PSL-week | March 4-8 2024

Lecture 3 (data mining and modeling for behavioral sciences)

The added value of behavioral research (over neuroscience) for understanding the mind

Valentin Wyart

Lab. de Neurosciences Cognitives et Computationnelles (LNC[2])

Institut National de la Santé et de la Recherche Médicale

Ecole Normale Supérieure, Université PSL

valentin.wyart@ens.psl.eu

PSL Data Science Program

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

Paris Artificial Intelligence Research Institute

https://prairie-institute.fr

Group project

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†*]

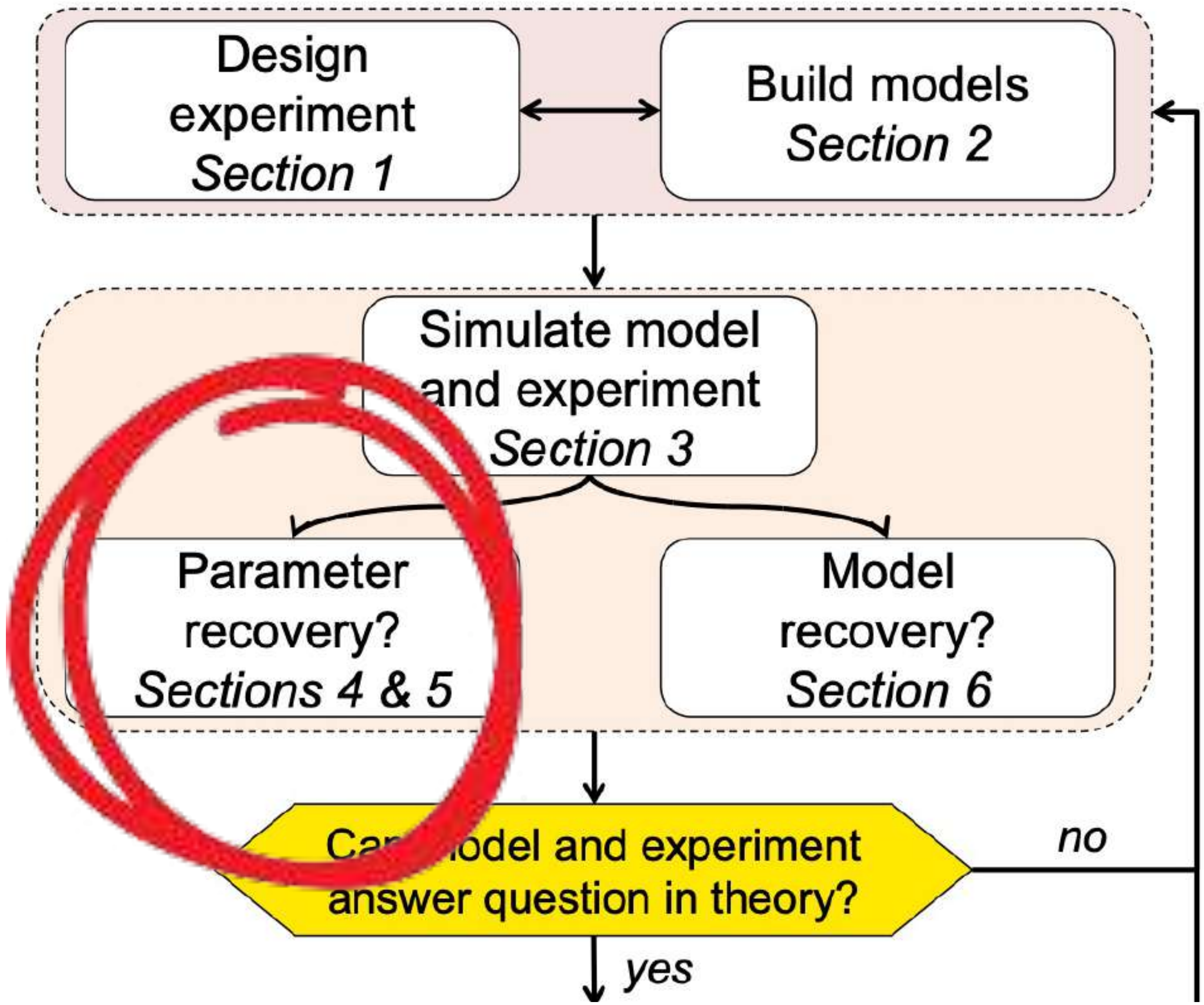
[1]Department of Psychology, University of Arizona, Tucson, United States; [2]Cognitive Science Program, University of Arizona, Tucson, United States; [3]Department of Psychology, University of California, Berkeley, Berkeley, United States; [4]Helen Wills Neuroscience Institute, University of California, 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 ... researchers will avoid many pitfalls and

Group project

- Could you open the behavioral dataset?
- Objective: identify the latent cognitive strategy that drives behavior (different for each group)
- Use data mining and modeling approaches:
  - ✓ describe behavior using data mining
  - ✓ identify strategy using data modeling
- Group presentation (15 min/group) on Friday
- Don't hesitate to ask for help or advice!

Modeling guidelines



Modeling guidelines

- Parameter recovery:
  Before reading too much into fitted parameter values, it is important to check whether the fitting procedure works, by fitting synthetic behavior from a known model whose true parameters are known.

- Model simulation code is needed:

  $>> \text{behavior} = f(\theta, s)$

- Model fitting code is needed as well:

  $>> \hat{\theta}_{\mathrm{MLE}} = \mathrm{argmax}_{\theta} \left(\log\left(p(\text{behavior} | \theta, s) \right) \right)$

Modeling guidelines

- Parameter recovery:

Box 4. Example: parameter recovery in the reinforcement learning model.

We performed parameter recovery with Model 3, the Rescorla Wagner model, on the two-armed bandit task. As before, we set the means of each bandit at $\mu_1 = 0.2$ and $\mu_2 = 0.8$ and the number of trials at $T = 1000$. We then simulated the actions of the model according to ***Equations 3 and 4***, with learning rate, α, and softmax temperature, β, set according to

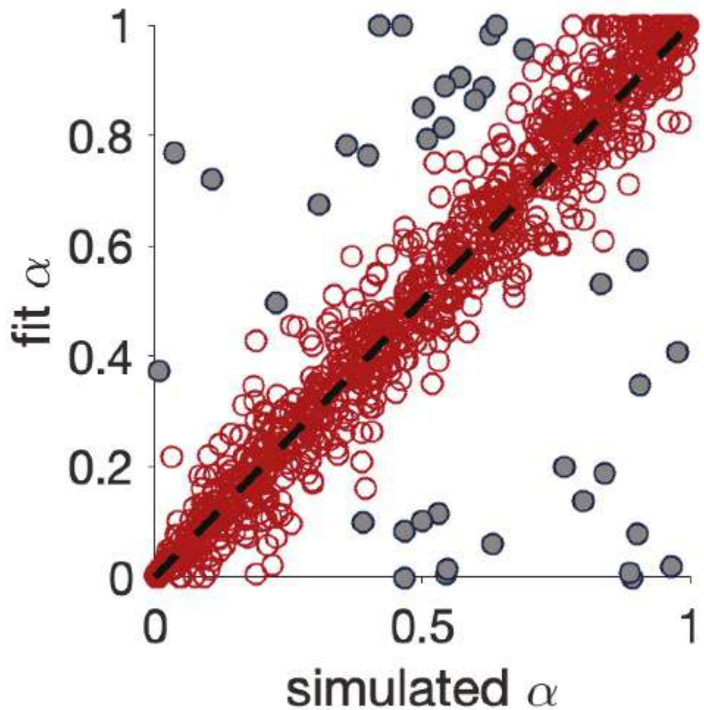
$$\alpha \sim U(0, 1) \quad \text{and} \quad \beta \sim \mathrm{Exp}(10) \tag{9}$$

After simulating the model, we fit the parameters using a maximum likelihood approach to get fit values of learning rate, α, and softmax parameter, β. We then repeated this process 1000 times using new values of α and β each time.

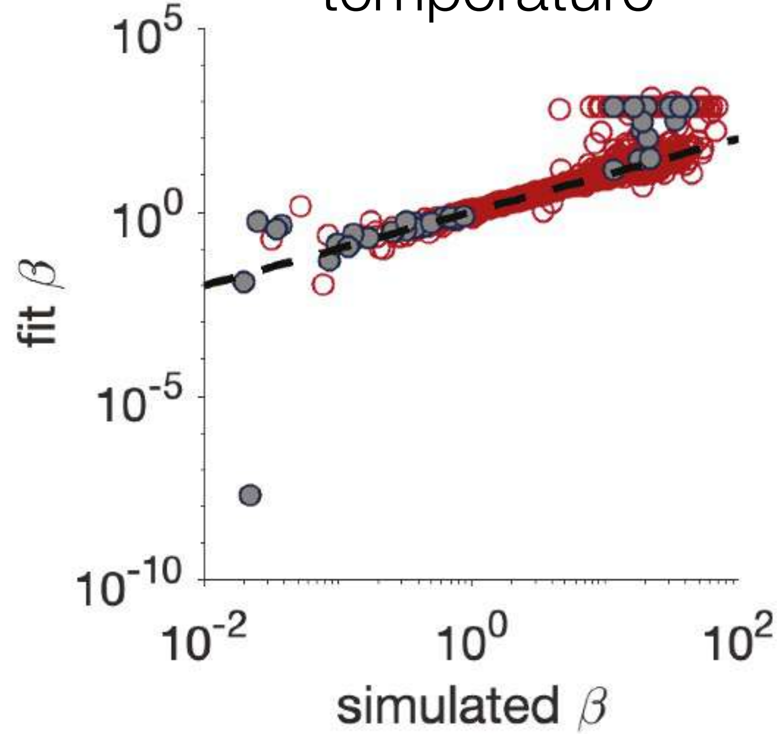
Modeling guidelines

- <u>Parameter recovery:</u>

learning rate



choice inverse temperature



Modeling guidelines

- Parameter recovery:
  output = parameter correlations
- Why is it important that all model parameters affect behavioral predictions?
- Would a model parameter that does not affect behavior in the tested task be recoverable? Why (or why not)?

Modeling guidelines

- Model recovery:
  Before reading too much into model comparison, it is important to check that the comparison procedure works, by comparing models fitted to synthetic behavior whose true model is known.

- Model simulation code is needed:

  $>> \mathrm{behavior} = f(\theta, s)$

- Model fitting code is needed as well:

  $>> \mathrm{MLE_M} = \max_\theta\Big(\log\big(p(\mathrm{behavior}|\theta, s)\big)\Big)$

Modeling guidelines

- Model recovery:

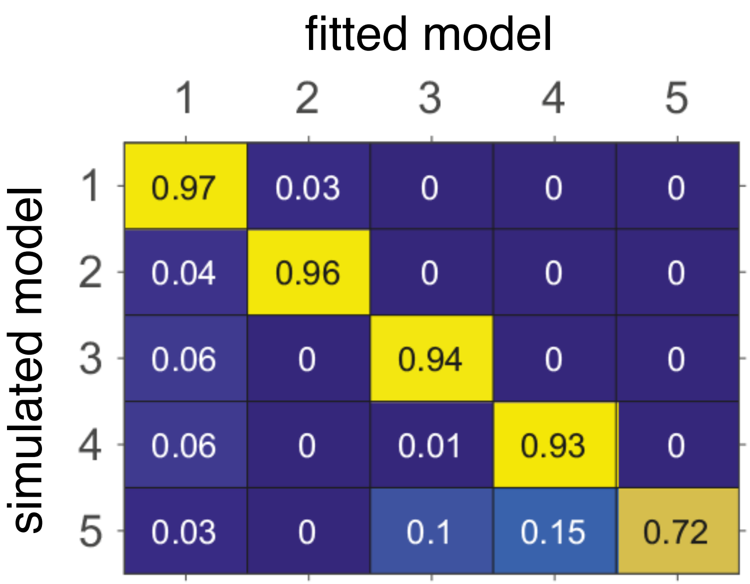
Box 5. Example: confusion matrices in the bandit task.

To illustrate model recovery, we simulated the behavior of the five models on the two-armed bandit task. As before, the means were set at $\mu_1 = 0.2$ and $\mu_2 = 0.8$, and the number of trials was set at $T = 1000$. For each simulation, model parameters were sampled randomly for each model. Each simulated data set was then fit to each of the given models to determine which model fit best (according to BIC). This process was repeated 100 times to compute the confusion matrices which are plotted below

Modeling guidelines

- <u>Model recovery:</u>

$p(\text{fitted} \mid \text{simulated})$



Modeling guidelines

- Model recovery:
  output = model confusion matrix
- Standard confusion matrix = p(fitted | simulated)
  Given behavior from a simulated model, probability of identifying each candidate model as the winning one.
- But what we want is p(simulated | fitted)!
  Given a winning model obtained by fitting, probability of each candidate model to have generated behavior.
- Use Bayes rule: $p(\text{simulated} \mid \text{fitted}) \propto p(\text{fitted} \mid \text{simulated})\, p(\text{simulated})$

Modeling guidelines

- <u>Model recovery:</u>

p(fitted | simulated)

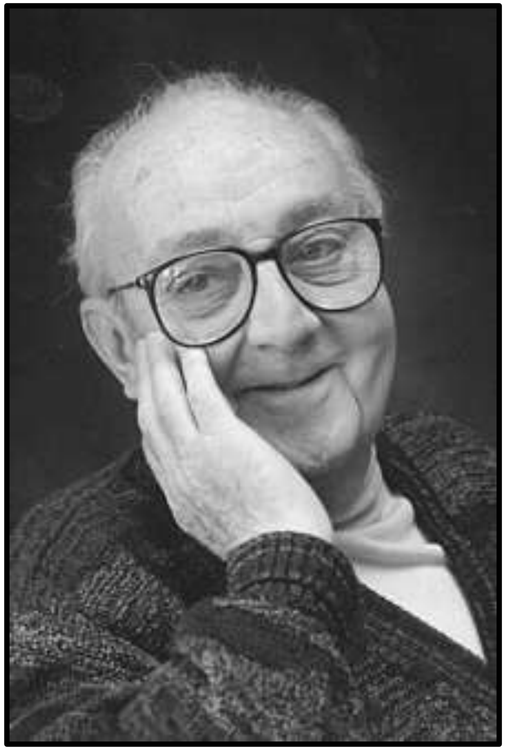
simulated model \ fitted model	1	2	3	4	5
1	0.97	0.03	0	0	0
2	0.04	0.96	0	0	0
3	0.06	0	0.94	0	0
4	0.06	0	0.01	0.93	0
5	0.03	0	0.1	0.15	0.72

p(simulated | fitted)

simulated model \ fitted model	1	2	3	4	5
1	0.84	0.03	0	0	0
2	0.03	0.97	0	0	0
3	0.05	0	0.9	0	0
4	0.05	0	0.01	0.86	0
5	0.03	0	0.1	0.14	1

Modeling guidelines

- *Essentially, all models are wrong, but some are useful.* (George Box, 1987)
- Scientific worries:
    - ✓ parsimony in theory and model building
    - ✓ wrong but preferably not importantly wrong



Modeling guidelines

- *Essentially, all models are wrong, but some are useful.* (George Box, 1987)
- But modeling unimportant model parameters can improve the fitting of important ones!
- Example of choice bias b in TD-based RL:

