

Reinforcement learning: From theory to implementation

Lei Zhang

Institute of Systems Neuroscience, University Medical Center Hamburg-Eppendorf

Outline

Two-armed bandit task
Rescorla-Wagner value update
Choice rule
The data: PRL
Trial-by-trial model fitting
Hierarchical Bayesian estimate

Goals:

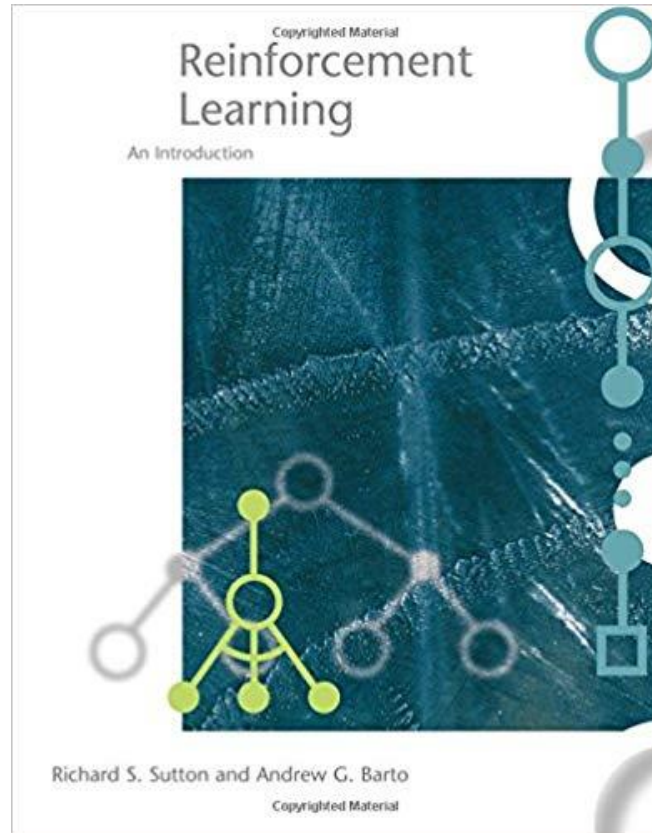
Understand how models are developed.
Start here, know where you can go from it.

The very short history

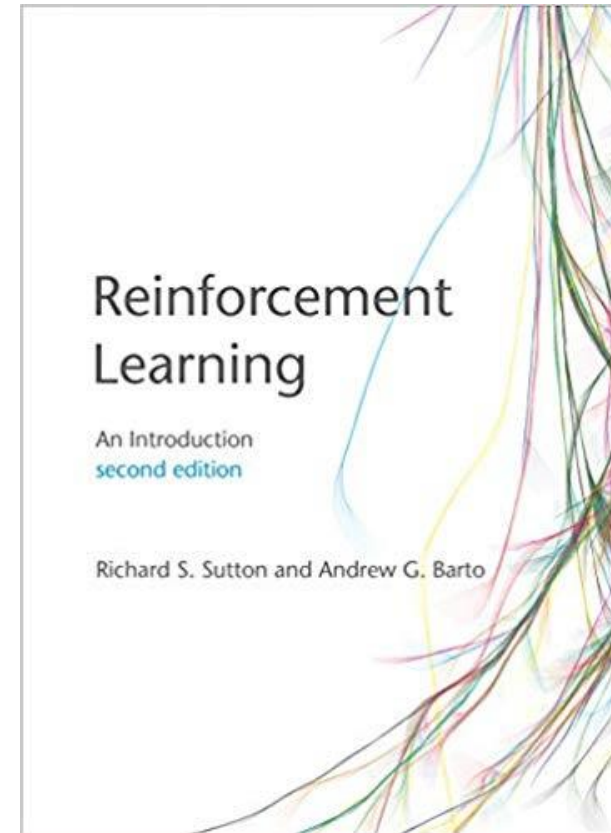
cognitive model

statistics

computing



1998



2018

Boom in Cognitive Modeling

cognitive model

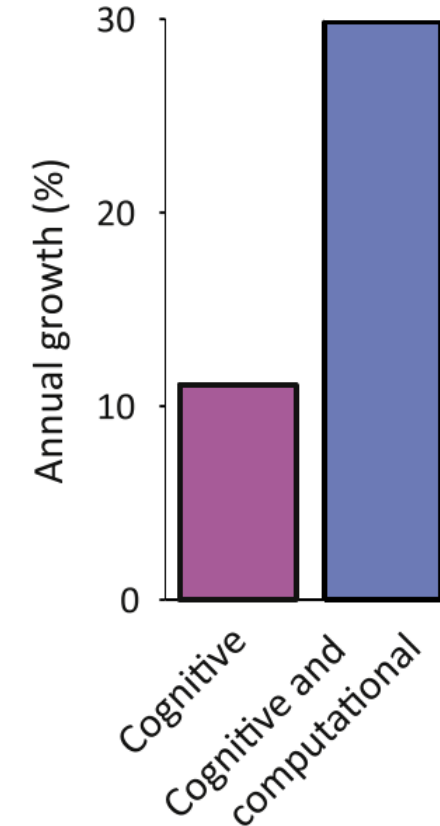
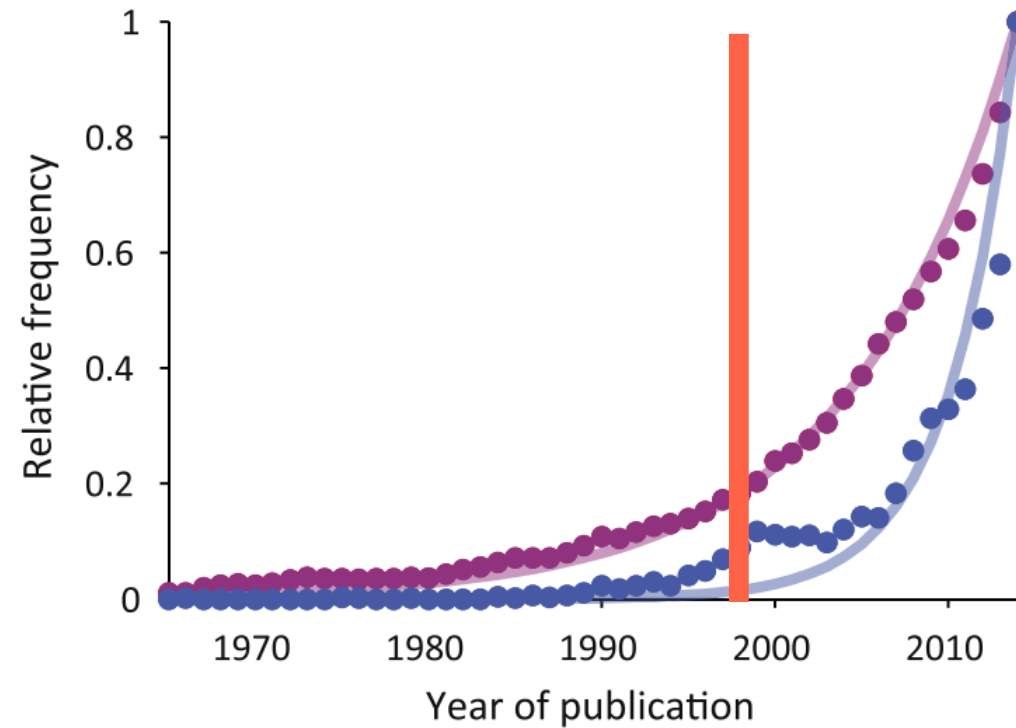
statistics

computing

(A)

■ Cognitive
■ Cognitive and computational

Source: PubMed



Icebreaker

Please describe what RL is to you in 1 – 2 sentences.

The keywords below may help:

reward

action

update

...

From RL to cognitive (neuro)science

cognitive model

statistics

computing

Reinforcement learning (RL) exemplifies **two (related) ways** that computer science informs cognitive (neuro)science

conceptual

- how to characterize hard problems (formally analyzable tasks)
- optimal (typically intractable) solution
- approximate algorithms and their properties
 - algorithms as **hypotheses**
 - common **process-level** explanation for different kinds of data

analytical

- algorithms as **likelihood functions** for inference from data
- data analysis as statistical machine learning

2-armed bandit task

cognitive model

statistics

computing



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

cognitive model

statistics

computing



What can be your strategies:

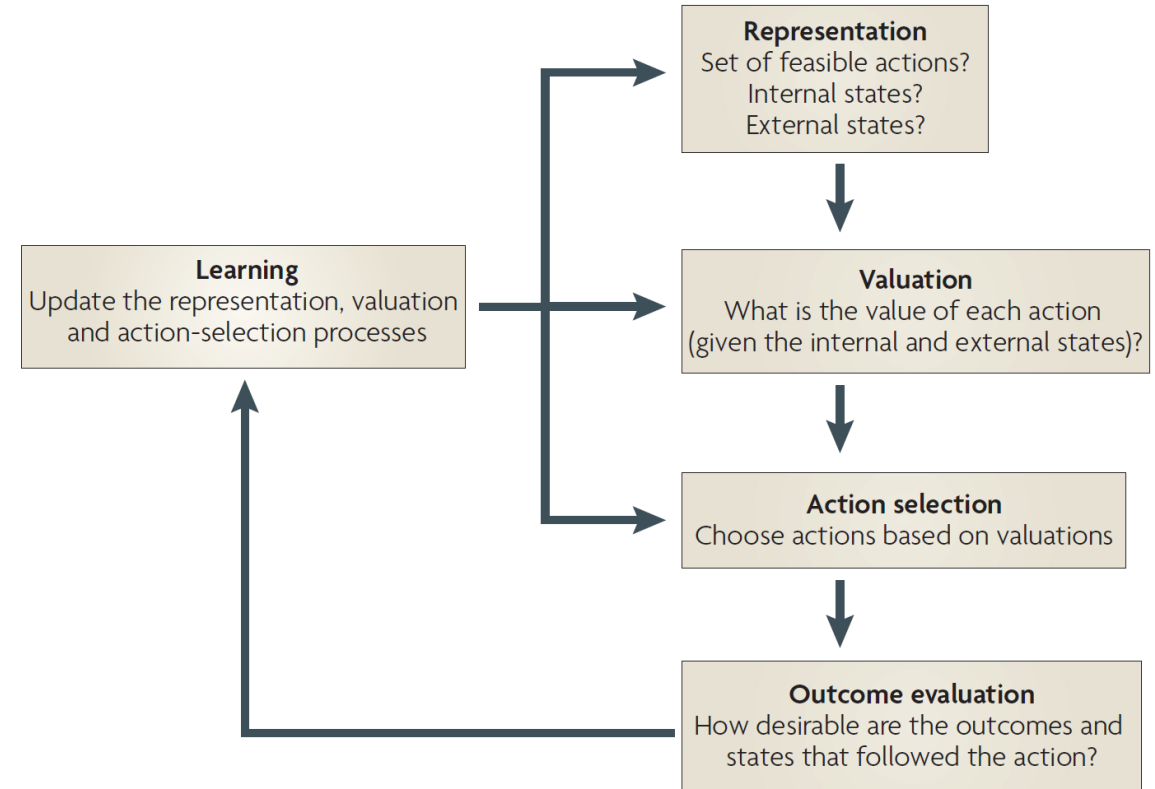
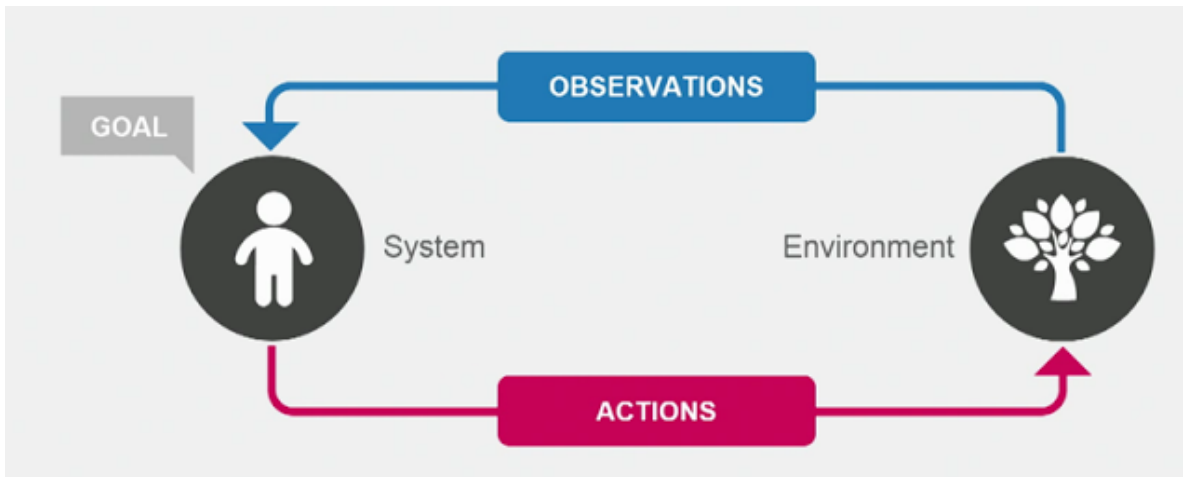
1. **predict** the value of each deck
2. **choose** the best
3. **learn** from outcome to update predictions (repeat)

How prediction is shaped by learning?

cognitive model

statistics

computing



Modeling the 2-armed bandit task



how do you suggest to model this learning process?

