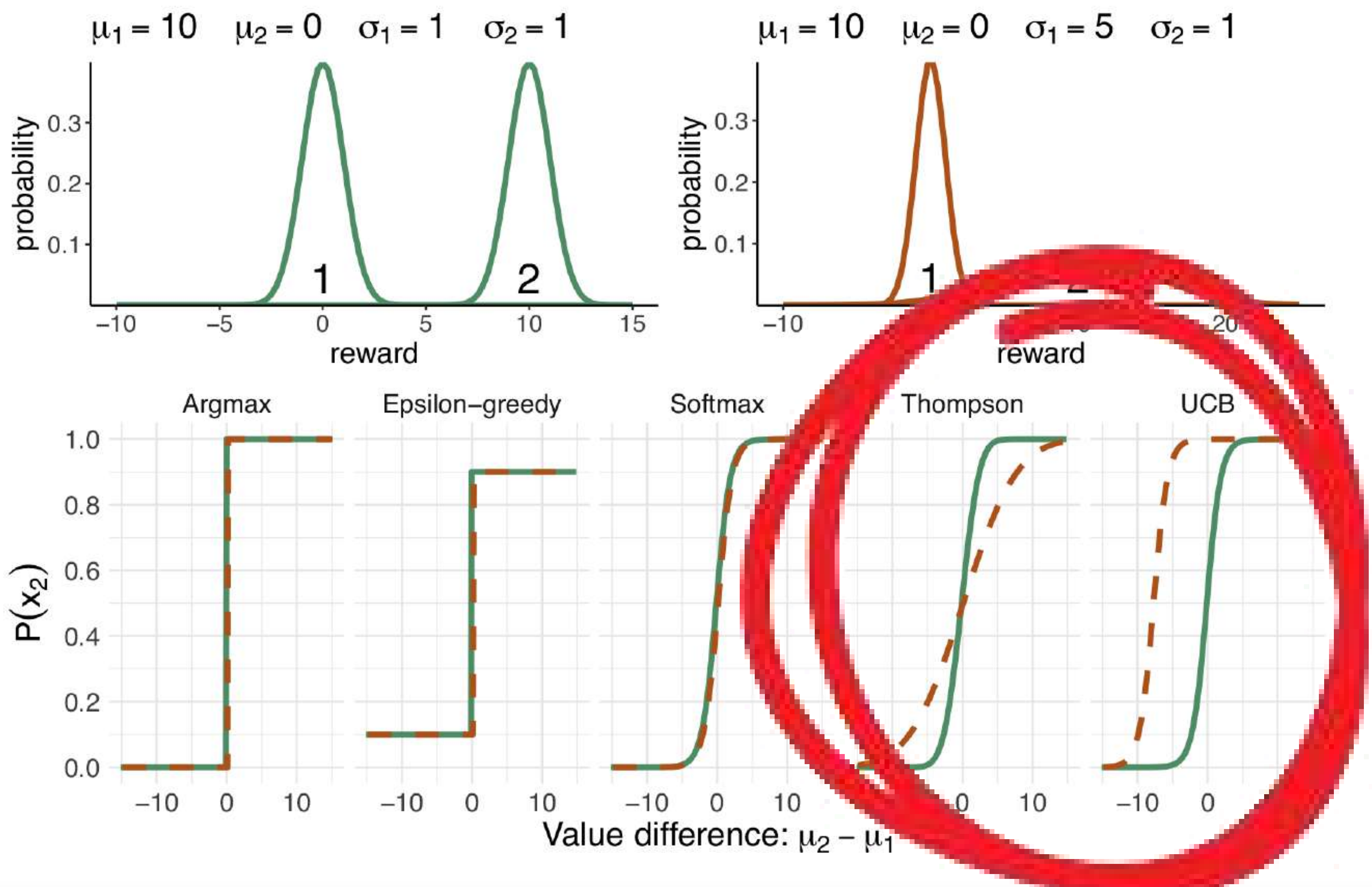
Address
Department of Psychology and Center for Brain Science, Harvard University, 52 Oxford Street, Cambridge, MA 02138, USA

an independent payoff distribution. It is then an agents goal to maximize rewards by repeatedly selecting an arm and observing and collecting the resulting reward.

We first summarize evidence that humans use two distinct exploration strategies [4,5]: *random exploration*, which increases choice stochasticity to the agent's uncertainty about the values of available actions, and *directed exploration* which adds a bonus to each action in proportion to the agent's uncertainty about each action's value. These two algorithms offer heuristic yet efficient solutions to the exploration–exploitation dilemma. Signatures of directed and random exploration can be observed in human choice behavior, develop differently across the lifespan, and recruit distinct neural mechanisms.

In addition to using uncertainty to guide exploration, ... evidence suggests that humans use structured

How to model cognition?

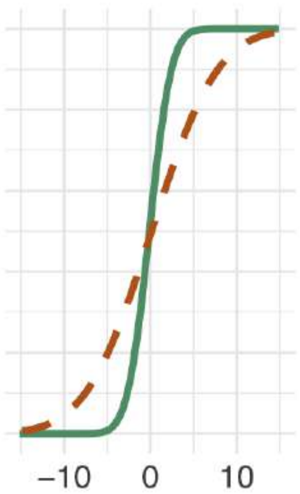


How to model cognition?

