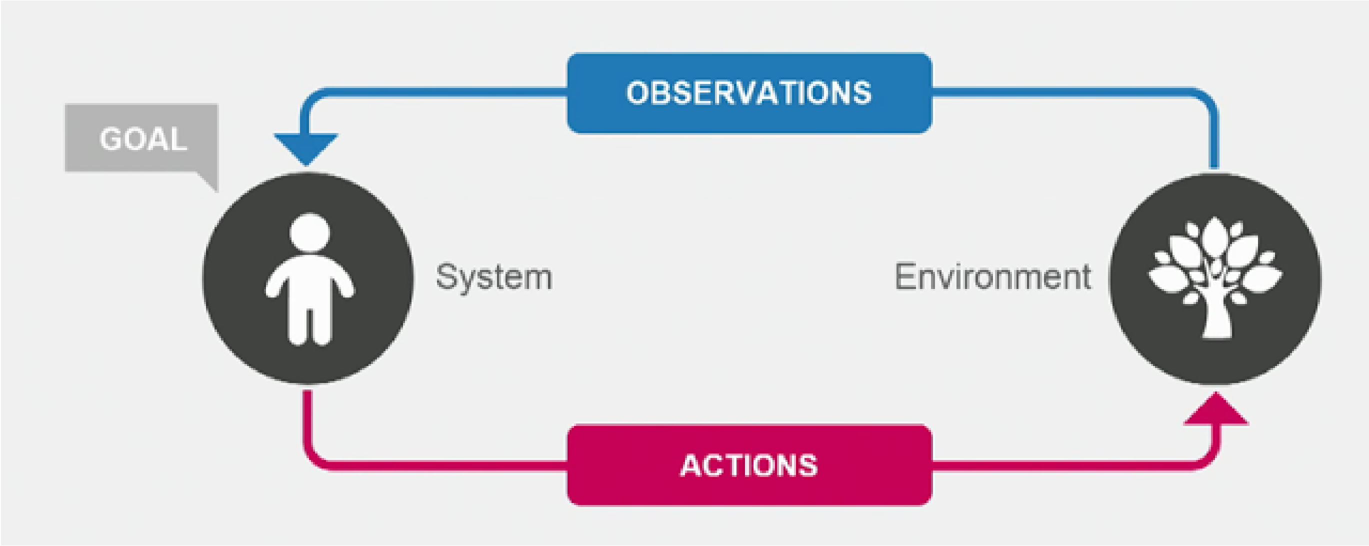


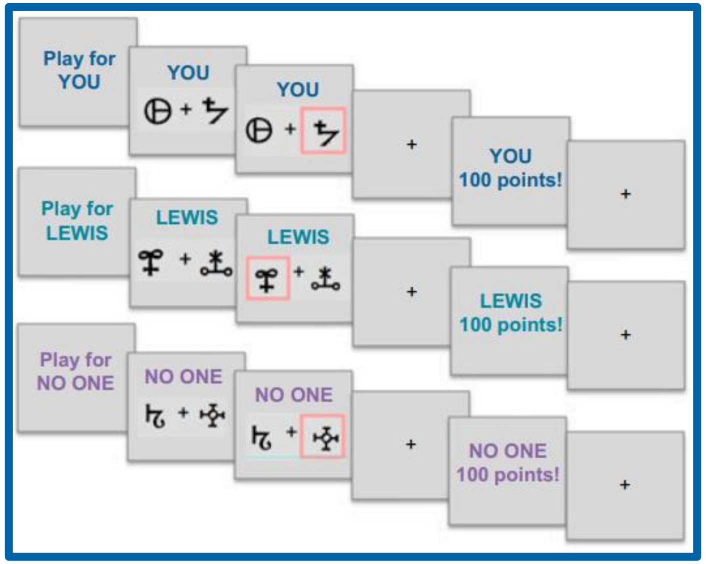
Using reinforcement learning models in social neuroscience: Frameworks, pitfalls, and suggestions

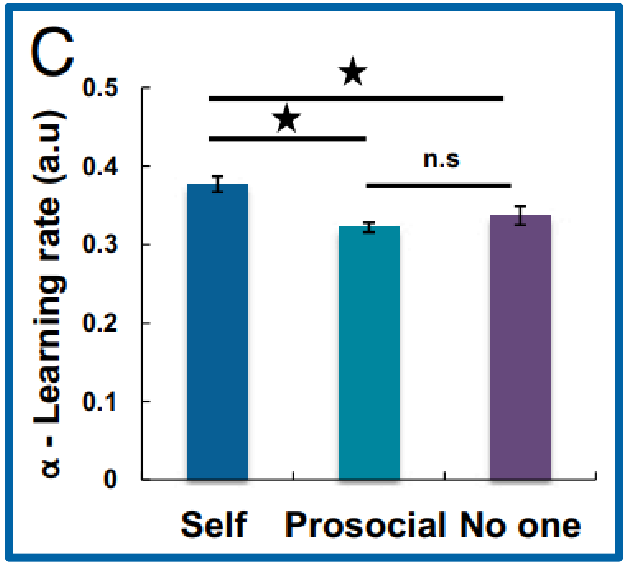
Lei Zhang*, Lukas Lengersdorff*, Nace Mikus, Jan Gläscher, Claus Lamm

lei.zhang@univie.ac.at
lei-zhang.net
@lei_zhang_lz

universität wien
Fakultät für Psychologie

Background

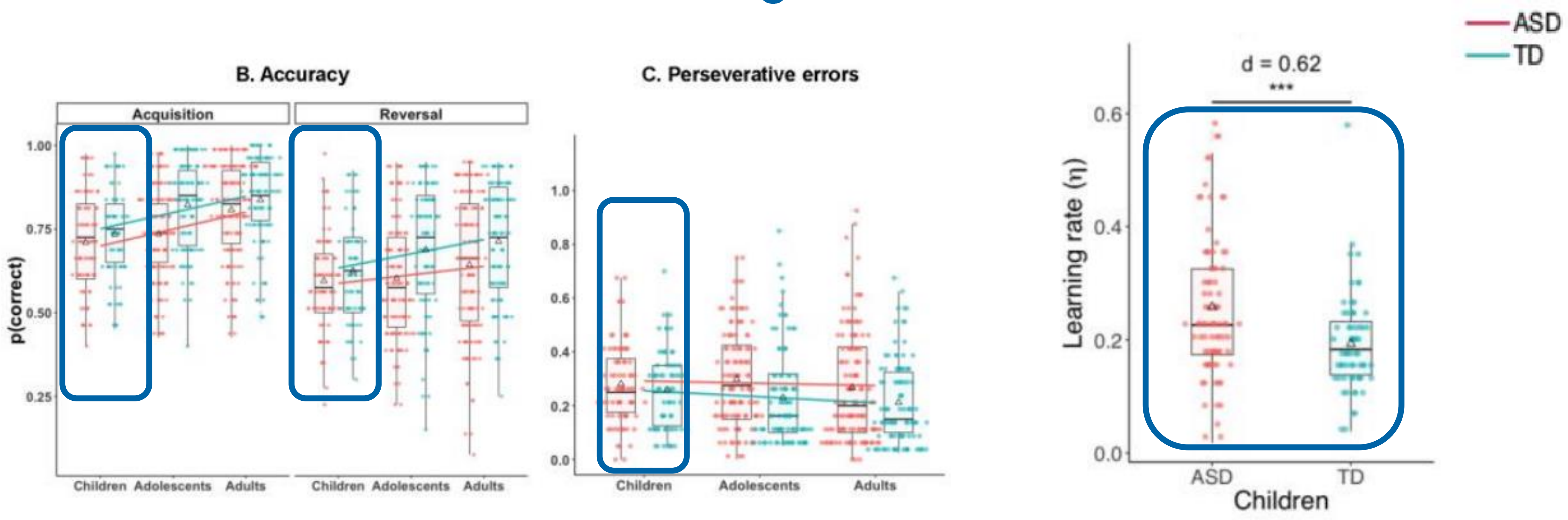




Good to have a large learning rate?

Lockwood et al. (2016)

Background

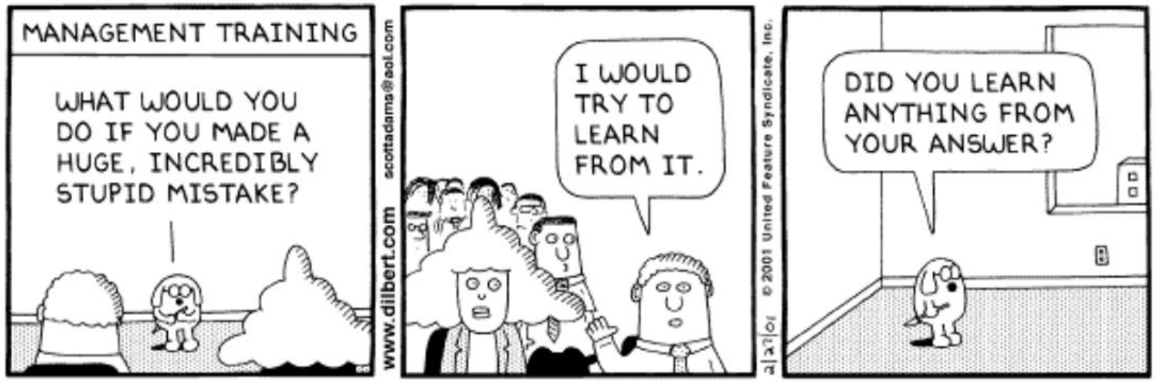


Good to have a small learning rate?