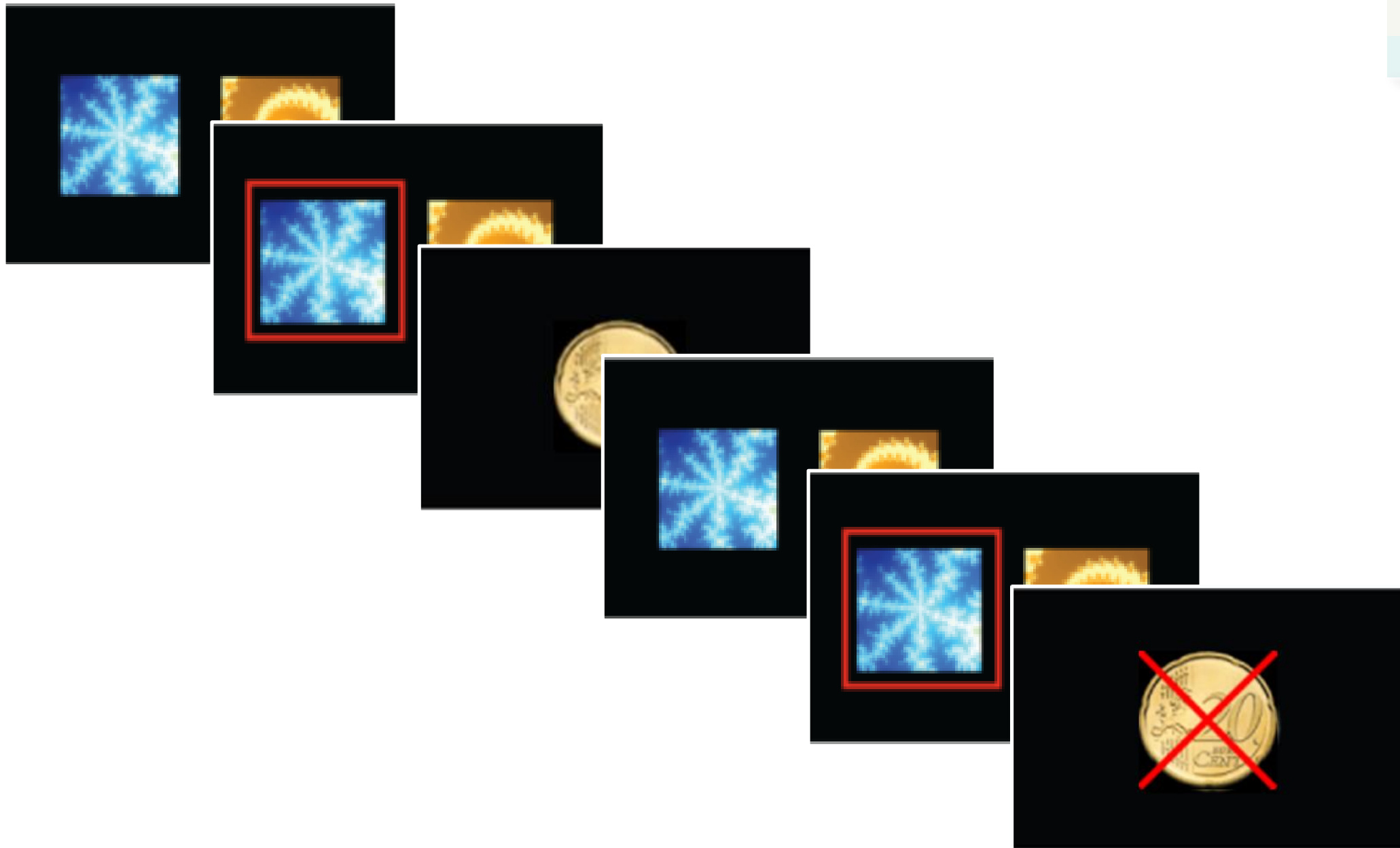
suppose we ran this experiment on a person

our models are basically detailed hypotheses about behavior and about the brain... we can test these hypotheses!

cognitive model

statistics

computing

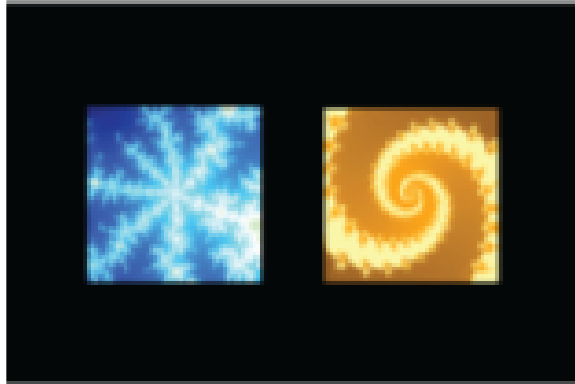


One simplified experiment

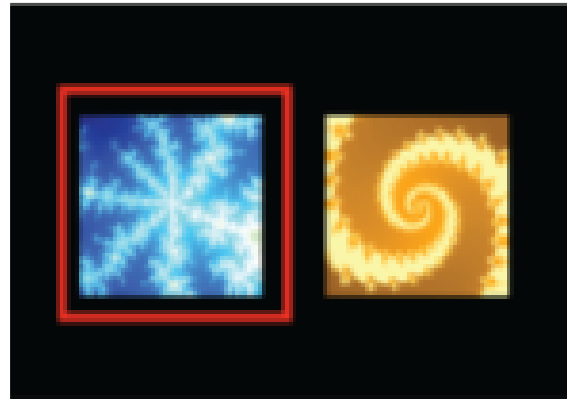
cognitive model

statistics

computing



choice
presentation



action
selection



outcome

reward contingency – 80:20

Elements

cognitive model

statistics

computing

what do we know?

Data: choice & outcome

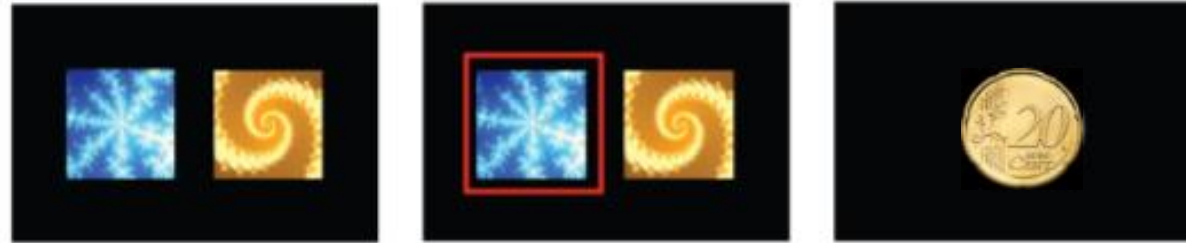
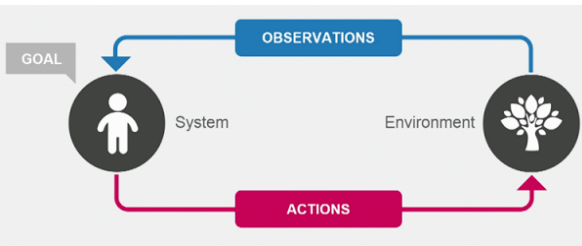
what can we measure?

Summary stats: choice accuracy

what do we not know?

Learning algorithm: RL update

Rescorla-Wagner Value Update



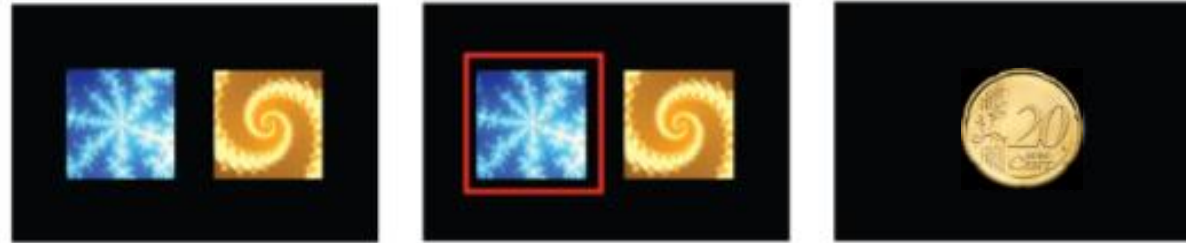
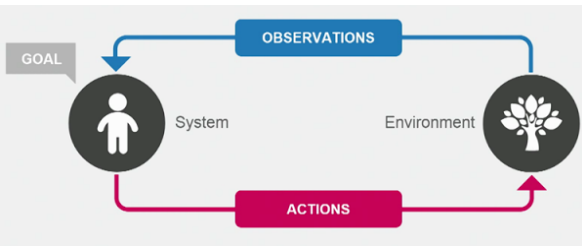
Cognitive Model

- cognitive process
- using internal variables and free parameters

Observation Model (Data Model)

- relate model to observed data
- has to account for noise

Rescorla-Wagner Value Update



Value update:

$$V_{t+1} = V_t + \alpha * PE$$

Prediction error:

$$PE = R_t - V_t$$

α - learning rate
PE - reward prediction error
V - value
R - reward

Understand the learning rate

cognitive model

statistics

computing

Value update:

$$V_{t+1} = V_t + \alpha * PE$$

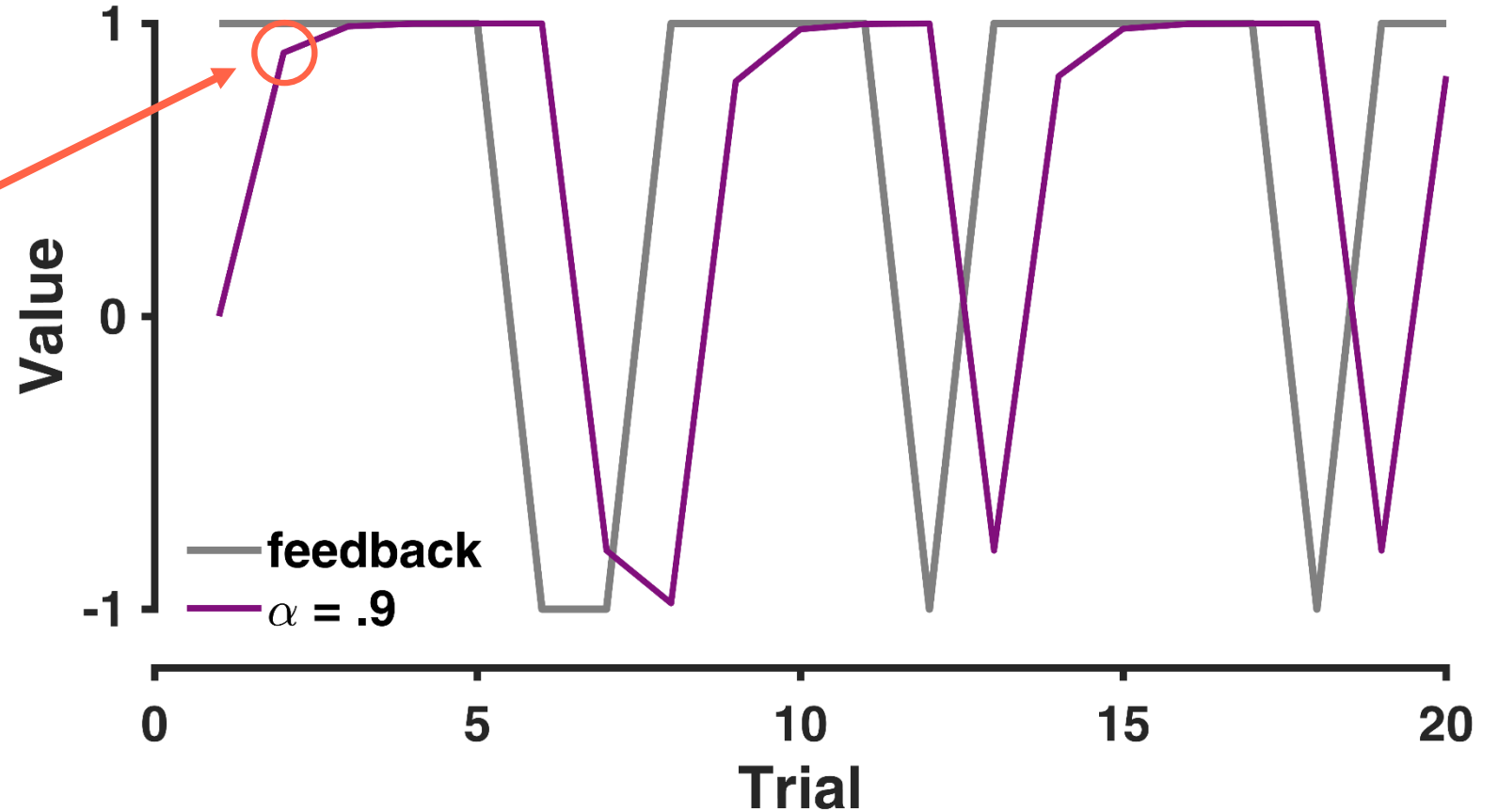
Prediction error:

$$PE = R_t - V_t$$

if $\eta = 0.9$

$$V_1 = 0$$

$$\begin{aligned} V_2 &= V_1 + 0.9 * (1 - 0) \\ &= 0 + 0.9 \\ &= 0.9 \end{aligned}$$