Thompson policy

random exploration

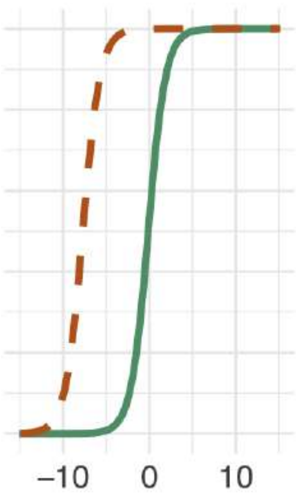
$$P(a=1) = \Phi\left(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}\right)$$



UCB policy

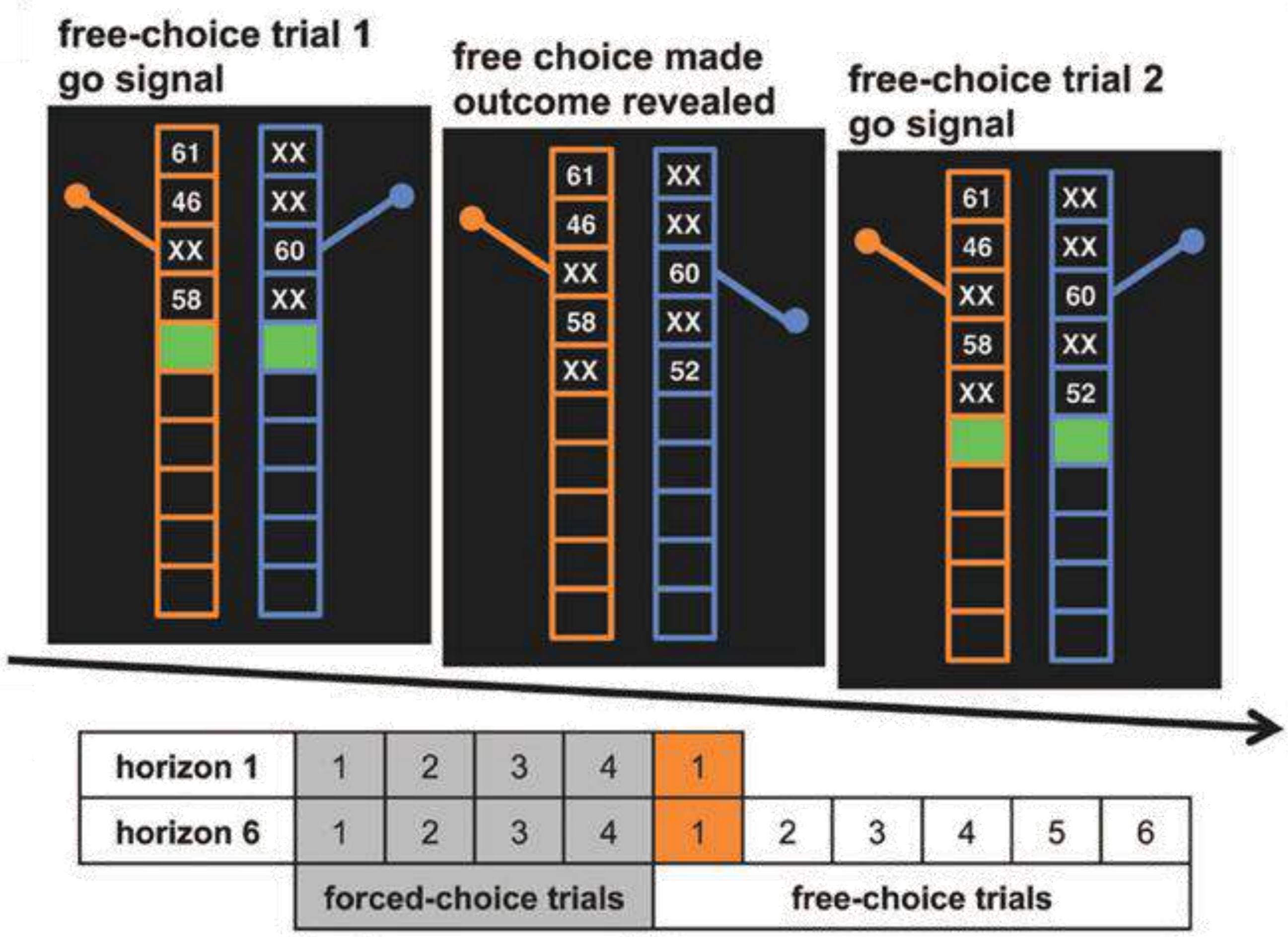
directed exploration

$$P(a=1) = \Phi\left(\frac{\mu_1 - \mu_2 + \gamma[\sigma_1 - \sigma_2]}{\tau}\right)$$



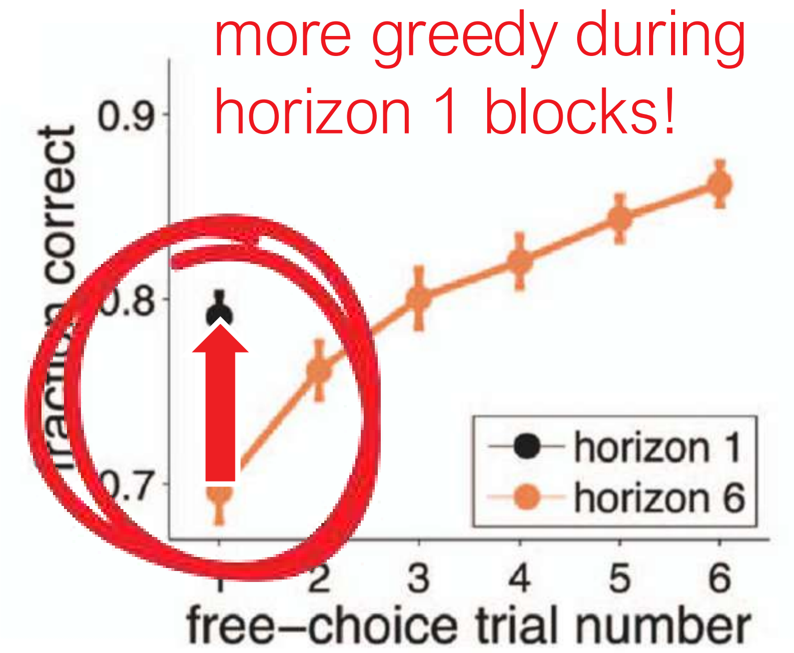
How to model cognition?

- Do humans engage in directed exploration?