Crawley, Zhang, …, EU-AIMS (2019)

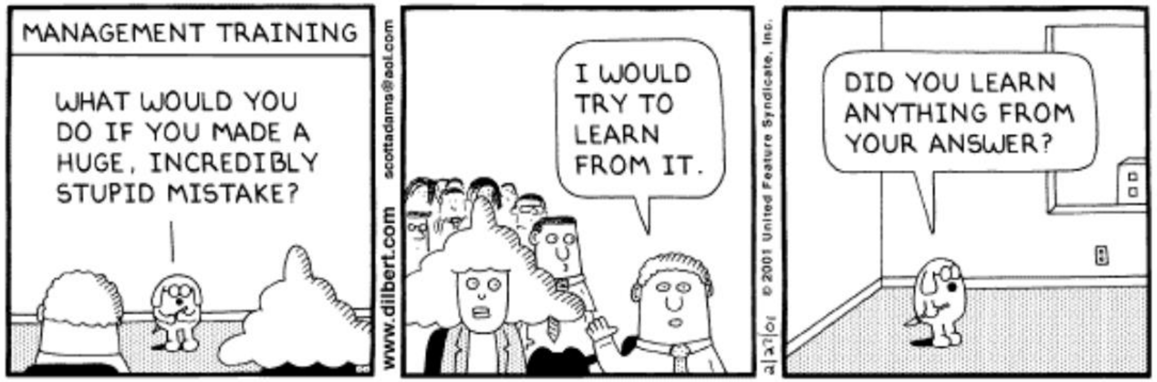
Outline

- the Reinforcement Learning framework
- the learning rate
  - what is it?
  - is there a optimal learning rate
- searching for prediction error signals in the brain
- model validation
- moving toward hierarchical estimation



Outline

- the Reinforcement Learning framework
- the learning rate
  - what is it?
  - is there a optimal learning rate
- searching for prediction error signals in the brain
- model validation
- moving toward hierarchical estimation



2-armed bandit task

a simple task often used in the laboratory:

- repeated choice between N options (N-armed bandit)
- …whose properties (reward amounts, probabilities) are learned through trial-and-error
- …with a goal in mind: maximize the overall reward

2-armed bandit task

What can be your strategies:

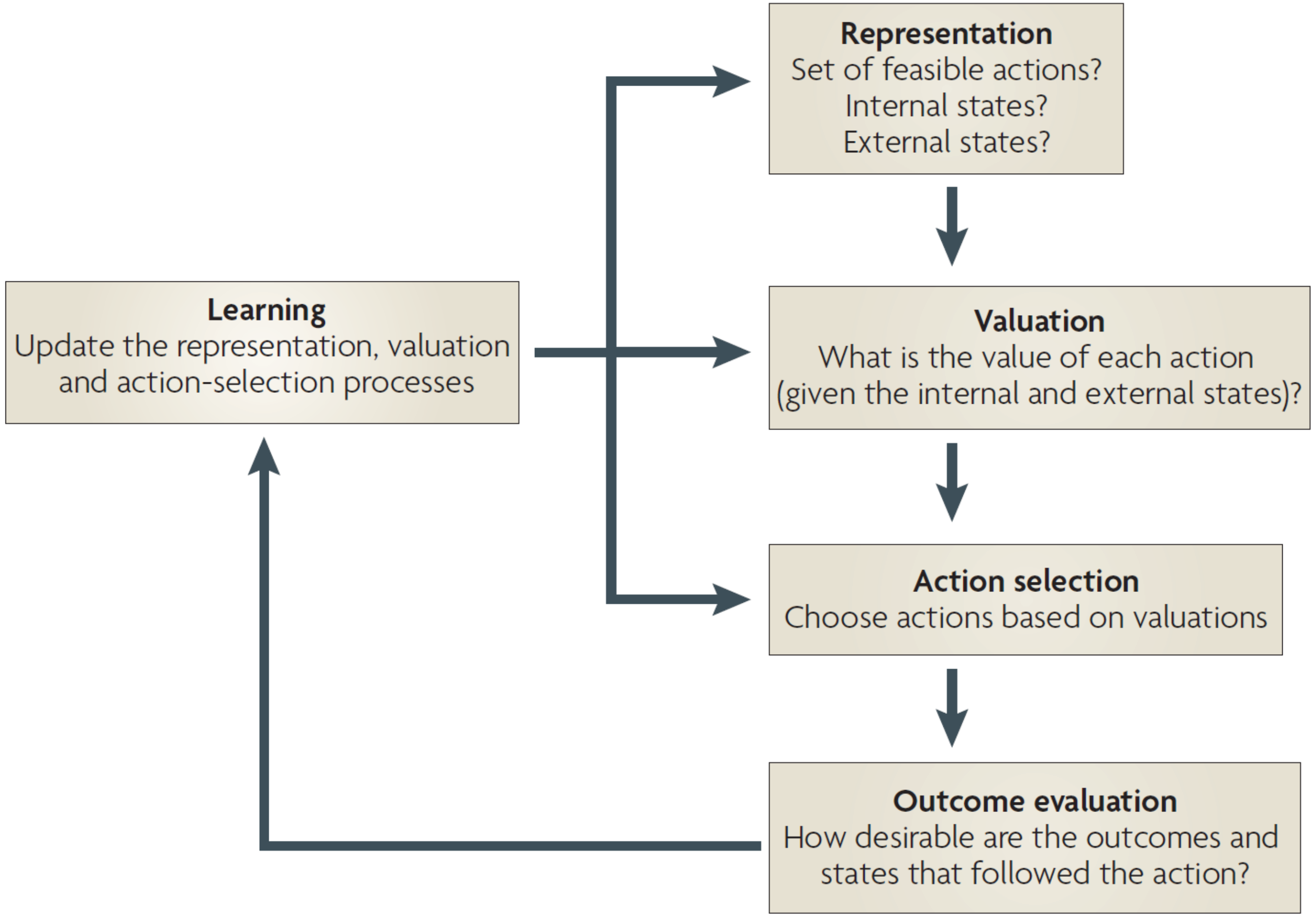
1. predict the value of each deck
2. choose the best
3. learn from outcome to update predictions

(repeat)

How prediction is shaped by learning?

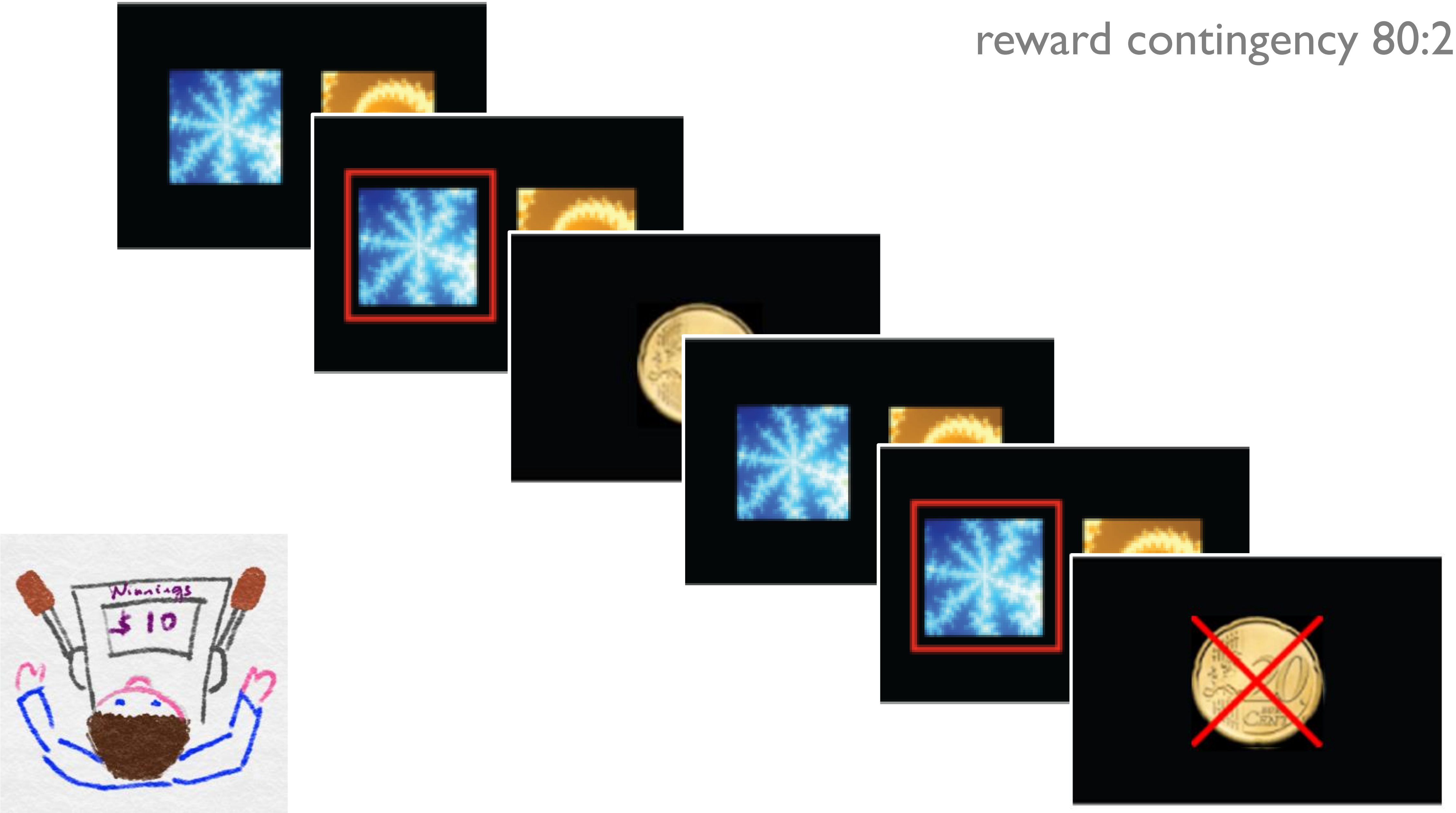
What can be your strategies:

1. predict the value of each deck
2. choose the best
3. learn from outcome to update predictions (repeat)

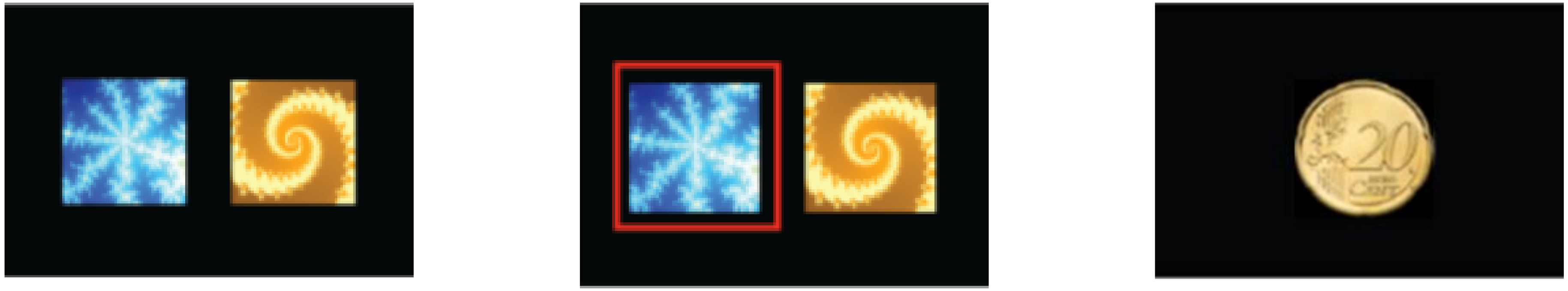


Sutton & Barto (1998); Rangel et al (2008); [Daw & Niv workshop 2015](); [Niv PSY338 2013]()

reward contingency 80:20



One simple experiment: two choice task



choice presentation | action selection | outcome

what do we know?

what can we measure?

what do we not know?