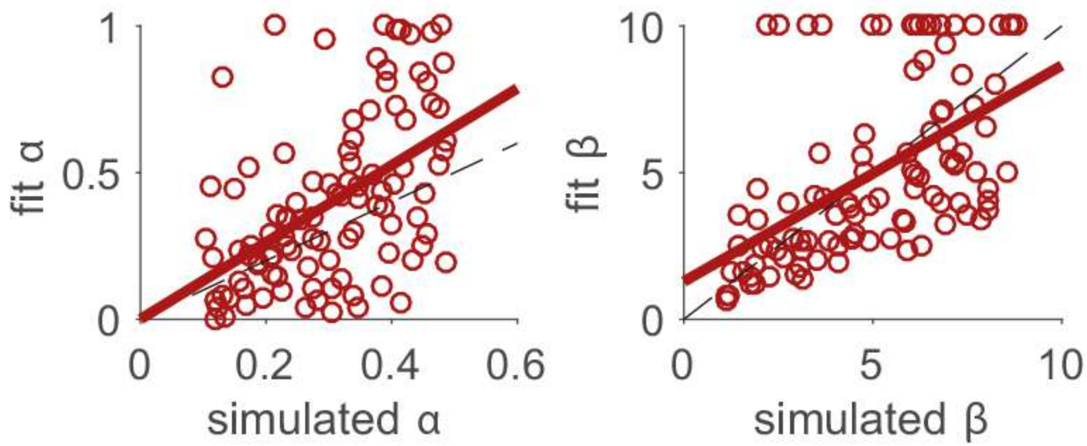
$$Q_{1,t} = Q_{1,t-1} + \alpha\left(r_t - Q_{1,t-1}\right)$$
$$p_t = 1/\left(1 + \exp\left(-\beta\left(Q_{1,t} - Q_{2,t} + b\right)\right)\right)$$

Modeling guidelines

- *Essentially, all models are wrong, but some are useful.* (George Box, 1987)

- <u>Simulated model:</u> M3 with choice bias
  <u>Fitted model:</u> M3 <u>without</u> choice bias

parameter recovery

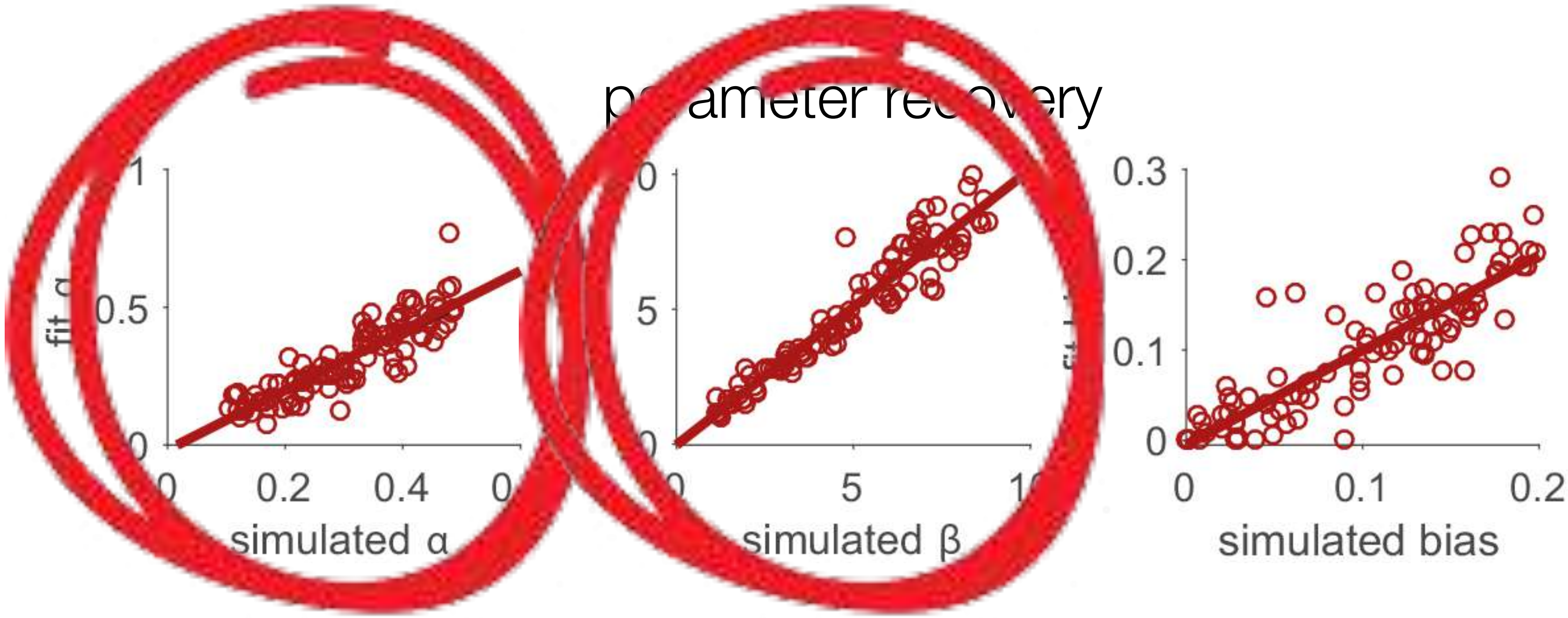


<u>without</u>
choice bias in
fitted model

Modeling guidelines

- *Essentially, all models are wrong, but some are useful.* (George Box, 1987)

- Simulated model: M3 with choice bias
  Fitted model: M3 with choice bias



Modeling guidelines

- *Essentially, all models are wrong, but some are useful.* (George Box, 1987)
- Simulated model: M3 with choice bias
  Fitted model: M3 with choice bias
- What differences between the results of the parameter recovery procedure? Why?
- Do these results conflict with the two worries identified by George Box? Why?

Modeling guidelines

When a good fit can be bad

Mark A. Pitt and In Jae Myung

How should we select among computational models of cognition? Although it is commonplace to measure how well each model fits the data, this is insufficient. Good fits can be misleading because they can result from properties of the model that have nothing to do with it being a close approximation to the cognitive process of interest (e.g. overfitting). Selection methods are introduced that factor in these properties when measuring fit. Their success in outperforming standard goodness-of-fit measures stems from a focus on measuring the generalizability of a model's data-fitting abilities, which should be the goal of model selection.

The explosion of interest in modeling cognitive processes over the past 20 years has fueled the cognitive sciences in many ways. Not only has it opened up new ways of thinking about research problems and possible solutions, but it has also enabled researchers to gain a better understanding of their theories by simulating a computational instantiation of it. Modeling is now sufficiently mainstream that one can get the impression that the

of it. A thorough evaluation of a model requires methods that are sensitive to its quantitative form. Criteria used for evaluating theories [1], such as testing their performance in an experimental setting, do not speak to the quality of the choices that are made in building their quantitative counterparts (i.e. choice of parameters, how they are combined) or their ramifications. The paucity of such model selection methods is surprising given the centrality of the problem itself. What could be more fundamental than deciding between two alternative explanations of a cognitive process?

How *not* to compare models

Mathematical model are frequently tested against one another by evaluating how well each fits the data generated in an experiment or simulation. Such a test makes sense given that one criterion of model performance is that it reproduce the data. A goodness-of-fit measure (GOF; see Glossary) is invariably used to measure their adequacy in achieving this goal. What is measured is how much a model's predictions deviate from the observed data [2,3]. The model that provides the best fit (i.e. smallest deviation) is favored. The logic of this choice rests on the assumption that the model that provides the best fit to all data must be a closer approximation to the cognitive process under investigation than its competitors [4].

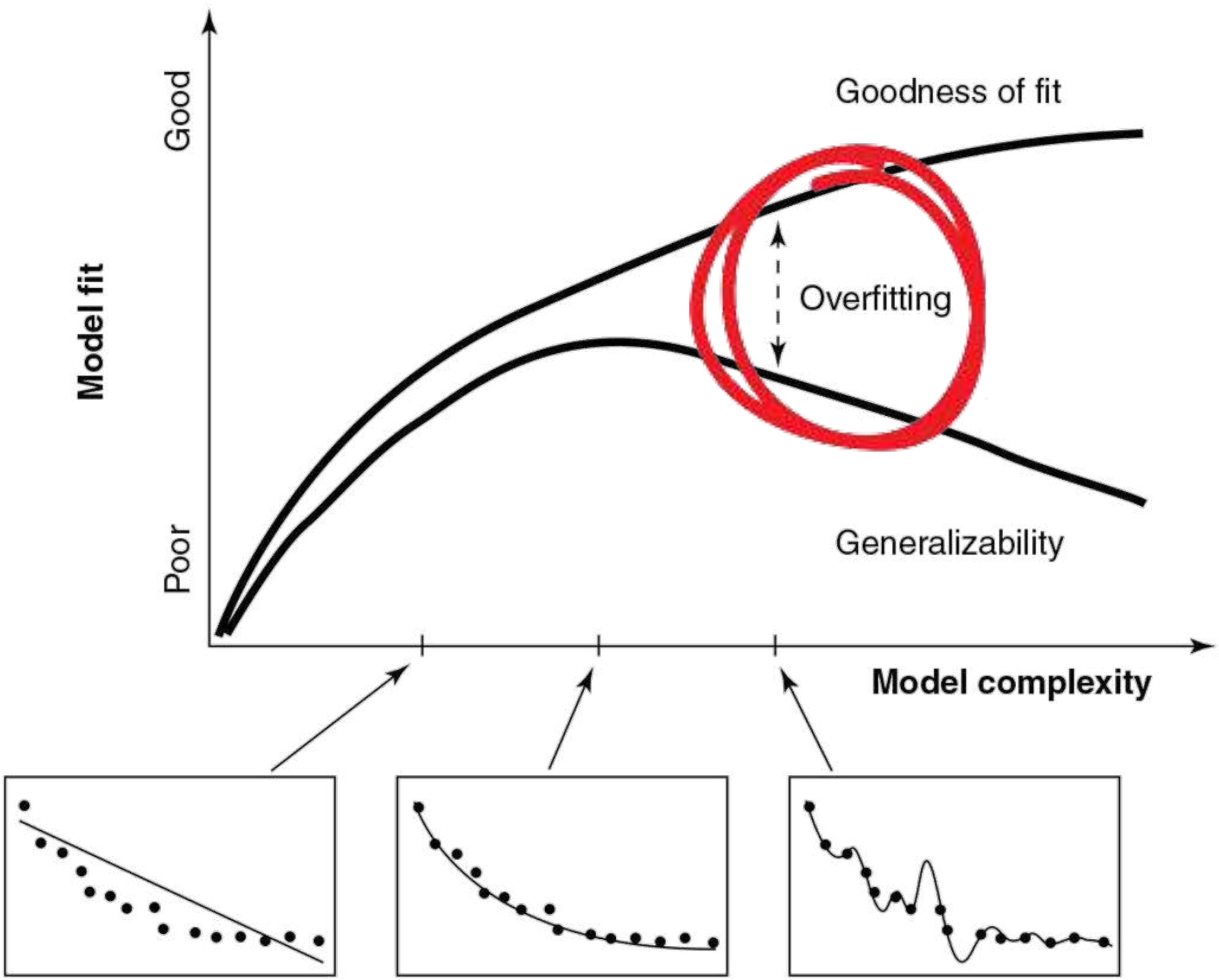
Such a conclusion is reasonable if measurements

Modeling guidelines

- Overfitting issue and Occam's razor
- Law of parsimony: "The simplest explanation is usually the best one."
- Why is this principle important for modeling data?

William of Ockham
(1287–1347)
medieval philosopher

Modeling guidelines



Modeling guidelines

- How to deal in practice with overfitting?
- Idea: use a complexity-penalizing metric of fit
- Which of these metrics penalize complexity?
  - ✓ RMSE (root mean squared error)
  - ✓ PVAF (percent variance accounted for)
  - ✓ AIC (Akaike information criterion)
  - ✓ BIC (Bayesian information criterion)

Modeling guidelines

- How to deal in practice with overfitting?
- Idea: use a complexity-penalizing metric of fit
- Which of these metrics penalize complexity?

Table II. Two GOF Measures, four generalizability measures, and the dimensions of complexity to which each is sensitive

Selection method	Criterion equation	Dimensions of complexity considered
Root Mean Squared Error	$RMSE = (SSE/N)^{1/2}$	None
Percent Variance Accounted For	$PVAF = 100(1 - SSE/SST)$	None
Akaike Information Criterion	$AIC = -2\ ln(f(y\mid\theta_0)) + 2k$	Number of parameters
Bayesian Information Criterion	$BIC = -2\ ln(f(y\mid\theta_0)) + k \cdot ln(n)$	Number of parameters, sample size
Bayesian Model Selection	$BMS = -ln \int f(y\mid\theta)\pi(\theta)d\theta$	Number of parameters, sample size, functional form
Minimum Description Length	$MDL = -ln(f(y\mid\theta_0)) + (k/2)ln(n/2\pi) + ln \int \sqrt{\det(I(\theta))}d\theta$	Number of parameters, sample size, functional form

In the equations above, y denotes observed data, θ is the model's parameter, θ_0 is the parameter value that maximizes the likelihood function $f(y\mid\theta)$, k is the number of parameters, n is the sample size, N is the number of data points fitted, SSE is the minimized sum of the squared errors between observations and predictions, SST is the sum of the squares total, $\pi(\theta)$ is the parameter prior density, $I(\theta)$ is the Fisher information matrix in mathematical statistics [a], *det* denotes the determinant of a matrix, and *ln* denotes the natural logarithm of base e.

Modeling guidelines

- How to deal in practice with overfitting?
- Example:
  - ✓ $M_A : y = (1 + x)^{-a}$
  - ✓ $M_B : y = (b + c \cdot x)^{-a}$

Table I. Results of a model recovery simulation in which a GOF measure (RMSE) was used to discriminate models when the source of the error was varied.

