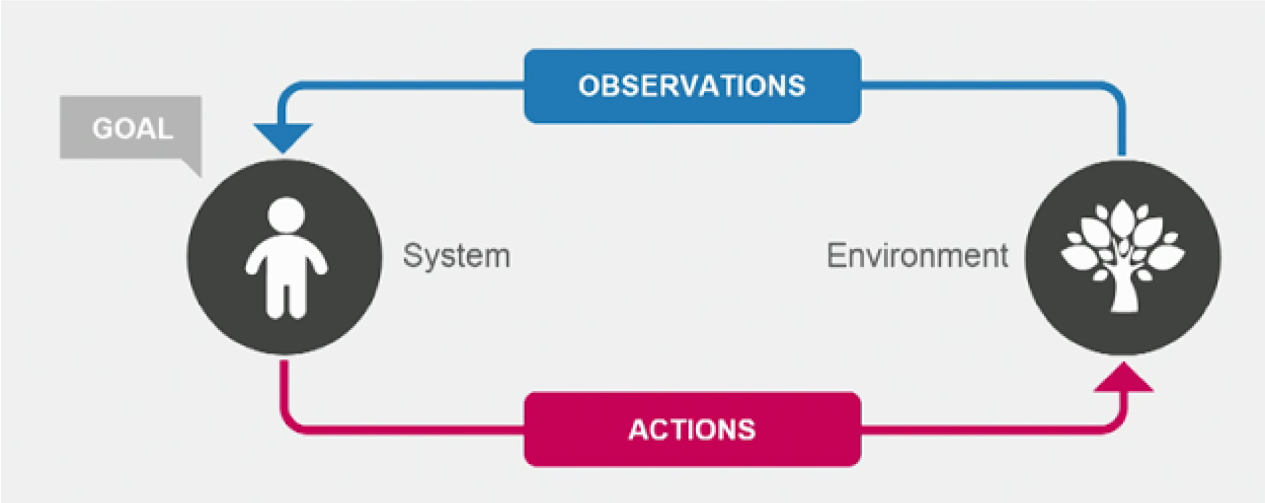


Reinforcement learning: From theory to implementation

Lei Zhang
Institute of Systems Neuroscience, University Medical Center Hamburg-Eppendorf

22-23 Nov. 2018, Tromsø

lei.zhang@uke.de
lei-zhang.net
@lei_zhang_lz

Universitätsklinikum Hamburg-Eppendorf

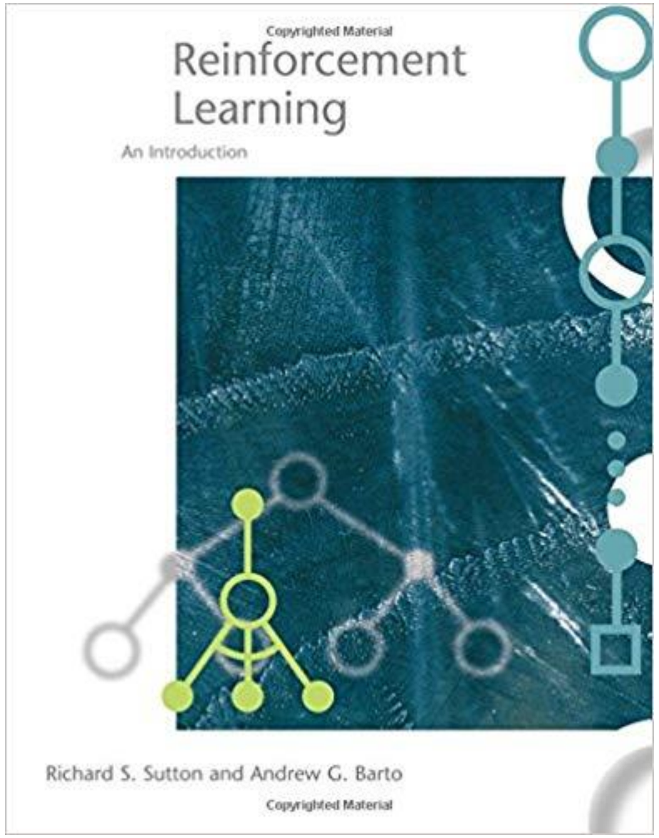
Outline

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

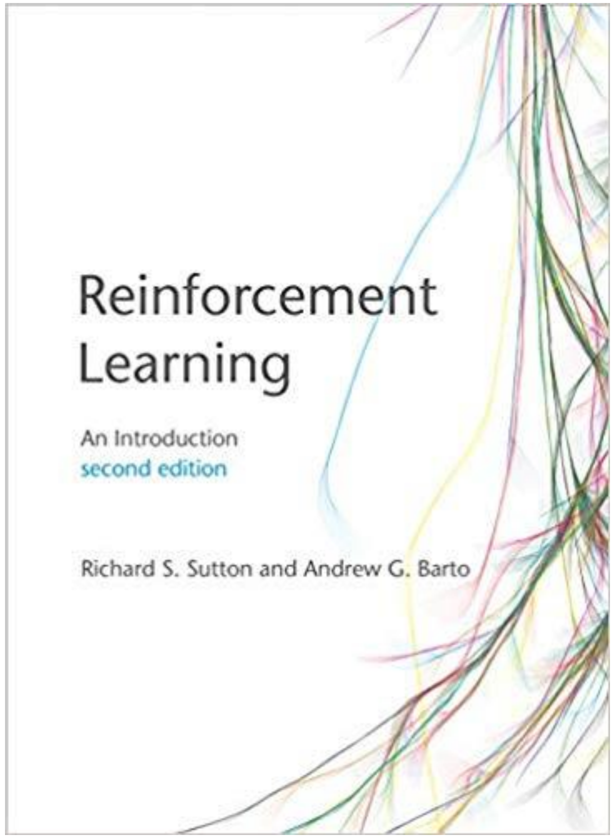
Goals:
- Understand how models are developed.
- Start here, know where you can go from it.

The teaching materials presented here are a combination between my previous workshop and the Daw & Niv workshop 2015.

The very short history

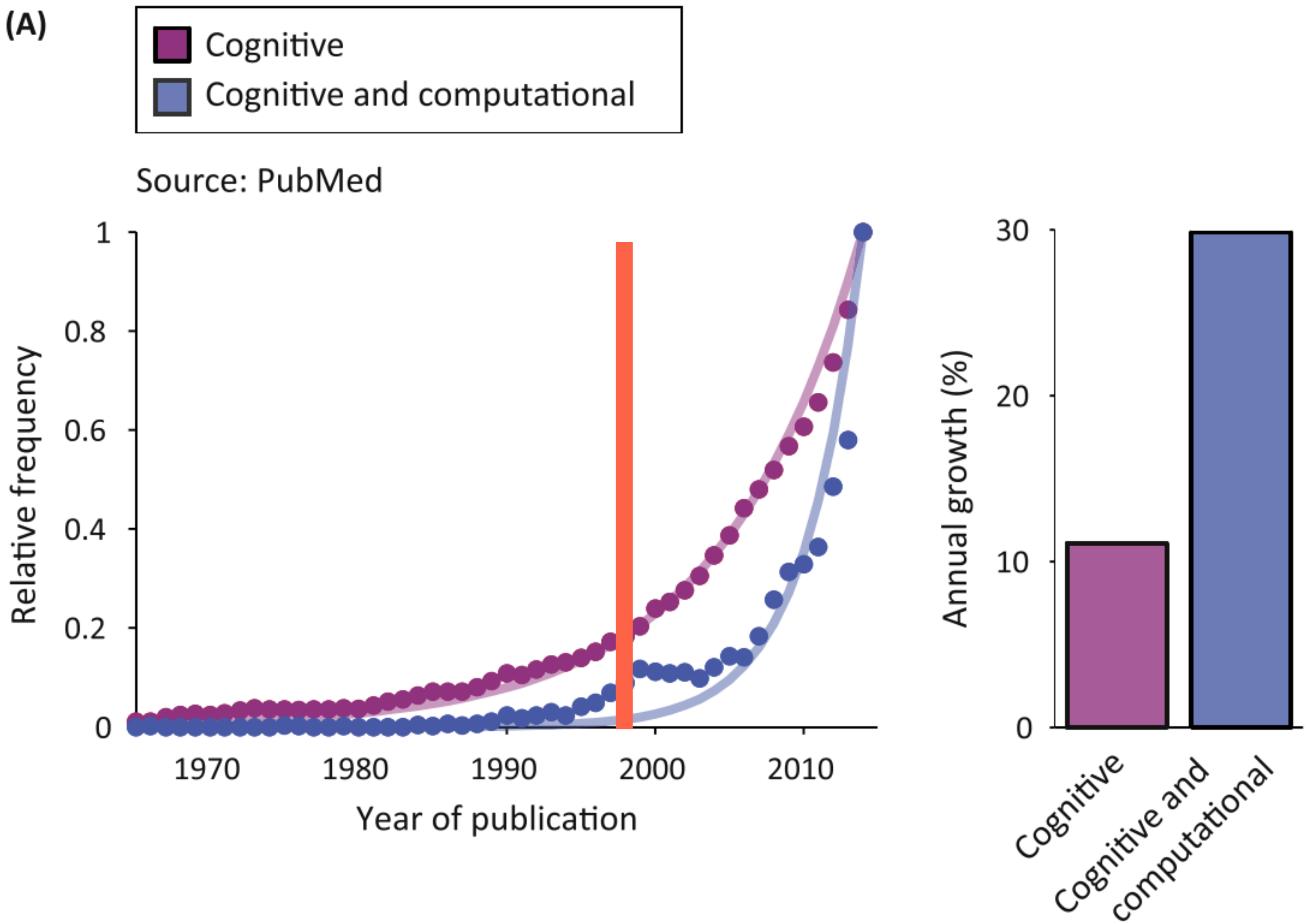


1998



2018

Boom in Cognitive Modeling



Palminteri et al. (2017)