reward contingency – 80:20

Understand the learning rate

cognitive model

statistics

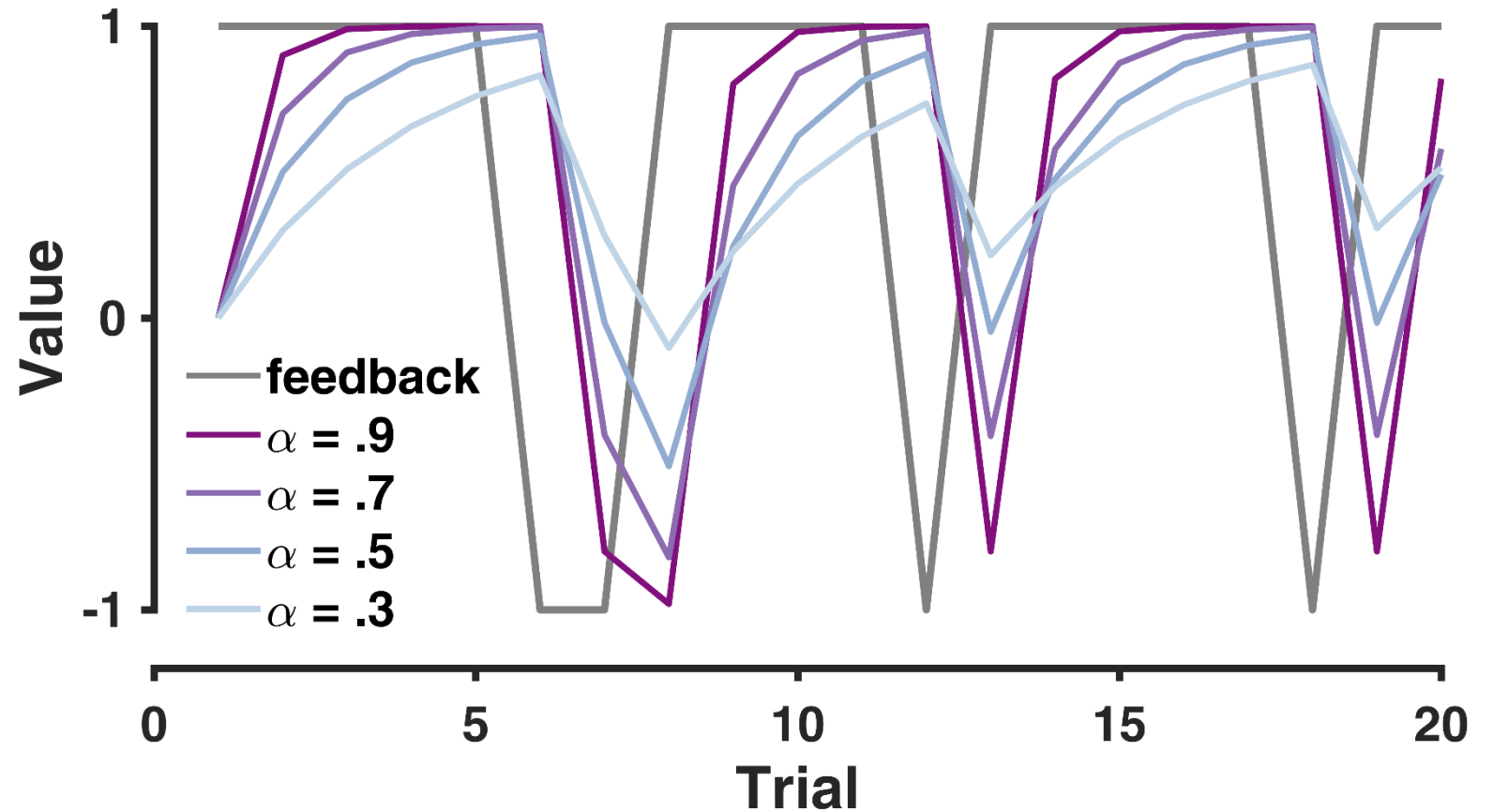
computing

Value update:

$$V_{t+1} = V_t + \alpha * PE$$

Prediction error:

$$PE = R_t - V_t$$



reward contingency – 80:20

Understand the learning rate

cognitive model

statistics

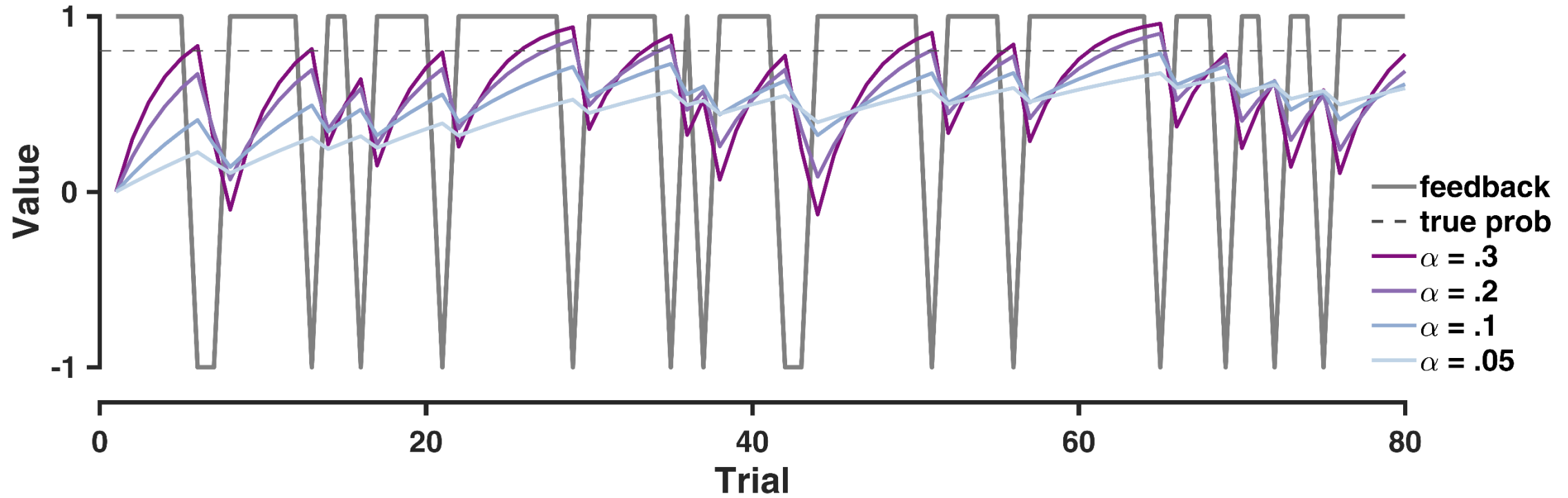
computing

Value update:

$$V_{t+1} = V_t + \alpha * PE$$

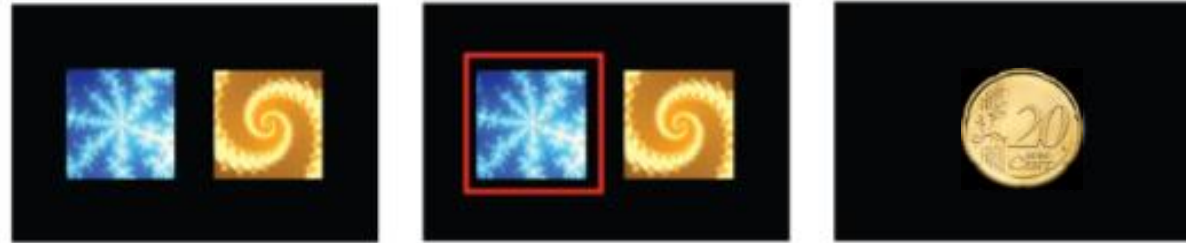
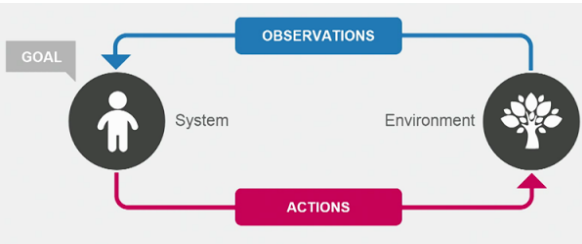
Prediction error:

$$PE = R_t - V_t$$



reward contingency – 80:20

Rescorla-Wagner Value Update



Value update:

$$V_{t+1} = V_t + \alpha * PE$$

Prediction error:

$$PE = R_t - V_t$$

choice rule:

greedy / ϵ -greedy / softmax

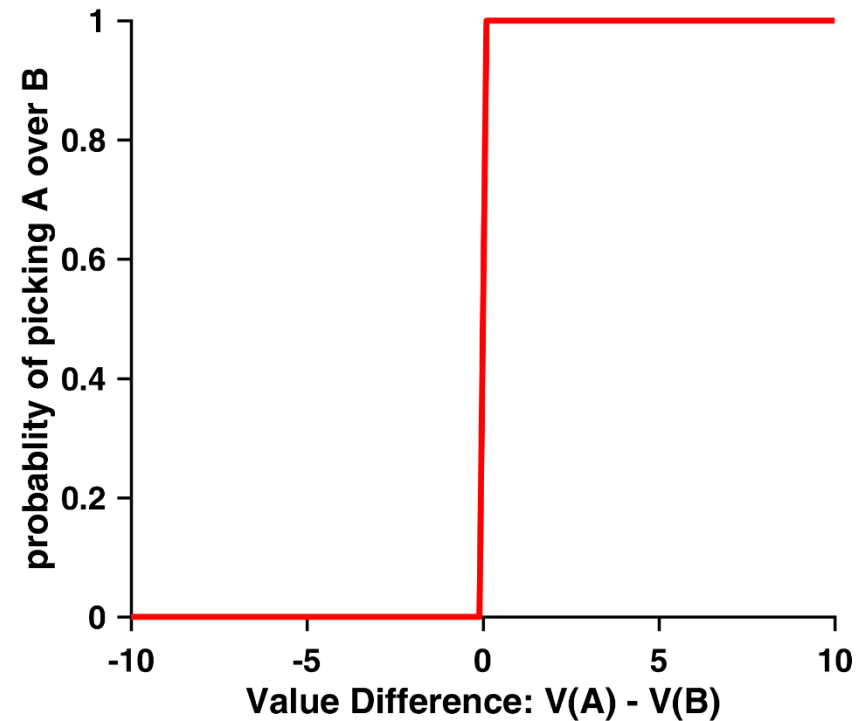
Choice rule: greedy

cognitive model

statistics

computing

$$p(C = a) = \begin{cases} 1, & V(a) > V(b) \\ 0, & V(a) < V(b) \end{cases}$$



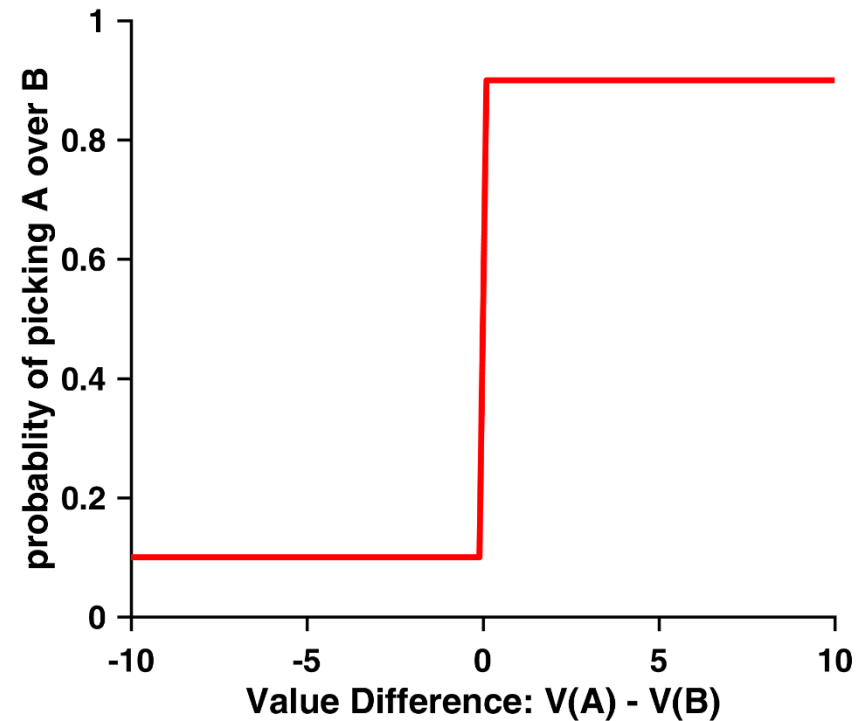
Choice rule: ϵ -greedy

cognitive model

statistics

computing

$$p(C = a) = \begin{cases} 1 - \epsilon, & V(a) > V(b) \\ \epsilon & , V(a) < V(b) \end{cases}$$



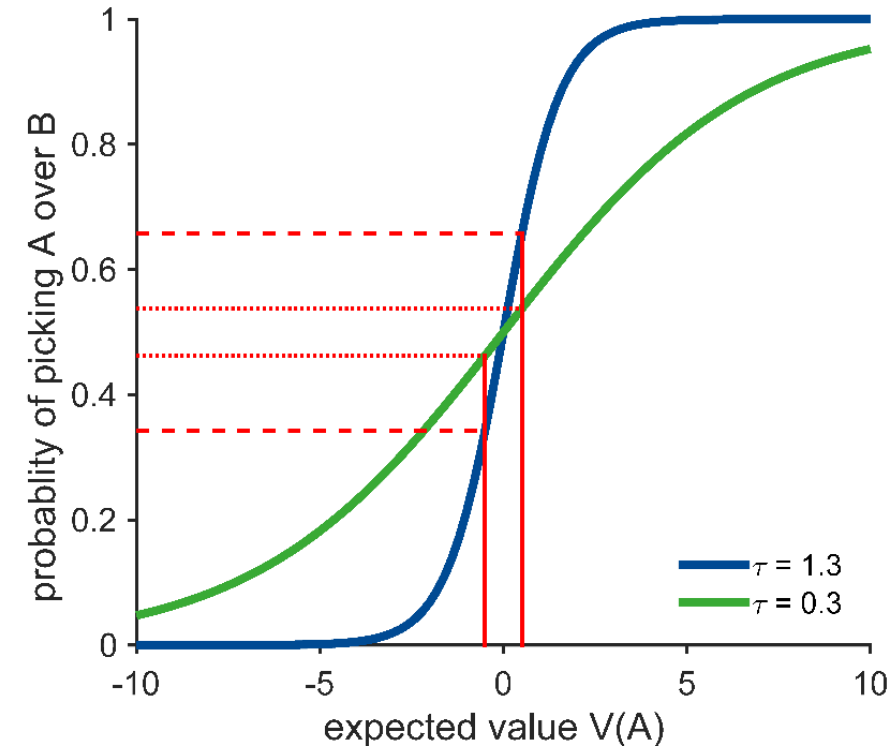
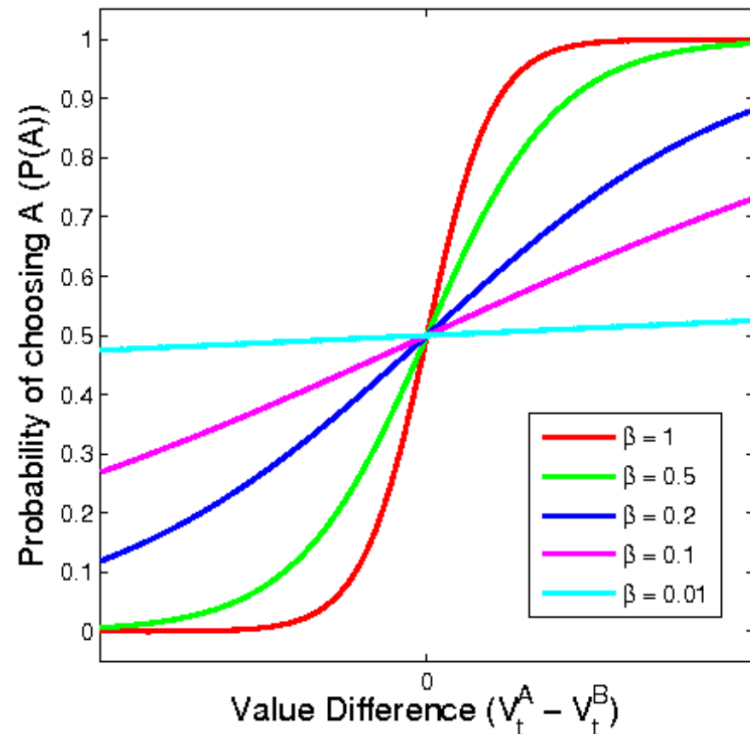
Choice rule: softmax

cognitive model

statistics

computing

$$p(C = a) = \frac{e^{\tau * V(a)}}{e^{\tau * V(a)} + e^{\tau * V(b)}} = \frac{1}{1 + e^{-\tau * (V(a) - V(b))}}$$