Condition (sources of variation)	Model the data were generated from			Model fitted	
	M_A a = 0.4	M_A a = 0.6	M_B	M_A	M_B
(1) Sampling error	100	–	–	0.040 (0%)	0.029 (100%)
(2) Sampling error + individual differences	50	50	–	0.041 (0%)	0.029 (100%)
(3) Different models	–	50	50	0.075 (0%)	0.029 (100%)
(4) Sampling error	–	–	100	0.079 (0%)	0.029 (100%)

Modeling guidelines

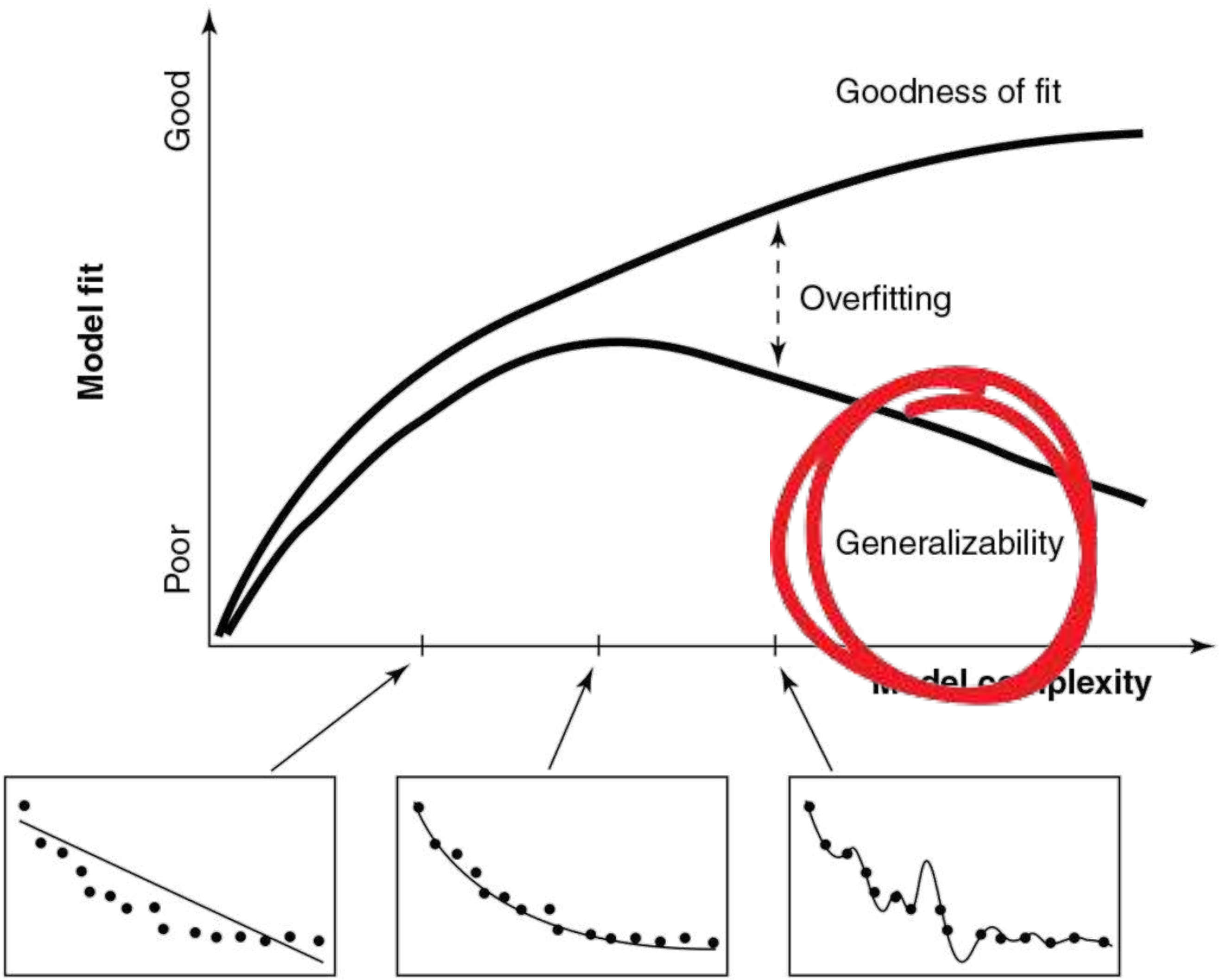
- How to deal in practice with overfitting?
- Other example:
  - ✓ $M_1 : y = (1 + x)^{-a}$
  - ✓ $M_2 : y = (b + x)^{-a}$
  - ✓ $M_3 : y = (1 + c \cdot x)^{-a}$

Modeling guidelines

- How to deal in practice with overfitting?

Selection method	Model fitted	Model the data were generated from		
		M_1	M_2	M_3
PVAF	M_1	0	0	0
	M_2	38	97	30
	M_3	62	3	70
AIC	M_1	79	0	0
	M_2	9	97	30
	M_3	12	3	70
MDL	M_1	86	0	0
	M_2	1	92	8
	M_3	13	8	92

Modeling guidelines



Modeling guidelines

- How to deal in practice with overfitting?
- Other idea: use a cross-validation approach
- General procedure:
  - ✓ Fit model on training set
  - ✓ Compute metric of fit on separate test set
- Why does it overcome overfitting?
- Why is it less arbitrary than using a complexity-penalized metric of fit?

Paper to read

AMERICAN PSYCHOLOGICAL ASSOCIATION

Behavioral Neuroscience

2021, Vol. 135, No. 5, 601–609
https://doi.org/10.1037/bne0000471

ISSN: 0735-7044

COMMENTARY

The Primacy of Behavioral Research for Understanding the Brain

Yael Niv
Department of Psychology and Neuroscience Institute, Princeton University

Understanding the brain requires us to answer both *what* the brain does, and *how* it does it. Using a series of examples, I make the case that behavior is often more useful than neuroscientific measurements for answering the first question.[1] Moreover, I show that even for "how" questions that pertain to neural mechanism, a well-crafted behavioral paradigm can offer deeper insight and stronger constraints on computational and mechanistic models than do many highly challenging (and very expensive) neural studies. I conclude that purely behavioral research is essential for understanding the brain—especially its cognitive functions—contrary to the opinion of prominent funding bodies and some scientific journals, who erroneously place neural data on a pedestal and consider behavior to be subsidiary.

Keywords: behavioral experiments, cognition, neuroscience, priorities

In an era of increasingly precise methods for measuring and perturbing neurons in the brain, it often seems that with more neural data, we will soon make untold breakthroughs in understanding the brain. Such anticipation has heralded neuroscience-data-focused projects such as the Brain Initiative and the Human Connectome

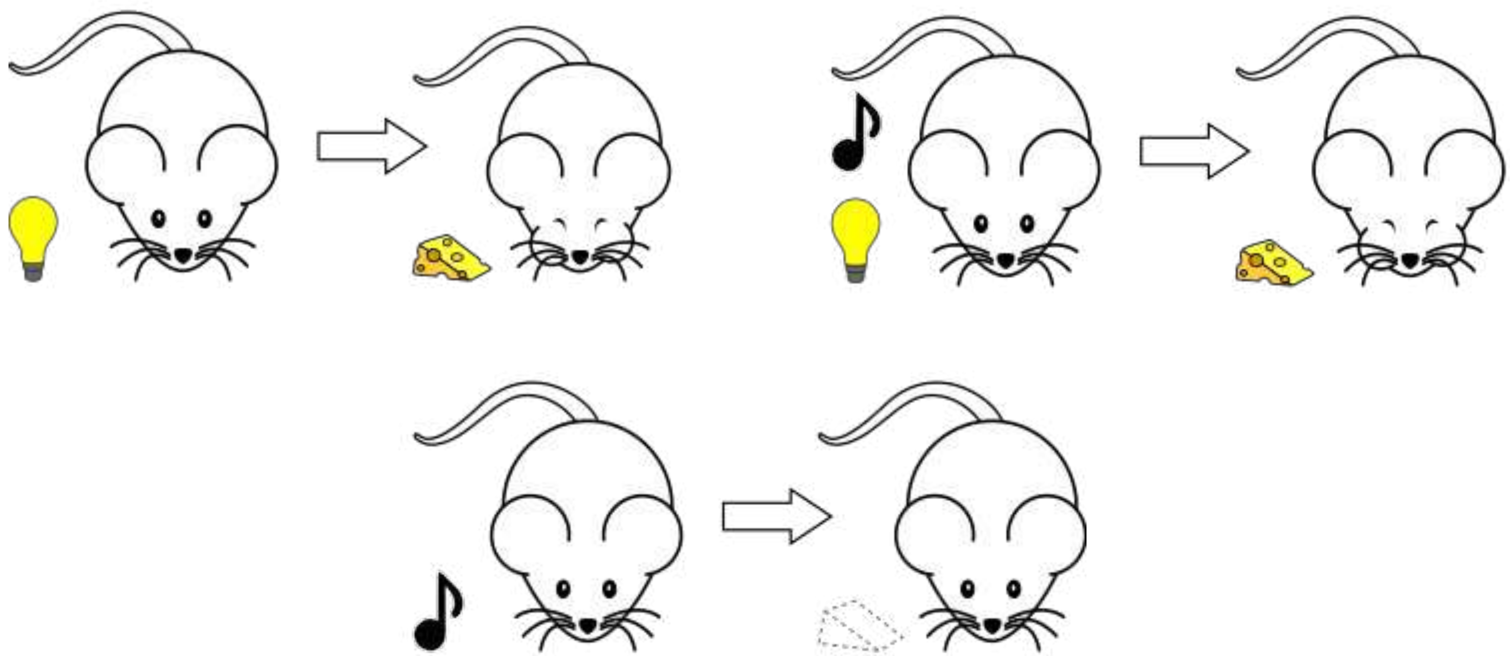
looking at single neurons, or even their ensembles, is like attempting to understand why people in Australia drive on the left side of the road from examination of their DNA. Neural firing patterns are the wrong level for investigating many pressing questions in neuroscience. Even if we could measure all the neurons in the brain with arbitrary precision, without an incisive behavioral paradigm we

Paper to read

- What are the main messages?
- Behavioral research better than neuroscience for studying what the mind does
  how the mind works
- Yael Niv: “a <u>well-crafted</u> behavioral experiment offers deeper insight and stronger constraints on cognitive models than very challenging and expensive neuroscience studies”

Paper to read

- Example: TD-based learning rule
- Blocking effect in classical conditioning
  Kamin (1968)



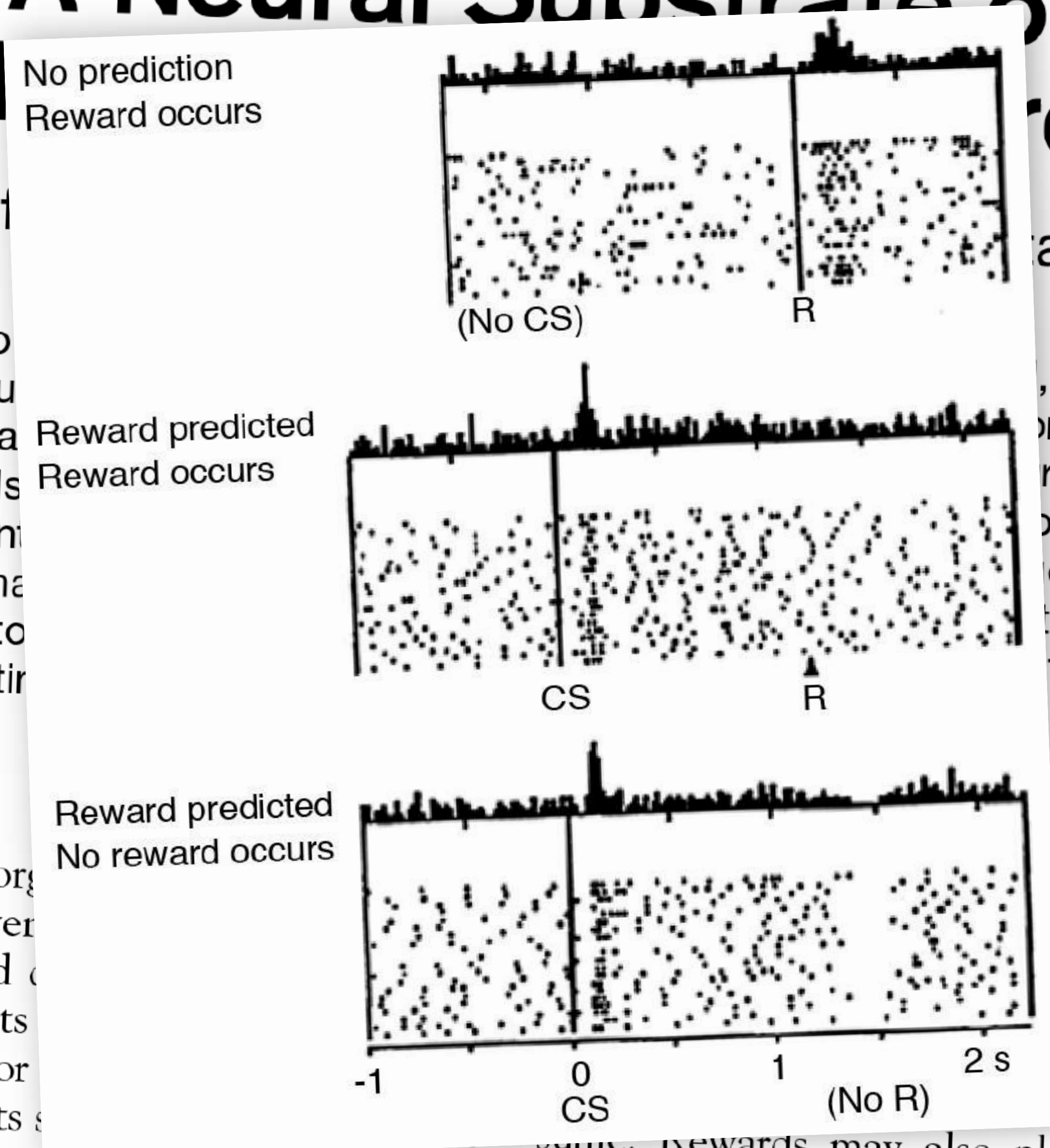
Paper to read

- Example: TD-based learning rule
- Blocking effect in classical conditioning
  Kamin (1968)
- Learning rule derived from behavioral experiments
  Rescorla and Wagner (1972)
  Barto et al. (1980), Sutton (1988), Sutton and Barto (1990)
- Neural recordings much later
  Montague et al. (1995, 1996), Schultz et al. (1997)

A Neural Substrate of
... rd

Wolf ... ague*

The capacity to ... the causal stru ... , and manipulate
suggest that lea ... ral experiments
such as rewards ... re salient events
studies by iden ... plemented these
apparently signa ... ctuating output
events. Taken to ... and rewarding
of adaptive optir ... titative theories



An adaptive or ... The function
predict future even ... ccording to the
mates, food, and ... ample, appeti-
the features of its ... duce approach
the time scales for ... nimal to con-
to be useful for its ... sume. Rewards may also play the role of
an animal time to prepare behavioral reac- positive reinforcers where they increase the
tions and can be used to ...

A Neural Substrate of Prediction and Reward

Wolfram Schultz, Peter Dayan, P. Read Montague*

The capacity to predict future events permits a creature to detect, model, and manipulate the causal structure of its interactions with its environment. Behavioral experiments suggest that learning is driven by changes in the expectations about future salient events such as rewards and punishments. Physiological work has recently complemented these studies by identifying dopaminergic neurons in the primate whose fluctuating output apparently signals changes or errors in the predictions of future salient and rewarding events. Taken together, these findings can be understood through quantitative theories of adaptive optimizing control.

An adaptive organism must be able to predict future events such as the presence of mates, food, and danger. For any creature, the features of its niche strongly constrain the time scales for prediction that are likely to be useful for its survival. Predictions give an animal time to prepare behavioral reactions and can be used to improve the choices

or an internal physical state. The function of reward can be described according to the behavior elicited (2). For example, appetitive or rewarding stimuli induce approach behavior that permits an animal to consume. Rewards may also play the role of positive reinforcers where they increase the

Paper to read

- Interesting/important claim from Niv:
  clever behavioral experiments afford causal conclusions despite correlative measures.
- Inferring causality from converging correlational measurements is not a ultimate sin.
- Isn't it how perception works?
  = inferring the cause of sensations from noisy
  and/or ambiguous sensory signals

Paper to read

- Why is neuroscience valued more than purely behavioral research?

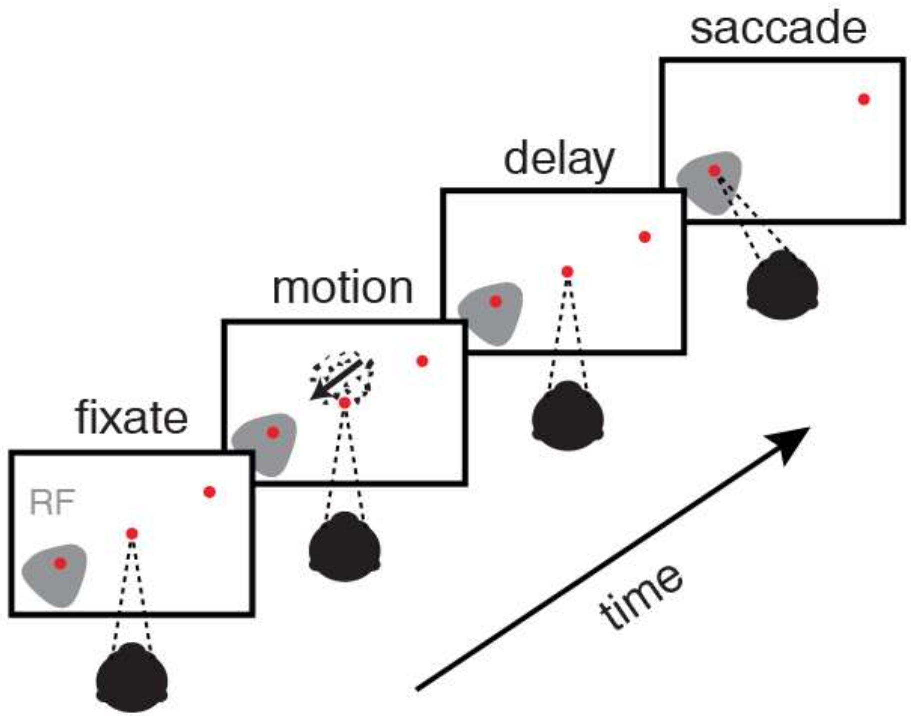
- Economic biases:
  1/ People value an expensive good more than a cheap one.
  2/ People prefer a reward earned through more effort to one obtained more easily.