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

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

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

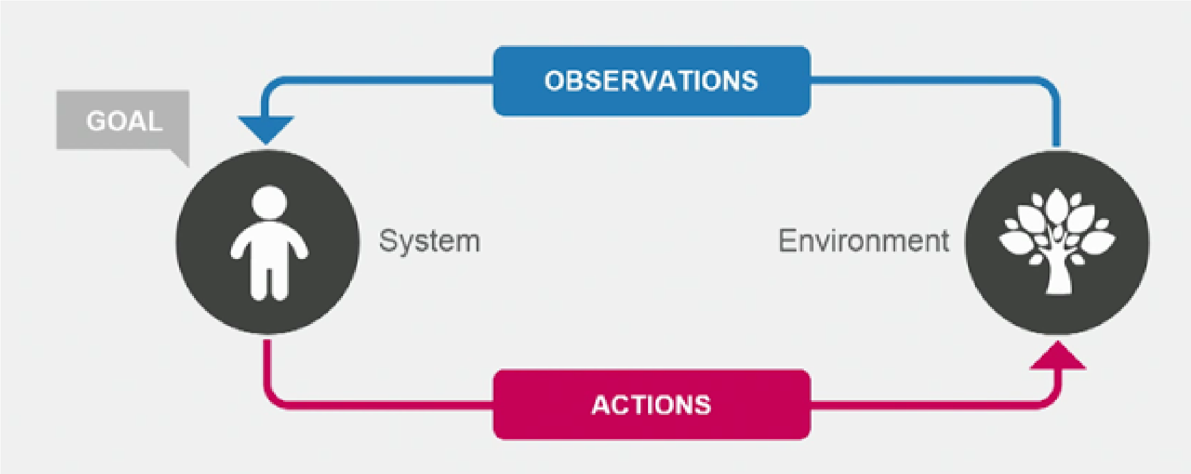
2-armed bandit task

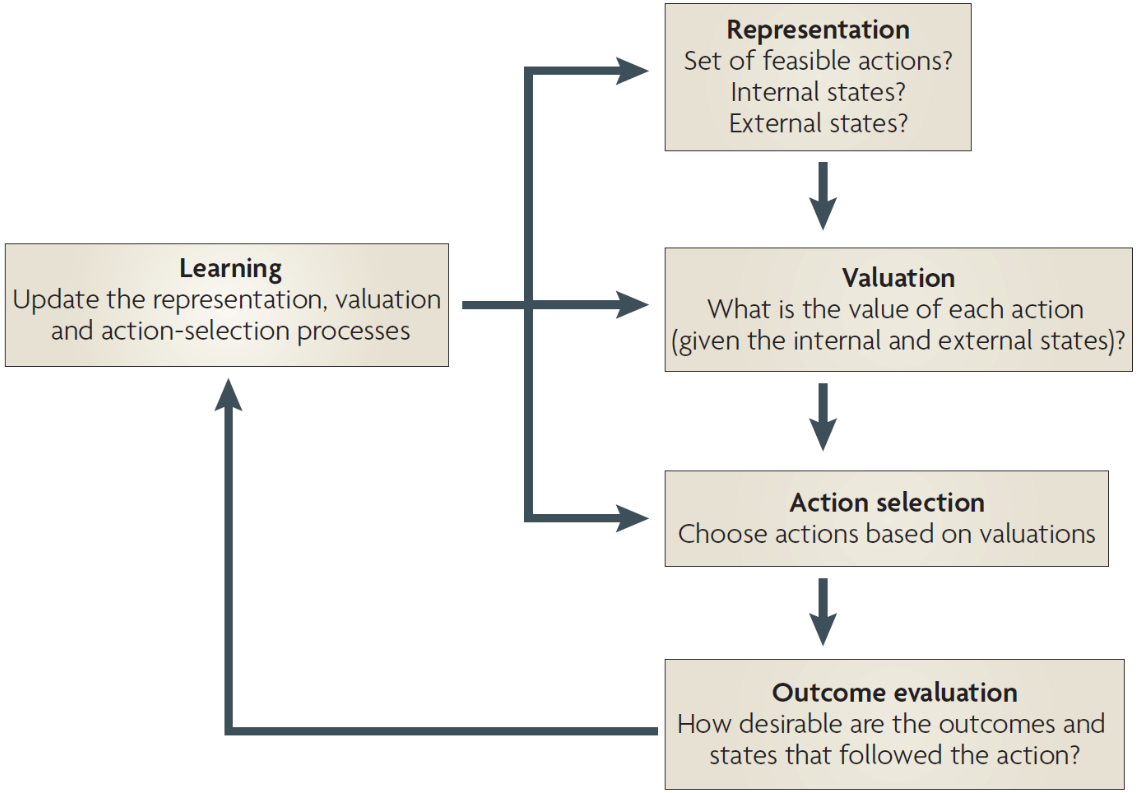
What can be your strategies:

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?





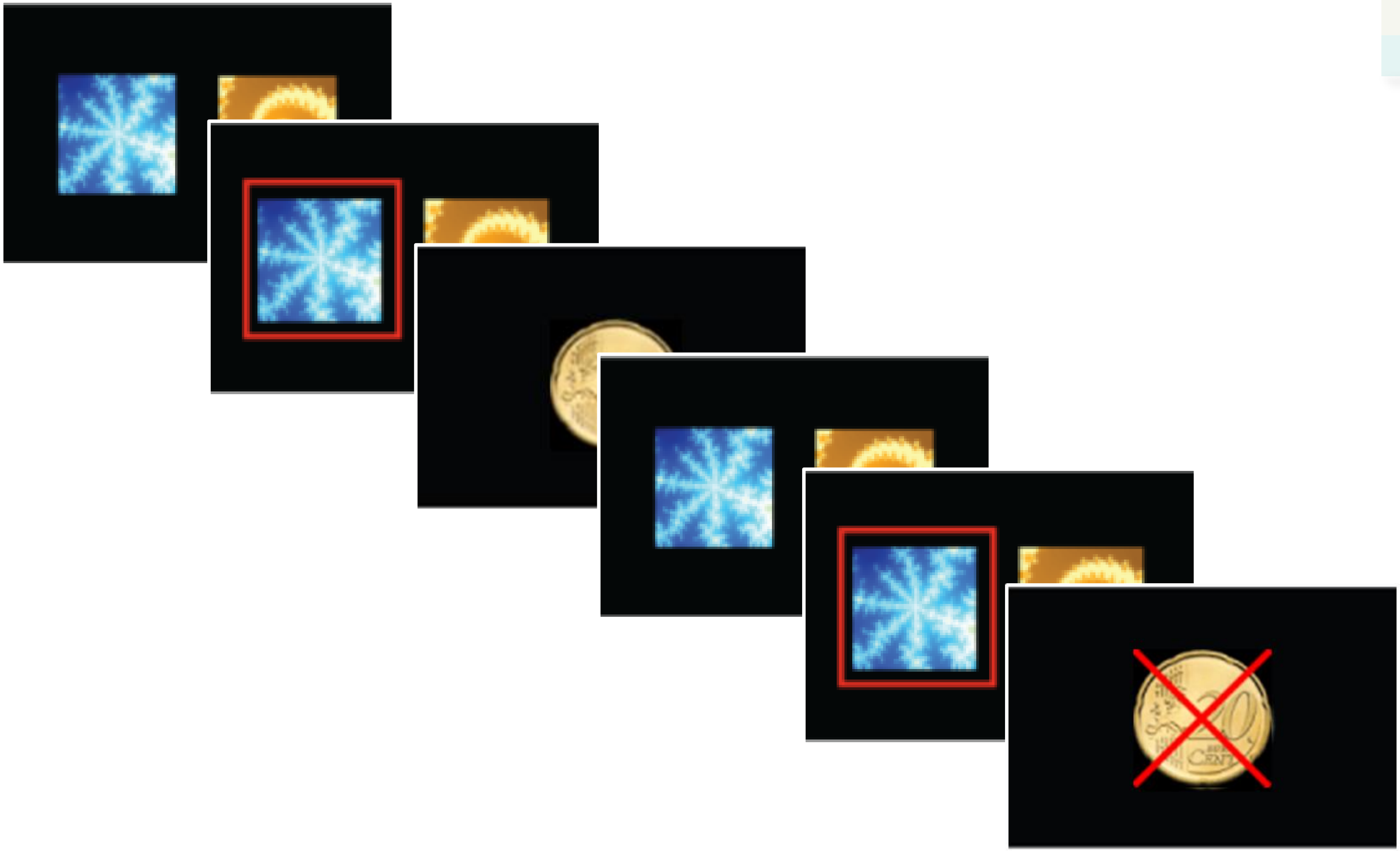
Sutton & Barto (1998, 2018); Rangel et al (2008)