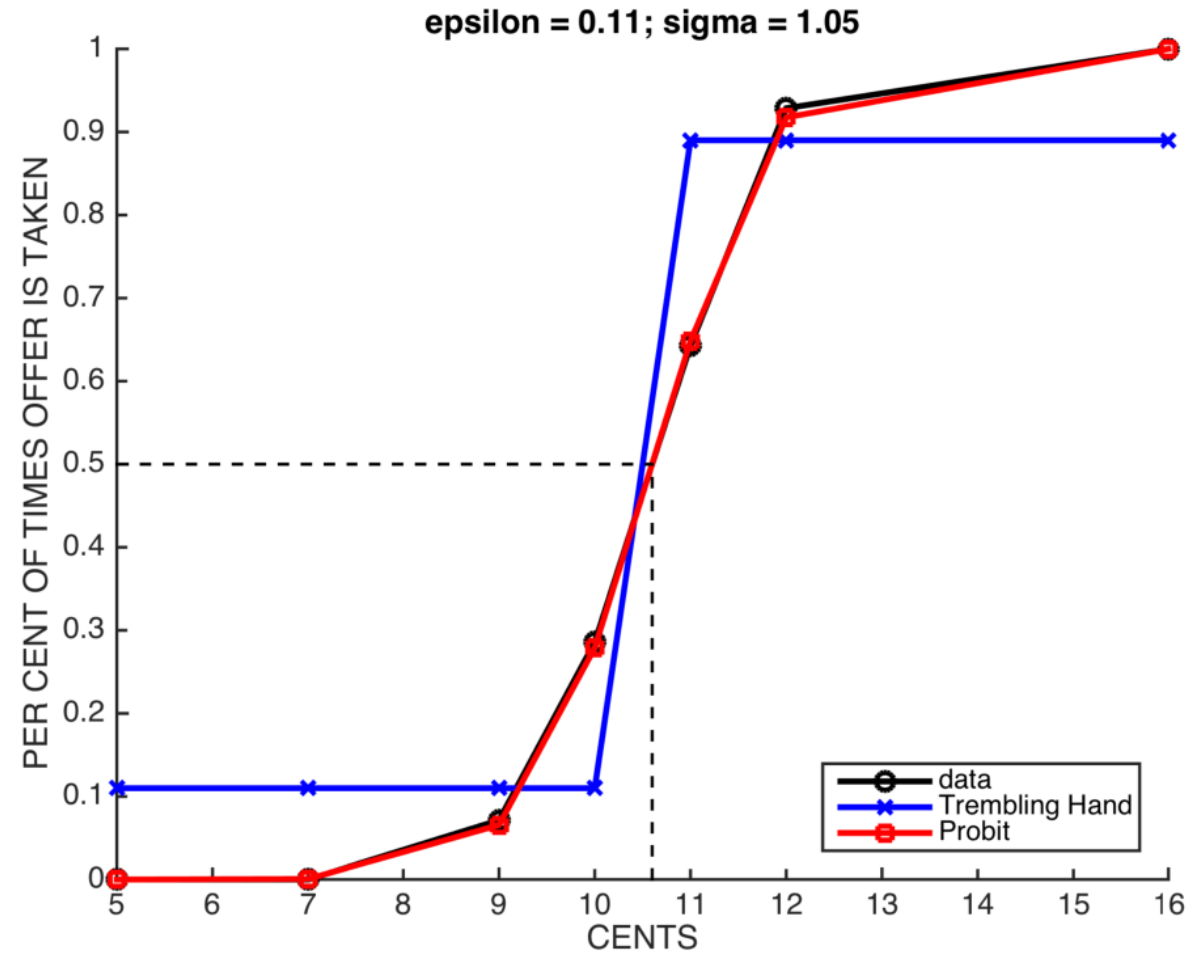


Choice rule: direct comparison

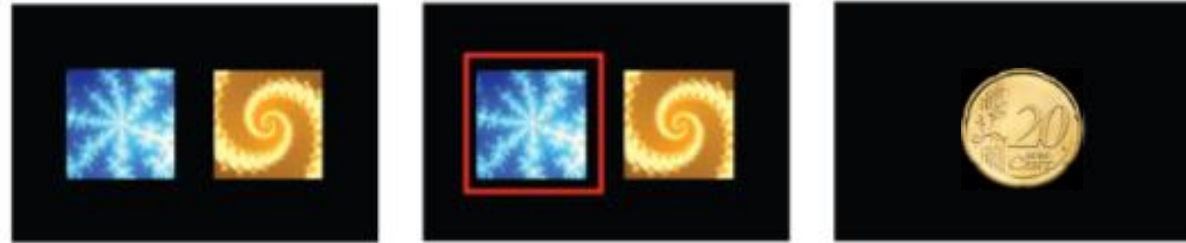
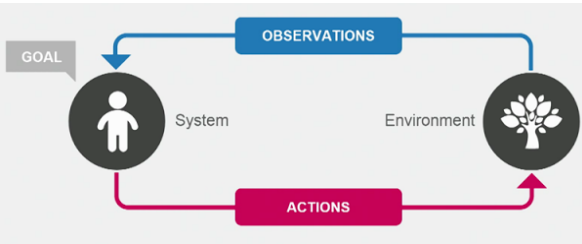
cognitive model

statistics

computing



Rescorla-Wagner Value Update



Value update:

$$V_{t+1} = V_t + \alpha * PE$$

Prediction error:

$$PE = R_t - V_t$$

choice rule (sigmoid /softmax):

$$p(C=a) = \frac{1}{1 + e^{\tau * (v(b) - v(a))}}$$

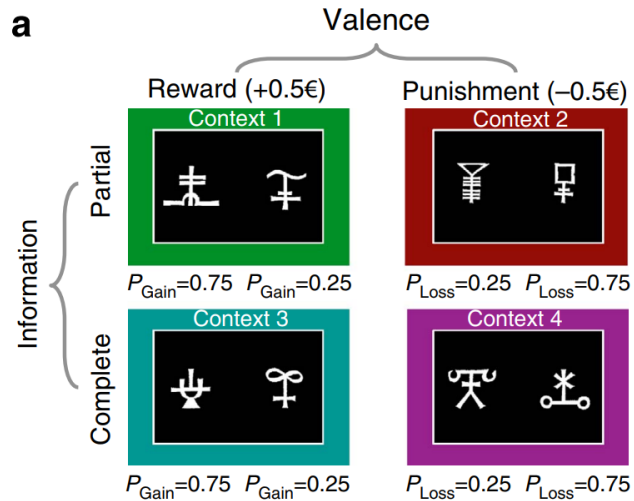
- α - learning rate
- PE - reward prediction error
- V - value
- R - reward
- τ - softmax temperature

Generalizing RL framework

cognitive model

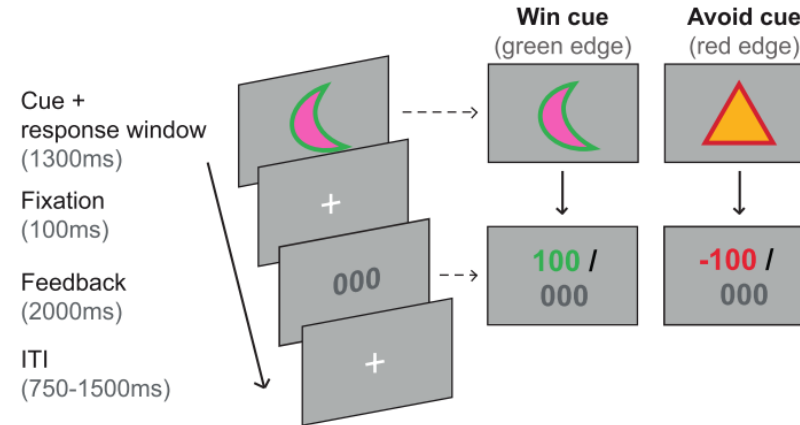
statistics

computing

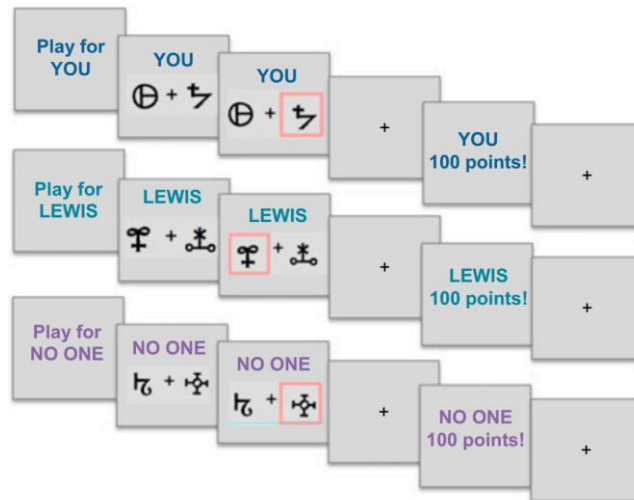


Palminteri et al. (2015)

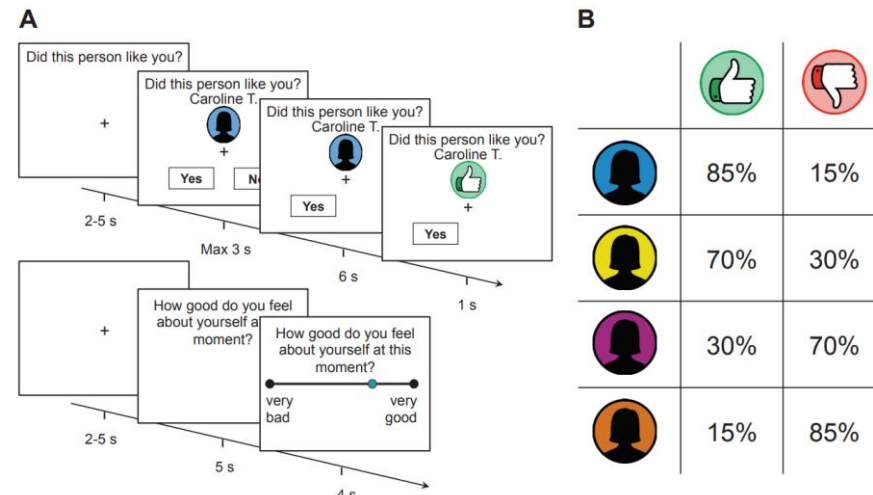
A. Trial details



Swart et al. (2017)



Lockwood et al. (2016)



Will et al. (2017)