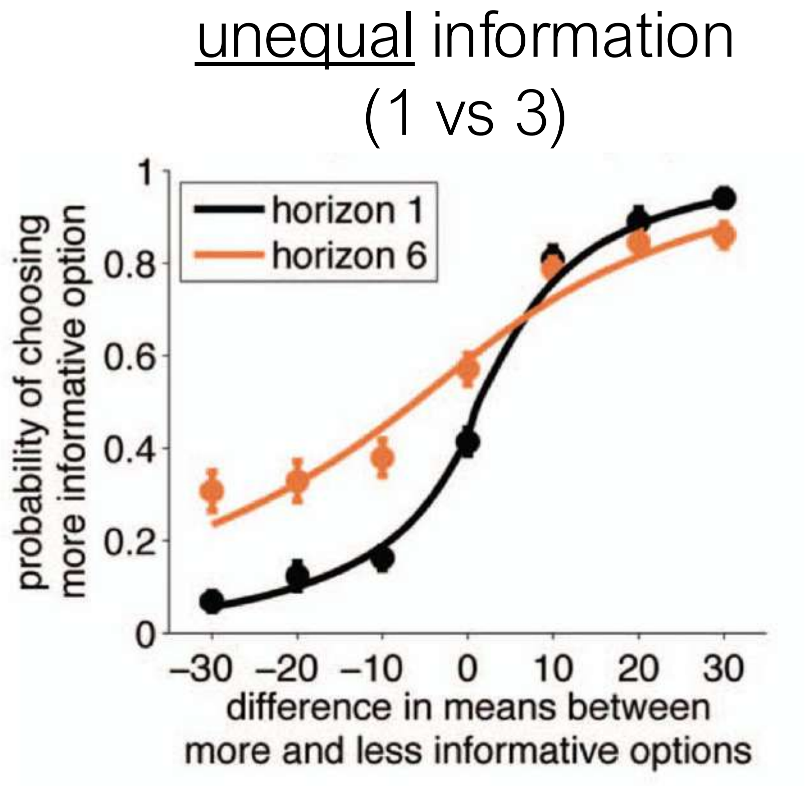
How to model cognition?

- Do humans engage in directed exploration?



How to model cognition?

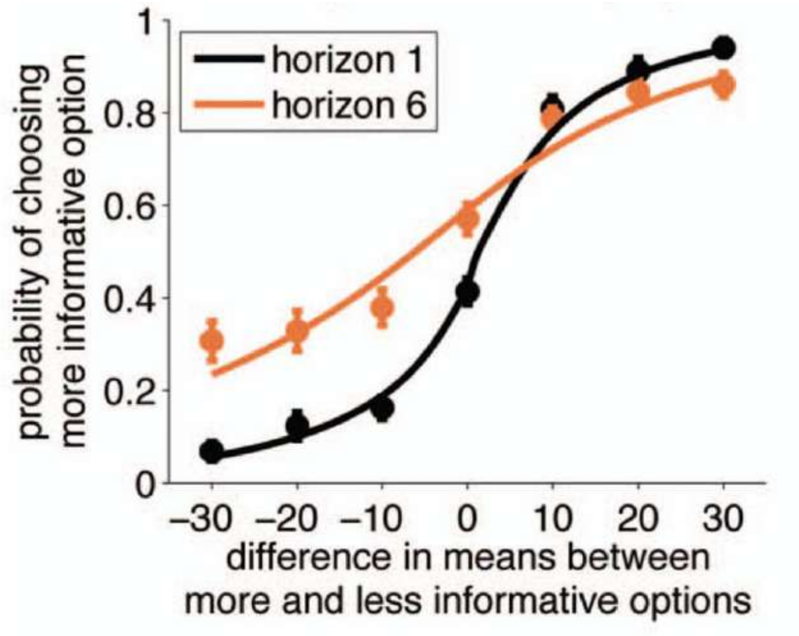
- Do humans engage in directed exploration?



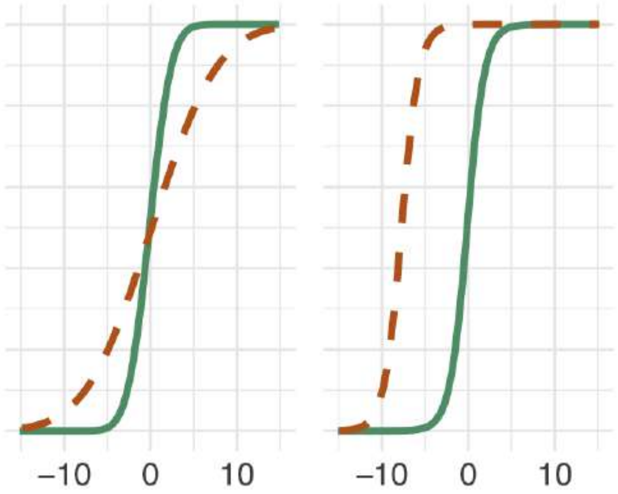
How to model cognition?

- Do humans engage in directed exploration?

unequal information
(1 vs 3)