Modeling the 2-armed bandit task

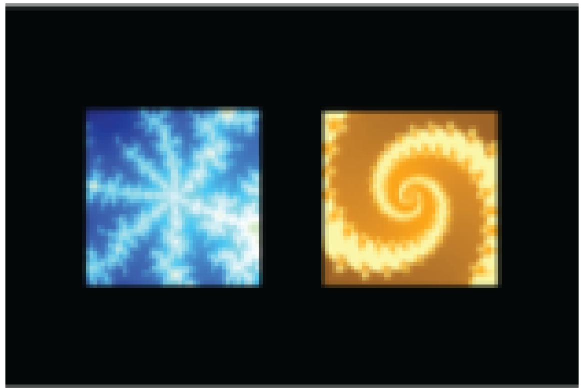
how do you suggest to model this learning process?

suppose we ran this experiment on a person

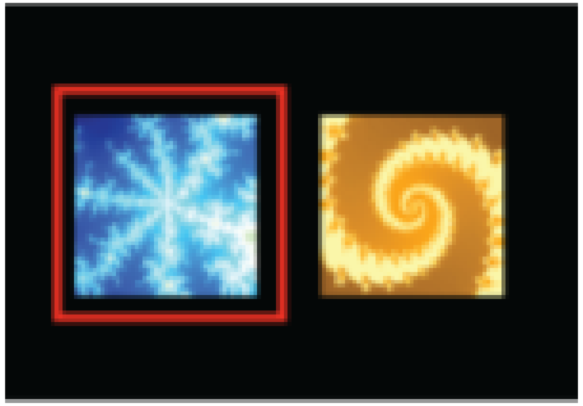
our models are basically detailed hypotheses about behavior and about the brain… we can test these hypotheses!



One simplified experiment



choice
presentation



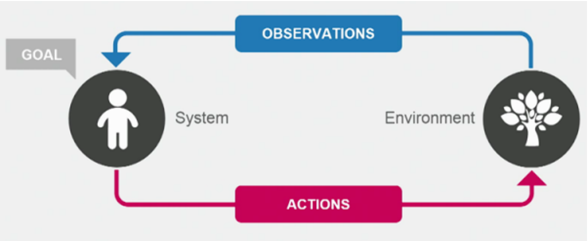
action
selection

outcome

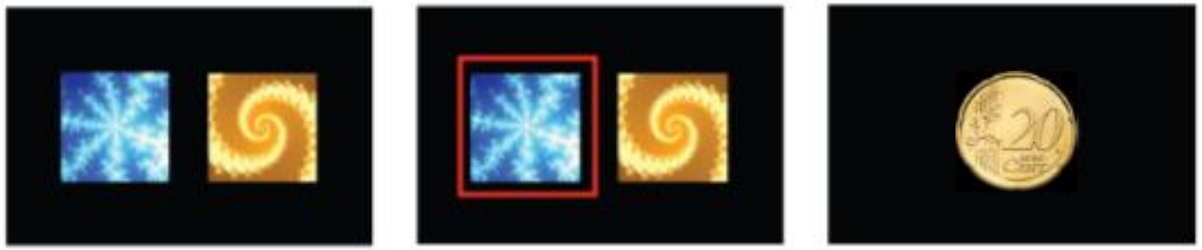
reward contingency – 80:20

Elements

what do we know?	Data: choice & outcome
what can we measure?	Summary stats: choice accuracy
what do we not know?	Learning algorithm: RL update



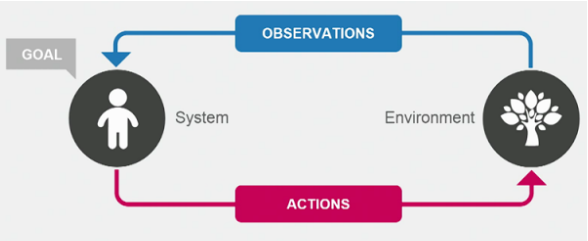
Rescorla-Wagner Value Update



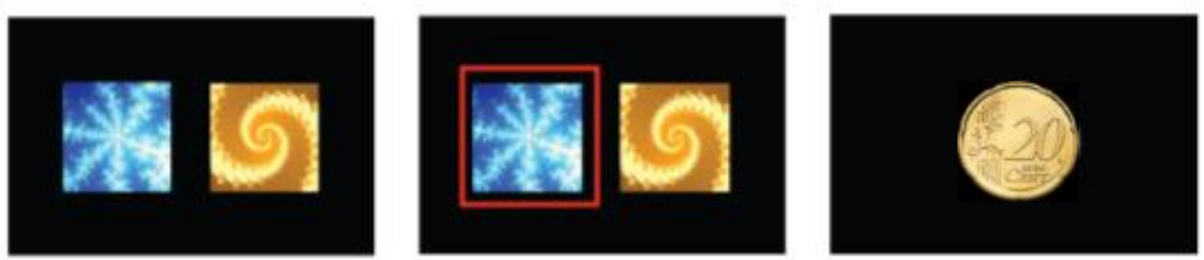
Cognitive Model
- cognitive process
- using internal variables and free parameters

Observation Model (Data Model)
- relate model to observed data
- has to account for noise

Rescorla & Wagner (1972)



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

Rescorla & Wagner (1972)

Understand the learning rate

Value update:

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

Prediction error:

$$PE = R_t - V_t$$

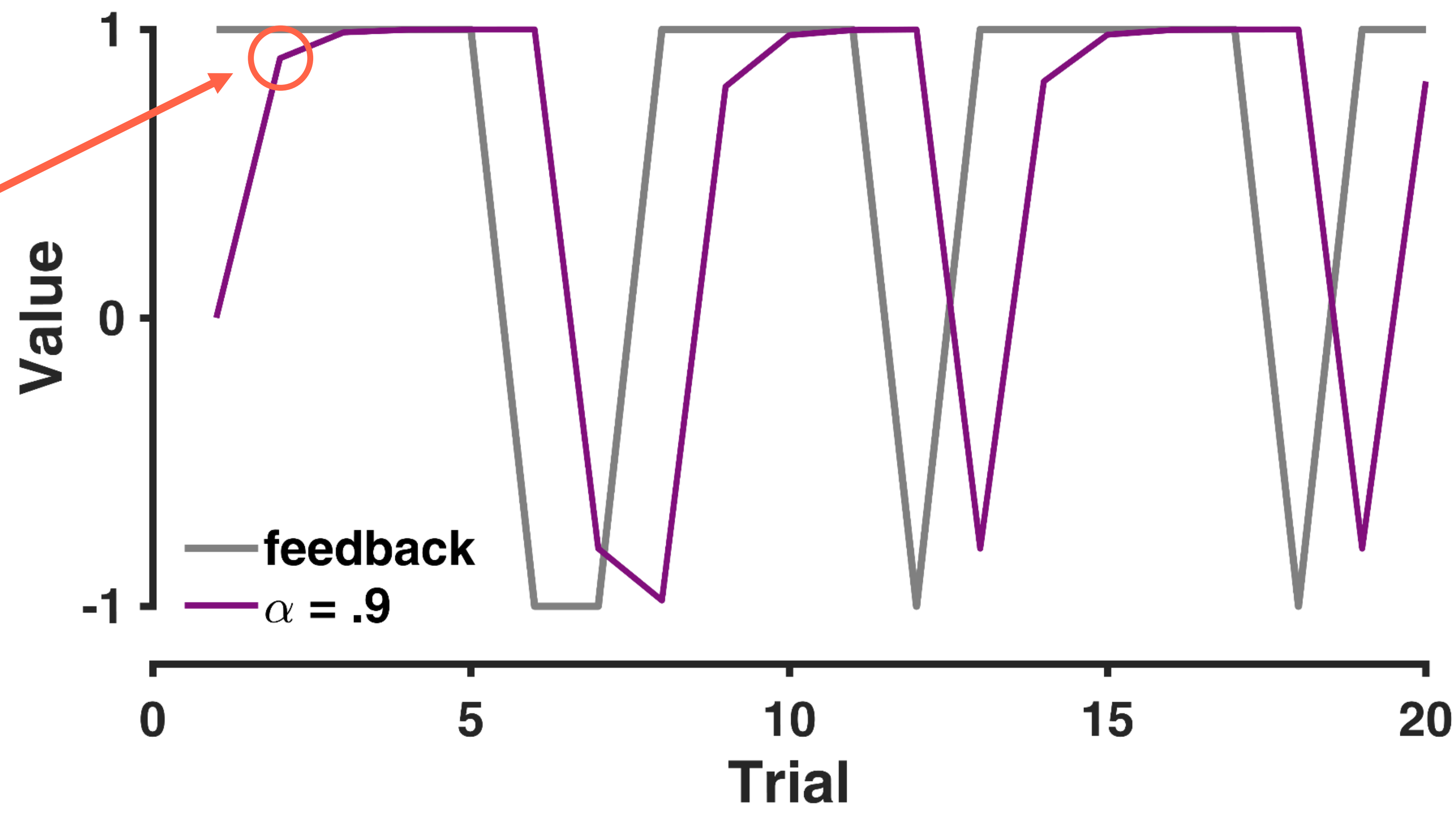
if $\eta = 0.9$

$v_1 = 0$

$v_2 = v_1 + 0.9 * (1 - 0)$

$= 0 + 0.9$

$= 0.9$



reward contingency – 80:20

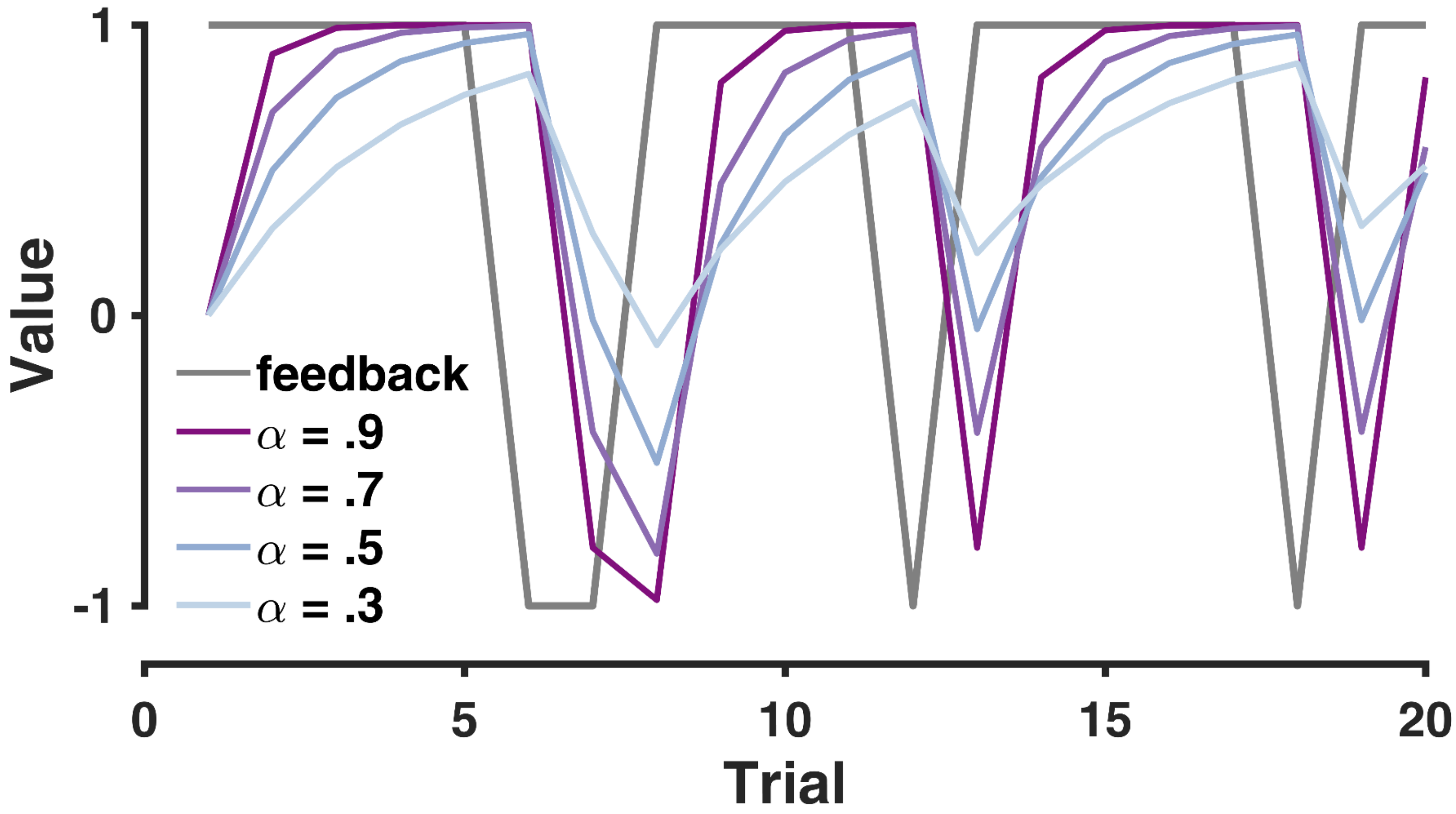
Understand the learning rate

Value update:

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

Prediction error:

$$PE = R_t - V_t$$



reward contingency – 80:20

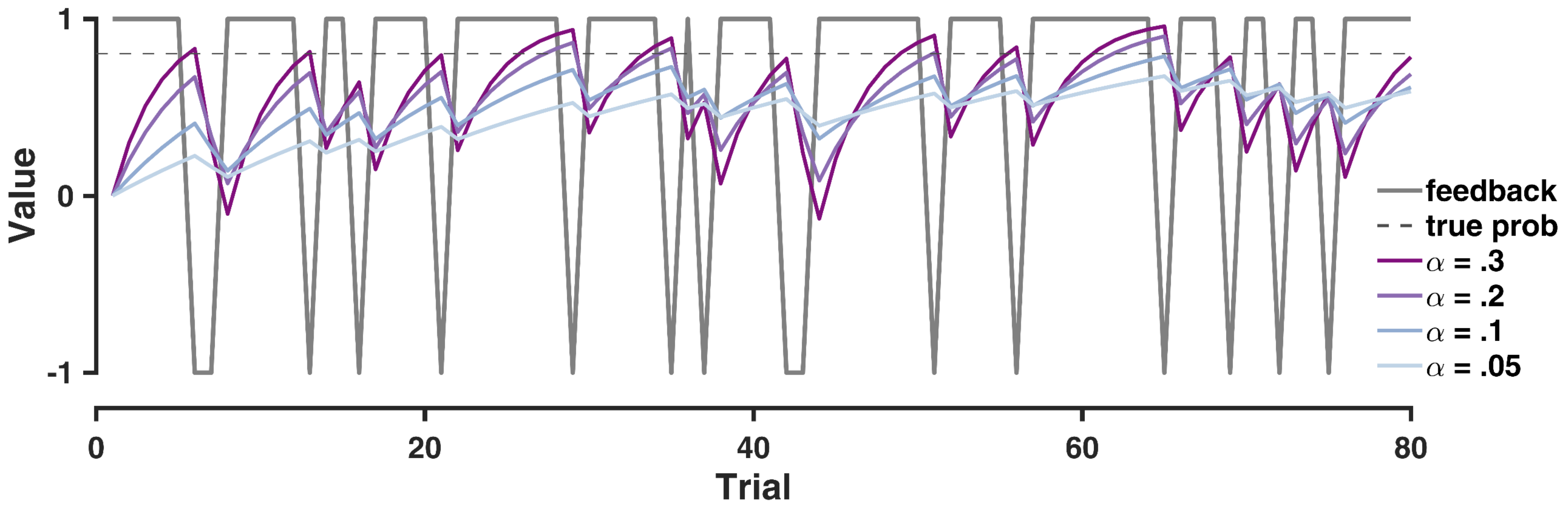
Understand the learning rate

Value update:

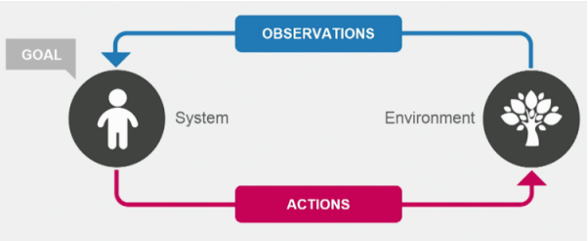
$$V_{t+1} = V_t + \alpha * PE$$

Prediction error:

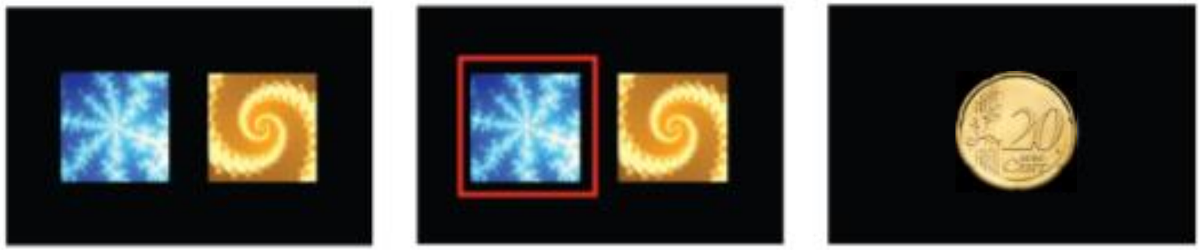
$$PE = R_t - V_t$$



reward contingency – 80:20



Rescorla-Wagner Value Update



Value update:

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

Prediction error:

$$PE = R_t - V_t$$

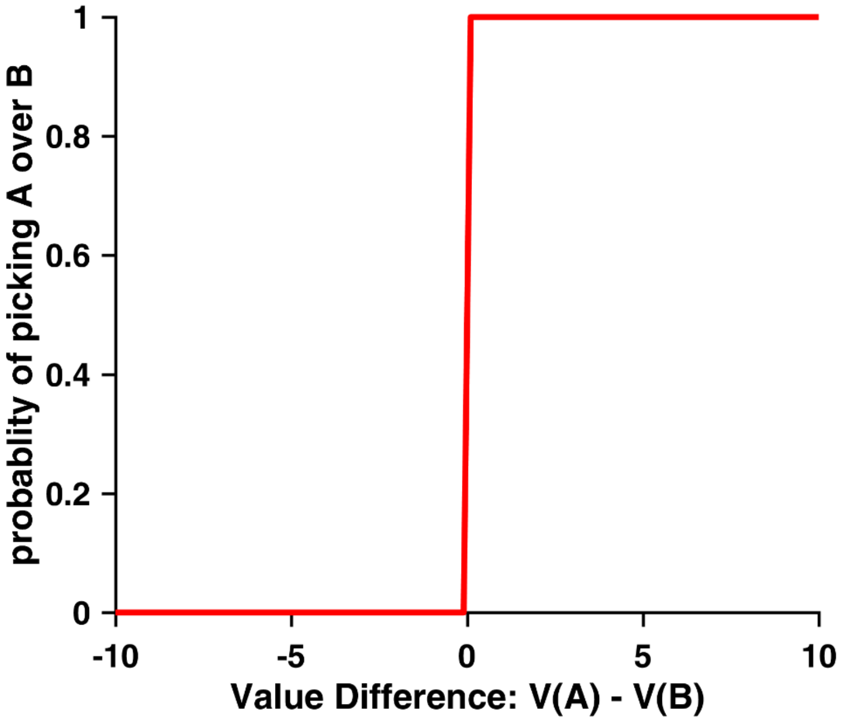
choice rule:

greedy / ε-greedy / softmax

Rescorla & Wagner (1972)