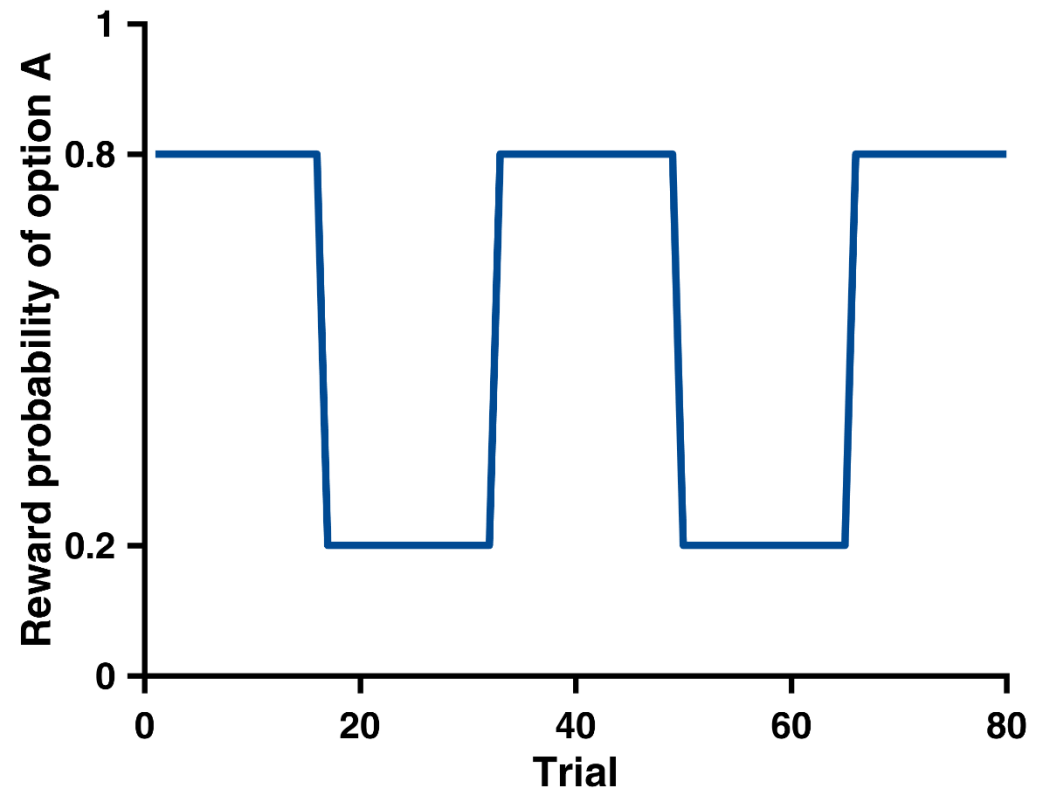
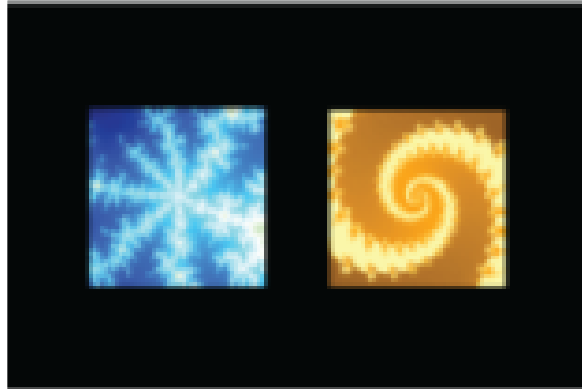
The real task

cognitive model

statistics

computing

PRL: probabilistic reversal learning



The data

cognitive model

statistics

computing

- PRL task
- nSub = 10
- nTrial = 80

./_data/_raw_data/sub01/raw
_data_sub01.txt

sub01
sub02
sub03
sub04
sub05
sub06
sub07
sub08
sub09
sub10

subjID, trialID, choice, outcome, correct
1,1,2,-1,1
1,2,1,1,1
1,3,1,1,1
1,4,1,1,1
1,5,2,-1,1
1,6,1,1,1
1,7,1,1,1
1,8,1,1,1
1,9,1,-1,1
1,10,2,-1,1
1,11,1,1,1
1,12,1,1,1
1,13,1,-1,2

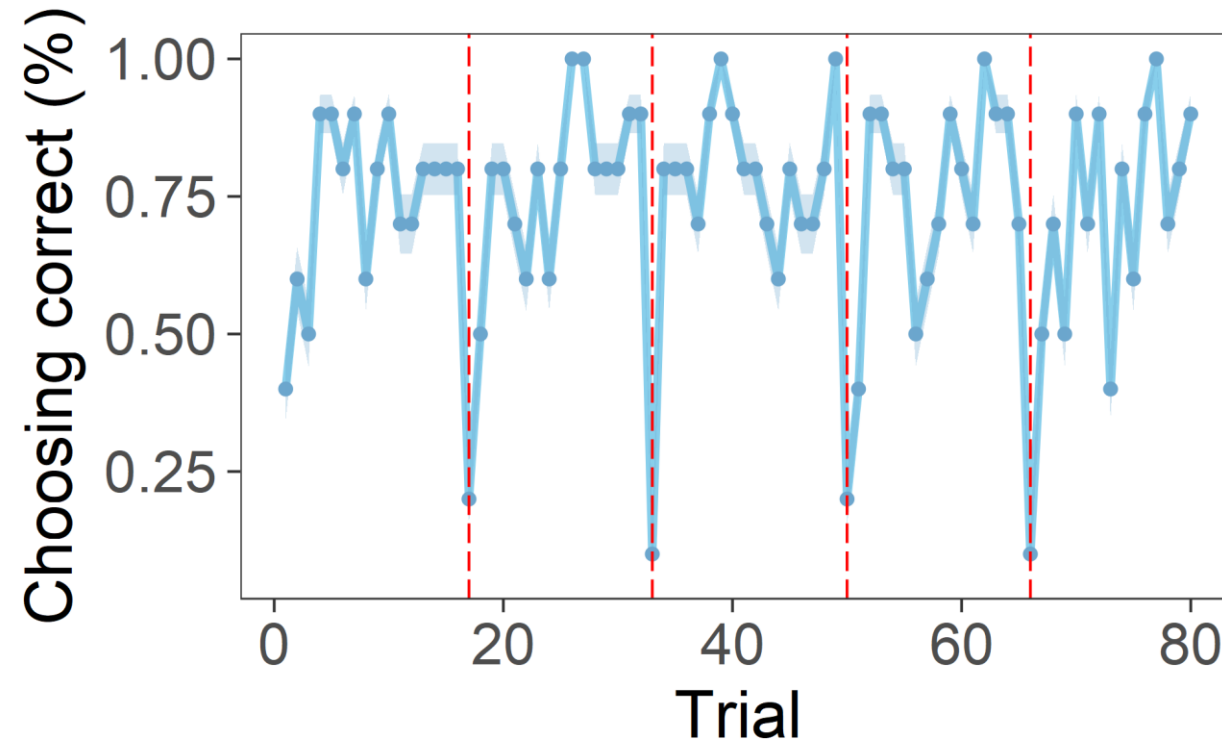
Model-free analysis: summary stats

cognitive model

statistics

computing

How often participants choose the more rewarding option?



Exercise 1

cognitive model

statistics

computing

```
.../RL_tutorial/_scripts/MF_analysis.R
```

TASK:

write a for loop

... which reads in each participant's raw data

... and reshape it in the “long format”

```
for ( j in 1:n) {  
  read.table(file, header = T, sep = ",")  
}
```

Model-based analysis: parameter estimate

cognitive model

statistics

computing

Why estimate parameters?

- May measure **quantities of interest** (learning rates in different populations, how variance in the task affects learning rate etc.)
- Have to use these to generate **hidden(latent) variables** of interest (eg. prediction errors) in order to look for these in the brain

Parameter in light of data

cognitive model

statistics

computing

Likelihood

How plausible is the data given our parameter is true?

Prior

How plausible is our parameter before observing the data?

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

Posterior

How plausible is our parameter given the observed data?

Evidence

How plausible is the data under all possible parameters?

Estimation technique

cognitive model

statistics

computing

$$p(\theta | D) = \frac{p(D | \theta) p(\theta)}{\int p(D | \theta^*) p(\theta^*) d\theta^*}$$

$$p(\theta | D) \propto p(D | \theta) p(\theta)$$

Deterministic
Approximation

→ Variational Bayes

Stochastic
Approximation

→ Sampling Methods

MCMC Sampling Algorithms

cognitive model

statistics

computing

- Rejection sampling
- Importance sampling
- Metropolis algorithm
- Gibbs sampling → JAGS
- HMC sampling*



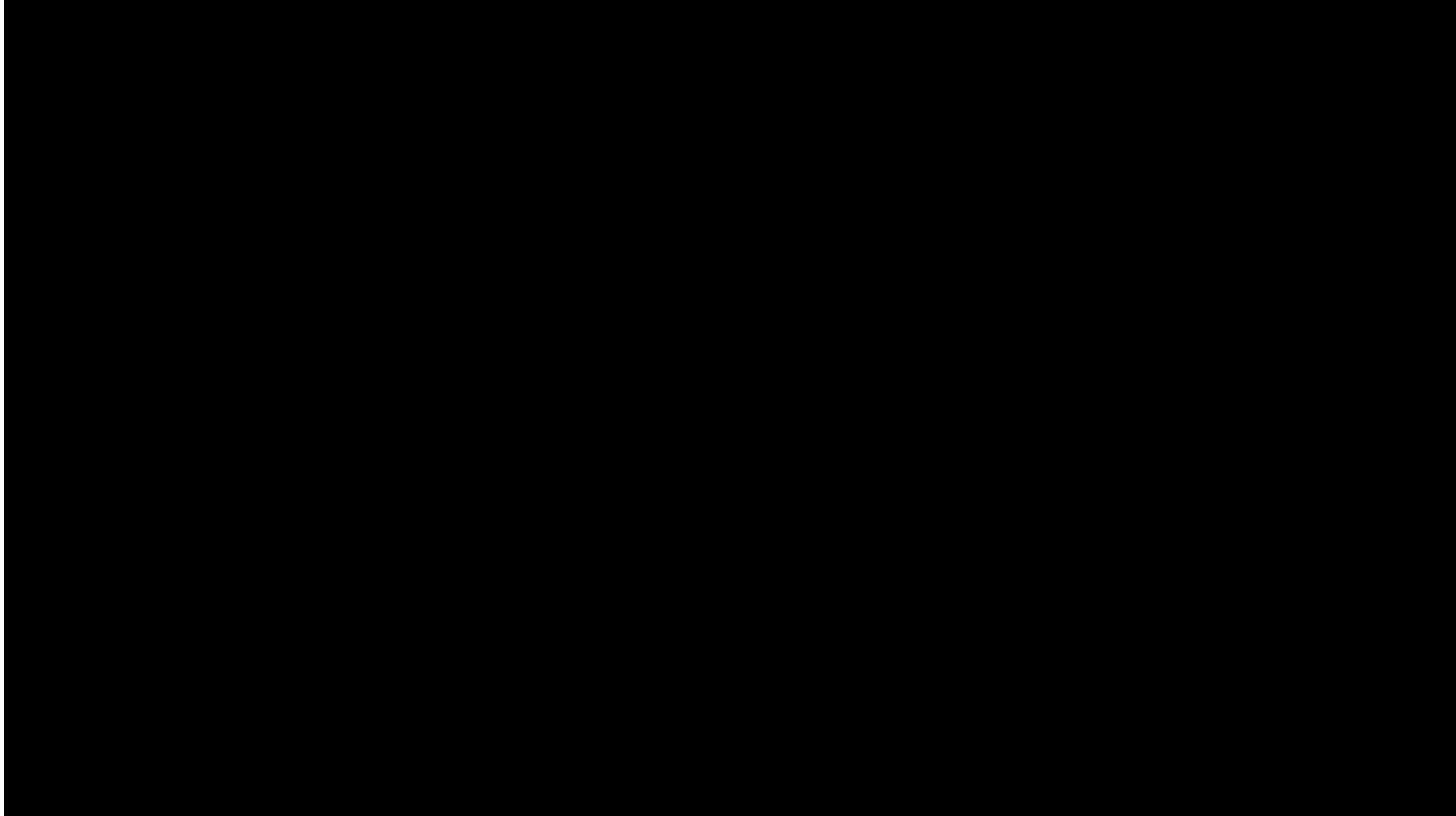
Stan!

Let's watch a video!