- Biases opposite to parsimony and rationality.

Paper to read

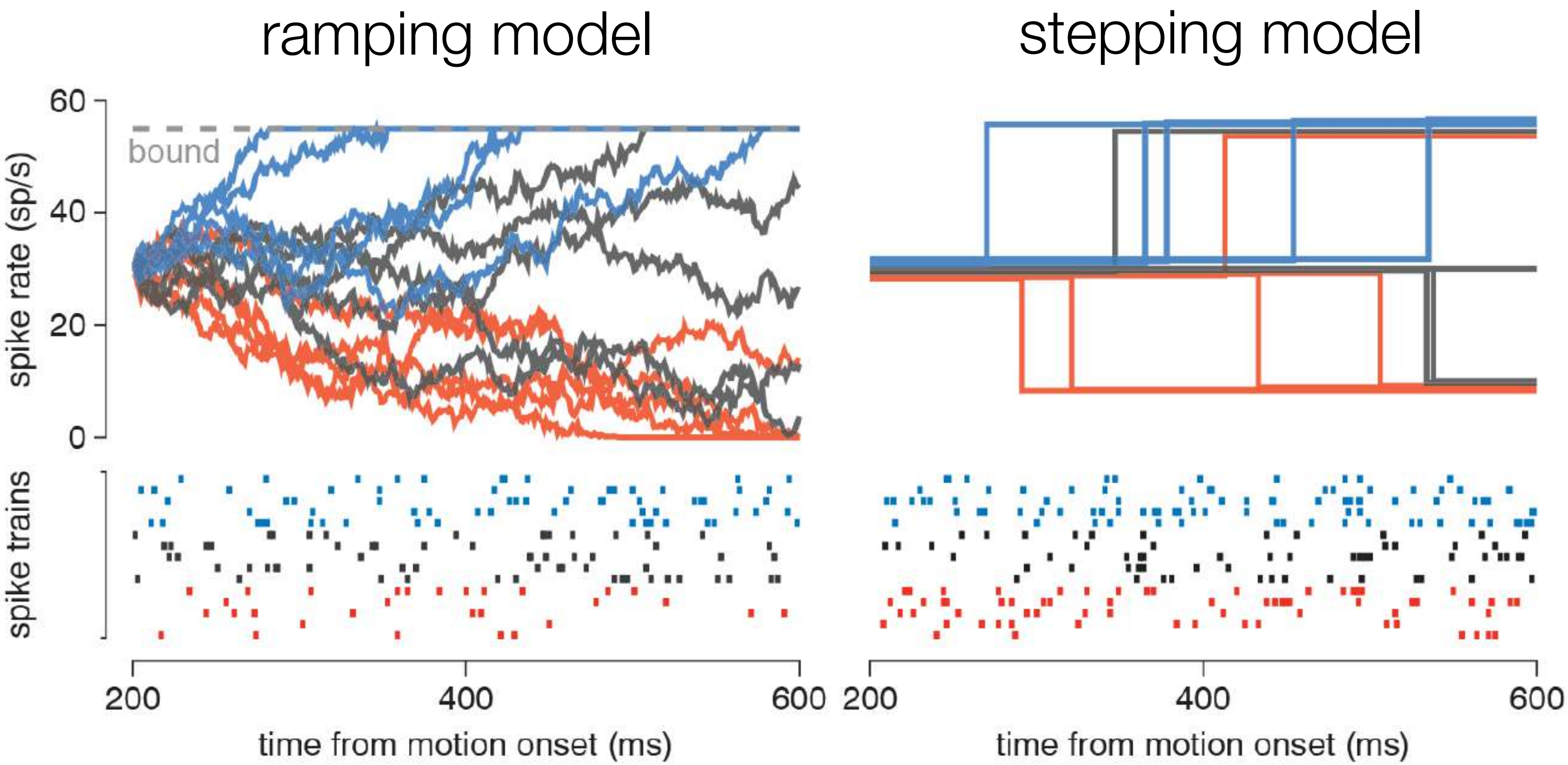
- Why is neuroscience valued more than purely behavioral research?
- Epistemological illusion:
  Neuroscience is seen as "*constraining theories of computation and representation*", but in practice neural data are weakly constraining!
- Example: ramping or stepping LIP neurons?

Neural modeling



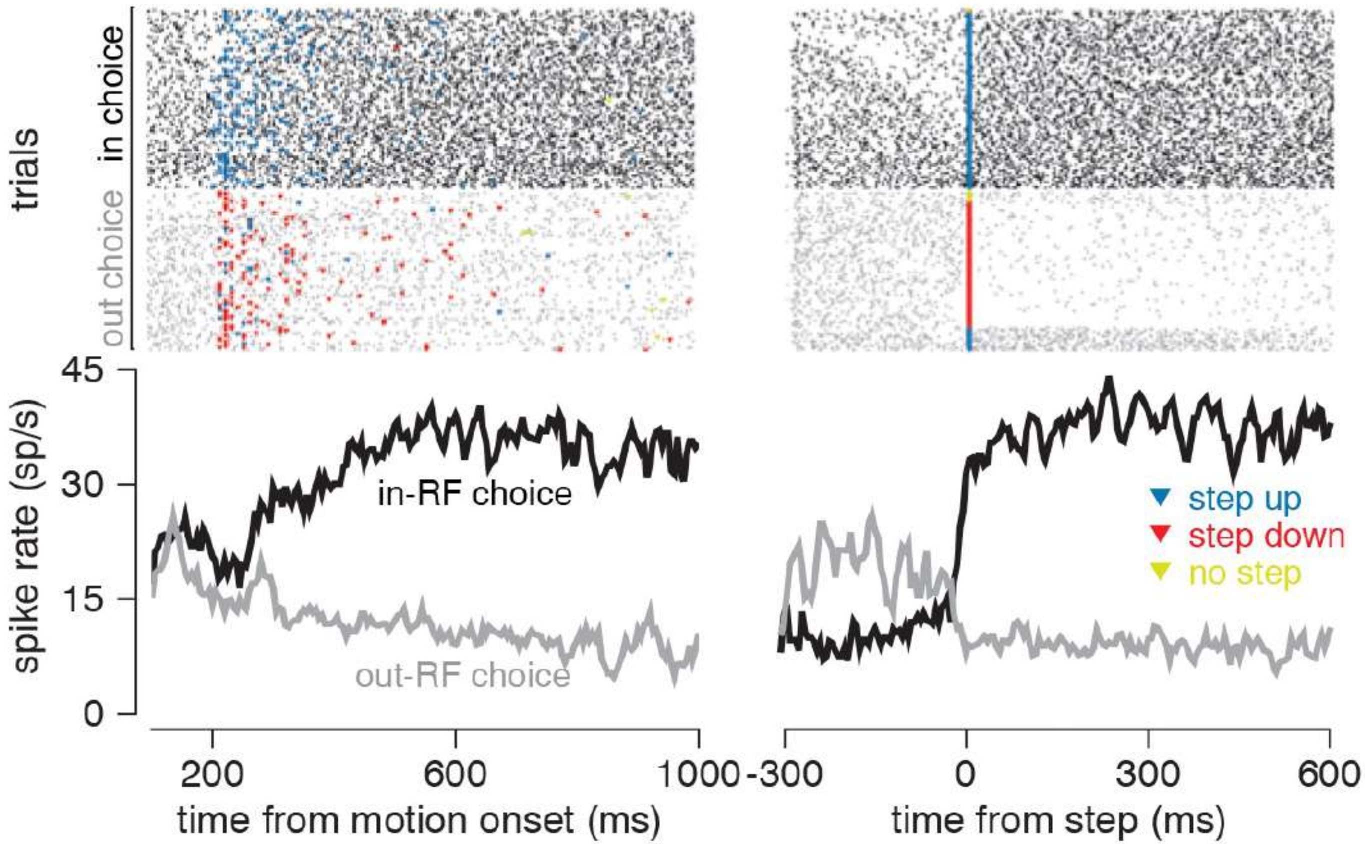
Latimer et al. (2015) Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science*

Neural modeling



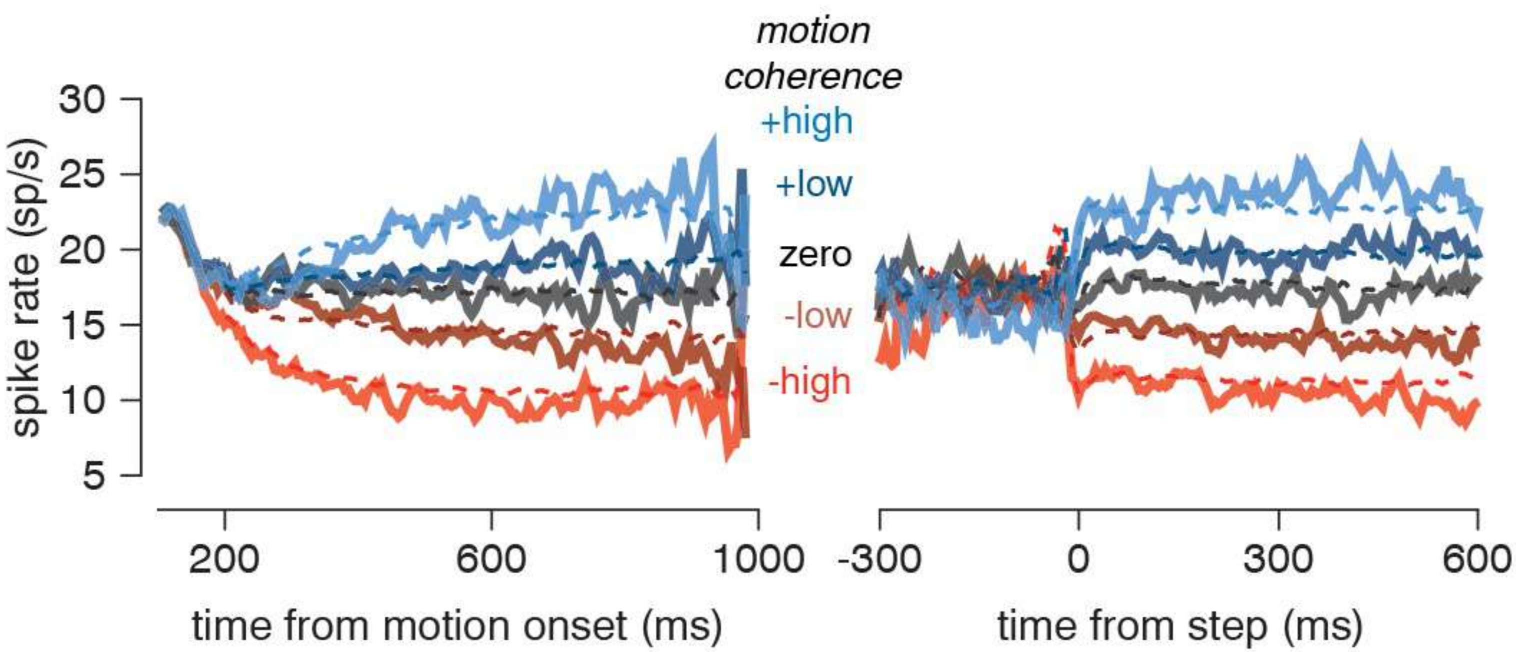
Latimer et al. (2015) Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science*

Neural modeling



Latimer et al. (2015) Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science*

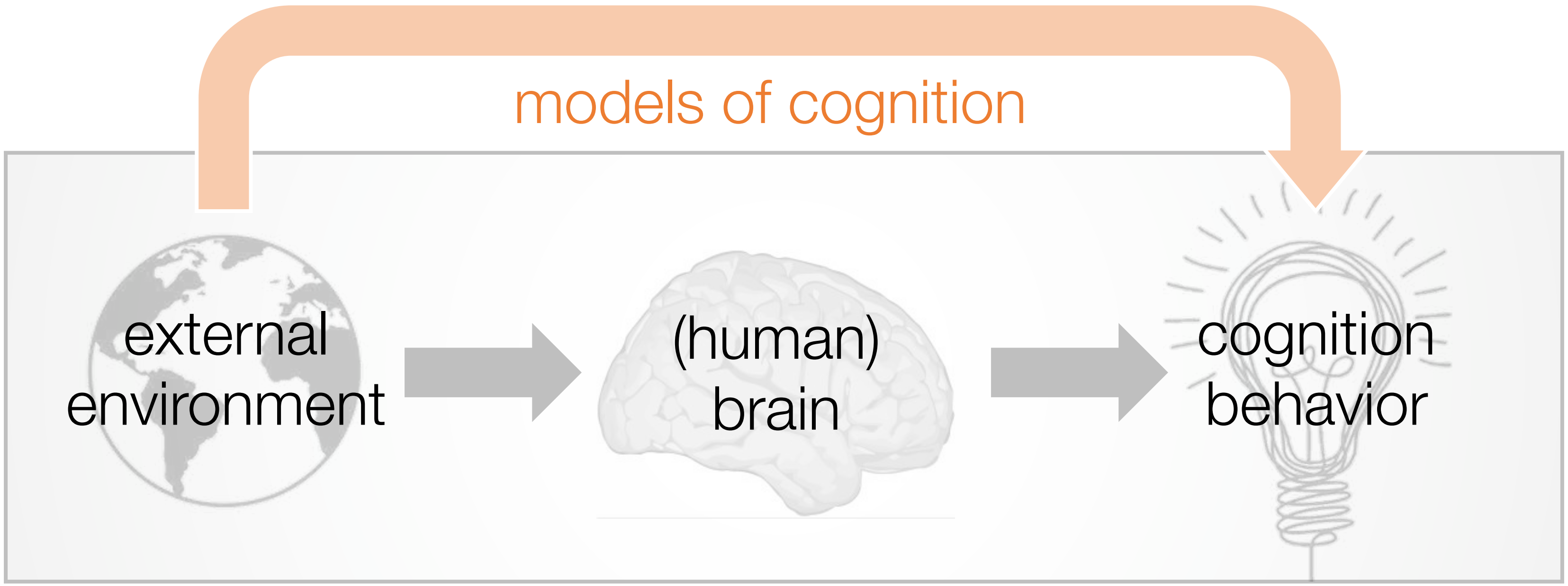
Neural modeling



Latimer et al. (2015) Single-trial spike trains in parietal cortex reveal discrete steps during decision-making. *Science*