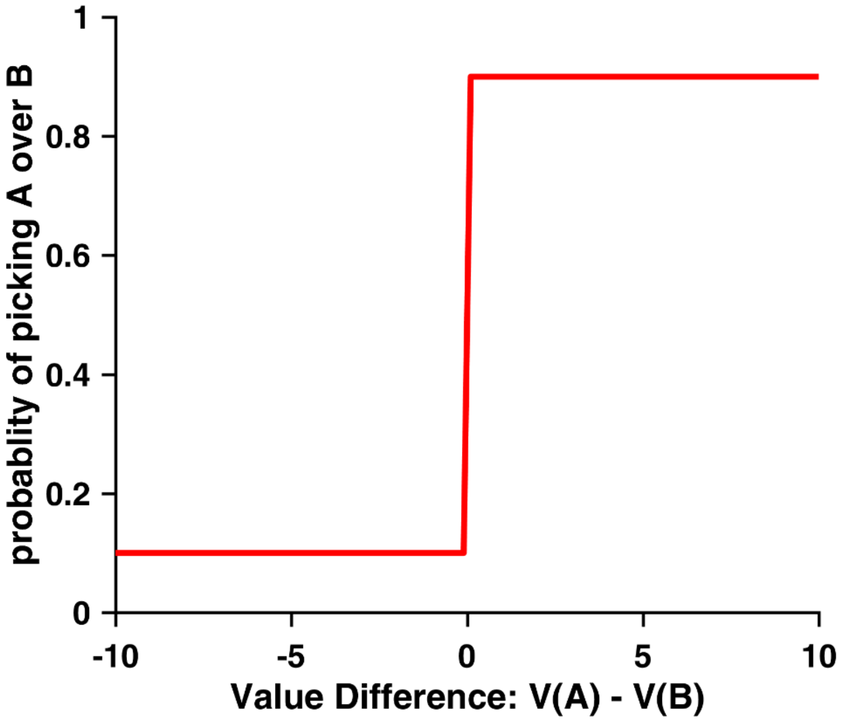
Choice rule: greedy

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



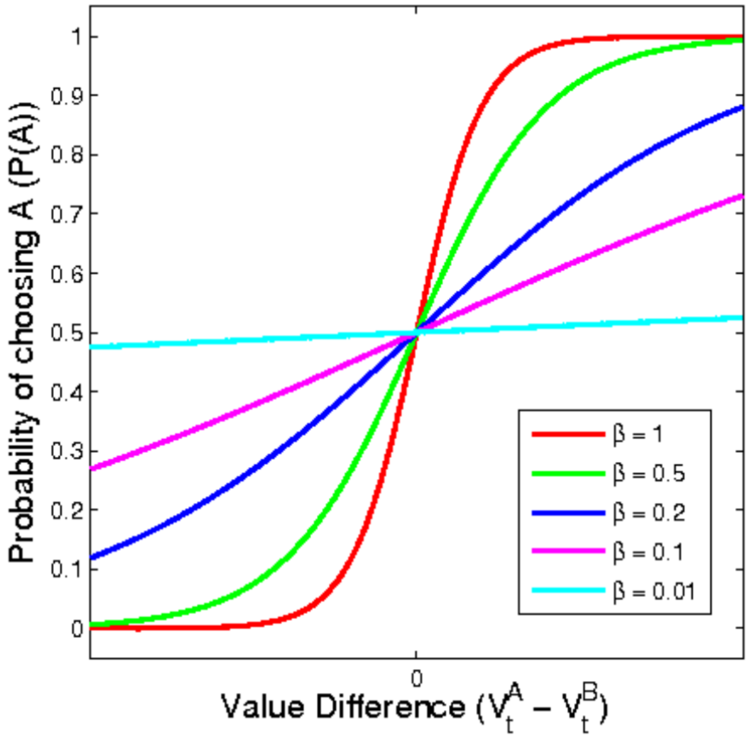
Choice rule: ε-greedy

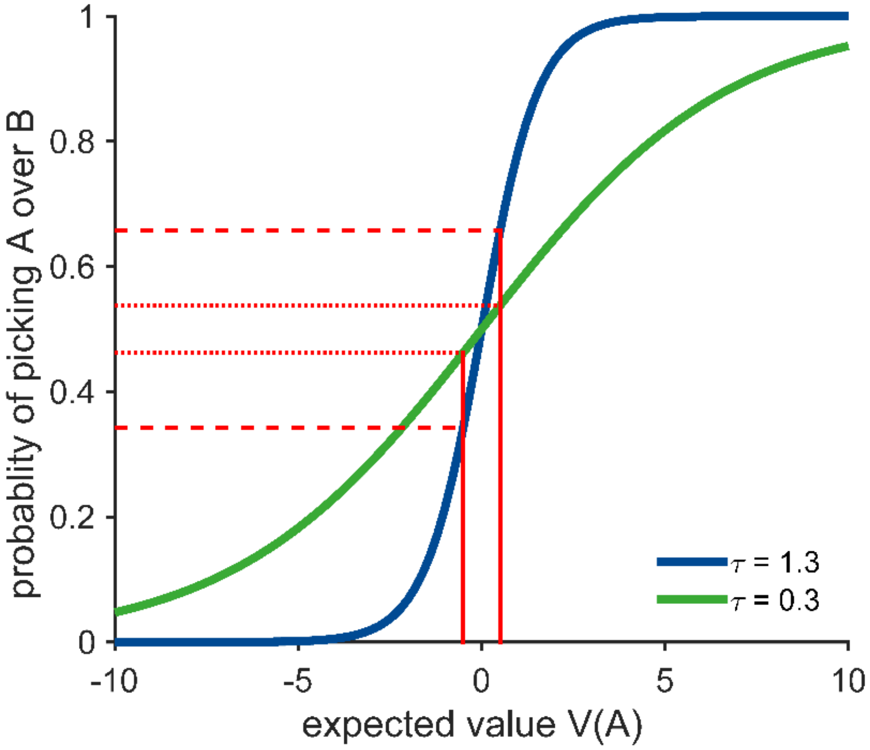
$$p(C=a) = \begin{cases} 1-\varepsilon, & V(a) > V(b) \\ \varepsilon, & V(a) < V(b) \end{cases}$$