random directed
exploration

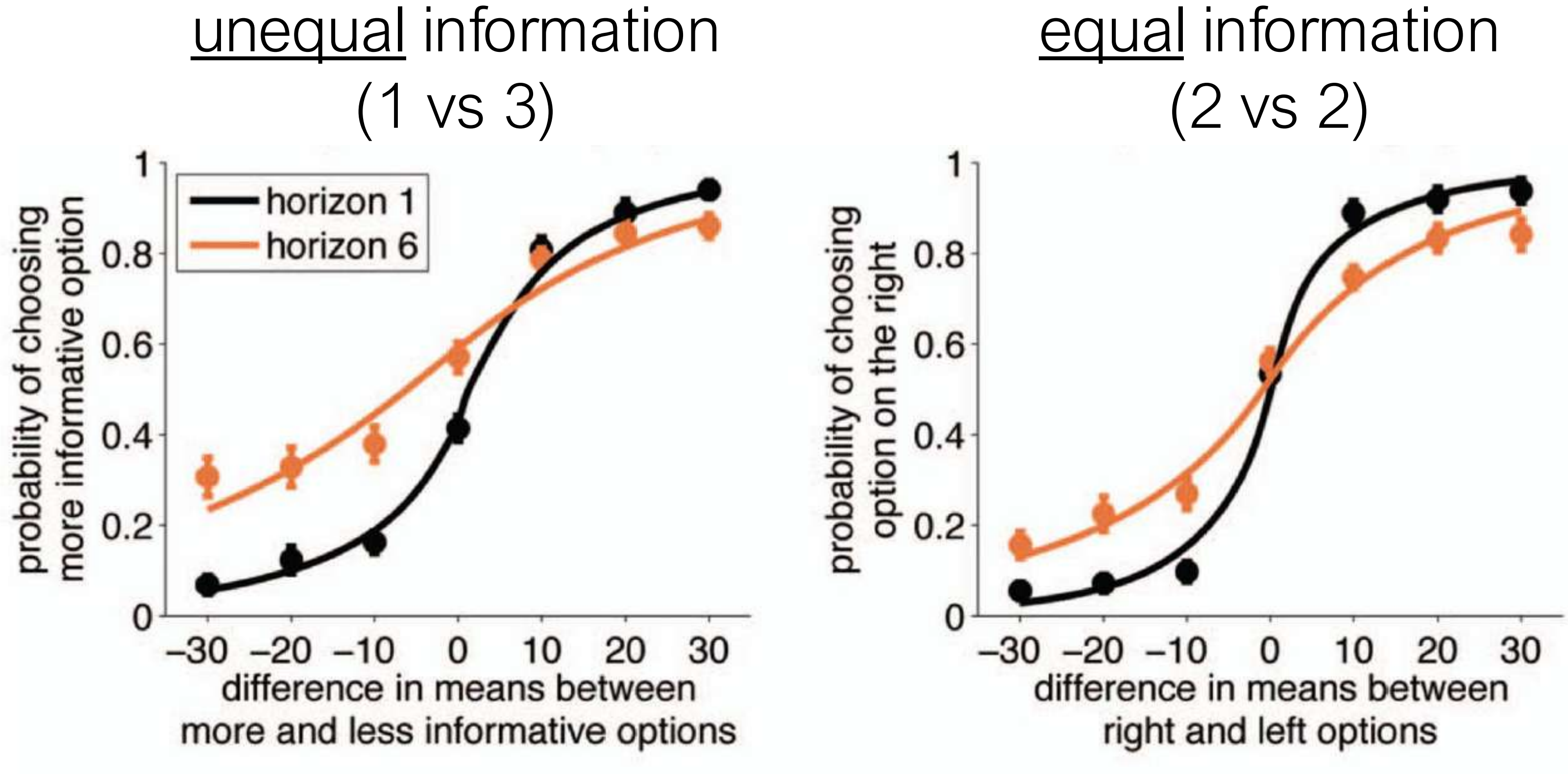


Thompson & softmax

UCB

How to model cognition?

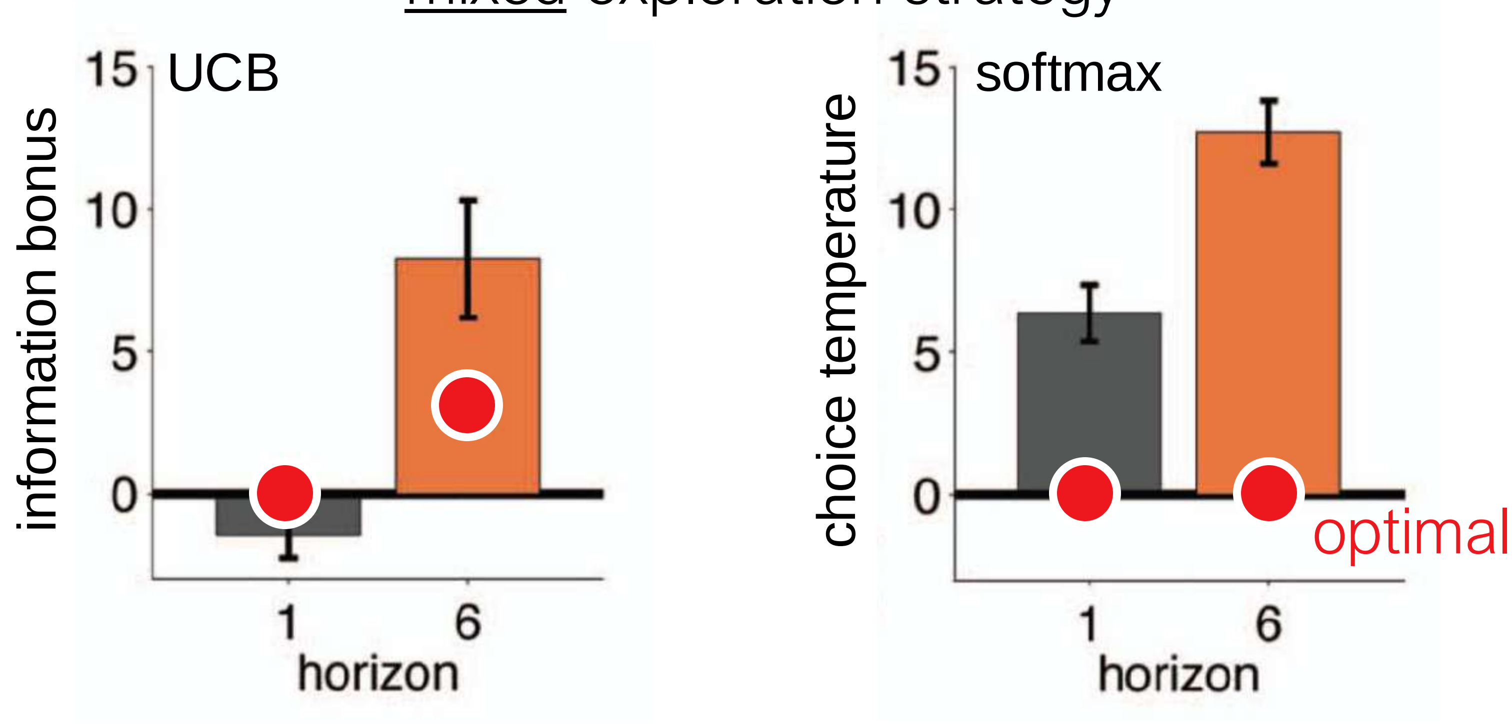
- Do humans engage in directed exploration?



How to model cognition?

- Do humans engage in directed exploration?

mixed exploration strategy



How to model cognition?