cognitive model

statistics

computing

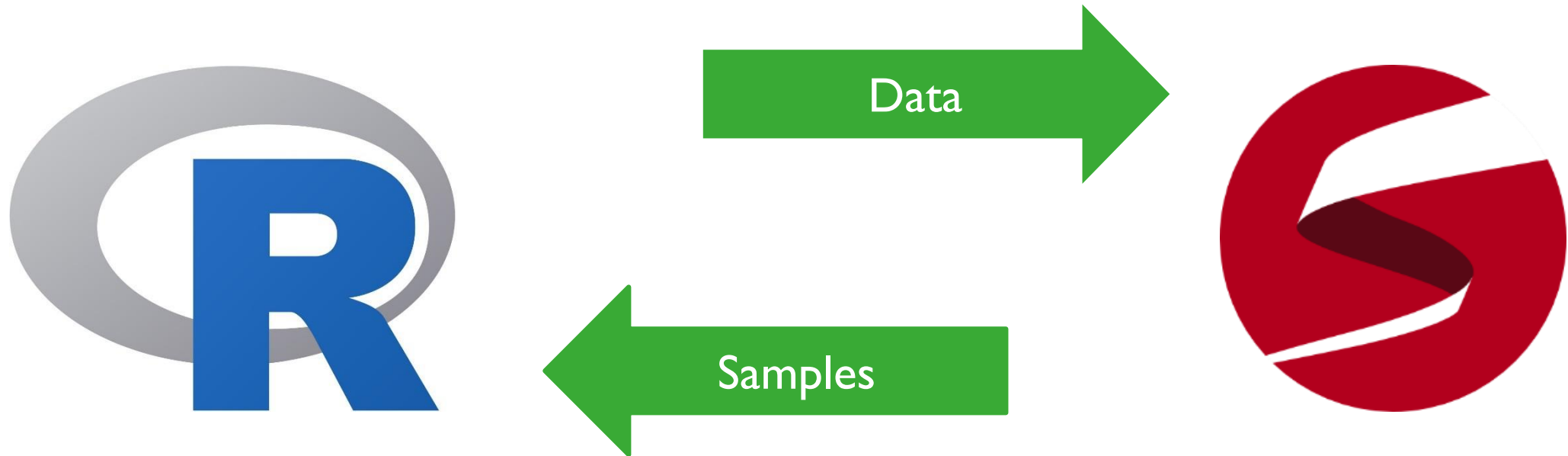


Stan and RStan

cognitive model

statistics

computing

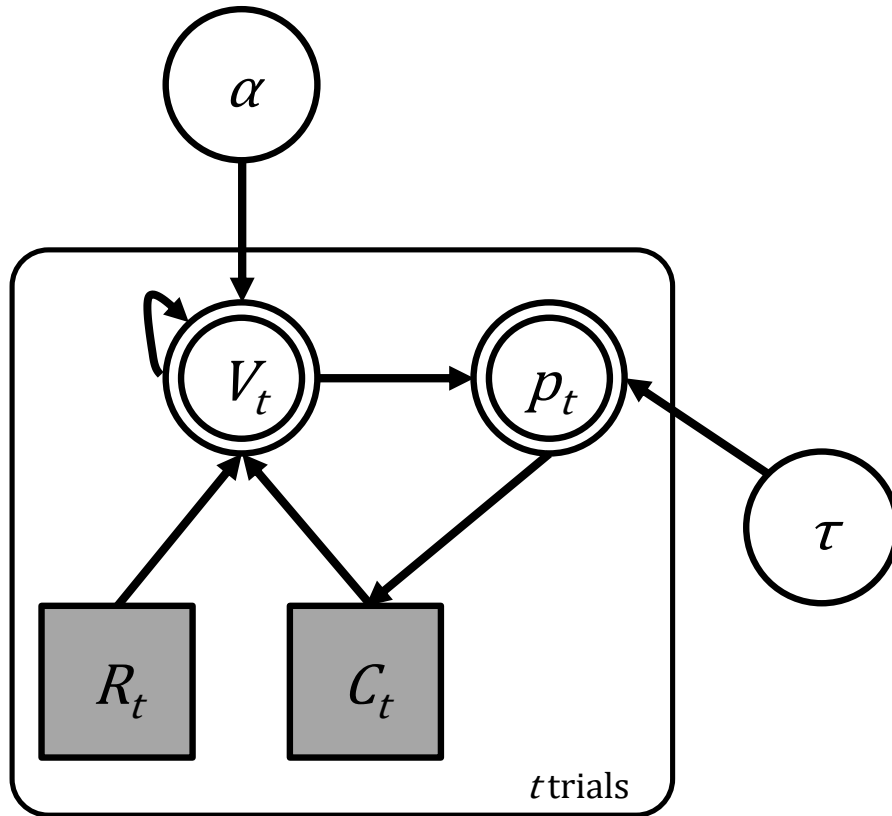


RL – Implementation

cognitive model

statistics

computing



$$\alpha \sim \text{Uniform}(0, 1)$$

$$\tau \sim \text{Uniform}(0, 3)$$

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

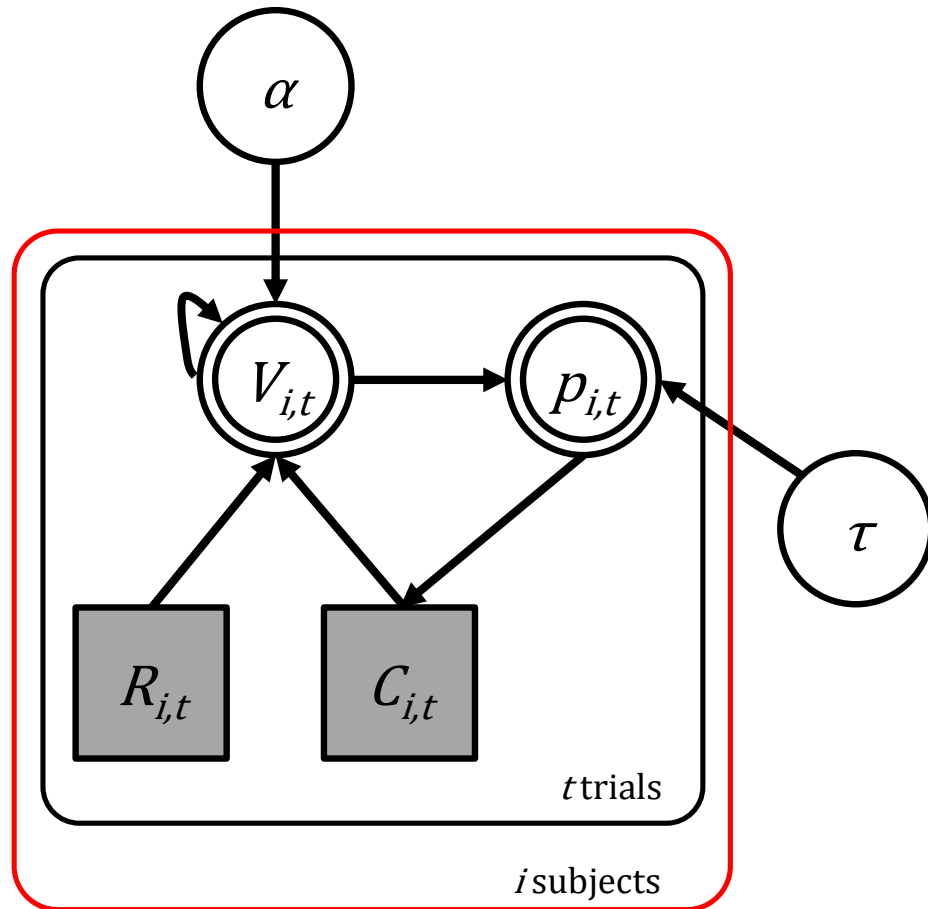
$$V_{t+1}^c = V_t^c + \alpha(R_t - V_t^c)$$

Fitting **Multiple** Participants as ONE

cognitive model

statistics

computing

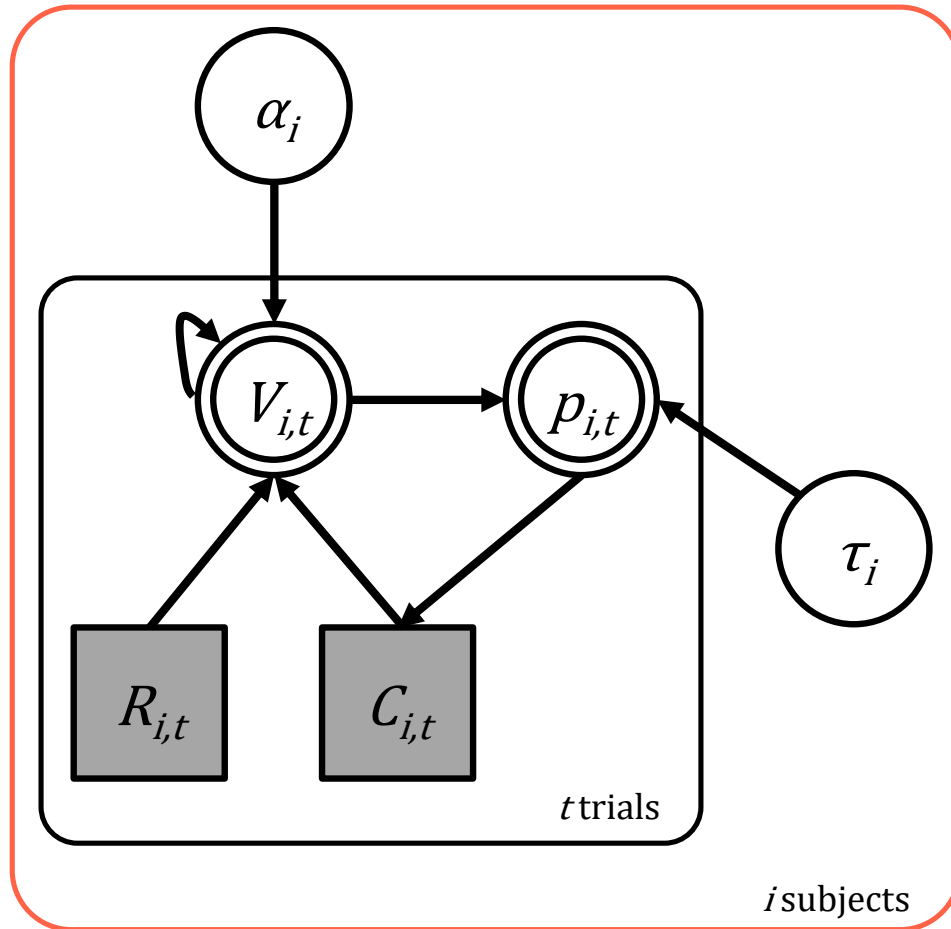


Fitting Multiple Participants **Independently**

cognitive model

statistics

computing



Why Hierarchical Bayesian Cognitive Modeling?

cognitive model

statistics

computing

Fixed effects

- all subjects are fitted with the same set of parameters
- worse model fit than “random effects”

Random effects

- each subject is fitted independently of the others
- best model fit for each subject
- parameter estimates can be noisy

Why Hierarchical Bayesian Cognitive Modeling?