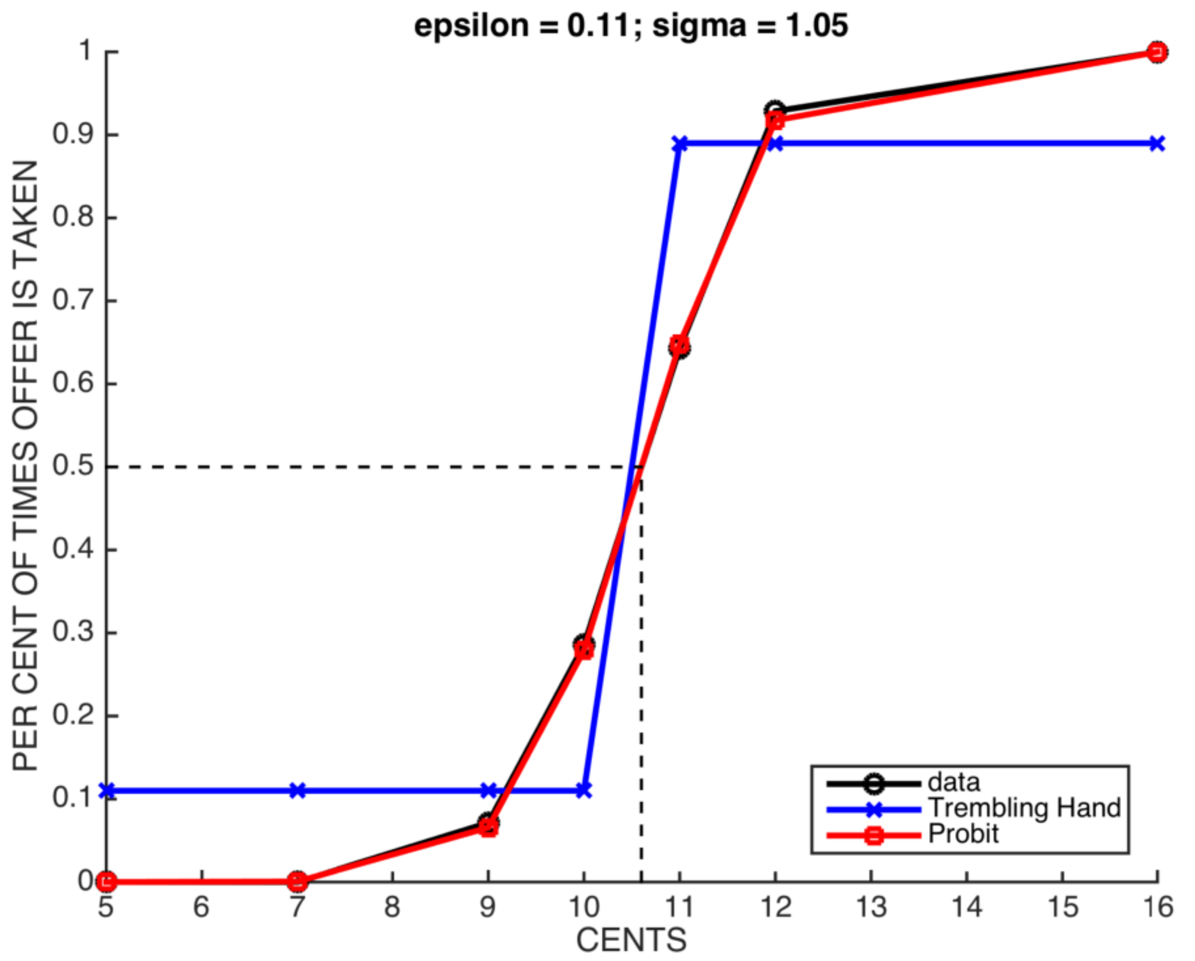
Choice rule: softmax

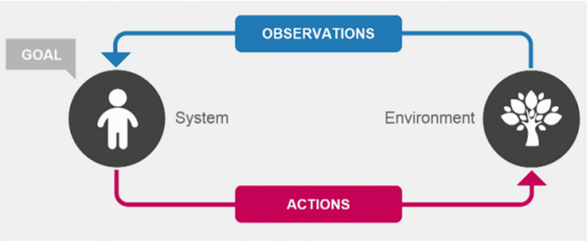
$$p(C=a) = \frac{e^{\tau * V(a)}}{e^{\tau * V(a)} + e^{\tau * V(b)}} = \frac{1}{1 + e^{-\tau * (V(a) - V(b))}}$$



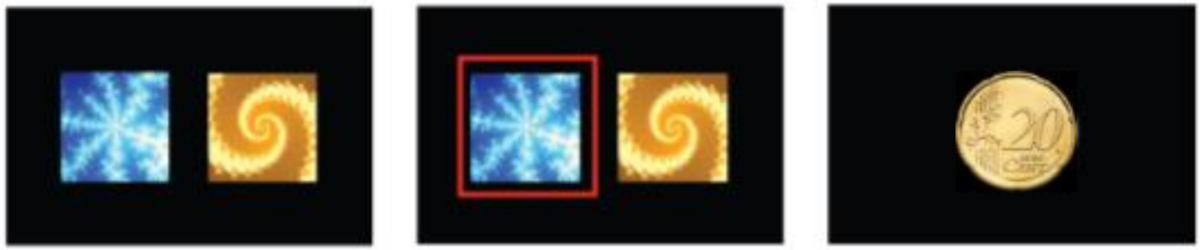


Choice rule: direct comparison





Rescorla-Wagner Value Update



Value update:

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

Prediction error:

$$PE = R_t - V_t$$

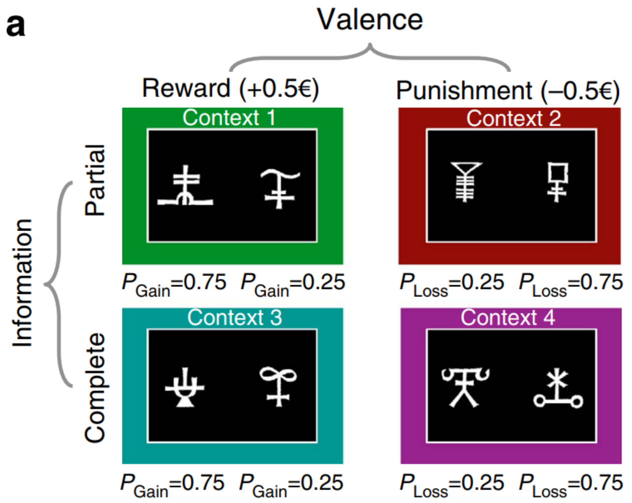
choice rule (sigmoid /softmax):

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

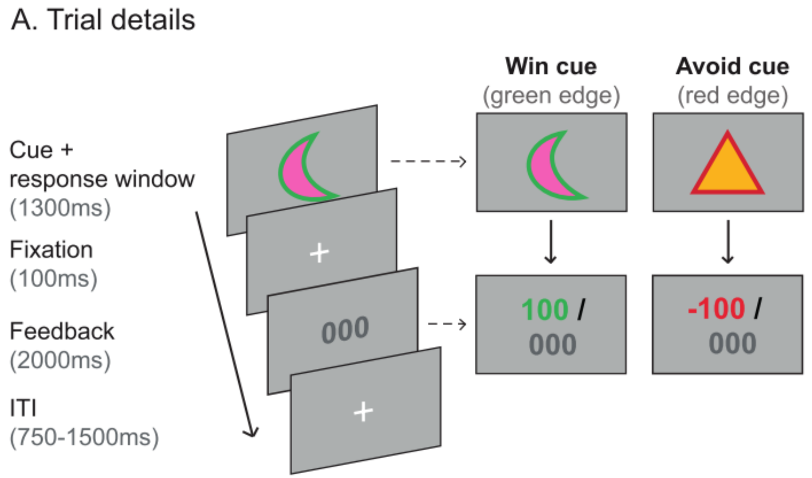
α - learning rate
PE - reward prediction error
V - value
R - reward
τ - softmax temperature

Rescorla & Wagner (1972)

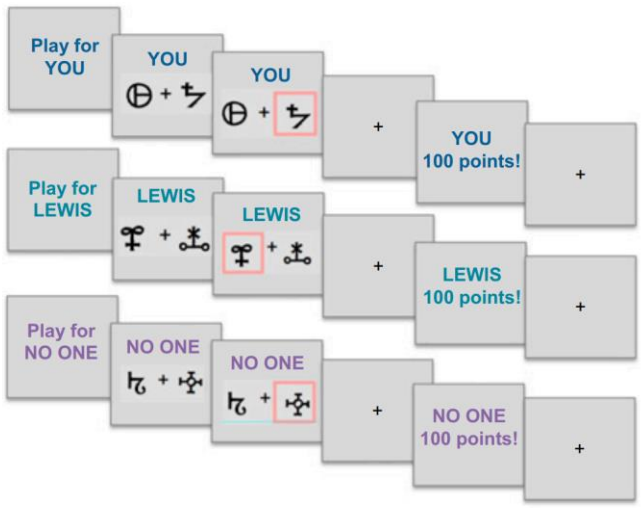
Generalizing RL framework



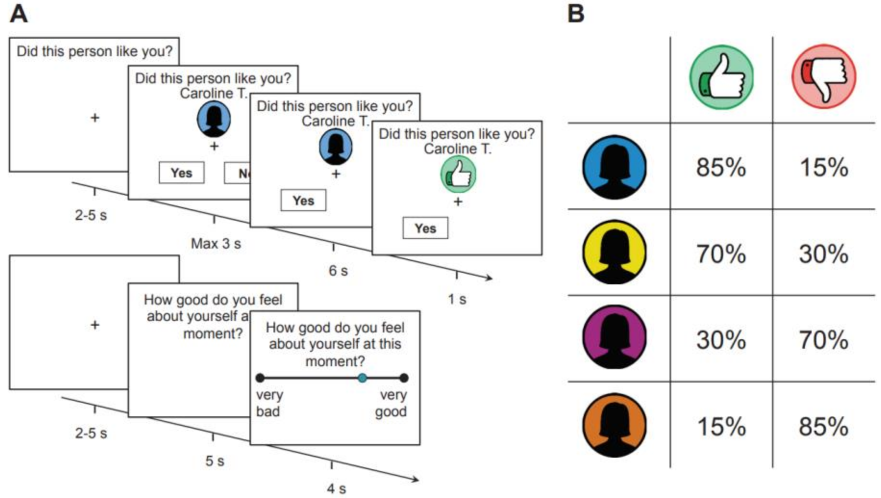
Palminteri et al. (2015)



Swart et al. (2017)



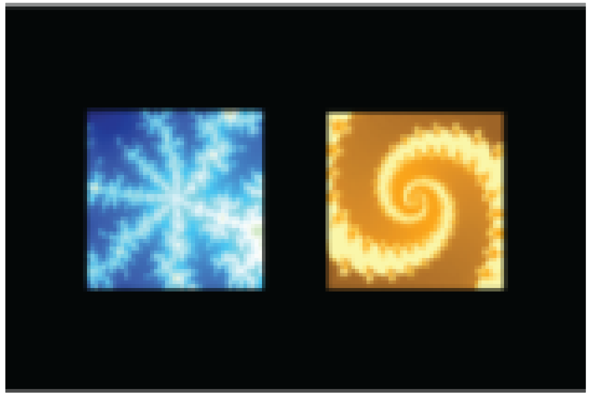
Lockwood et al. (2016)