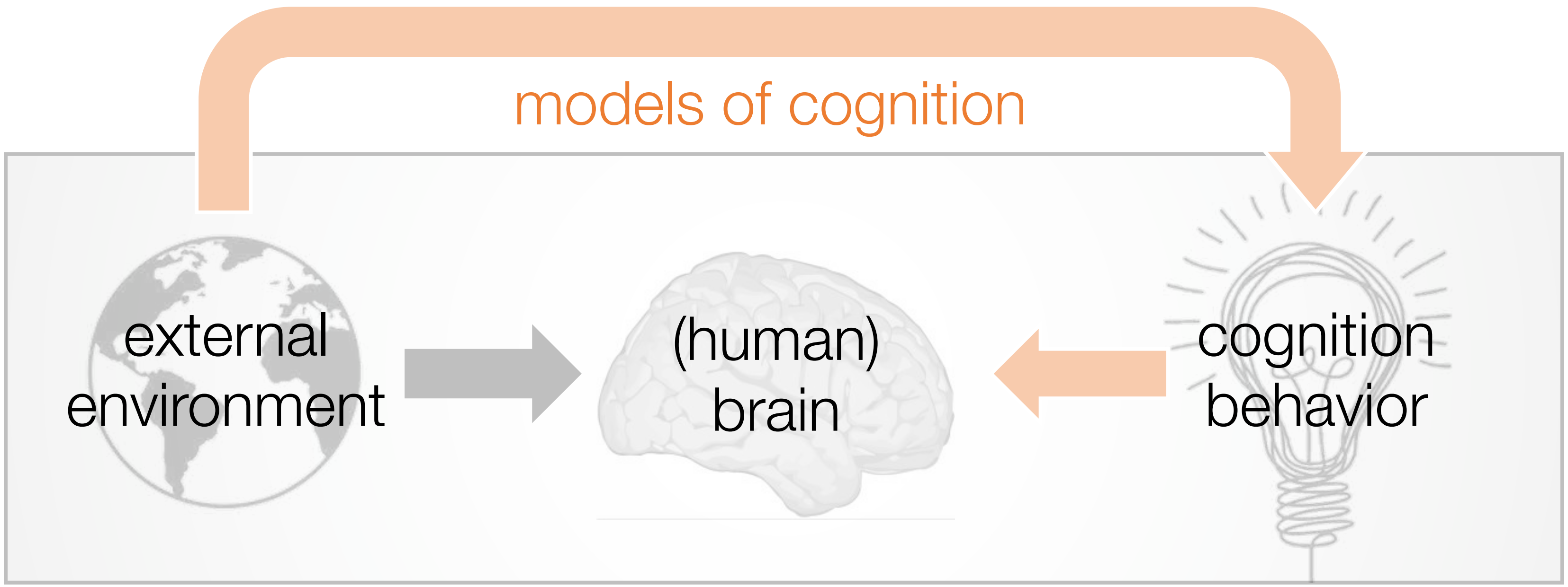
Neural modeling

- Even if the brain mediates the relation between the world and behavior, models of cognition are <u>needed</u> to guide neuroscience research.



Beyond the paper

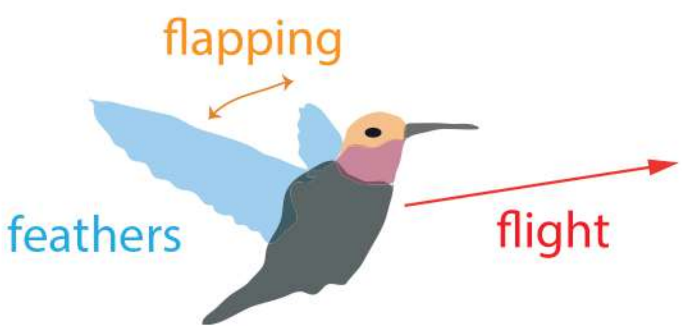
- Even if the brain mediates the relation between the world and behavior, models of cognition are needed to guide neuroscience research.

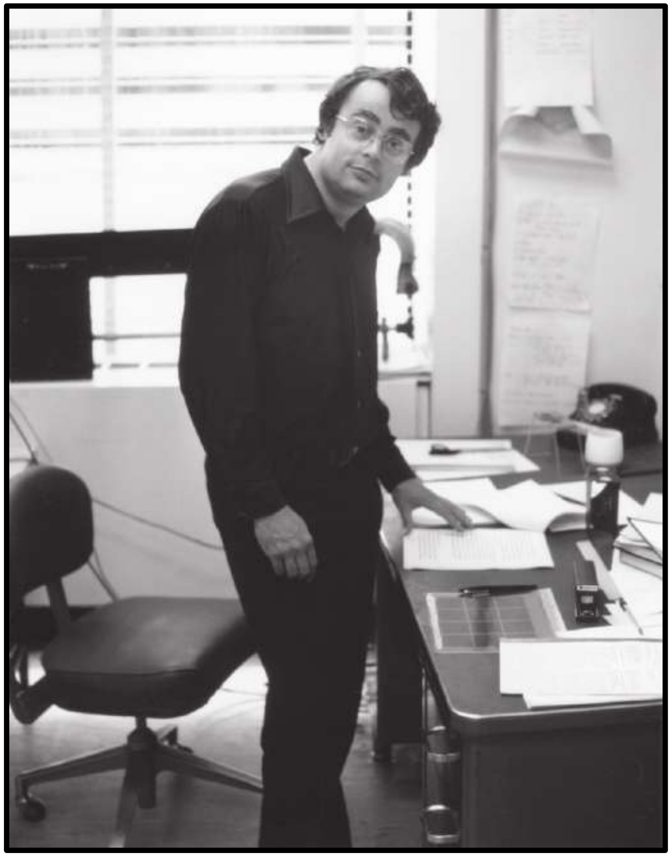


Beyond the paper

David Marr (again)

	LEVELS	
Computation	1	why (problem)
Algorithm	2	what (rules)
Implementation	3	how (physical)

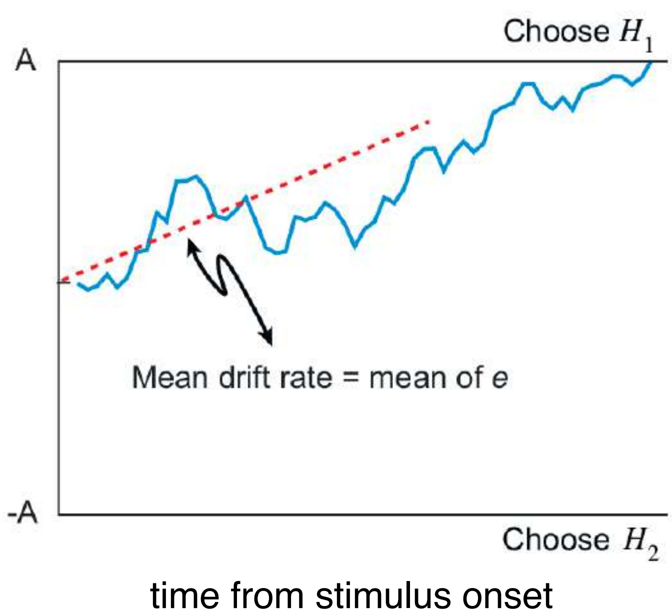




Beyond the paper

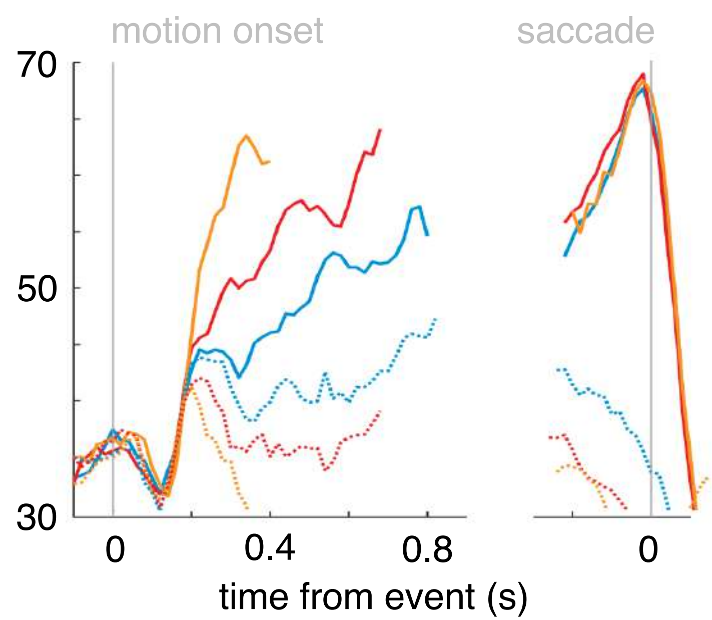
drift-diffusion model
(algorithm)

2



LIP spiking activity
(implementation)

3



Beyond the paper

If you are a neuroscientist, then you want 3 but you need to know 2 first.

If you are a cognitive scientist, then you want 2 but you can use 3 to validate 2.

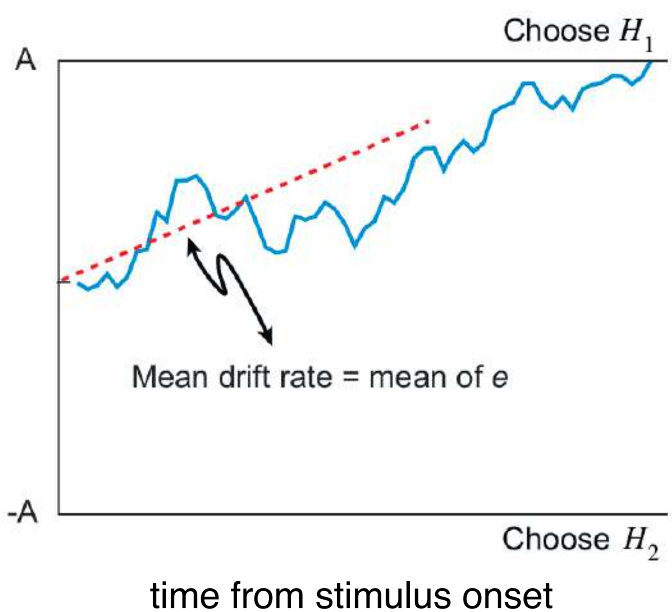
In both cases, studying which **algorithm** is used is necessary to look for its **implementation**.

Krakauer et al. (2016) Neuroscience needs behavior: correcting a reductionist bias. *Neuron*

Beyond the paper

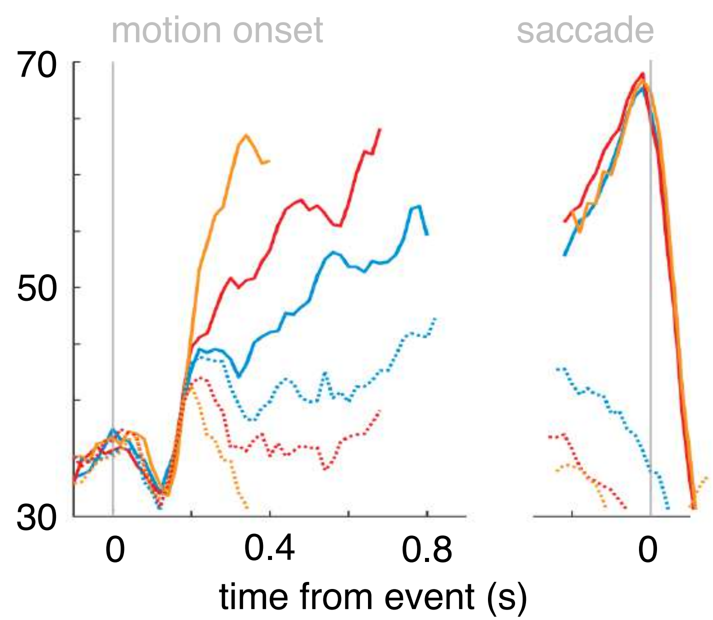
algorithm

Ratcliff (1978)
Psychological Review



implementation

Roitman and Shadlen (2002)
Journal of Neuroscience



Beyond the paper

- Translation across species using modeling
- Why translation across species?

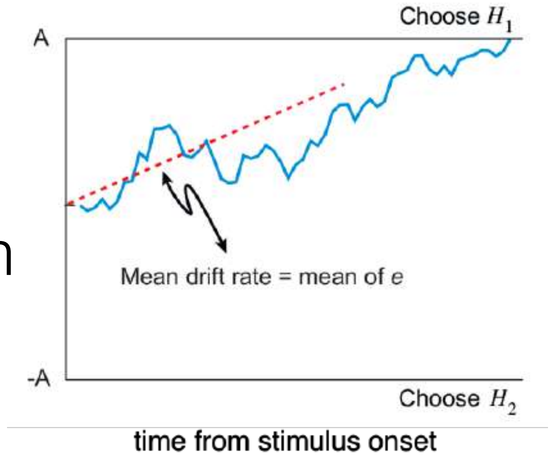
1/ Non-human animals are used to provide *models* of human abilities, diseases, etc.

2/ Animal models are often limited in terms of possible behavioral experiments…

3/ but can afford unique genetic manipulations!