Data: choice & outcome

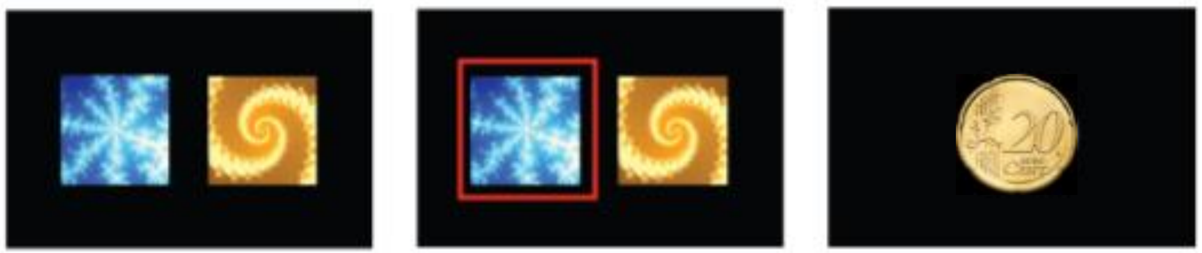
Summary stats: choice accuracy

Learning algorithm: RL update

p(choosing the better option)

Rescorla-Wagner (1972)

- The idea: error-driven learning
- Change in value is proportional to the difference between actual and predicted outcome



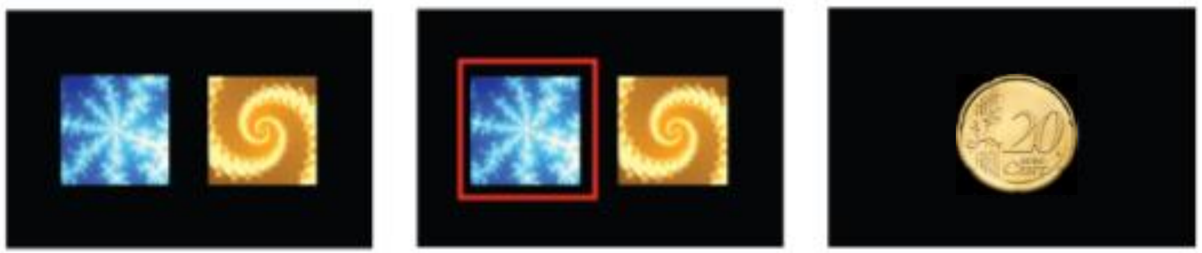
Value update: $V_t = V_{t-1} + \alpha * \mathrm{PE}_{t-1}$

Prediction error: $\mathrm{PE}_{t-1} = R_{t-1} - V_{t-1}$

Rescorla & Wagner (1972)

Rescorla-Wagner (1972)

- The idea: error-driven learning
- Change in value is proportional to the difference between actual and predicted outcome

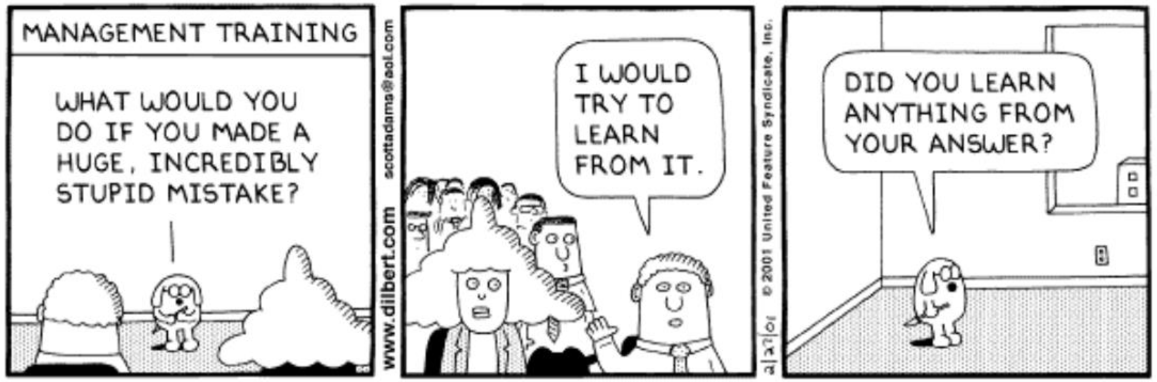


$$\text{Value update:} \quad V_t = V_{t-1} + \alpha * \mathrm{PE}_{t-1}$$

$$\text{Prediction error:} \quad \mathrm{PE}_{t-1} = R_{t-1} - V_{t-1}$$

Outline

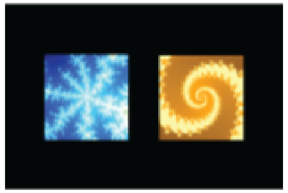
- the Reinforcement Learning framework
- the learning rate
  - what is it?
  - is there a optimal learning rate
- searching for prediction error signals in the brain
- model validation
- moving toward hierarchical estimation