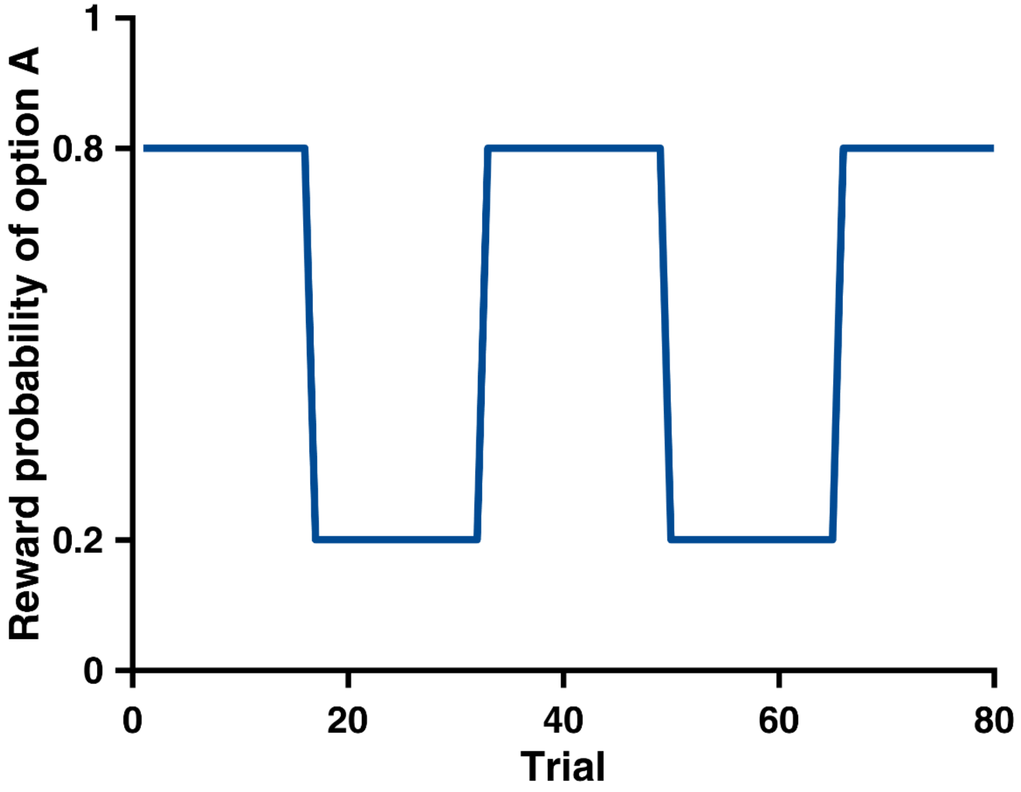


Will et al. (2017)

The real task

PRL: probabilistic reversal learning

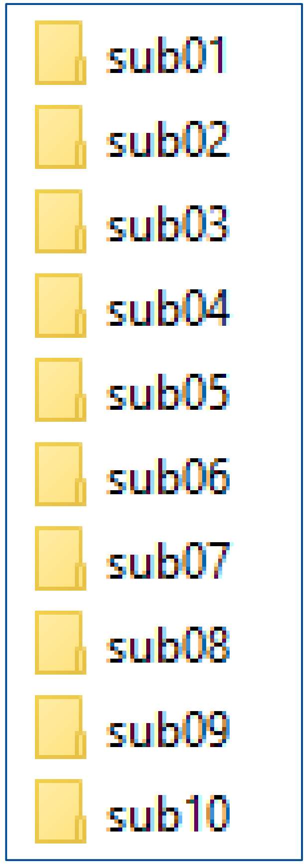




The data

- PRL task
- nSub = 10
- nTrial = 80

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

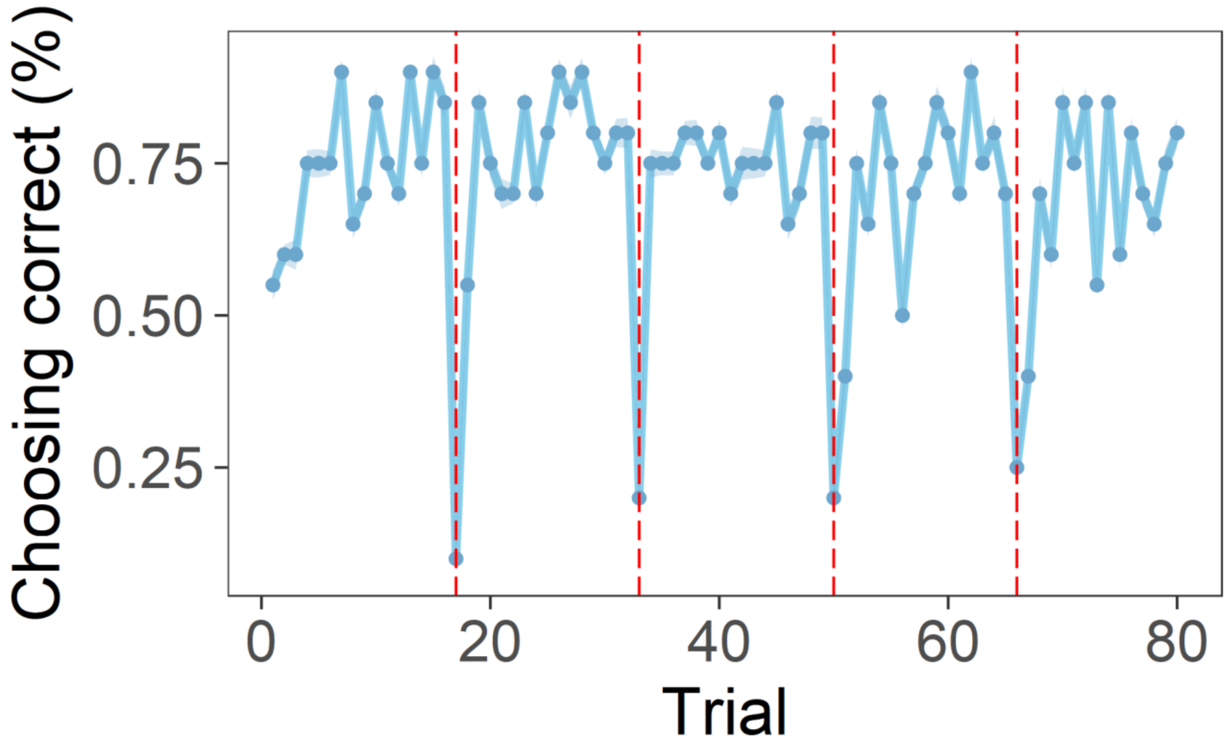



```
subjID, trialID, choice, outcome, correct
1,1,2,-1,1
1,2,1,1,1
1,3,1,1,1
1,4,1,1,1
1,5,2,-1,1
1,6,1,1,1
1,7,1,1,1
1,8,1,1,1
1,9,1,-1,1
1,10,2,-1,1
1,11,1,1,1
1,12,1,1,1
1,13,1,-1,2
```

Model-free analysis: summary stats

cognitive model
statistics
computing

How often participants choose the more rewarding option?