- Do humans engage in directed exploration?
- Humans use a mixture of directed and random exploration in the laboratory.
- How much can we generalize from this finding?
- What does ‘generalization’ mean?
  - ✓ as a cognitive ability
  - ✓ as an induction process in science

How to model cognition?

- What are the pros and cons of studying human exploration (or anything) in laboratory tasks?
- Advantages: tightly-controlled conditions, easy and fast to loop between hypothesis generation and test, ability to derive generative models…
- Disadvantage: laboratory tasks do not reflect real-world conditions… Cite some!
- Does exploration in the lab reflect exploration outside the lab?

How to model cognition?

PNAS PNAS PNAS

Check for updates

Structured, uncertainty-driven exploration in real-world consumer choice

Eric Schulz[a,1,2], Rahul Bhui[a,1], Bradley C. Love[b,c], Bastien Brier[d], Michael T. Todd[d], and Samuel J. Gershman[a]

[a]Department of Psychology, Harvard University, Cambridge, MA 02138; [b]Department of Experimental Psychology, University College London, London WC1H 0AP, United Kingdom; [c]The Alan Turing Institute, London NW1 2DB, United Kingdom; and [d]Data Science Team, Deliveroo, London EC4R 3TE, United Kingdom

Edited by Richard M. Shiffrin, Indiana University, Bloomington, IN, and approved May 23, 2019 (received for review December 10, 2018)

Making good decisions requires people to appropriately explore their available options and generalize what they have learned. While computational models can explain exploratory behavior in constrained laboratory tasks, it is unclear to what extent these models generalize to real-world choice problems. We investigate the factors guiding exploratory behavior in a dataset consisting of 195,333 customers placing 1,613,967 orders from a large online food delivery service. We find important hallmarks of adaptive exploration and generalization, which we analyze using computational models. In particular, customers seem to engage in uncertainty-directed exploration and use feature-based generalization to guide their exploration. Our results provide evidence that people use sophisticated strategies to explore complex, real-world environments.

exploration | generalization | reinforcement learning | decision making

When facing a vast array of new opportunities, a decision maker has two key tasks: to acquire information (often through direct experience) about available options and to apply that information to assess options not yet experienced. These

it is unclear whether these theories can successfully predict real-world choices.

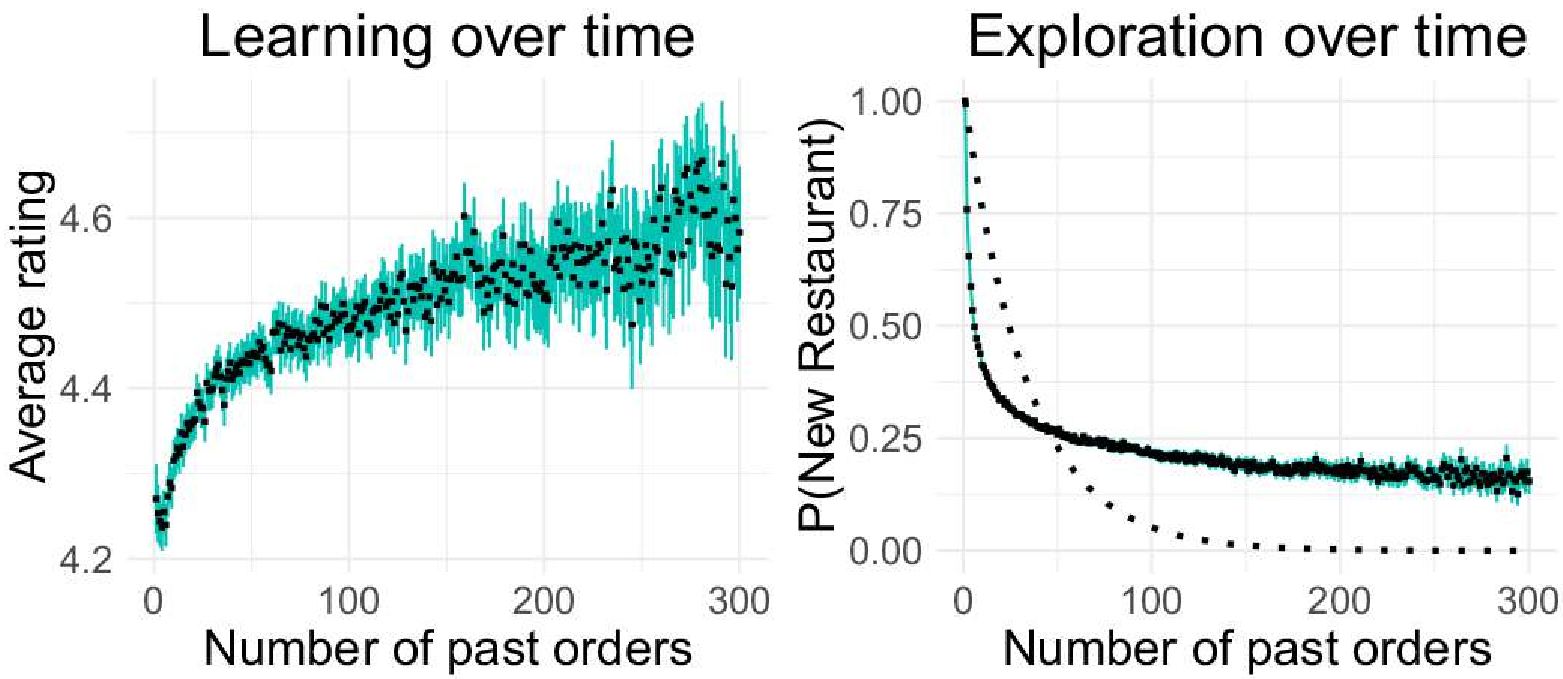
Our results suggest that customers explore (i.e., order from unexperienced restaurants) adaptively based on signals of restaurant quality and make better choices over time. Exploration is indeed risky and leads to worse outcomes on average, but people are more likely to explore in cities where this downside is lower due to higher mean restaurant quality. Moreover, we show that customers' exploratory behavior might take into account not only the prospective reward from choosing a restaurant, but also the degree of uncertainty in their reward estimates. Consistent with an optimistic uncertainty-directed exploration policy, they preferentially sample lesser-known options and are more likely to reorder from restaurants with higher uncertainties.