Learning rate

Value update: $V_t = V_{t-1} + \alpha * \mathrm{PE}_{t-1}$

Prediction error: $\mathrm{PE}_{t-1} = R_{t-1} - V_{t-1}$



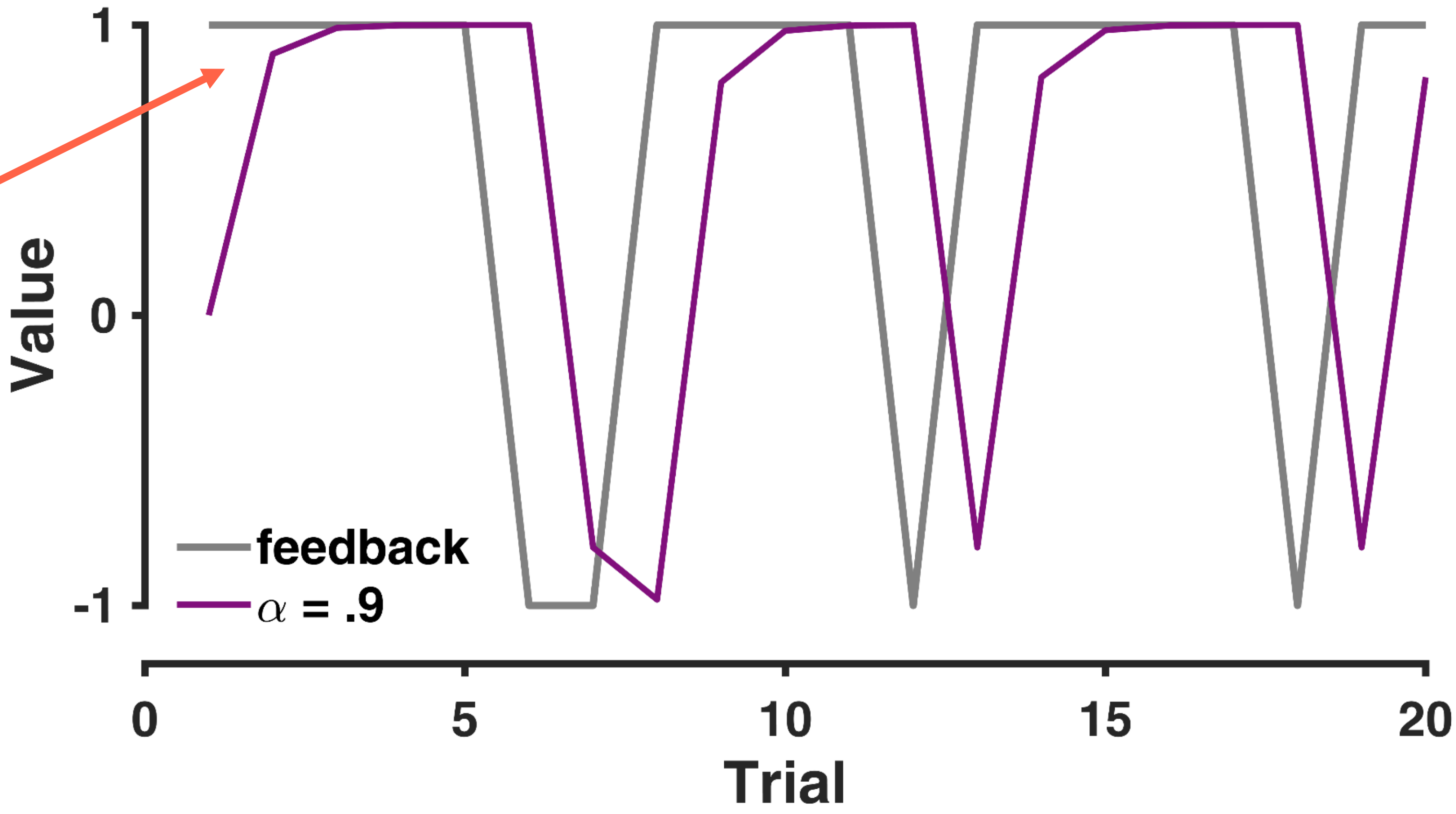
if $\alpha = 0.9$

$V_1 = 0$

$V_2 = V_1 + 0.9 * (1 - V_1)$

$= 0 + 0.9 * (1 - 0)$

$= 0.9$

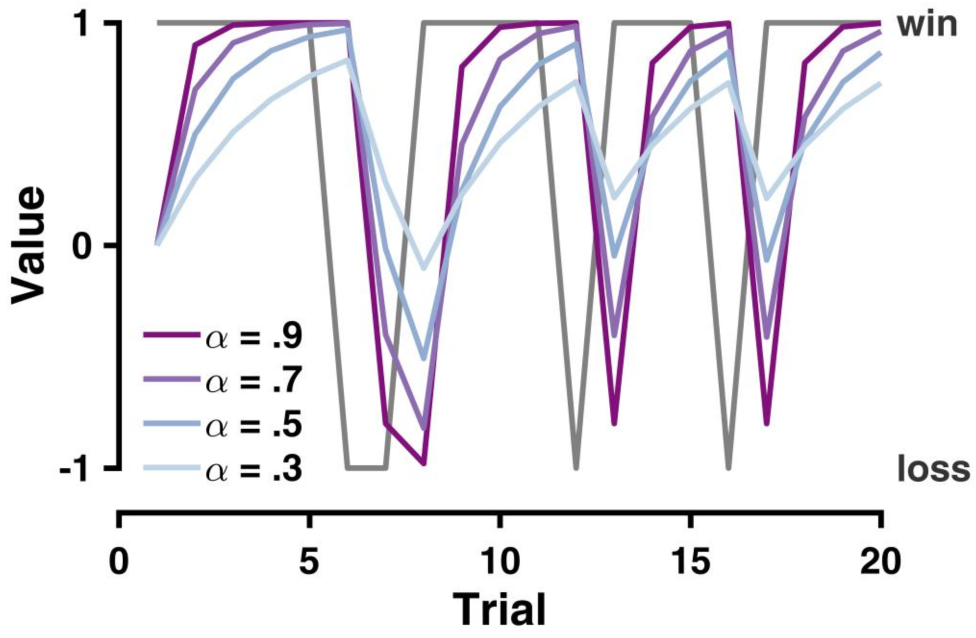


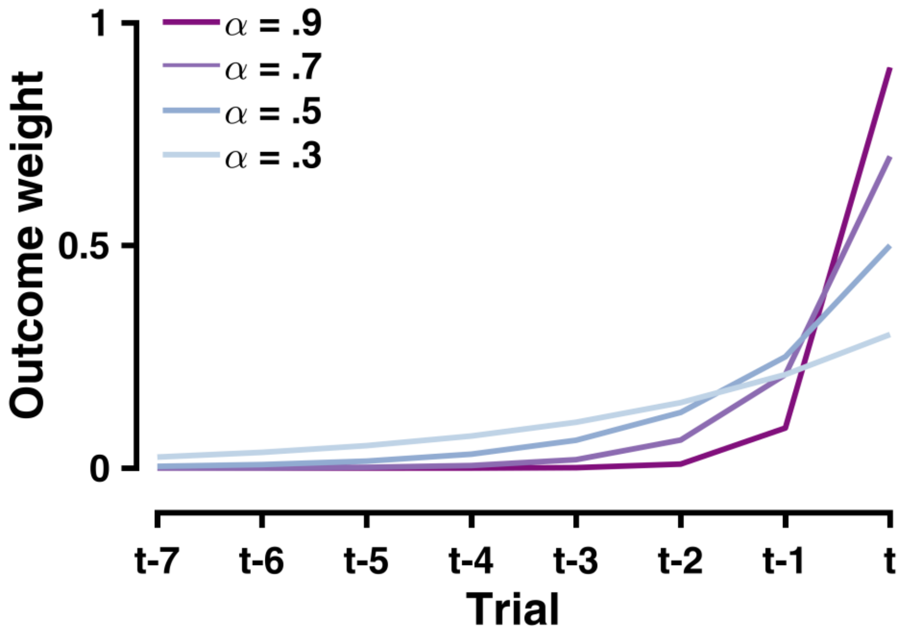
Learning rate

Value update: $V_t = V_{t-1} + \alpha * \mathrm{PE}_{t-1}$

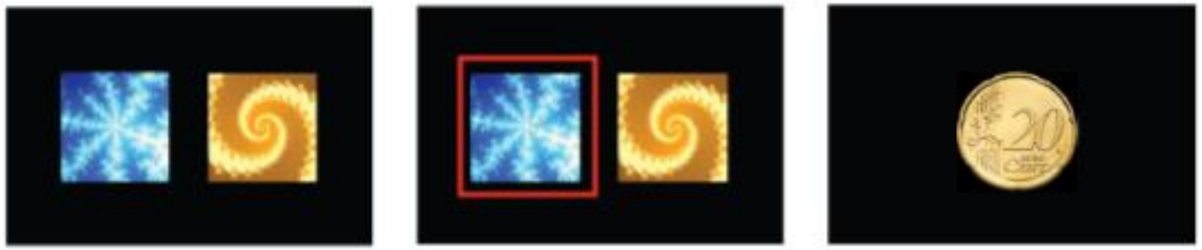
Prediction error: $\mathrm{PE}_{t-1} = R_{t-1} - V_{t-1}$

$$\begin{aligned} V_t &= (1-\alpha)\,V_{t-1} + \alpha R_{t-1} \\ &= (1-\alpha)\,(V_{t-2} + \alpha\,(R_{t-2} - V_{t-2})) + \alpha R_{t-1} \\ &= (1-\alpha)^t V_0 + \boxed{\sum_{i=1}^{t} (1-\alpha)^{t-i}\,\alpha} R_i \end{aligned}$$





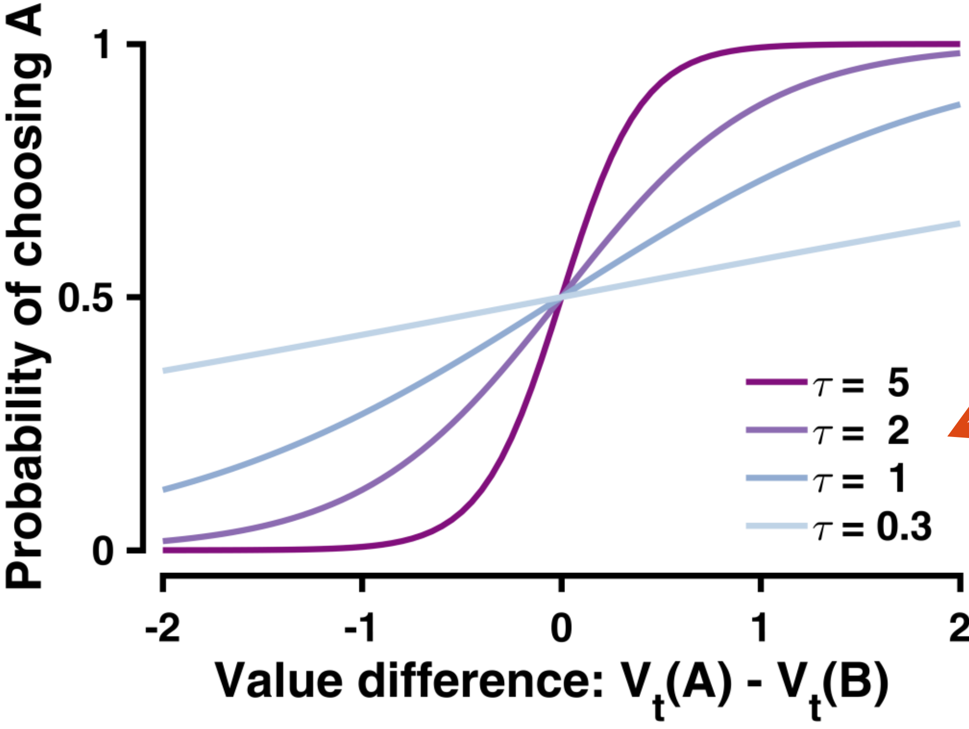
Choice rule



V()
V()