cognitive model

statistics

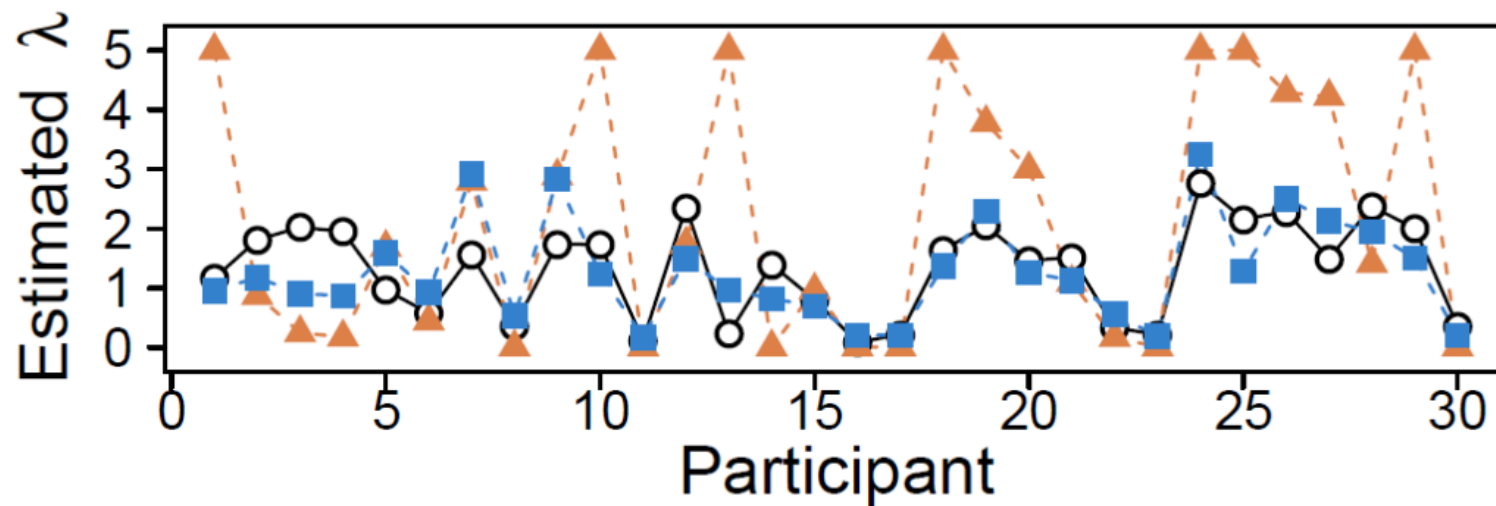
computing

Simulation study

Hierarchical Bayesian ■

Maximum likelihood ▲

Actual values ○

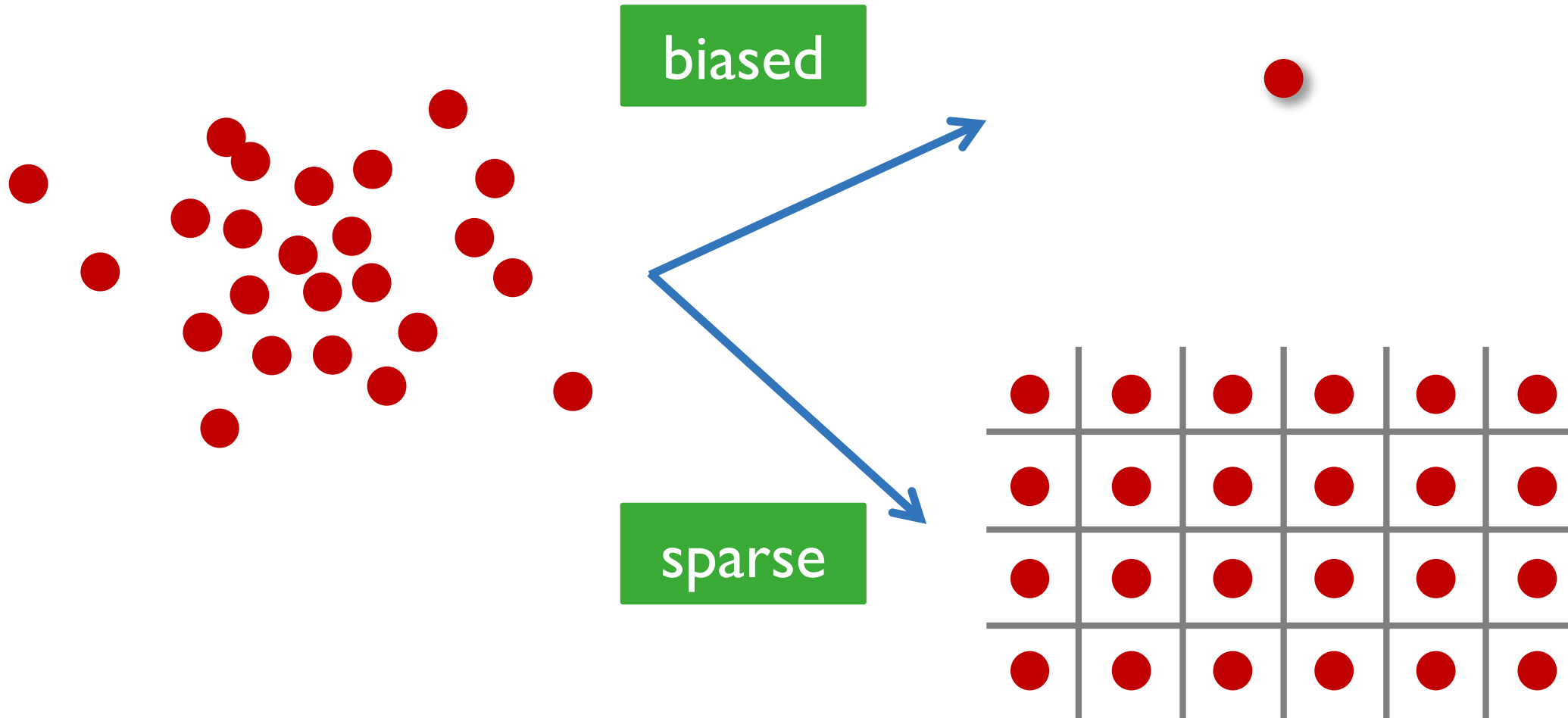


Fitting Multiple Participants

cognitive model

statistics

computing

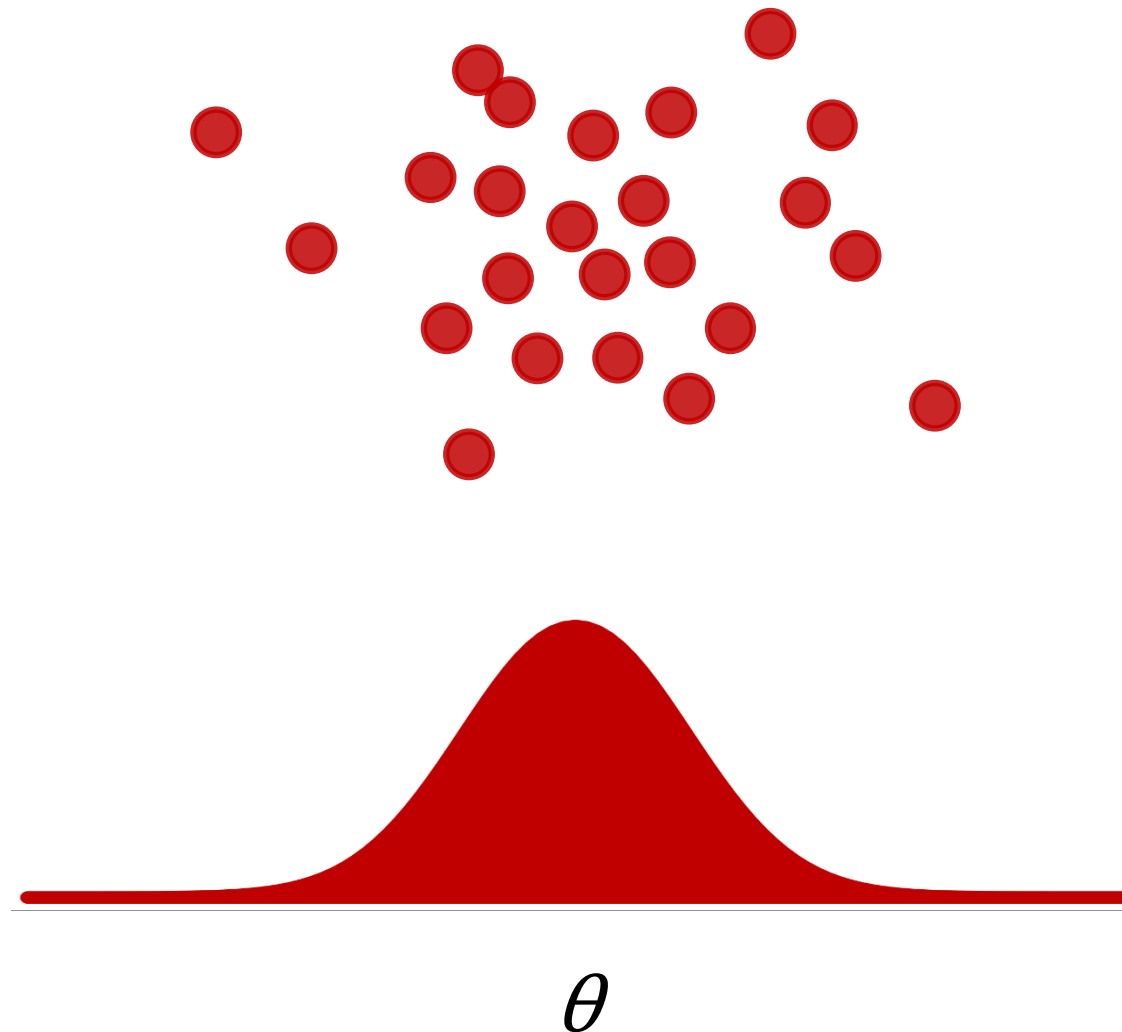


Fitting Multiple Participants

cognitive model

statistics

computing

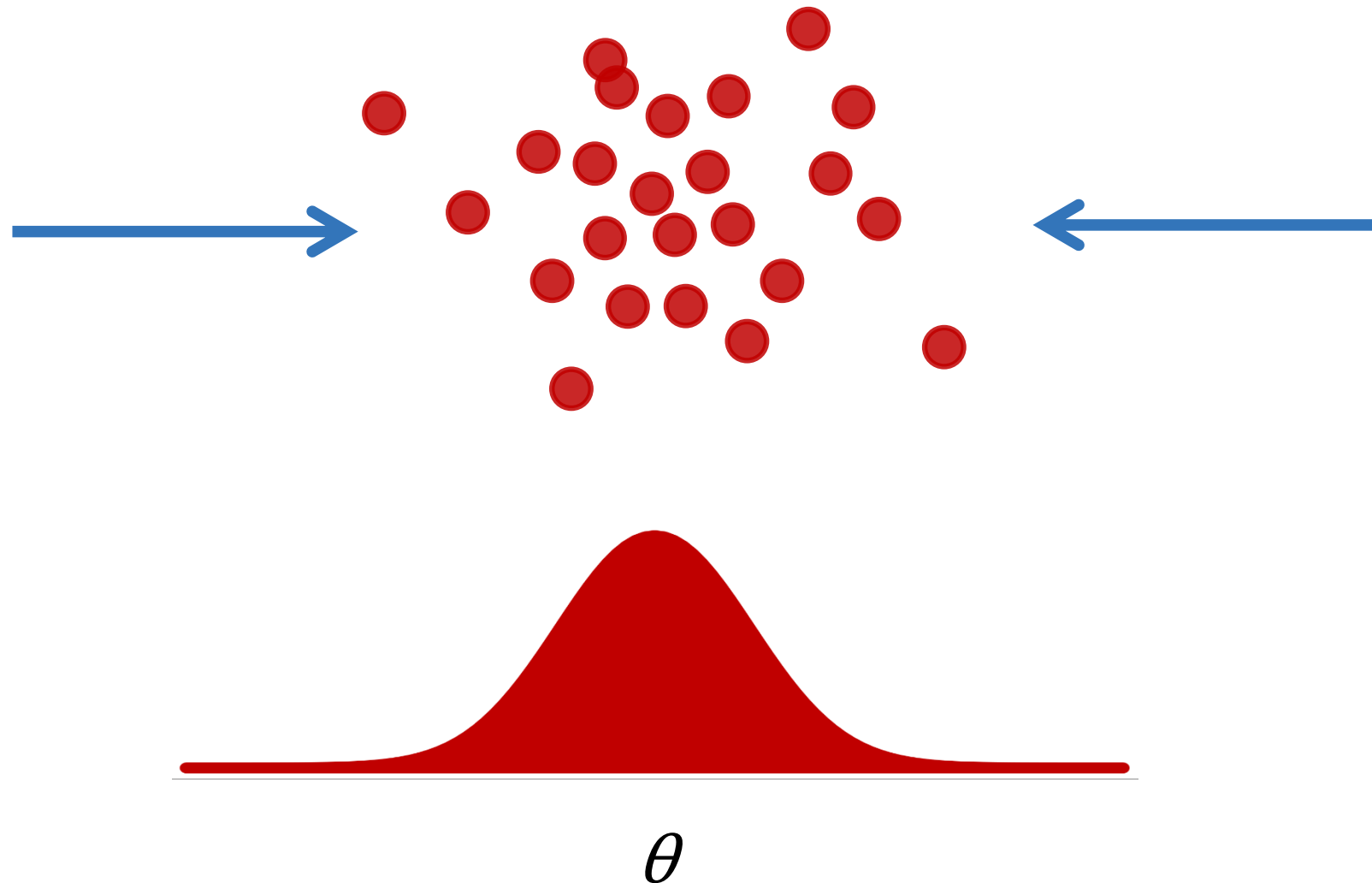


Fitting Multiple Participants

cognitive model

statistics

computing

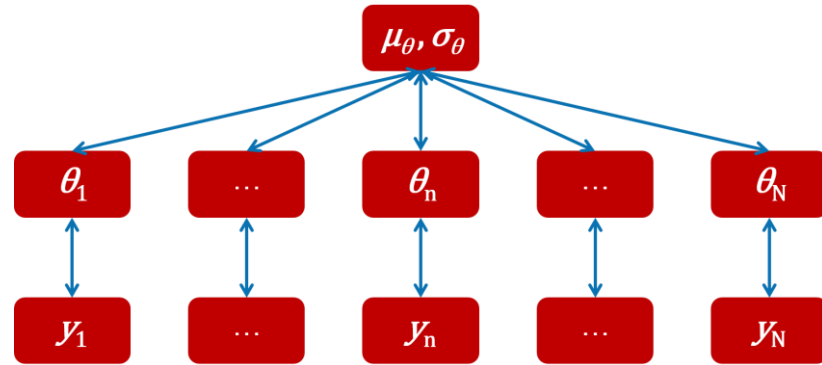


Hierarchical Structure

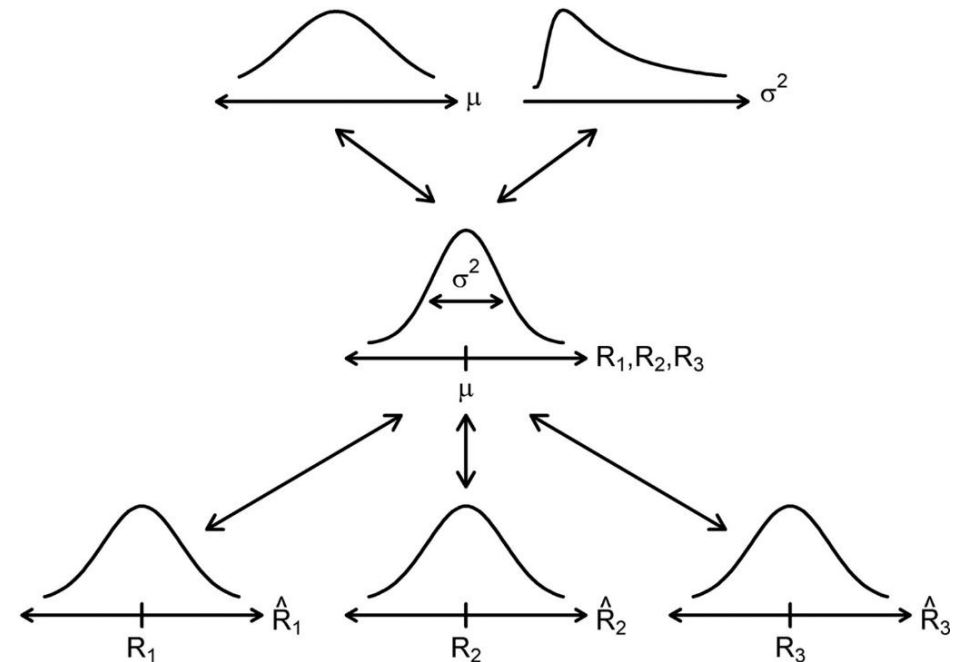
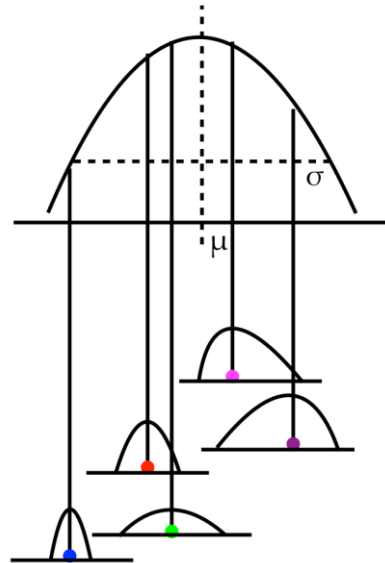
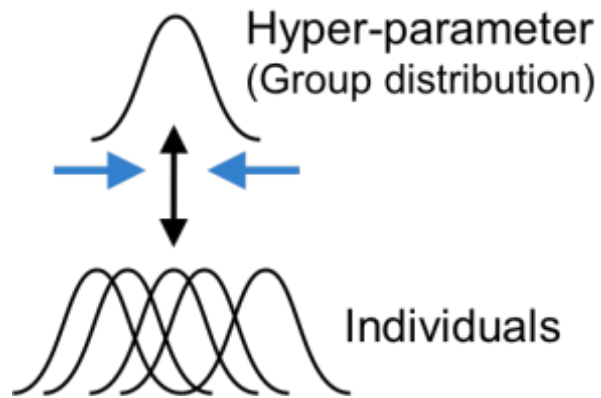
cognitive model

statistics

computing



$$P(\Theta, \Phi | D) = \frac{P(D | \Theta, \Phi) P(\Theta, \Phi)}{P(D)} \propto P(D | \Theta) P(\Theta | \Phi) P(\Phi)$$