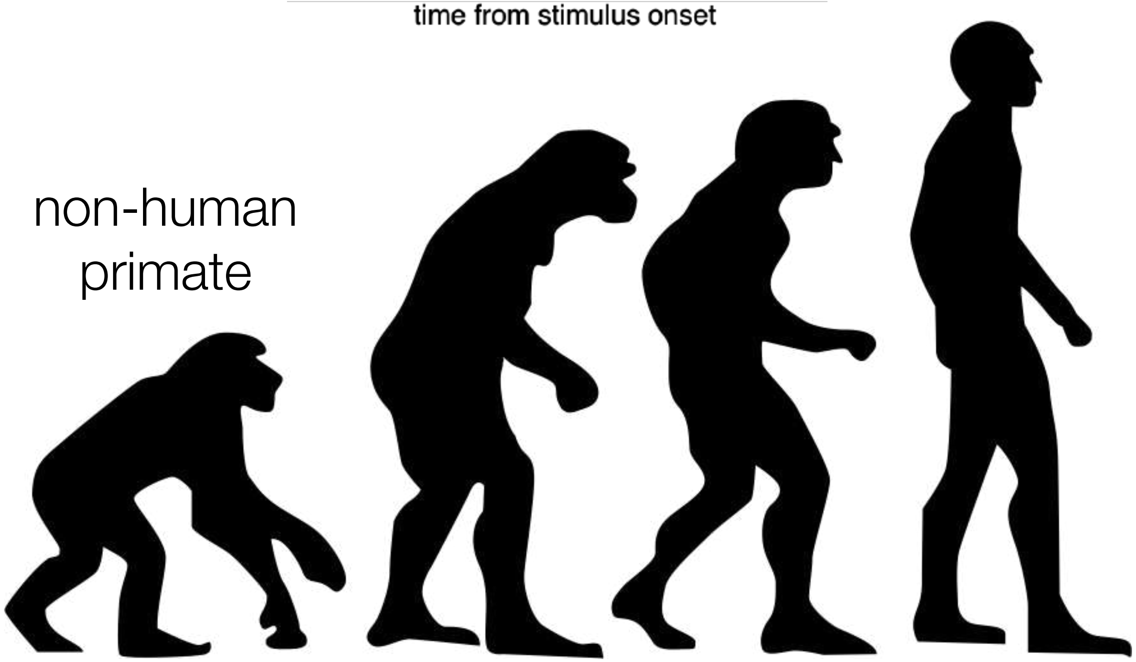
2

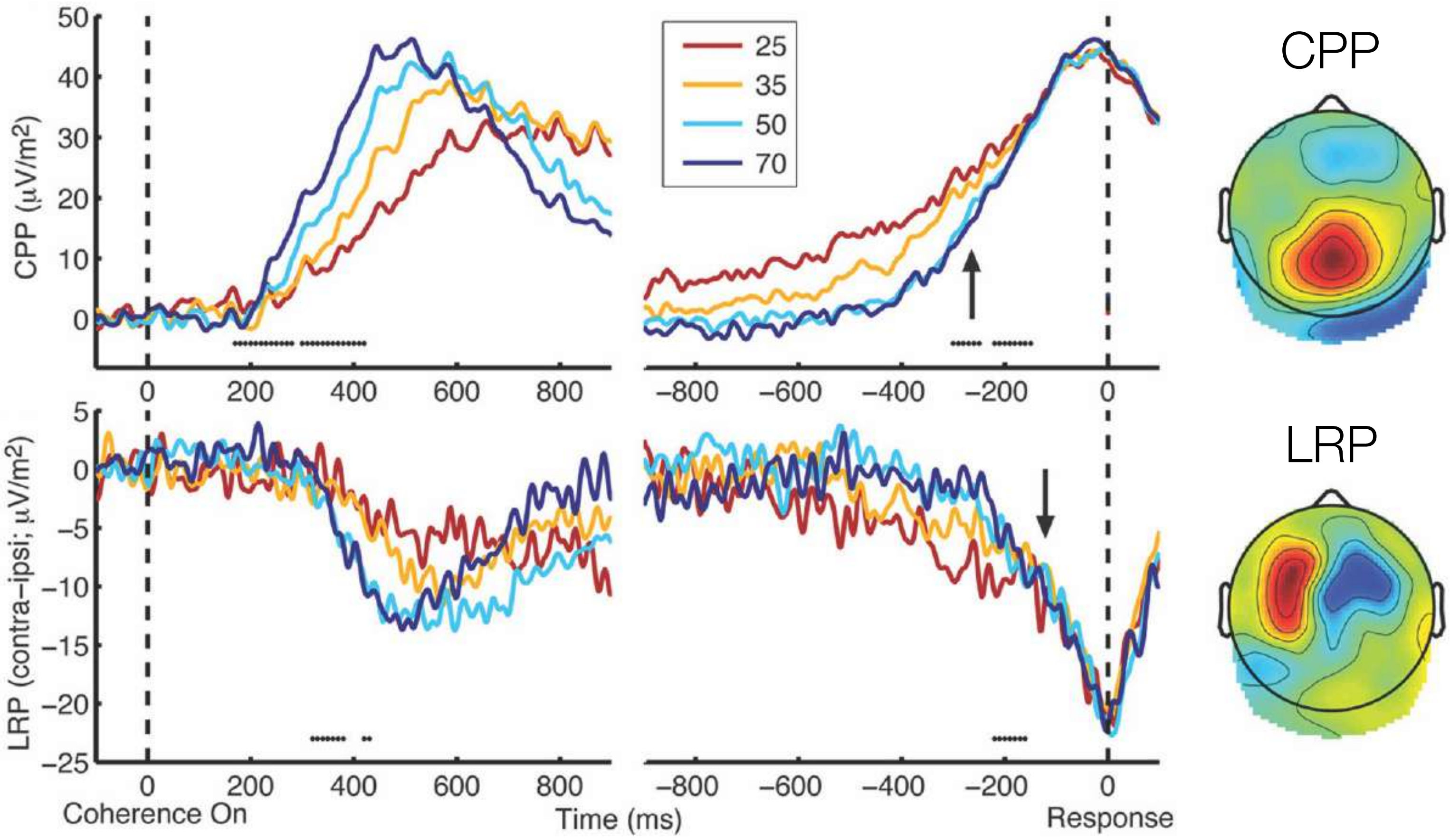
algorithm



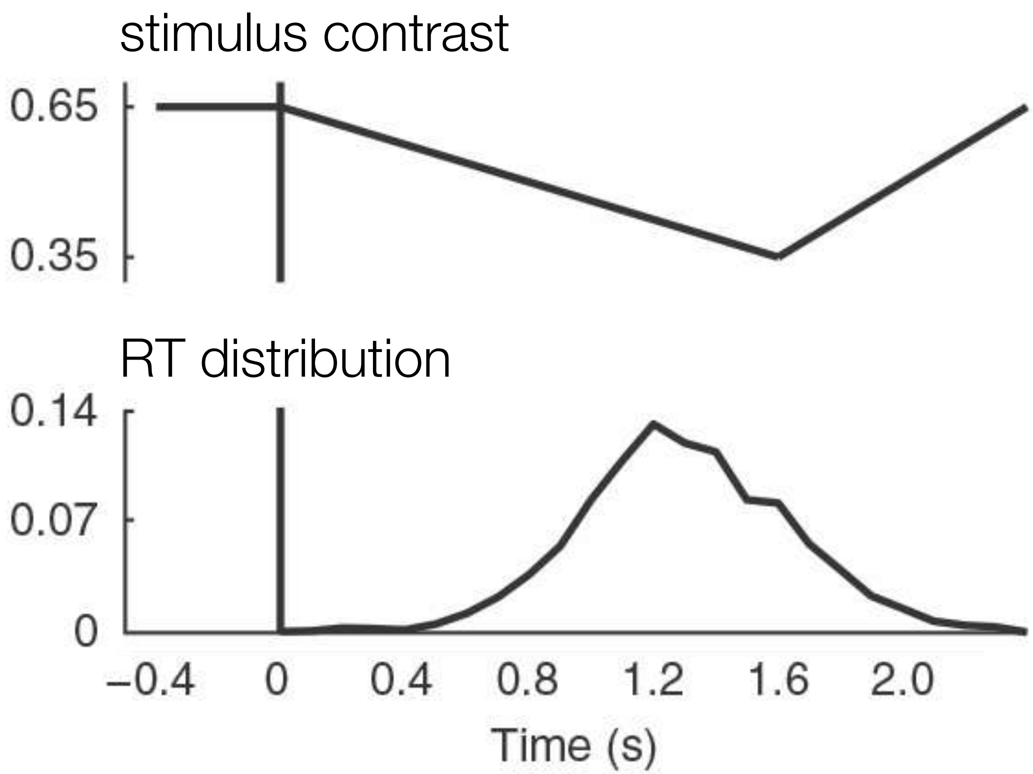
human

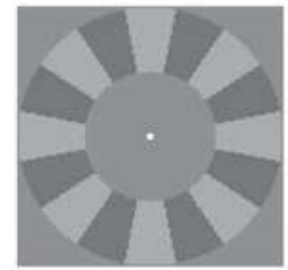
non-human
primate





Kelly and O'Connell (2013) Internal and external influences on the rate of sensory evidence accumulation. *Journal of Neuroscience*

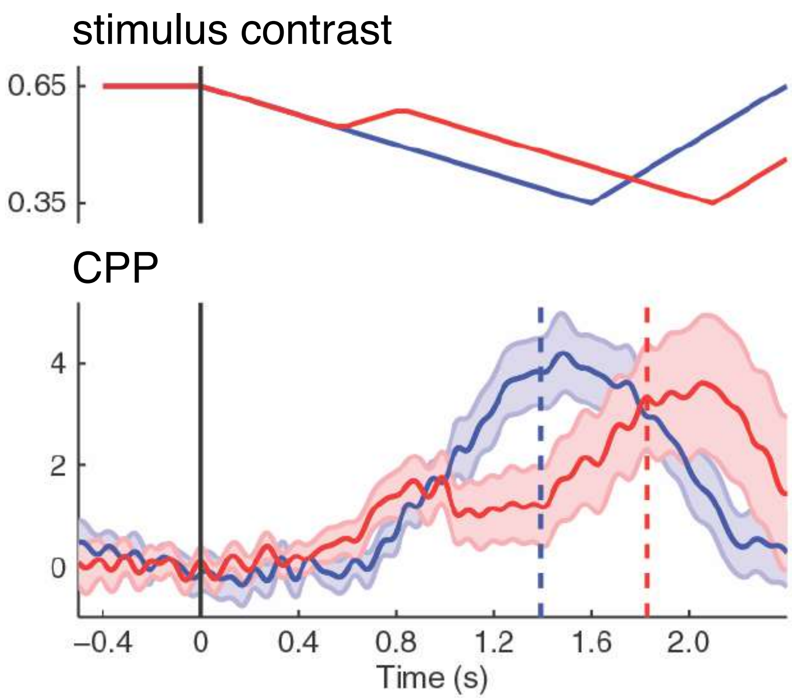




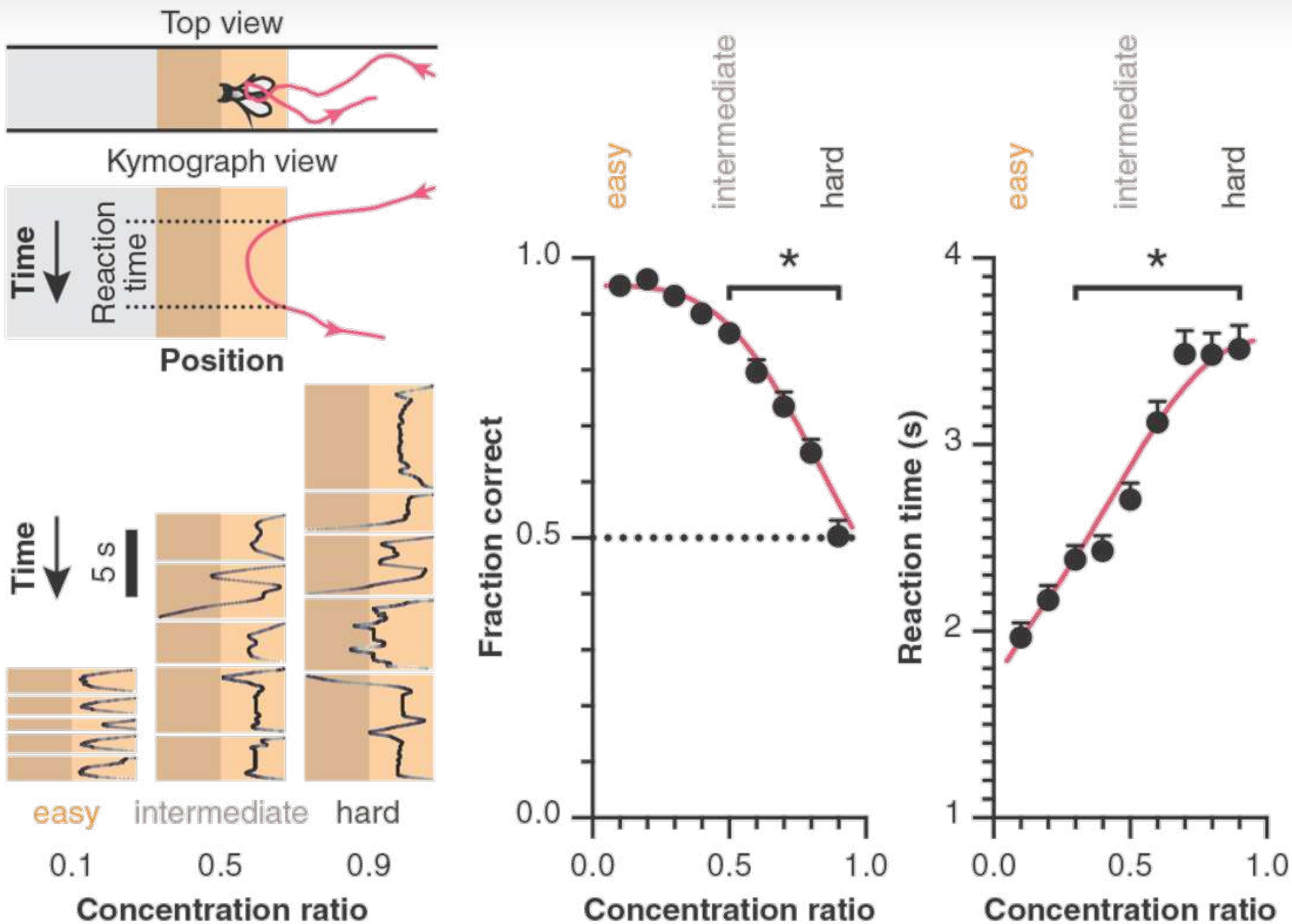
new stimulus
(contrast change)

Kelly and O’Connell (2013) Internal and external influences on the rate of sensory evidence accumulation. *Journal of Neuroscience*

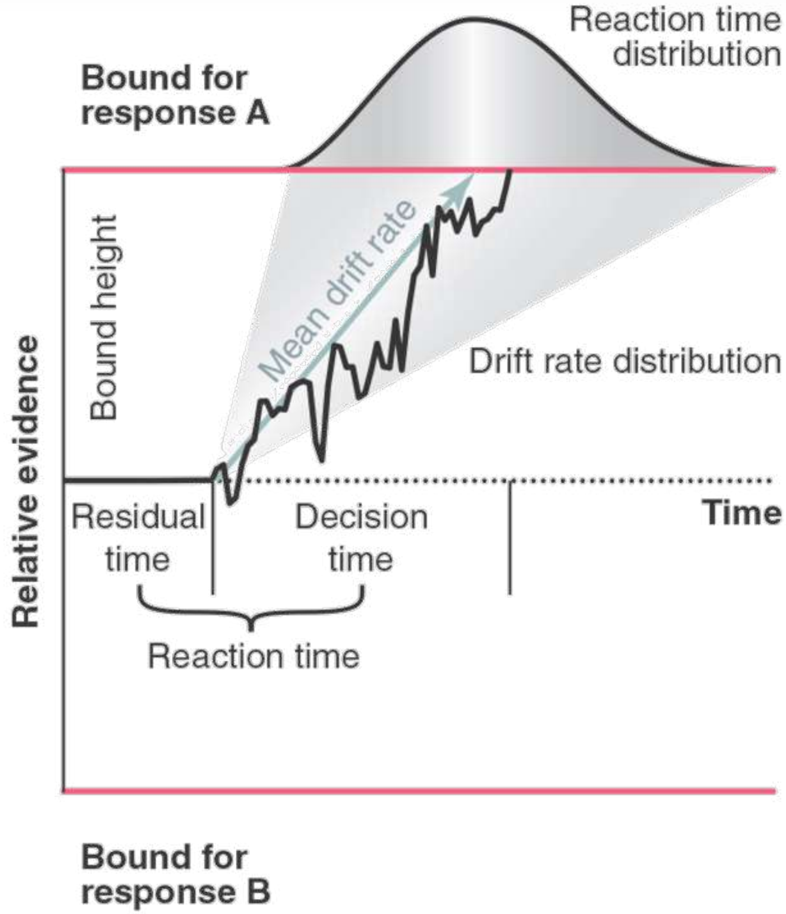
button press
<u>overt</u> decision bound