$$p_t(A) = \frac{e^{\tau * V_t(A)}}{e^{\tau * V_t(A)} + e^{\tau * V_t(B)}}$$

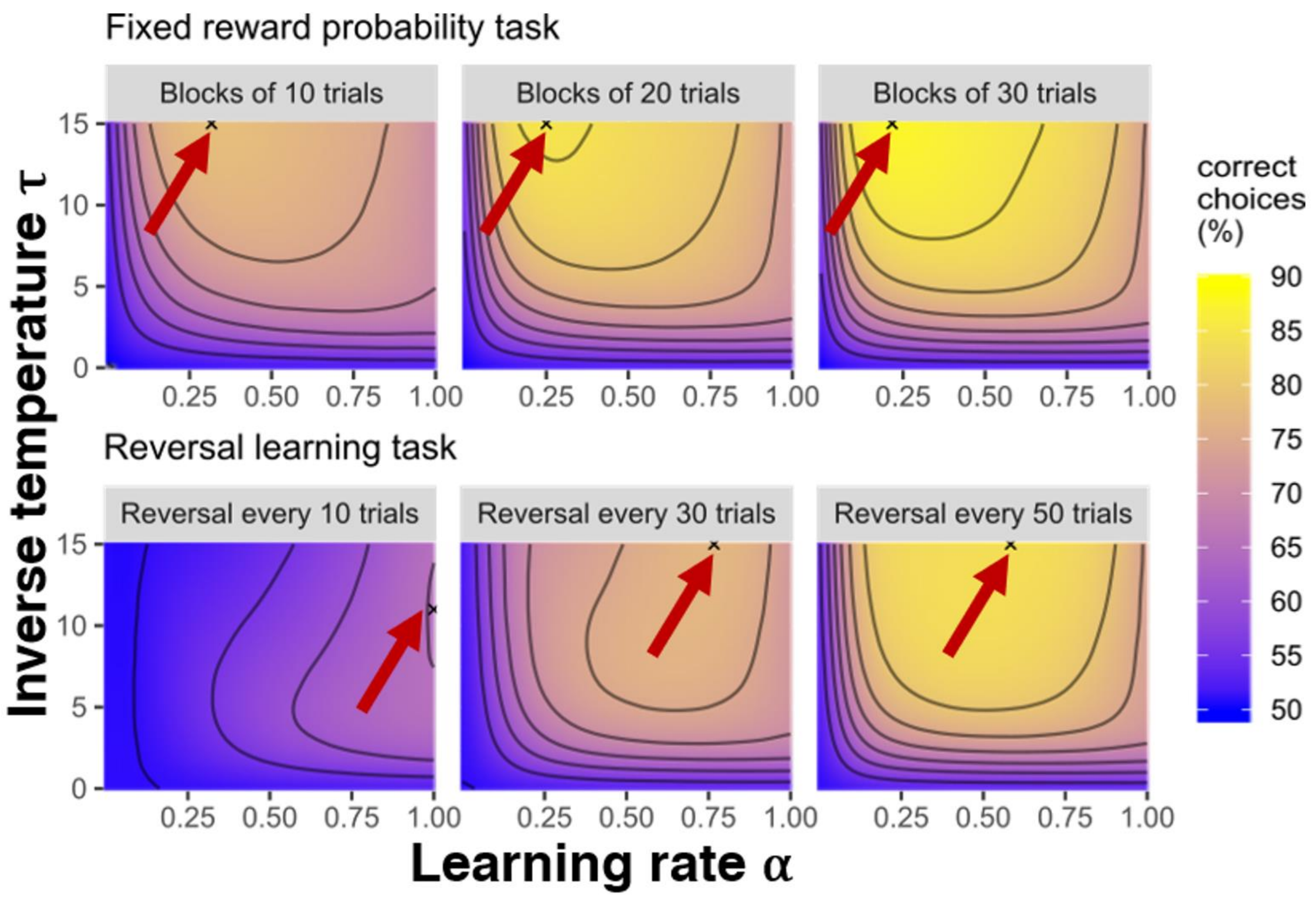
$$= \frac{1}{1 + e^{-\tau * (V_t(A) - V_t(B))}}$$



Optimal learning rate?

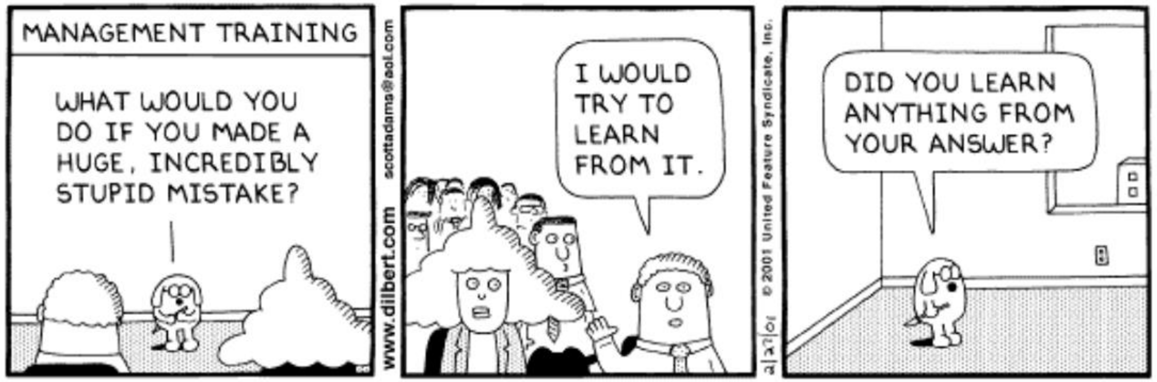
$$p(\mathrm{C} = \text{better option})$$

$$= f(\alpha, \tau)$$

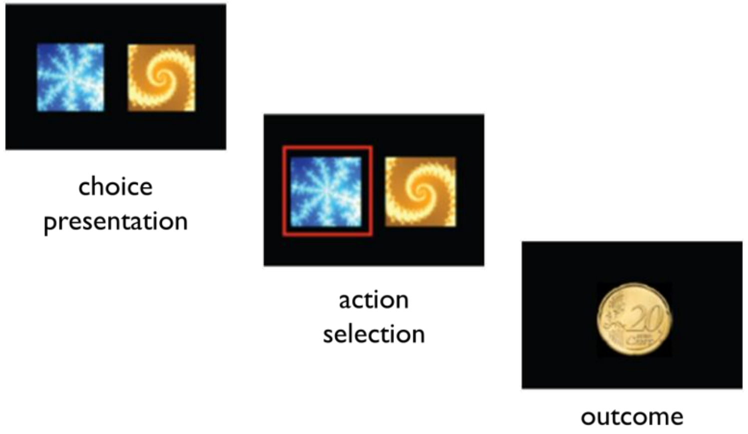


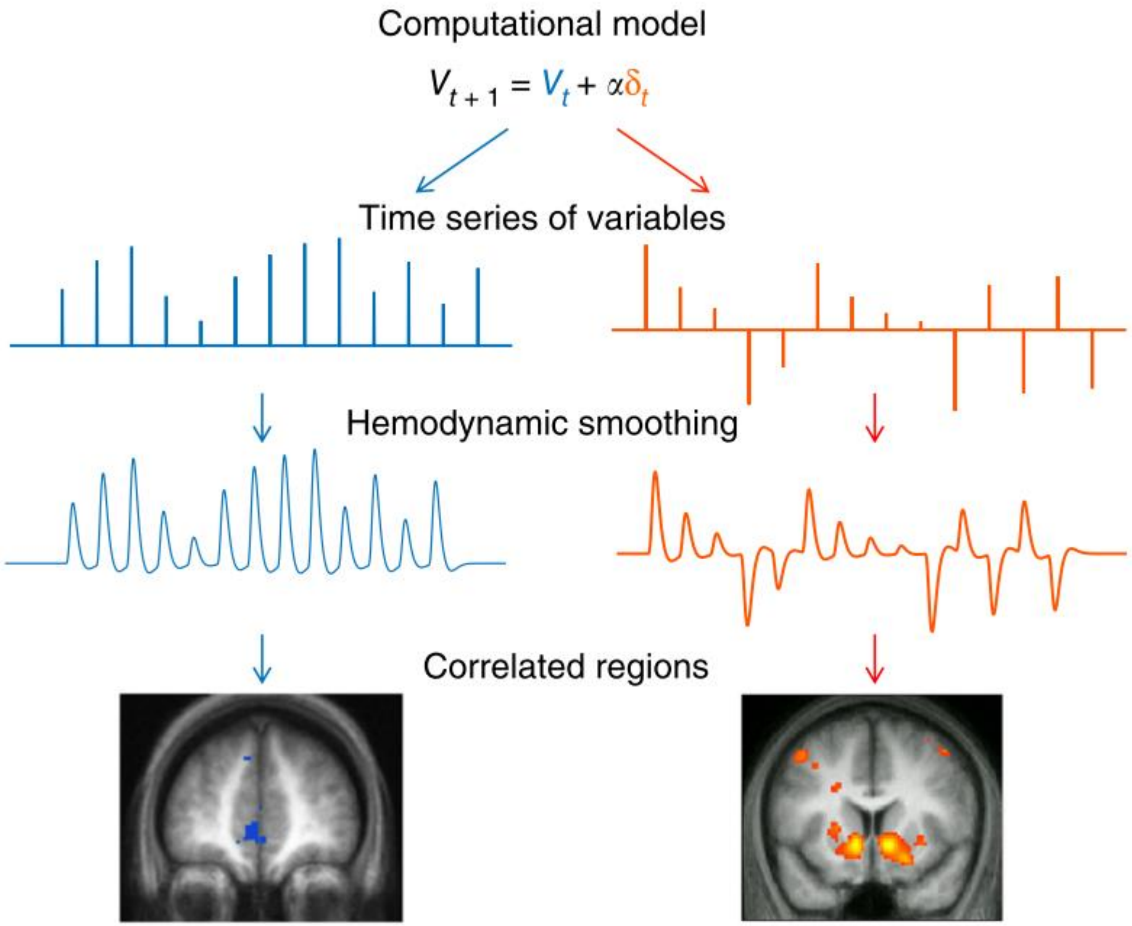
Outline

- the Reinforcement Learning framework
- the learning rate
  - what is it?
  - is there a optimal learning rate
- searching for prediction error signals in the brain
- model validation
- moving toward hierarchical estimation



Perform Model-based fMRI





O'Doherty et al. (2007); Gläscher & O'Doherty (2010); Forstmann & Wagenmarkers (2015); Cohen et al. (2017)

A closer look at PE

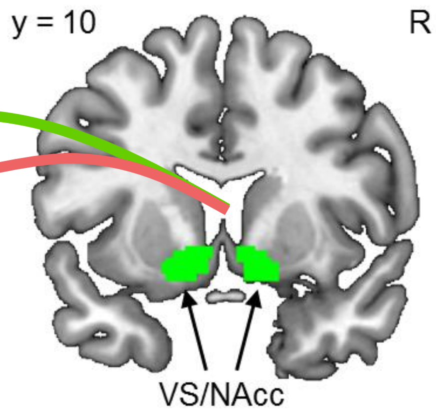
Prediction error:

$$PE = R_t - V_t$$

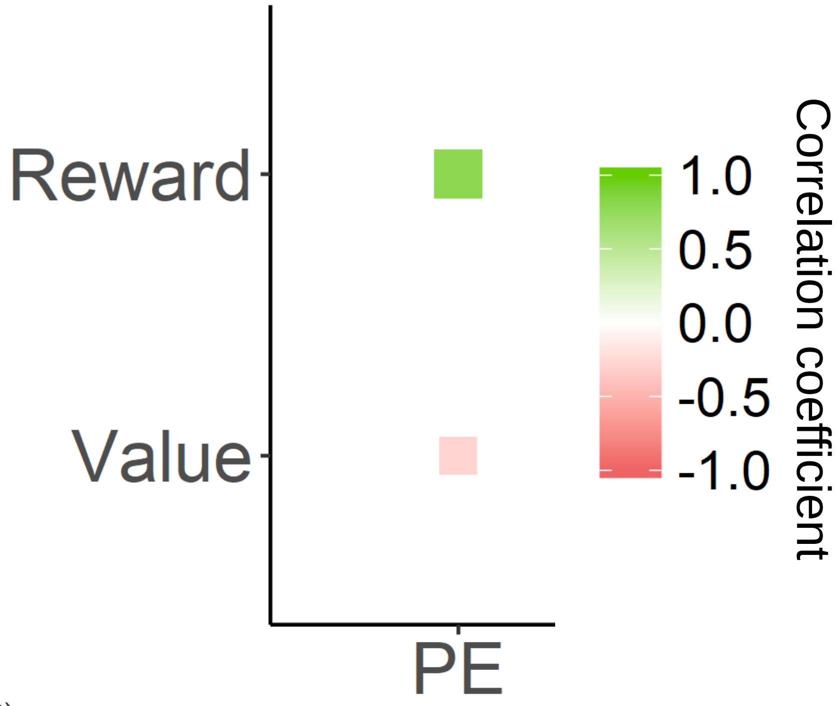
outcome

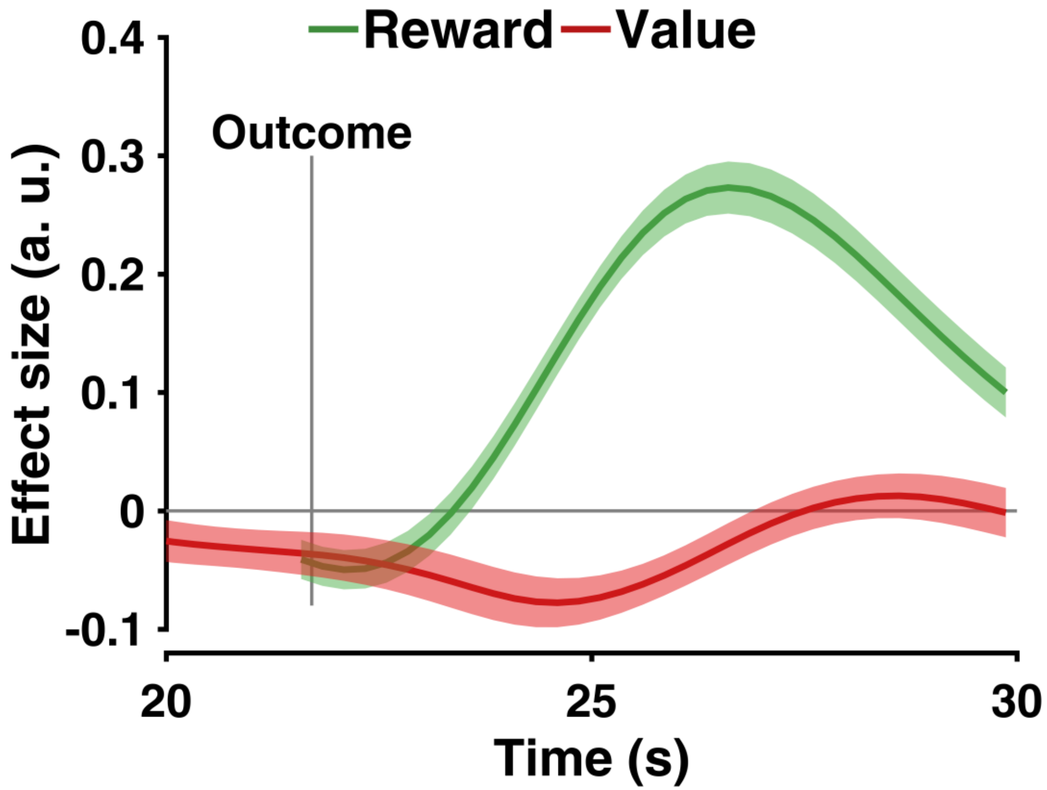
Q: how to justify the striatal activity is indeed associated with PE, rather than reward?

$$\mathrm{PE}_t = R_t - V_t$$



Justify PE signal

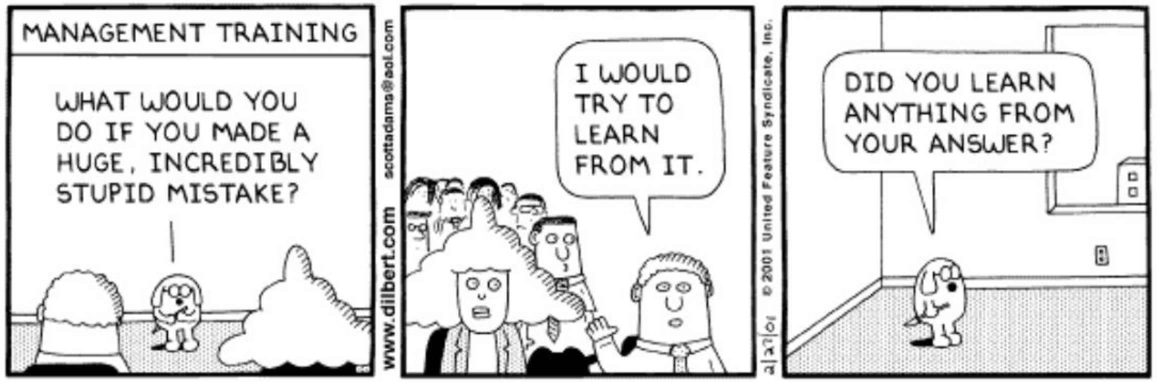




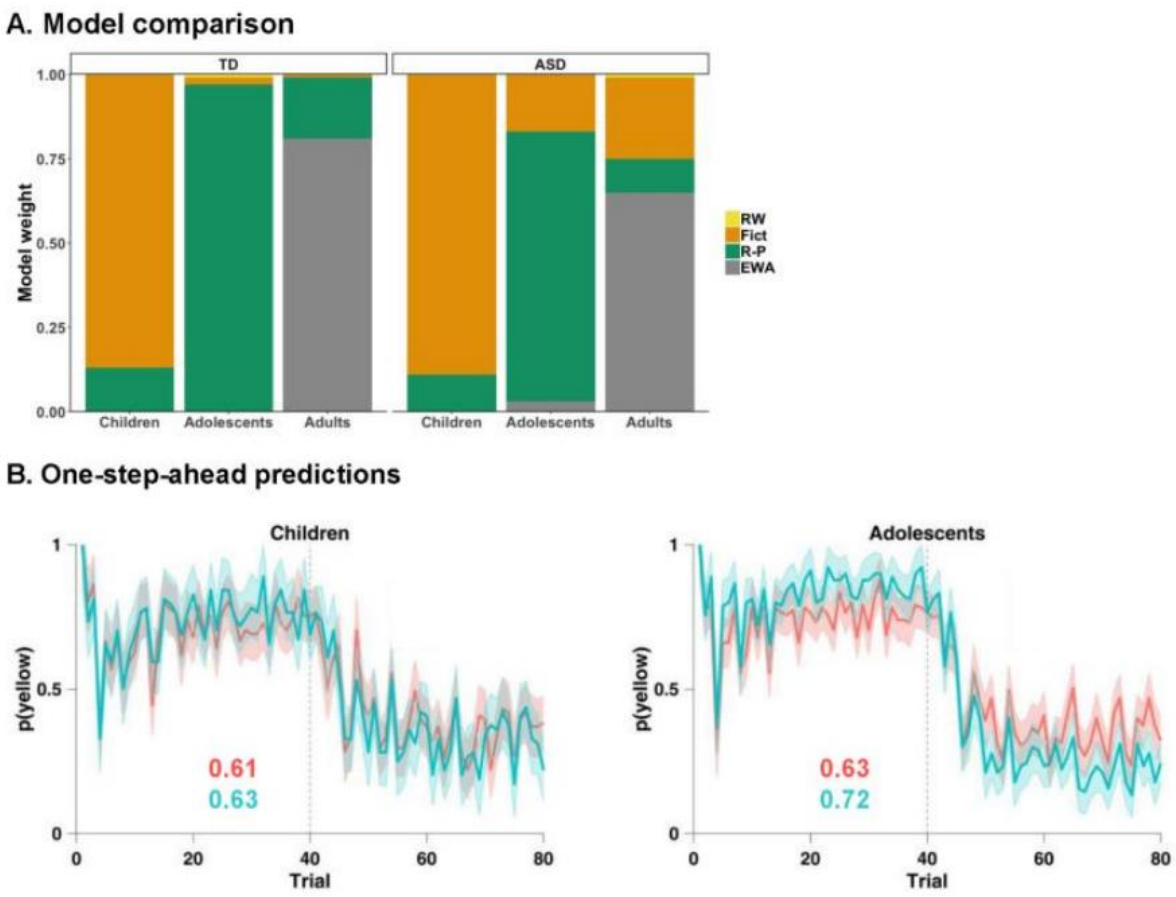
Zhang & Gläscher (2019)

Outline

- the Reinforcement Learning framework
- the learning rate
  - what is it?
  - is there a optimal learning rate
- searching for prediction error signals in the brain
- model validation
- moving toward hierarchical estimation