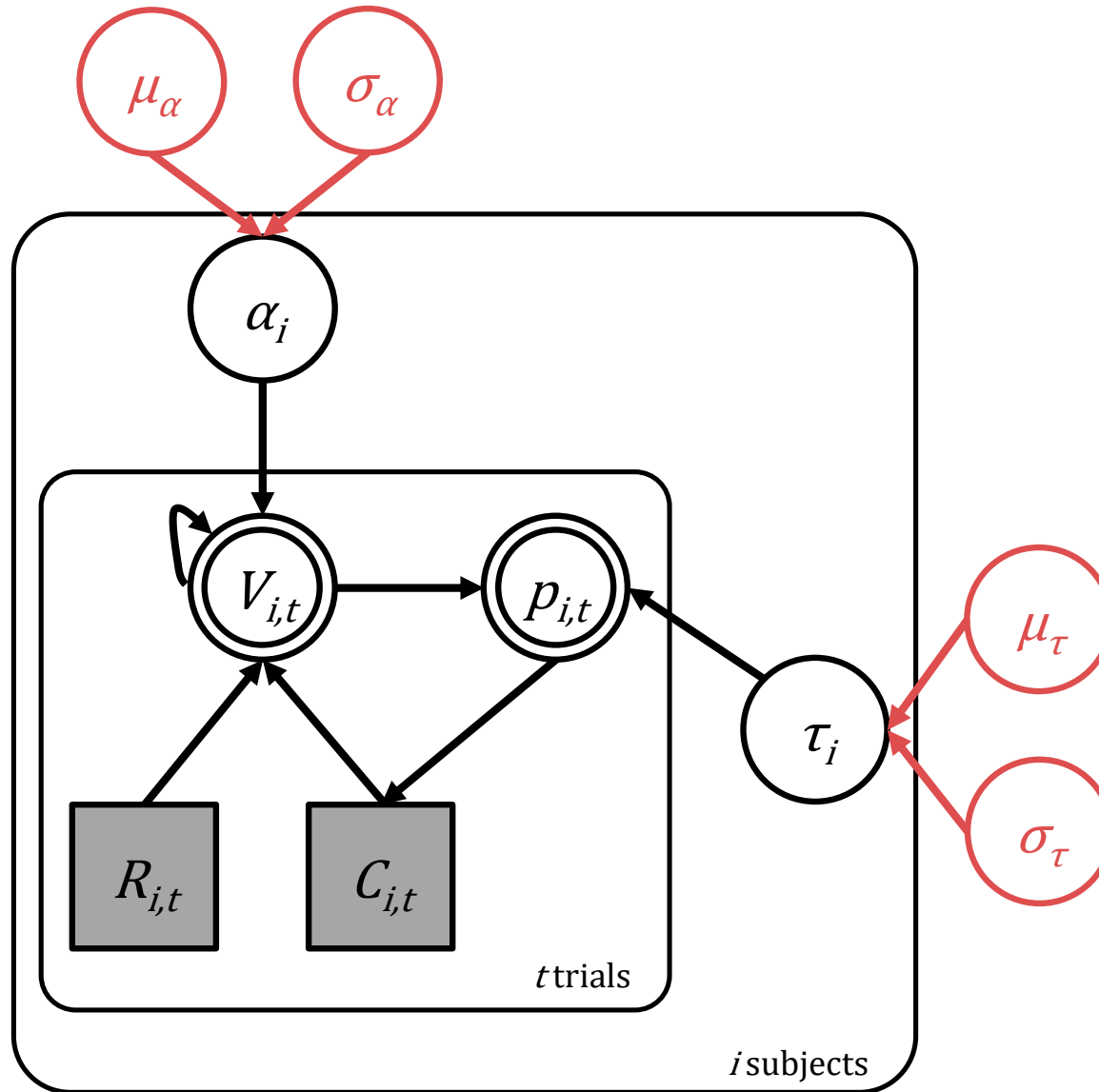


Hierarchical RL Model

cognitive model

statistics

computing



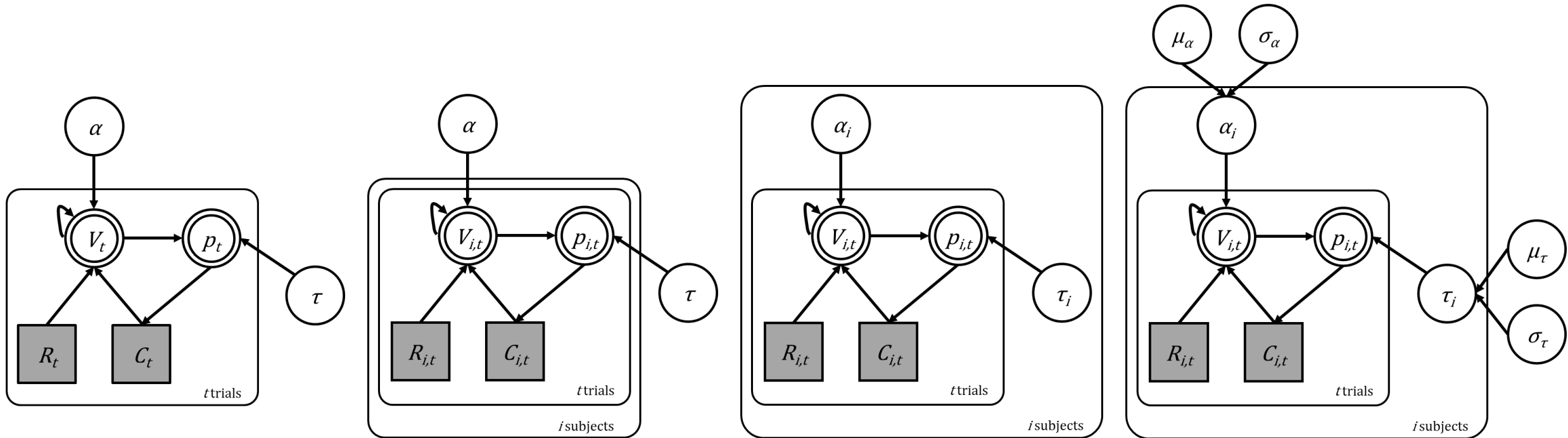
Noia!

HOW DID WE GET HERE?

cognitive model

statistics

computing



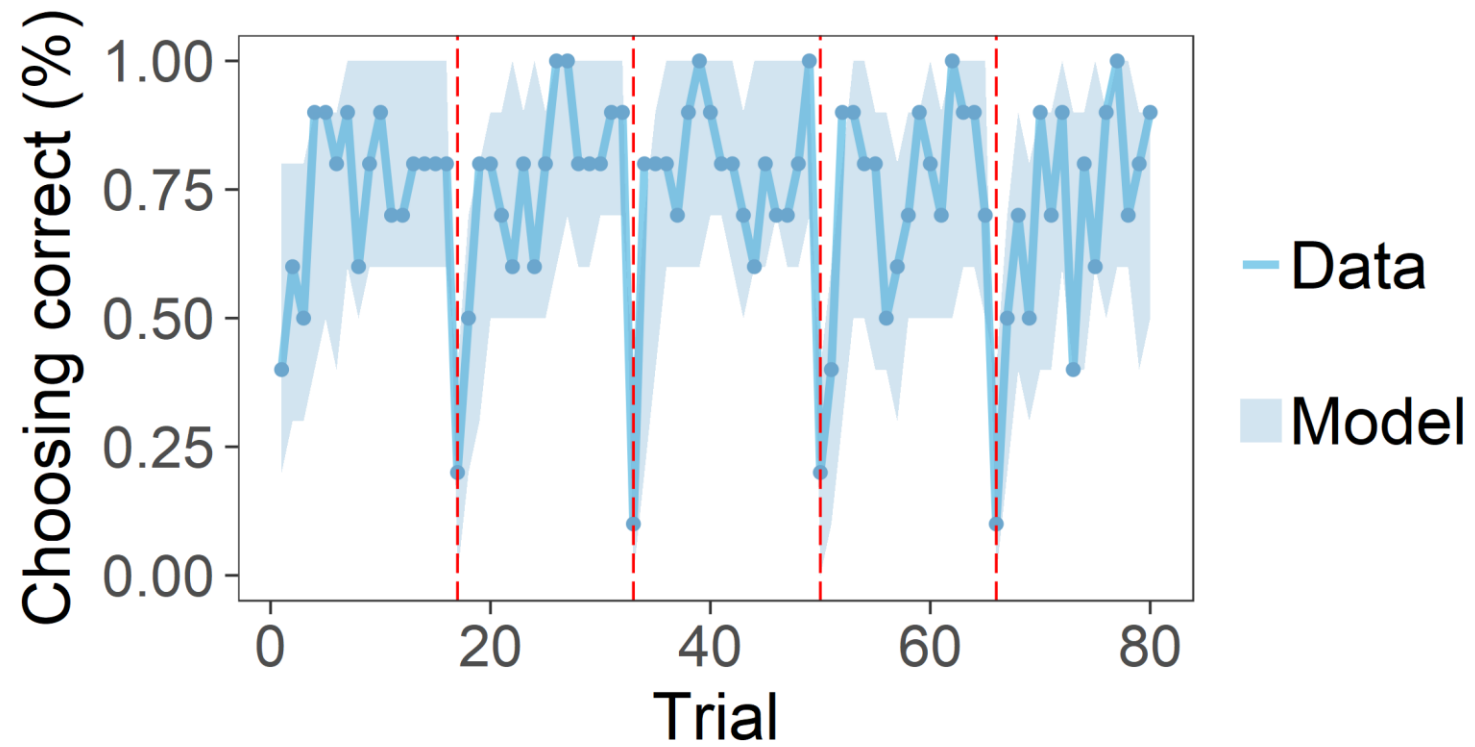
The cognitive model *per se* is the same!

Posterior Predictive Check

cognitive model

statistics

computing



Exercise 2

cognitive model

statistics

computing

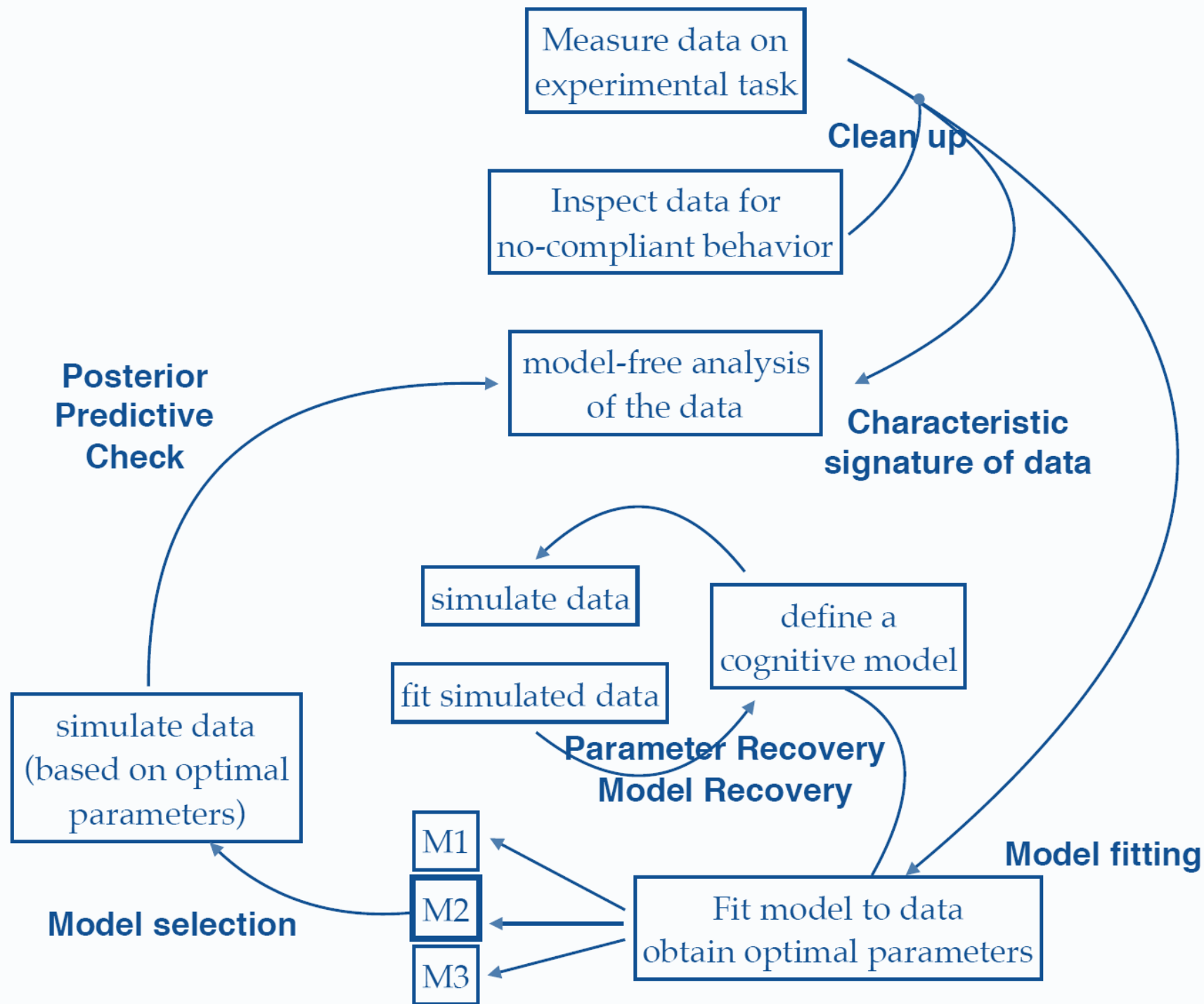
```
.../RL_tutorial/_scripts/MB_analysis.R
```

TASK:

fit the simple RL model

... and plot the model predictions on top of data

```
library(rstan)  
y_pred = extract(f$fit, pars='y_pred')$y_pred
```

Workshops / Summer schools

cognitive model

statistics

computing

- JAGS and WinBUGS Workshop @ Amsterdam, NL (annual)
- Model-based Neuroscience Summer School @ Amsterdam, NL (annual)
- European Summer School on Computational and Mathematical Modeling of Cognition @ multiple EU sites (biannual)
- Computational Psychiatry Course @ Zürich, CH (annual)
- London Computational Psychiatry Course @ London, UK (annual?)
- Methods in Neuroscience at Dartmouth Computational Summer School @ Dartmouth, US (annual)
- Brains, Minds & Machines Summer Course @ MIT, US (annual)

ANY
QUESTIONS
?

Happy Computing!