Importantly, we apply cognitive and statistical modeling to customers' choice behavior and find that their choices are best fitted by a model that includes both an "uncertainty bonus" for unfamiliar restaurants and a mechanism for generalization by function learning (based on restaurant features). People appear to benefit from such generalization, as exploration yields better realized outcomes in cities where features have more predictive

PSYCHOLOGICAL AND COGNITIVE SCIENCES

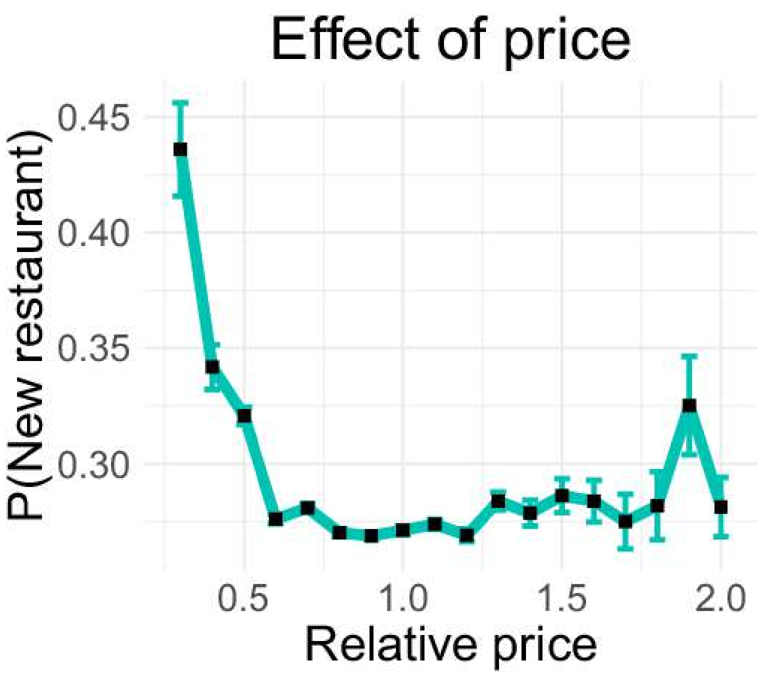
How to model cognition?

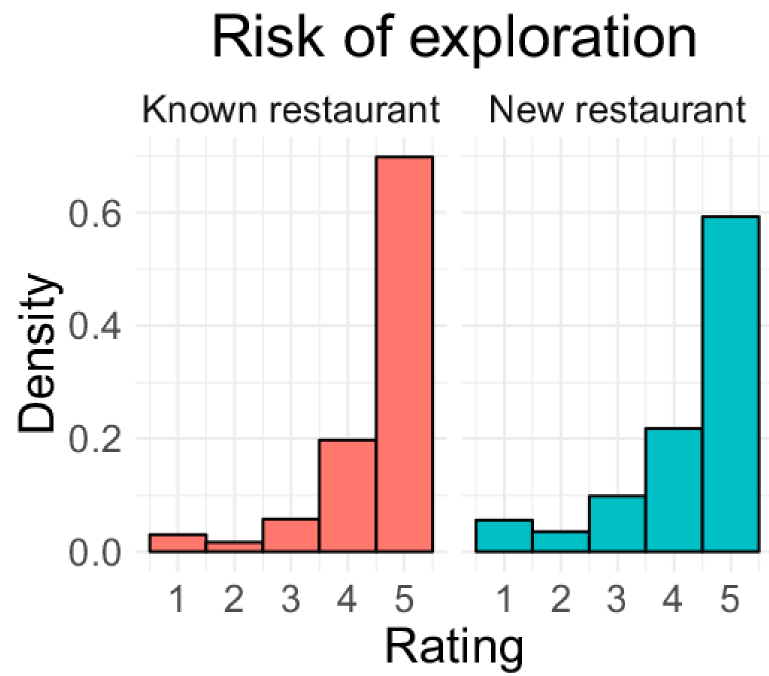
- The Deliveroo dataset: 195,333 customers
  1,613,967 unique orders



How to model cognition?

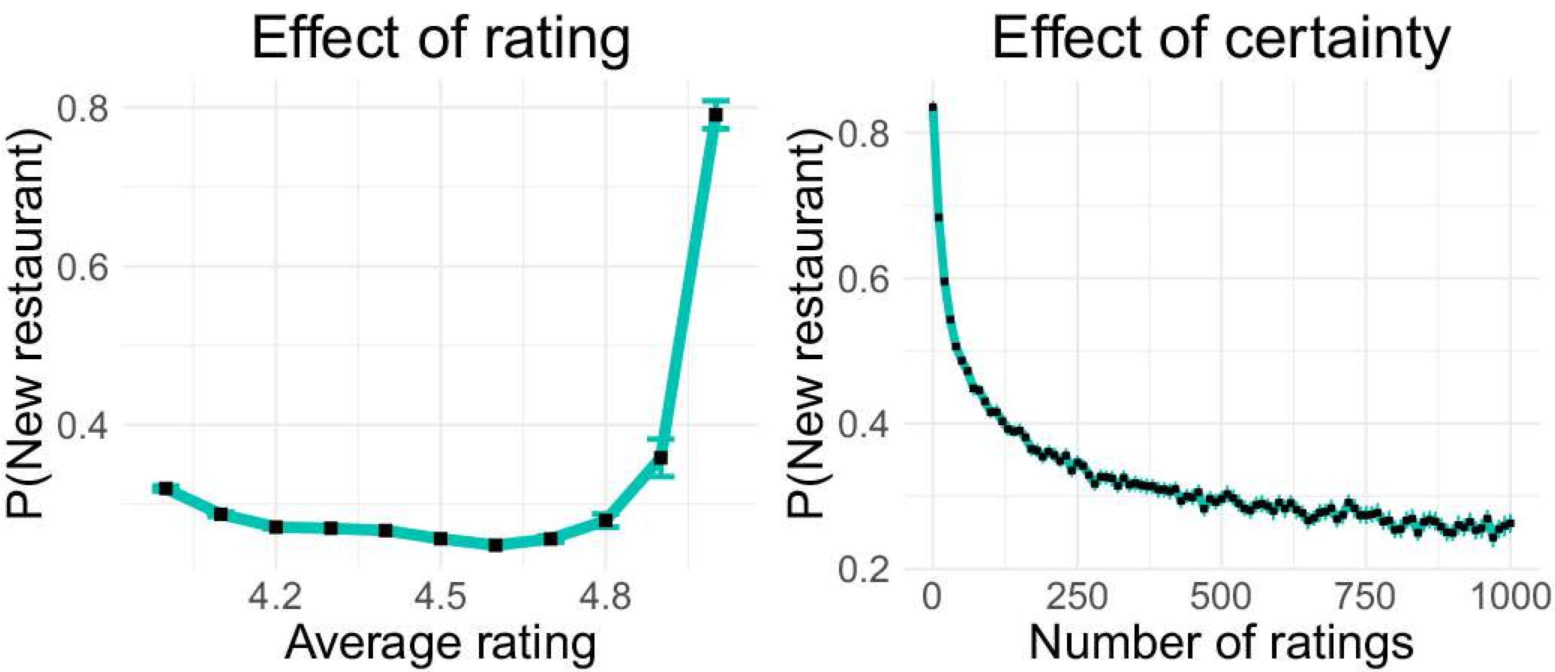
- Which factors influence real-life exploration? Signatures of directed exploration?