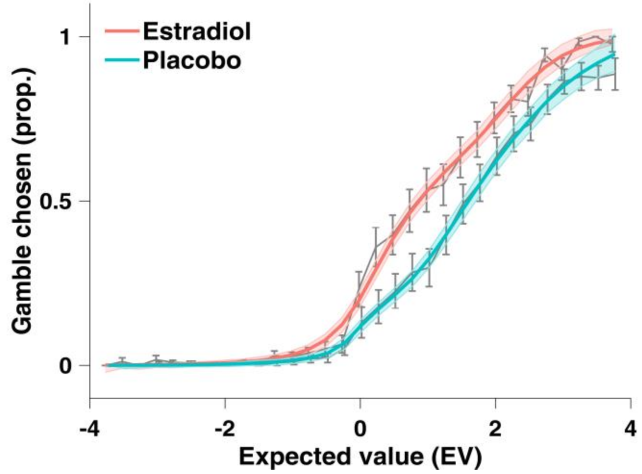


Model comparison + model validation



Model comparison

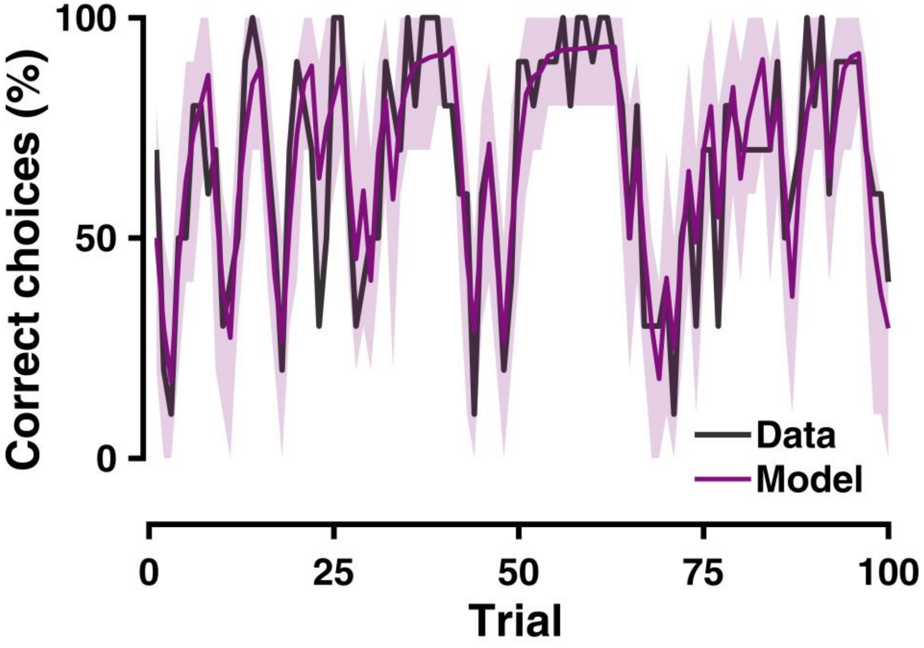
model	Estradiol		Placebo	
	LOOIC	weight	LOOIC	weight
1rho	4258	0.057	3650	0.074
2rho	4000	0.943	3442	0.926



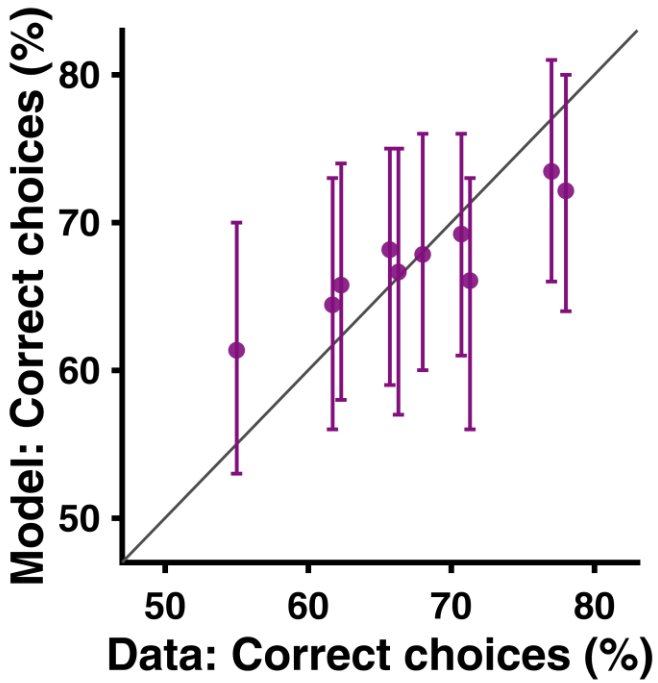
Crawley, Zhang, …, EU-AIMS (2019)

Losecaat Vermeer et al. (in prep.)

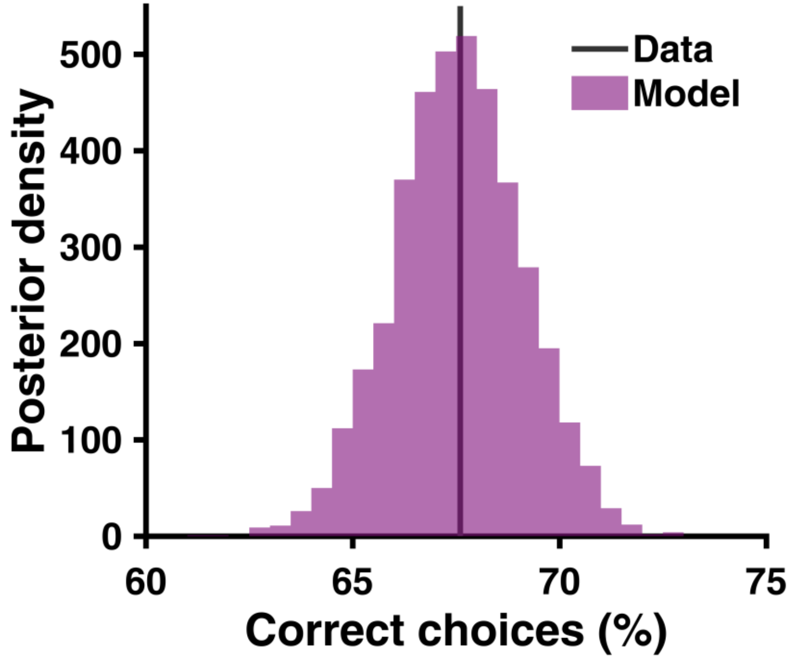
validation with posterior predictive check



trial level



subject level



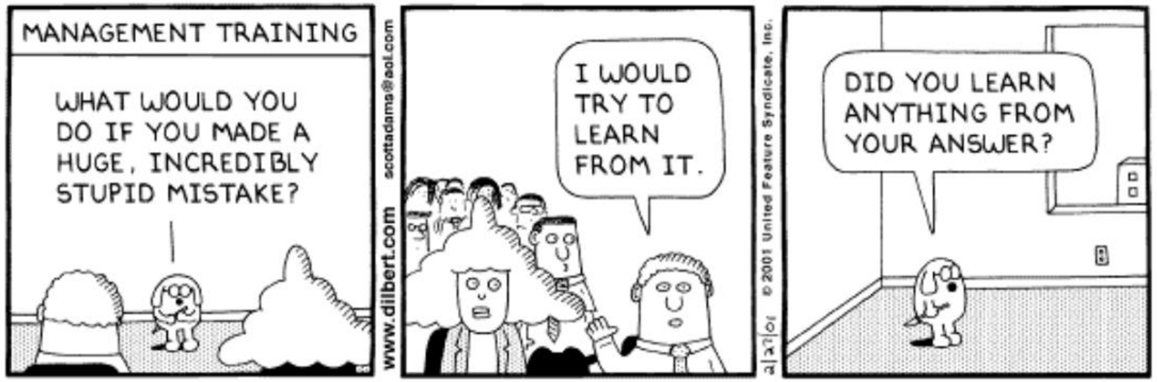
overall level

$$p(y_{\text{rep}} \mid y) = \int p(y_{\text{rep}} \mid \theta) \boxed{p(\theta \mid y)} d\theta$$