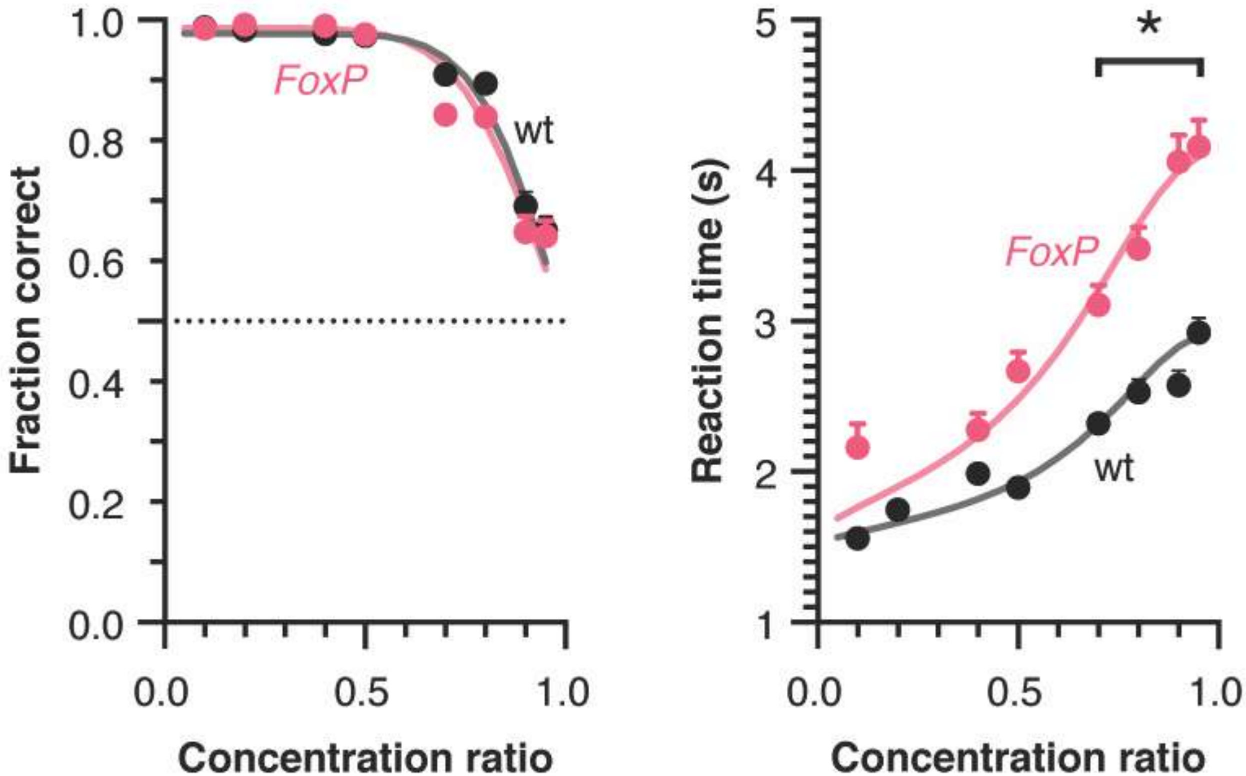
Kelly and O'Connell (2013) Internal and external influences on the rate of sensory evidence accumulation. *Journal of Neuroscience*



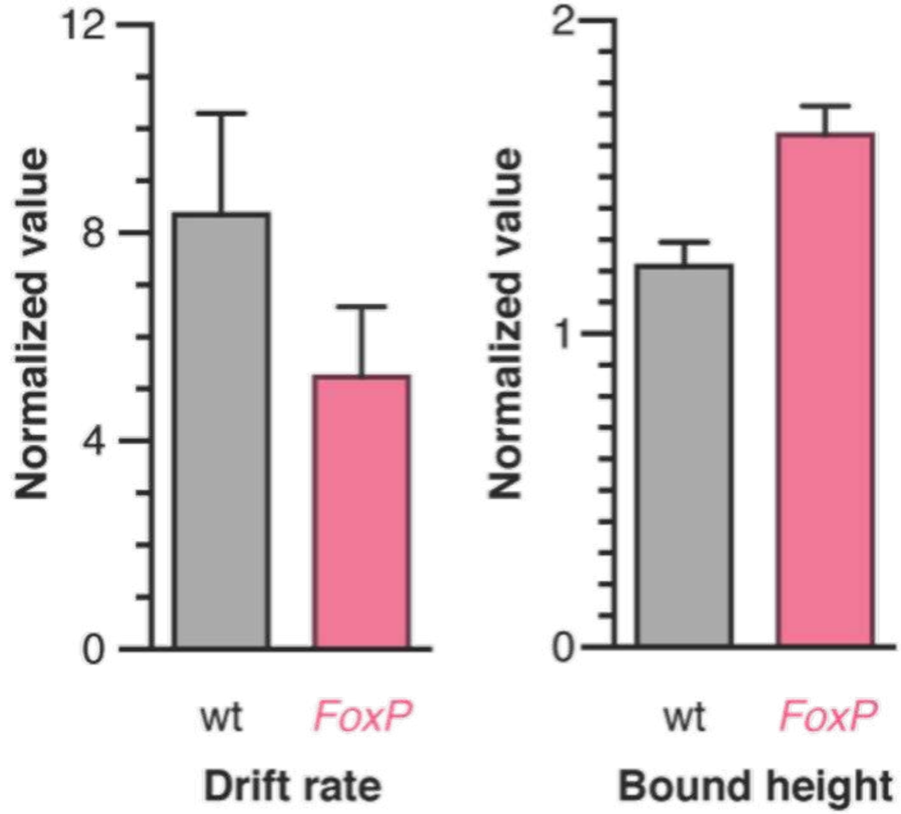
DasGupta et al. (2014) FoxP influences the speed and accuracy of a perceptual decision in Drosophila. *Science*



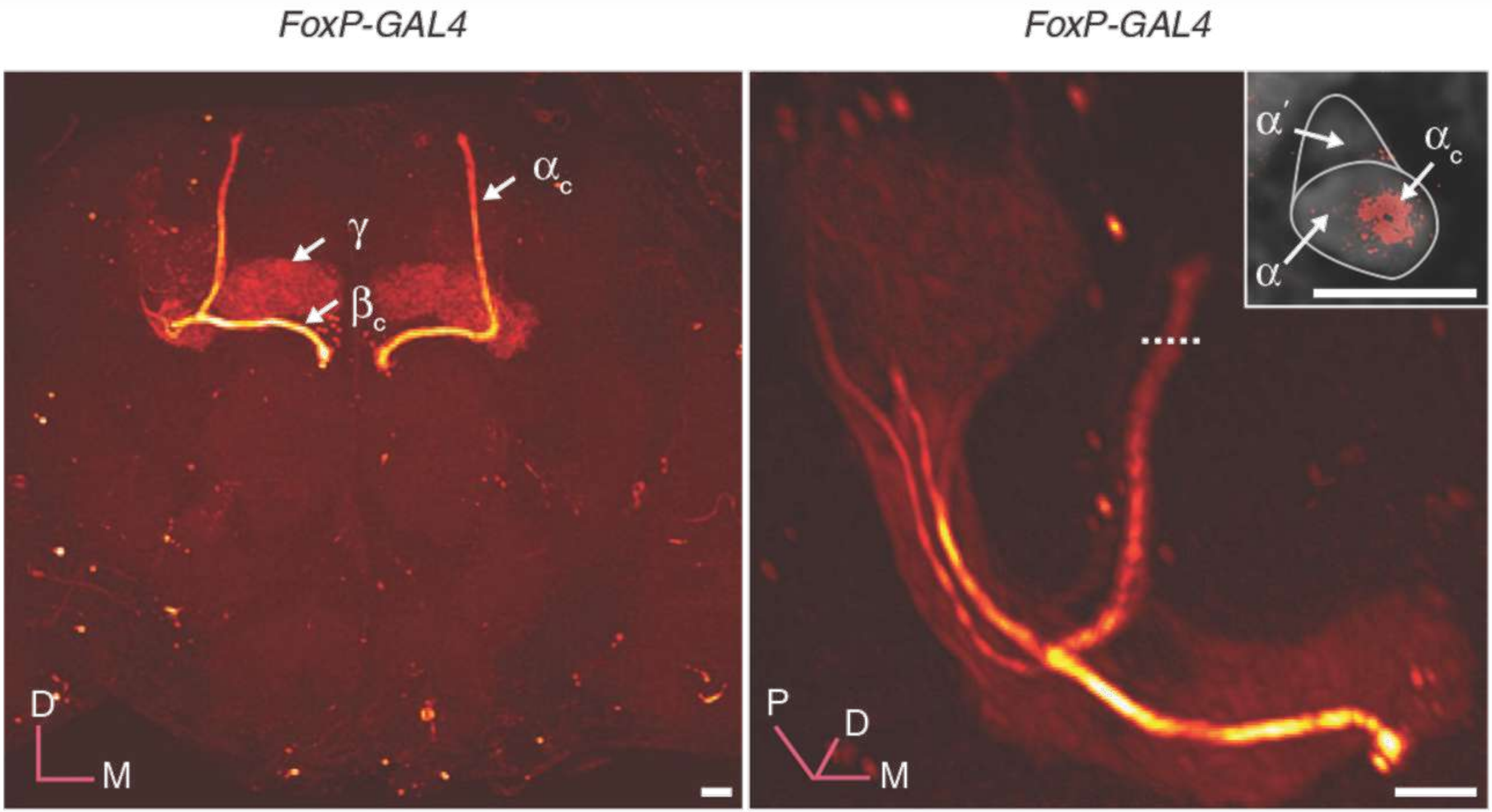
DasGupta et al. (2014) FoxP influences the speed and accuracy of a perceptual decision in Drosophila. *Science*



DasGupta et al. (2014) FoxP influences the speed and accuracy of a perceptual decision in Drosophila. *Science*



DasGupta et al. (2014) FoxP influences the speed and accuracy of a perceptual decision in Drosophila. *Science*

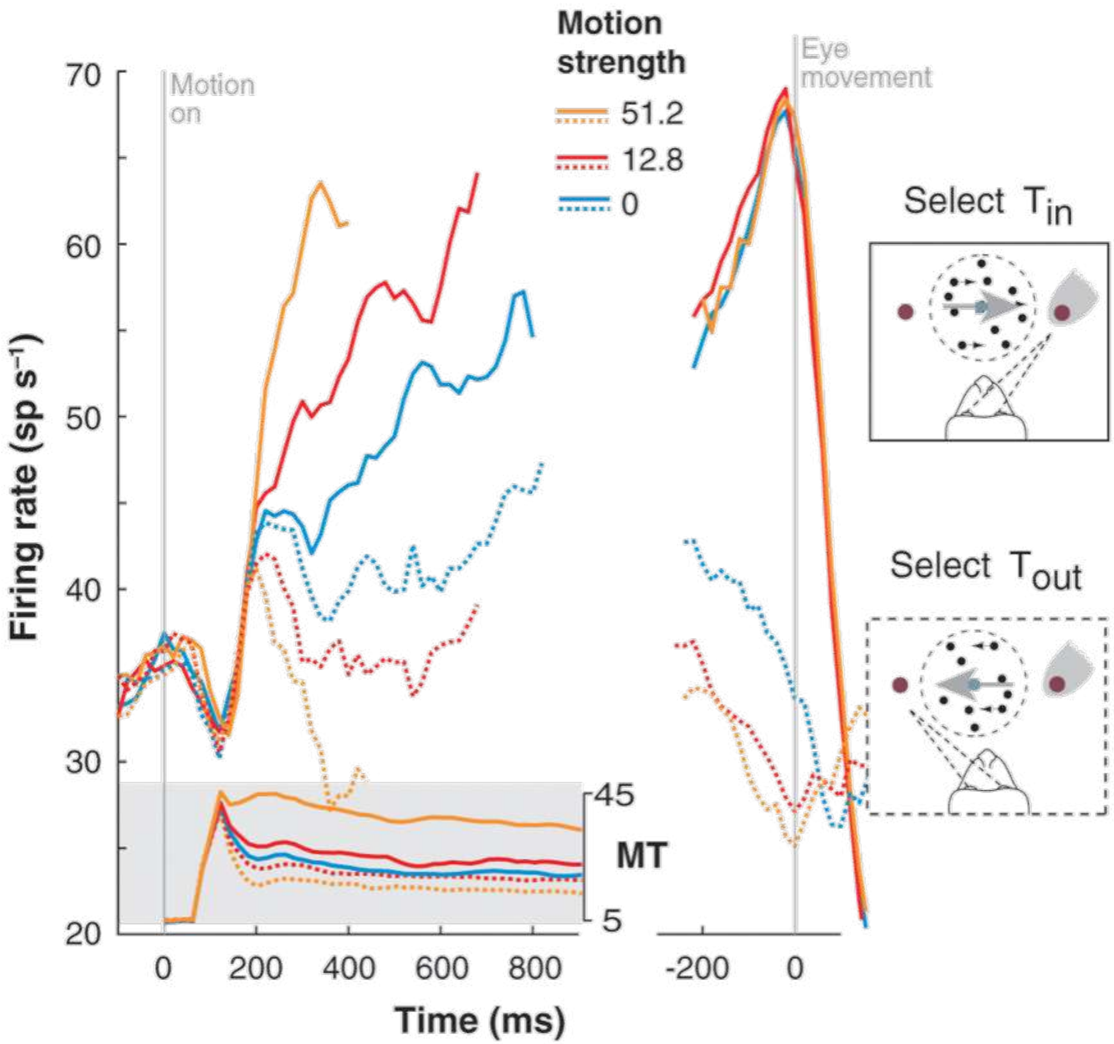


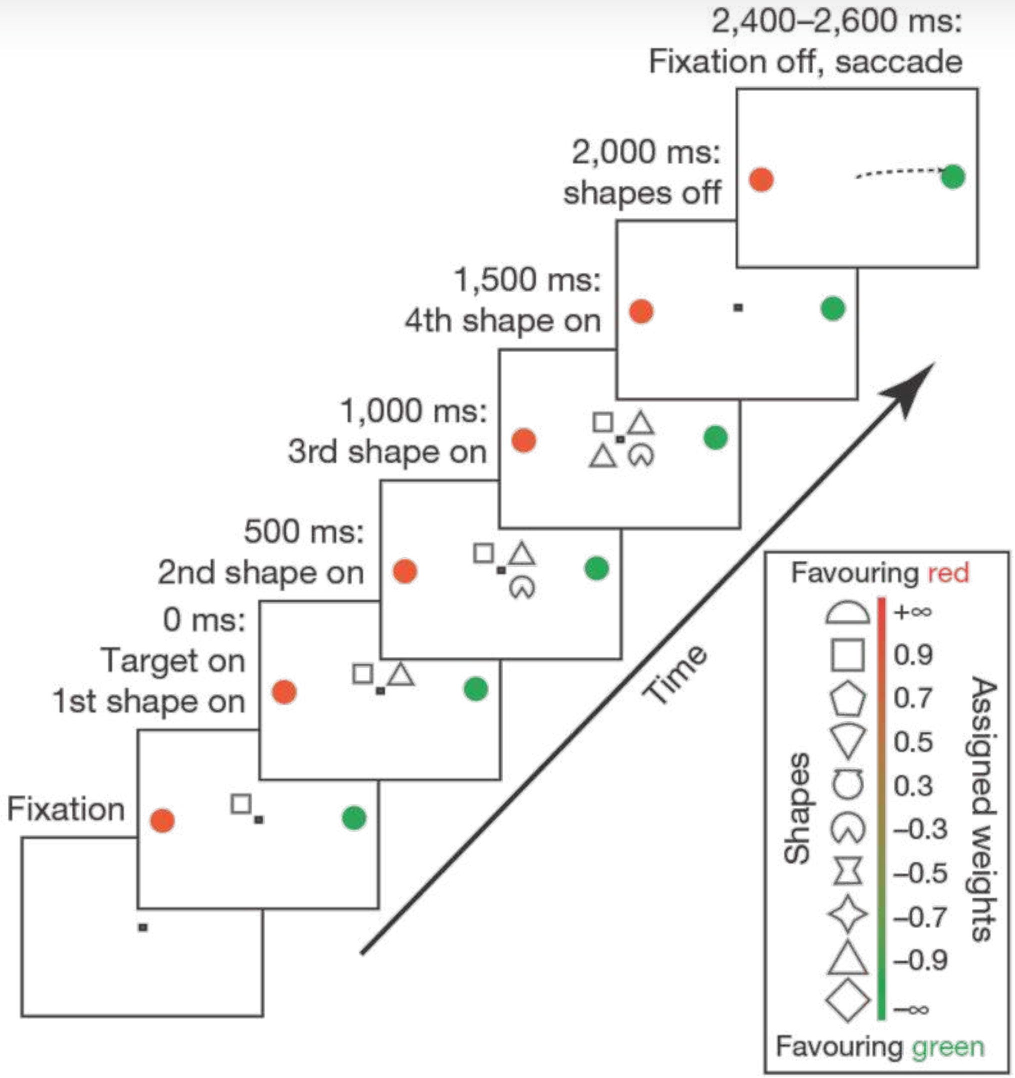
DasGupta et al. (2014) FoxP influences the speed and accuracy of a perceptual decision in Drosophila. *Science*