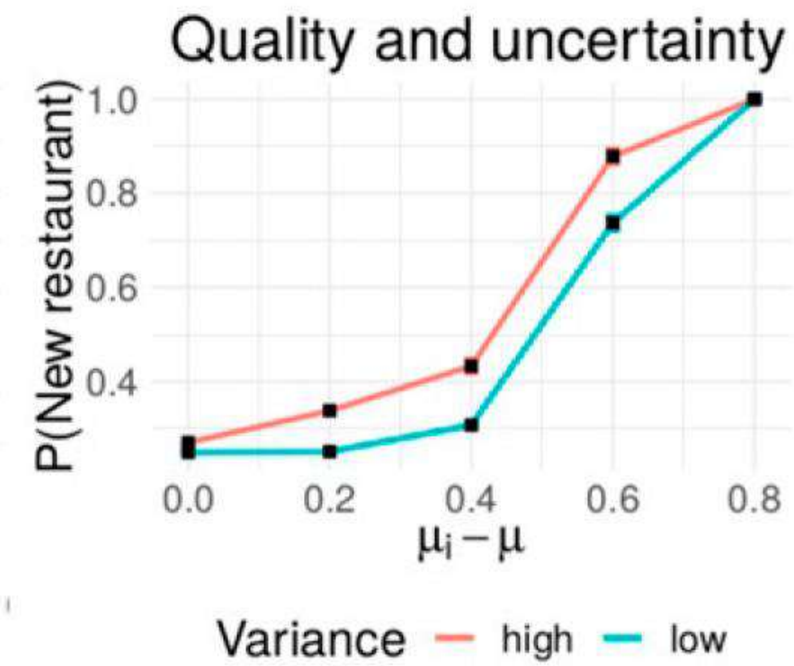
How to model cognition?

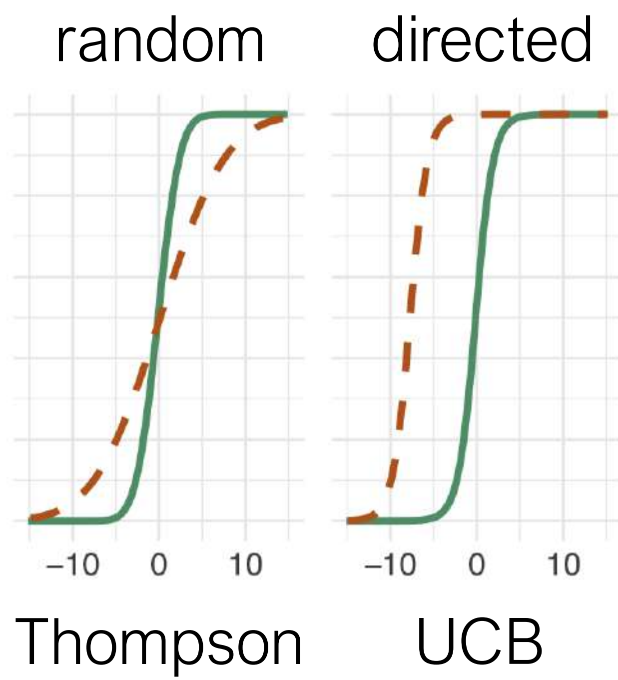
- Which factors influence real-life exploration? Signatures of directed exploration?



How to model cognition?

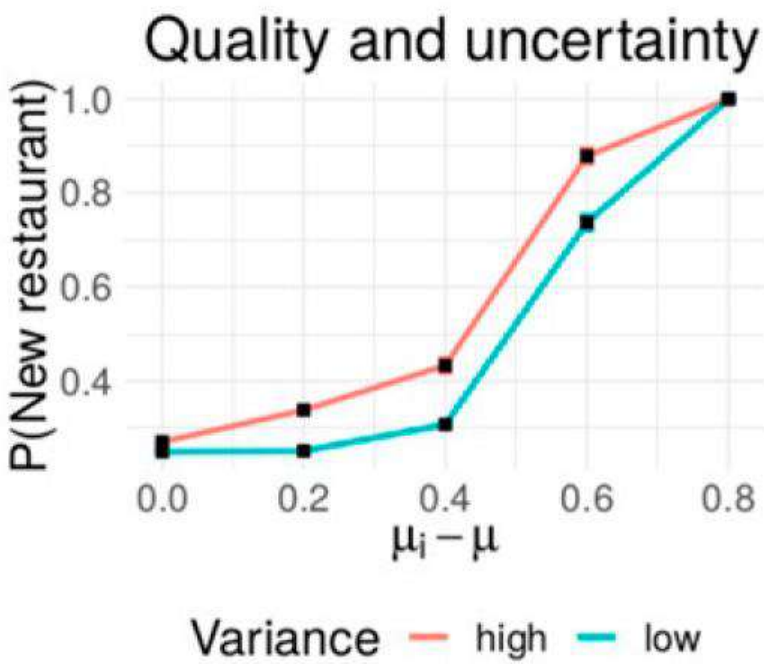
- Which factors influence real-life exploration? Signatures of directed exploration?