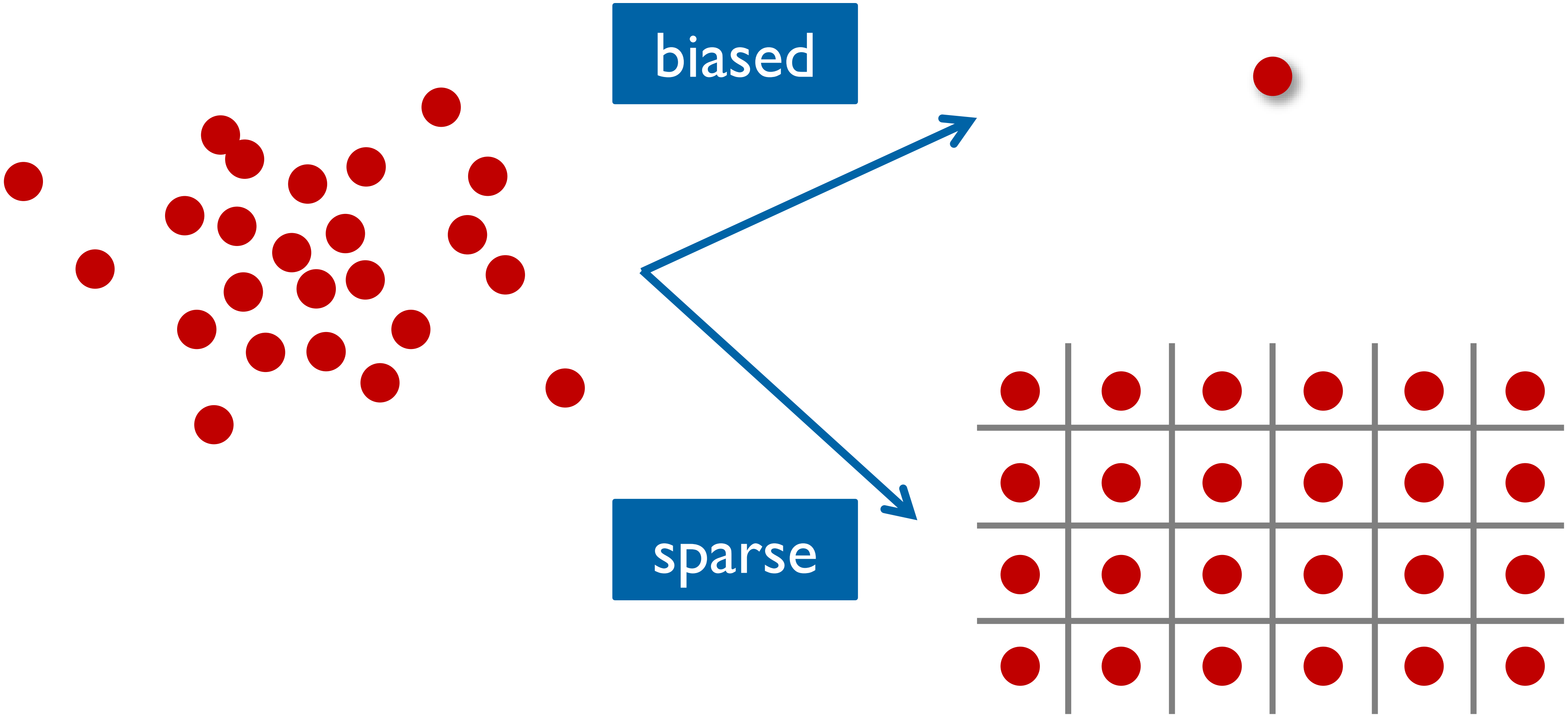
parameter estimates

Outline

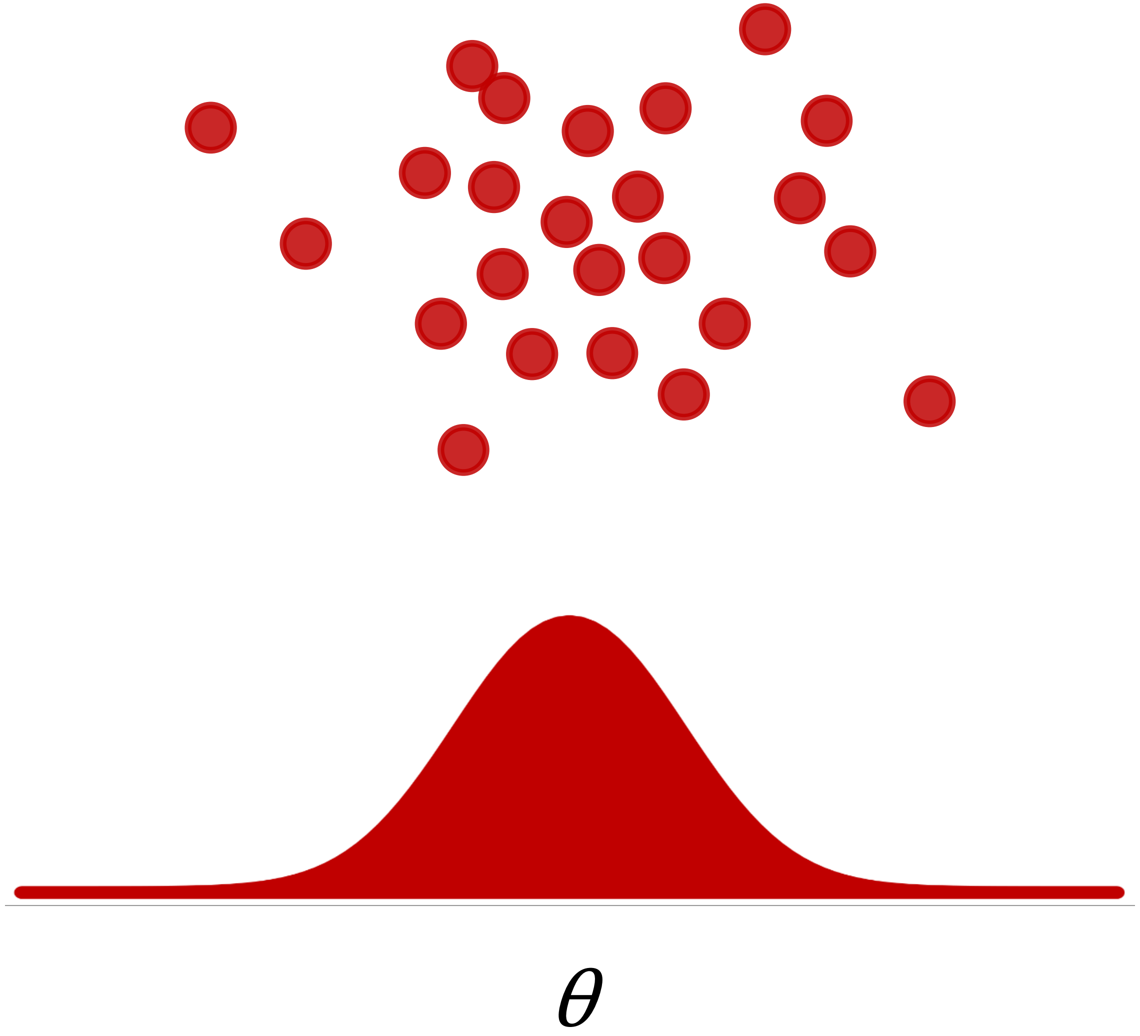
- the Reinforcement Learning framework
- the learning rate
  - what is it?
  - is there a optimal learning rate
- searching for prediction error signals in the brain
- model validation
- moving toward hierarchical estimation

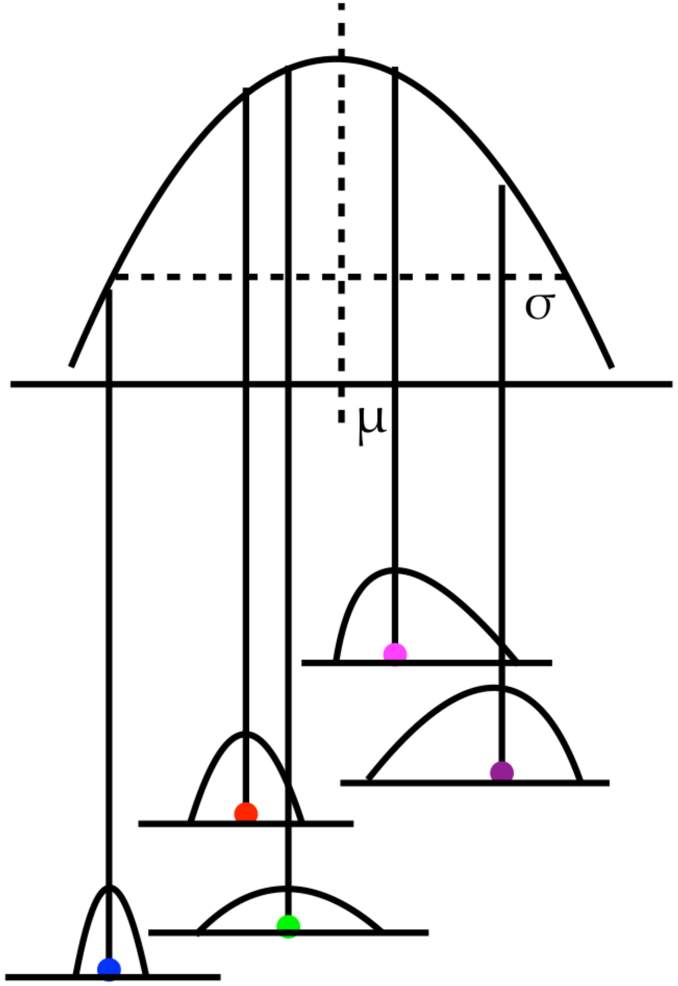


Fitting Multiple Participants



Fitting Multiple Participants





Toolboxes for hierarchical modeling

hBayesDM

hBayesDM (hierarchical Bayesian modeling of Decision-Making tasks) is a user-friendly pa
analysis of various computational models on an array of decision-making tasks. hBayesDM

repo status Active
build passing
CRAN 1.0.1 – 2019-09-01
downloads 20K
DOI 10.1162/CPSY_a_00002

HDDM 0.7.1

```
pip install HDDM
```

payampiray / cbm

<> Code | Issues 0

VBA (Variational Bayesian Analysis)

Interpreting experimental data through computational models

sjgershm / mfit

<> Code | Issues 0

TAPAS

Translational Algorithms for Psychiatry-Advancing Science

PyMC3

Summary

Box 2. Pitfalls and Suggestions

Pitfall 1: High learning rate infers fast learning, hence is more optimal than low learning rate.

Suggestion 1: The learning rate (α) quantifies the extent to which the prediction error is integrated into the value update in reinforcement learning (RL) models. High learning rate indicates vast value update that relies on only recent reward history, whereas low learning rate suggests graduate value update that carries long-lasting effect of outcomes. An "optimal" learning rate can be identified only in combination with the inverse temperature (τ). However, there is no generically optimal combination between α and τ; instead, the optimal combination is affected by the reward schedule, number of trials, the presence of reversals, and so on.

Pitfall 2: Nucleus accumbens (NAcc) encodes both reward prediction error and outcome valence.

Suggestion 2: In RL models, reward (R) and prediction errors (PE) are by definition positively correlated. However, the negative correlation between PE and value signal (*V*) is often overlooked, and these two theoretical subcomponents (i.e., R and *V*) of PE are, in fact, crucial to assess the neural substrates of PE. To qualify as a region encoding the PE signal, activities in NAcc ought to covary positively with the actual outcome (i.e., R) and negatively with the expectation (i.e., *V*).

Pitfall 3: Model comparison selects the winning model and validates model performance.

Suggestion 3: Model comparison is helpful in picking the best model, but it provides merely relative performance among candidate models. To validate model performance, one needs to examine whether the winning model's posterior prediction is able to replicate key features of the observed data. This procedure is called posterior predictive check (PPC). To perform PPC, let the model generate observations (e.g., choices) from the joint posterior densities of model parameters, and then assess whether the generated data could reproduce the behavioral pattern (e.g., choice accuracy) as in the behavioral analysis. Unsuccessful PPC is as valuable as successful ones, because they may help falsify a model construction and eventually facilitate model development.

well, how could I do it?

29 commits | 1 branch | 0 releases | 1 contributor

Branch: master | New pull request | Create new file | Upload files | Find file | Clone or download

lei-zhang update — Latest commit 5401250 39 seconds ago

code	update	2 days ago
data	update	2 days ago
README.md	update	39 seconds ago

README.md

Code and data for Zhang^, Lengersdorff^, Mikus, Gläscher, & Lamm (2019, PsyArXiv). Frameworks, pitfalls, and suggestions of using reinforcement learning models in social neuroscience. (^Equal contributions)

This repository contains:

```
root
  ├─ code      # Matlab & R code to run the analyses and produce figures
  ├─ data      # behavioral & fMRI data
```

Note 1: to properly run all scripts, you may need to set the root of this repository as your work directory.
Note 2: to reproduce the Matlab figures, you may need the color brewer toolbox and the offsetAxes function.

Happy Computing!