Beyond the paper

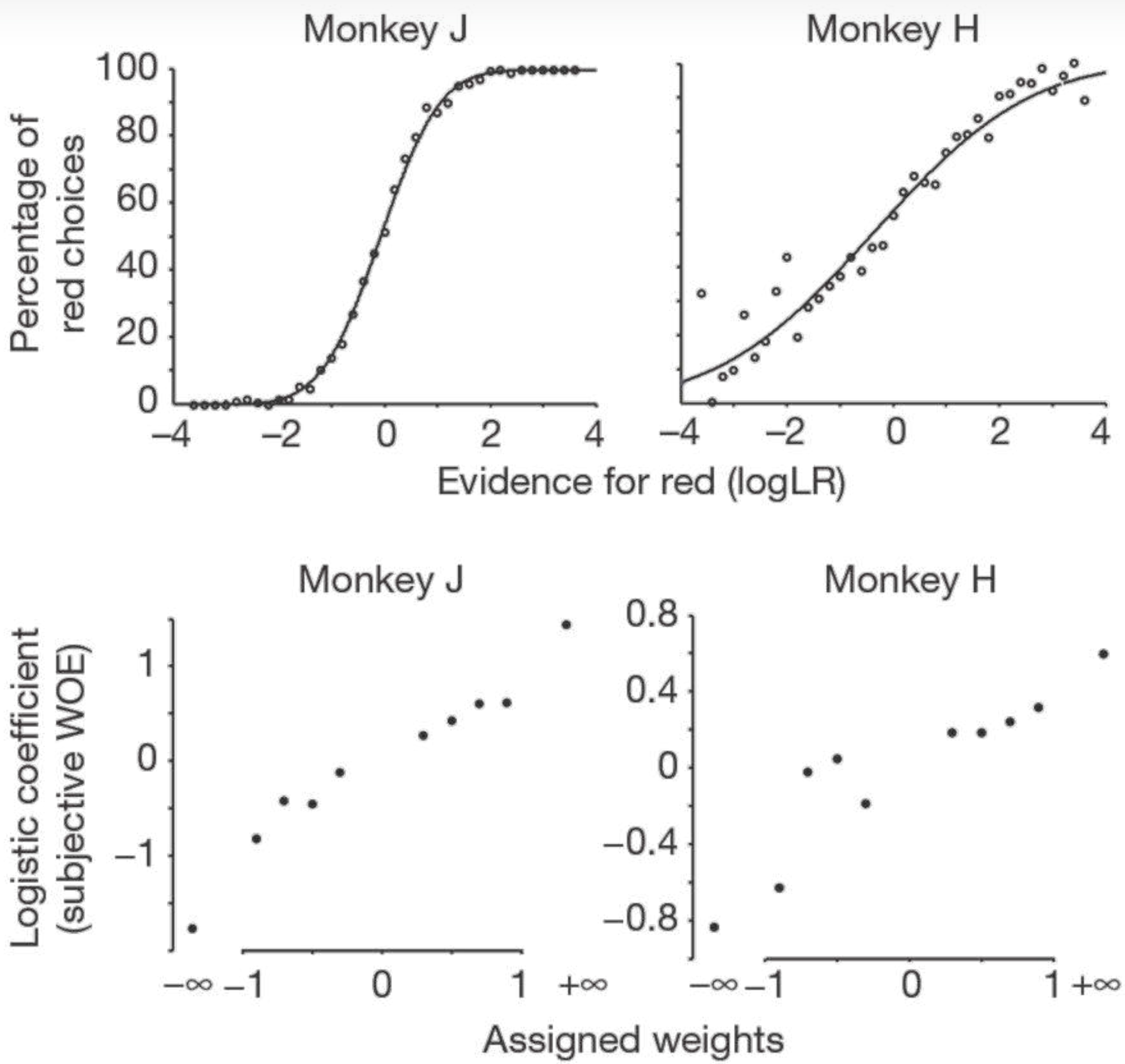
- Translation across tasks using modeling
- From random-dot motion perception…
  to associative learning and reasoning!
- Roitman and Shadlen (2002)
  neural correlates of motion-based evidence accumulation and decision-making in primate LIP
- Yang and Shadlen (2007)
  neural correlates of symbolic evidence accumulation and decision-making in the same brain region

Beyond the paper

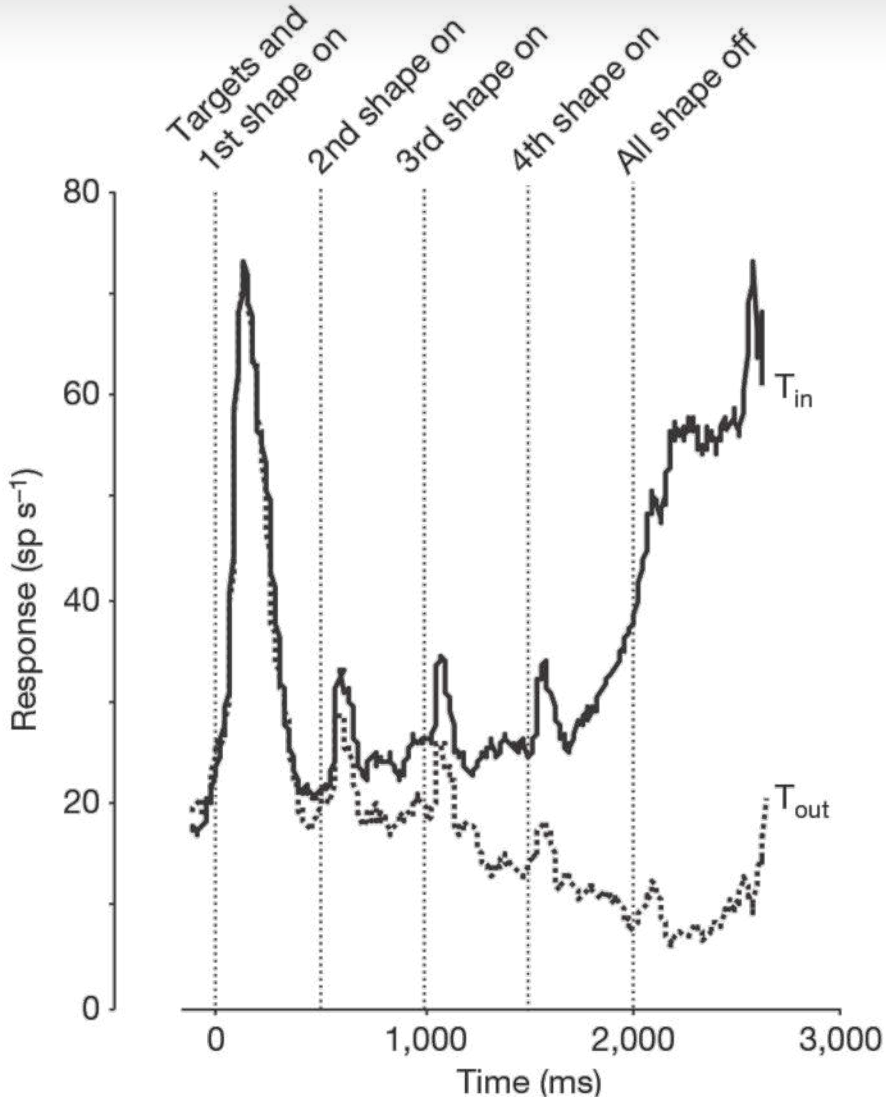




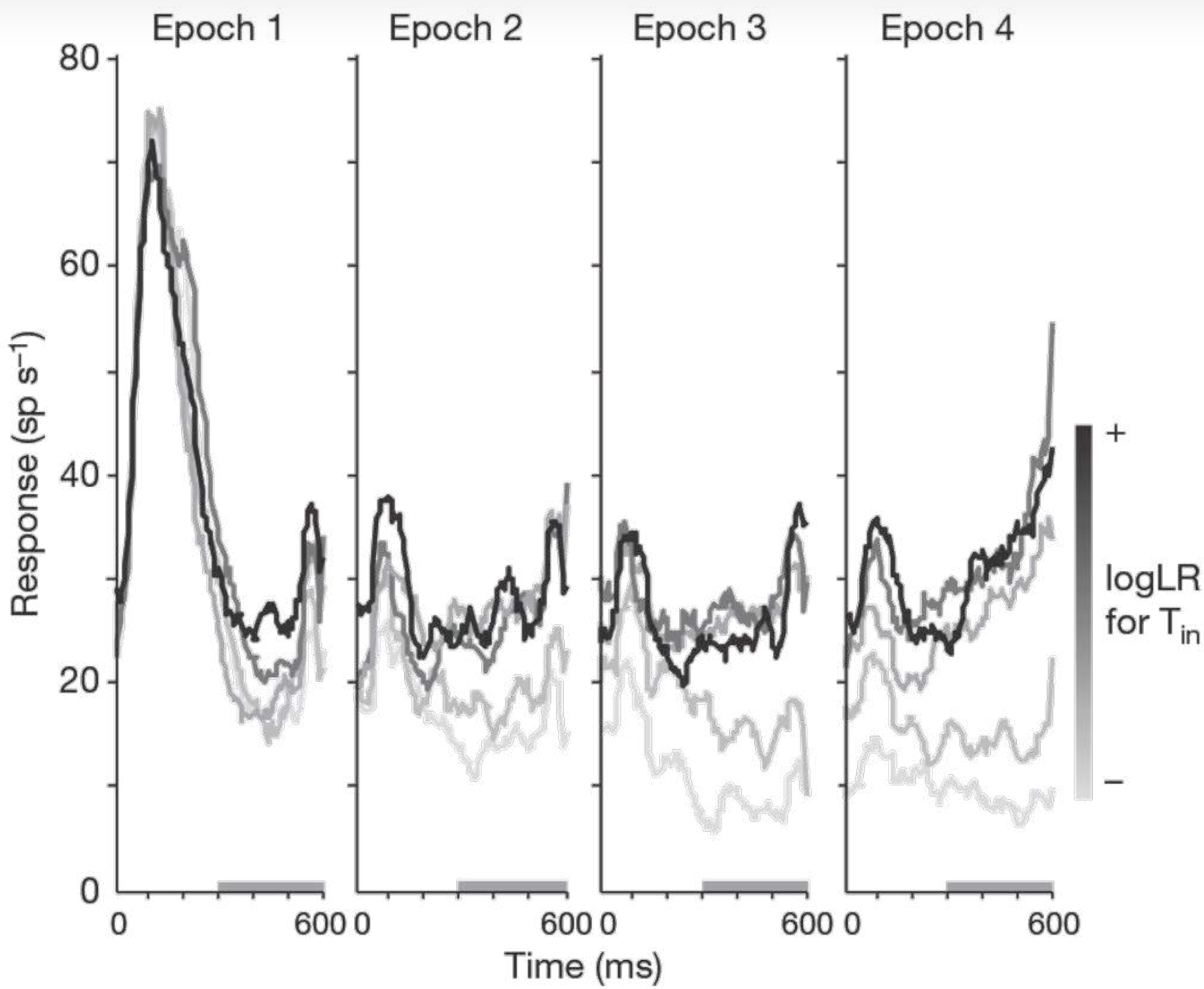
Yang and Shadlen (2007) Probabilistic reasoning by neurons. *Nature*



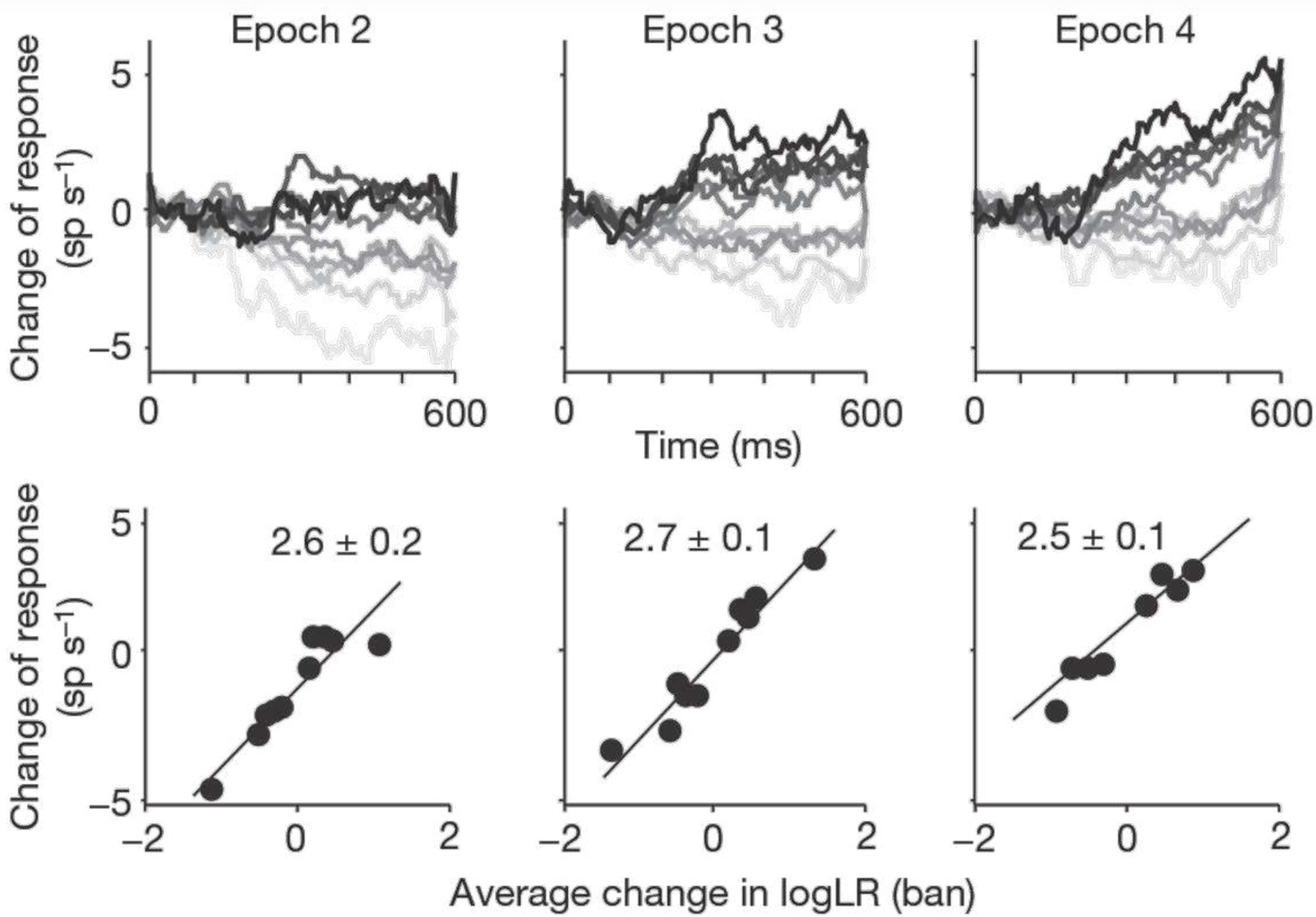
Yang and Shadlen (2007) Probabilistic reasoning by neurons. *Nature*



Yang and Shadlen (2007) Probabilistic reasoning by neurons. *Nature*



Yang and Shadlen (2007) Probabilistic reasoning by neurons. *Nature*



Yang and Shadlen (2007) Probabilistic reasoning by neurons. *Nature*

Coming next

- Next class: 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