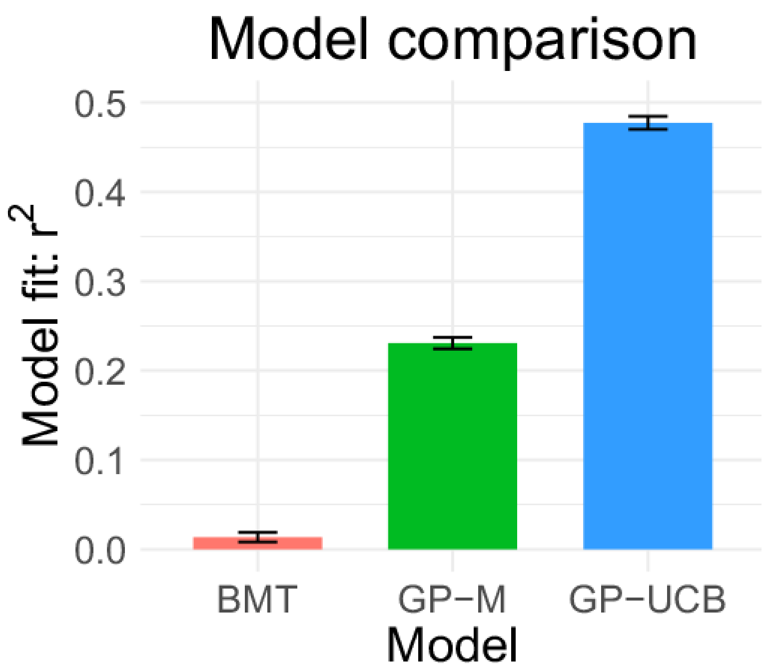




How to model cognition?

- Which factors influence real-life exploration? Signatures of directed exploration?





How to model cognition?

- Which factors influence real-life exploration? Signatures of directed exploration?
- Humans engage in directed exploration and similarity-based generalization when making restaurant choices.
- Hallmarks of sophisticated human strategies for exploring complex, real-world environments!

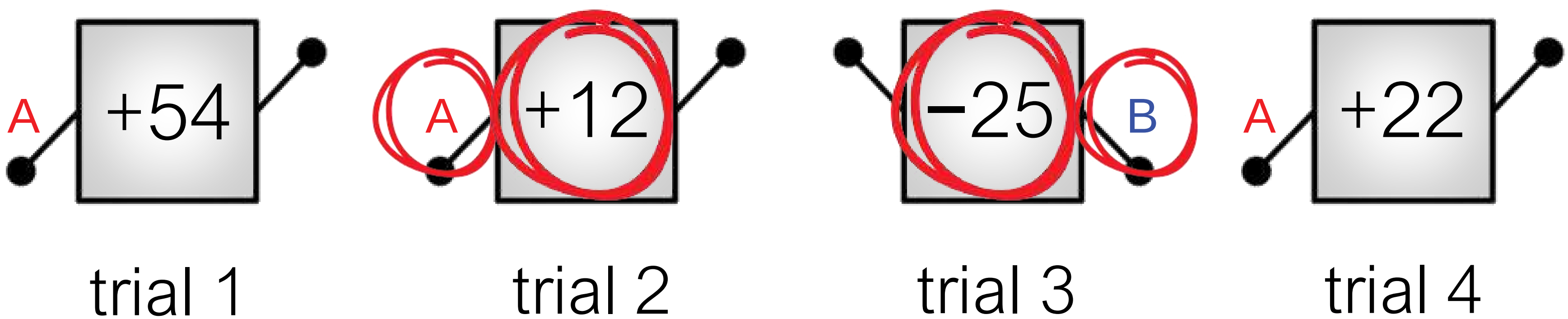
Class evaluation

- Create 5 groups of 5 students each
- In-depth analysis of behavioral data collected in a slot machine game (two-armed bandit)
- Use data mining <u>and</u> modeling approaches seen during lectures <u>and</u> practical sessions
- <u>Objective:</u> identify the latent cognitive strategy that drives behavior (<u>different</u> for each group)
- Group presentation (15 min/group) on Friday

Class evaluation

Class evaluation

two-armed bandit task



Guidelines for modeling

eLIFE
elifesciences.org

REVIEW ARTICLE

Ten simple rules for the computational modeling of behavioral data

Robert C Wilson[1,2†*], Anne GE Collins[3,4†*]

[1]Department of Psychology, University of Arizona, Tucson, United States; [2]Cognitive Science Program, University of Arizona, Tucson, United States; [3]Department of Psychology, University of California, Berkeley, Berkeley, United States; [4]Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, United States

Abstract Computational modeling of behavior has revolutionized psychology and neuroscience. By fitting models to experimental data we can probe the algorithms underlying behavior, find neural correlates of computational variables and better understand the effects of drugs, illness and interventions. But with great power comes great responsibility. Here, we offer ten simple rules to ensure that computational modeling is used with care and yields meaningful insights. In particular, we present a beginner-friendly, pragmatic and details-oriented introduction on how to relate models to data. What, exactly, can a model tell us about the mind? To answer this, we apply our rules to the simplest modeling techniques most accessible to beginning modelers and illustrate them with examples and code available online. However, most rules apply to more advanced

Coming next

- Practical session: today, 2.00pm, same room
- Guidelines for cognitive modeling:
  Wilson and Collins (2019) Ten simple rules for the computational modeling of behavioral data. *eLife*
  https://doi.org/10.7554/eLife.49547 (open-access)
- Contact:

Valentin Wyart valentin.wyart@ens.psl.eu
Lucas Benjamin lucas.benjamin78@gmail.com

Lab. de Neurosciences Cognitives et Computationnelles (LNC^2)
Institut National de la Santé et de la Recherche Médicale
Ecole Normale Supérieure, Université PSL