Exercise 1


```
…/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

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

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

$$p(\theta \mid D) = \frac{p(D \mid \theta)\, p(\theta)}{\boxed{\int p(D \mid \theta^*)\, p(\theta^*)\, d\theta^*}}$$

$$\boxed{p(\theta \mid D) \propto p(D \mid \theta)\, p(\theta)}$$

Deterministic Approximation	Stochastic Approximation
→ Variational Bayes	→ Sampling Methods

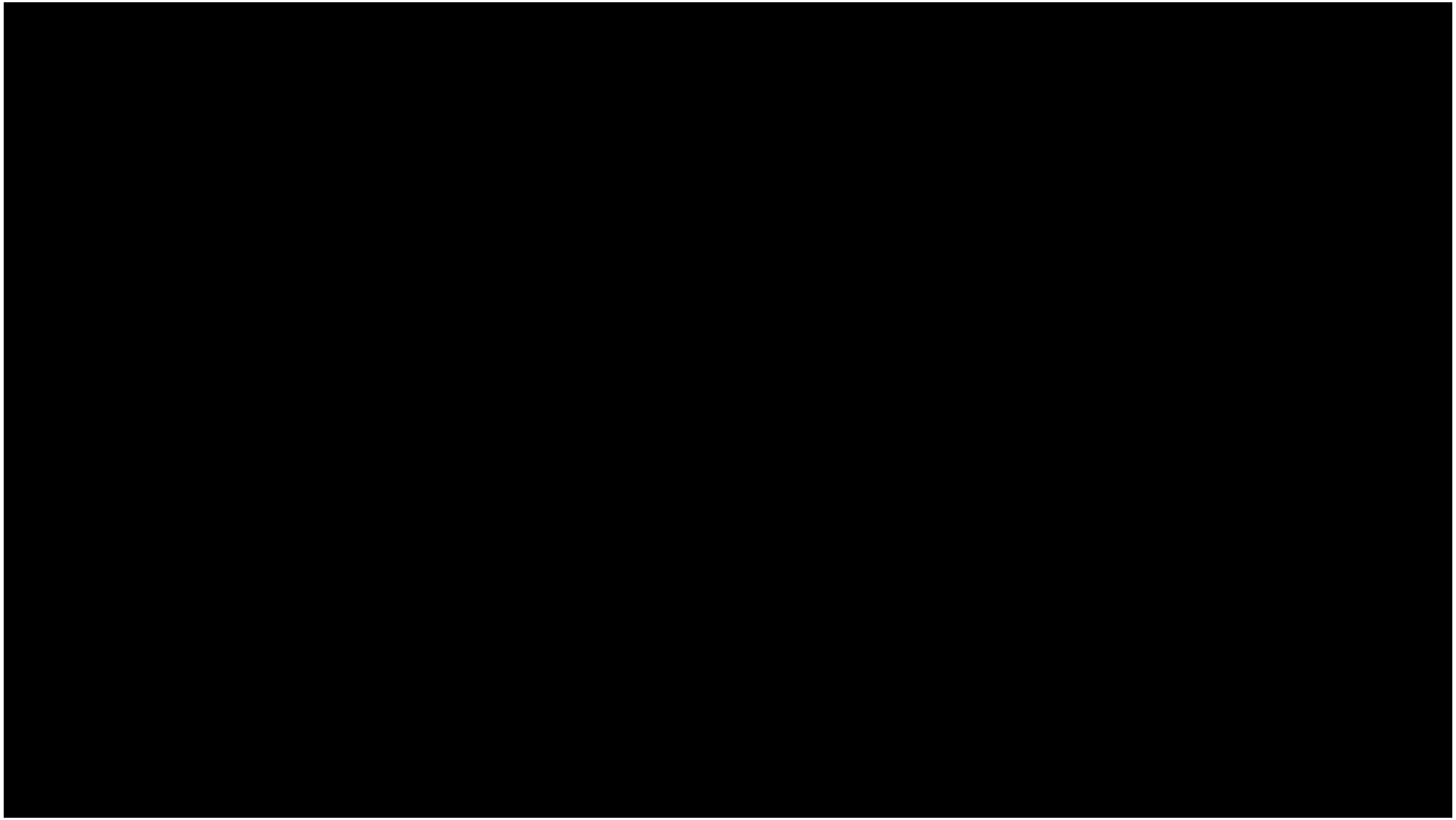
MCMC Sampling Algorithms

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

Stan!

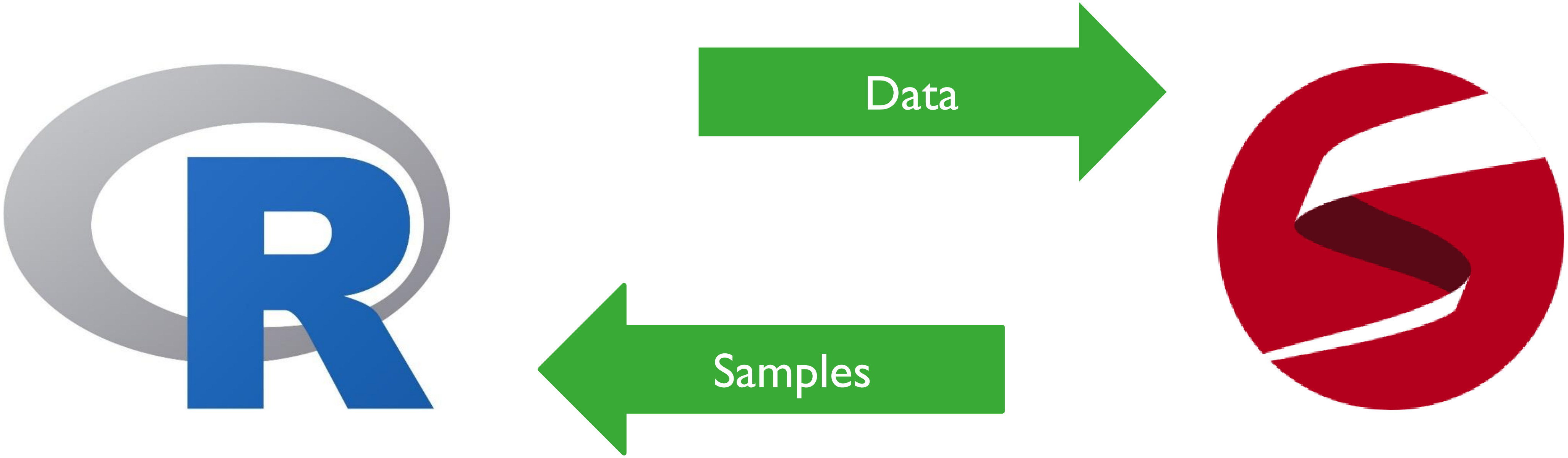
*: also PyMC3 PyMC3

Let's watch a video!



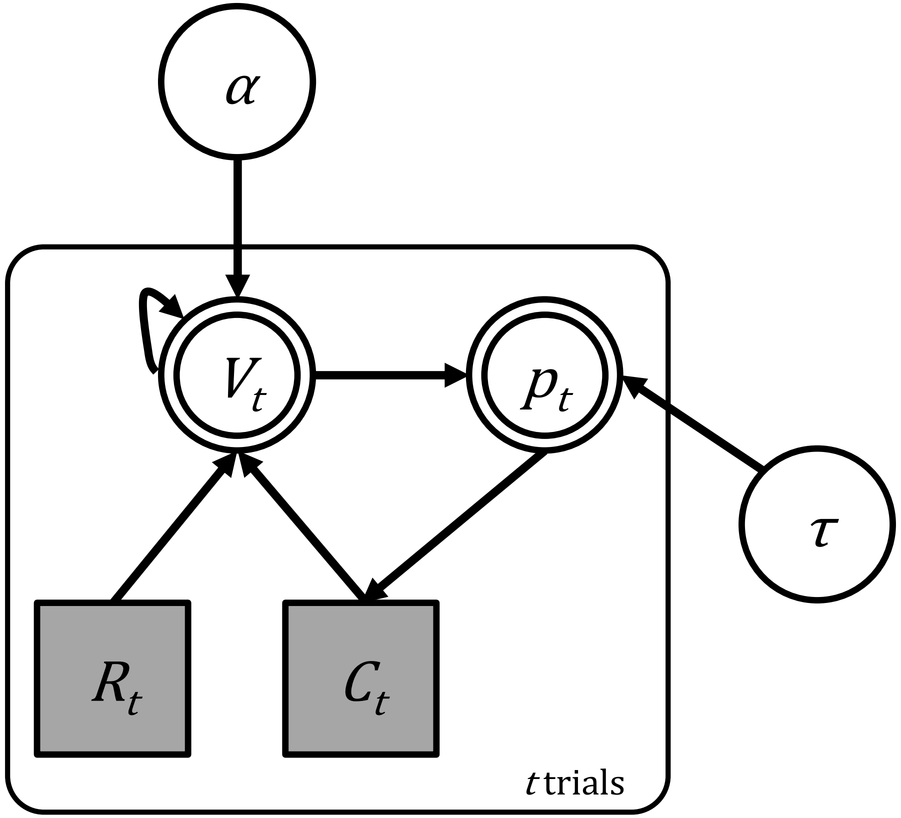
Credit: Thomas Wiecki

Stan and RStan



RL – Implementation

cognitive model
statistics
computing



$$\alpha \sim Uniform(0,1)$$

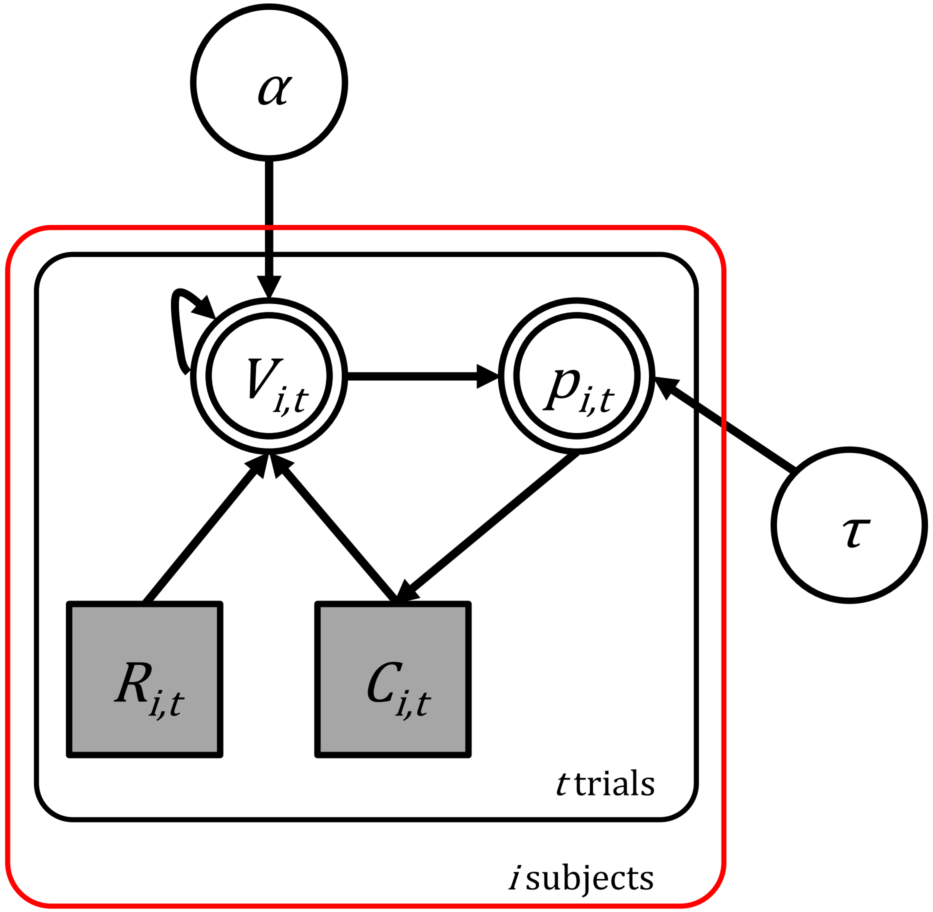
$$\tau \sim Uniform(0,3)$$

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

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

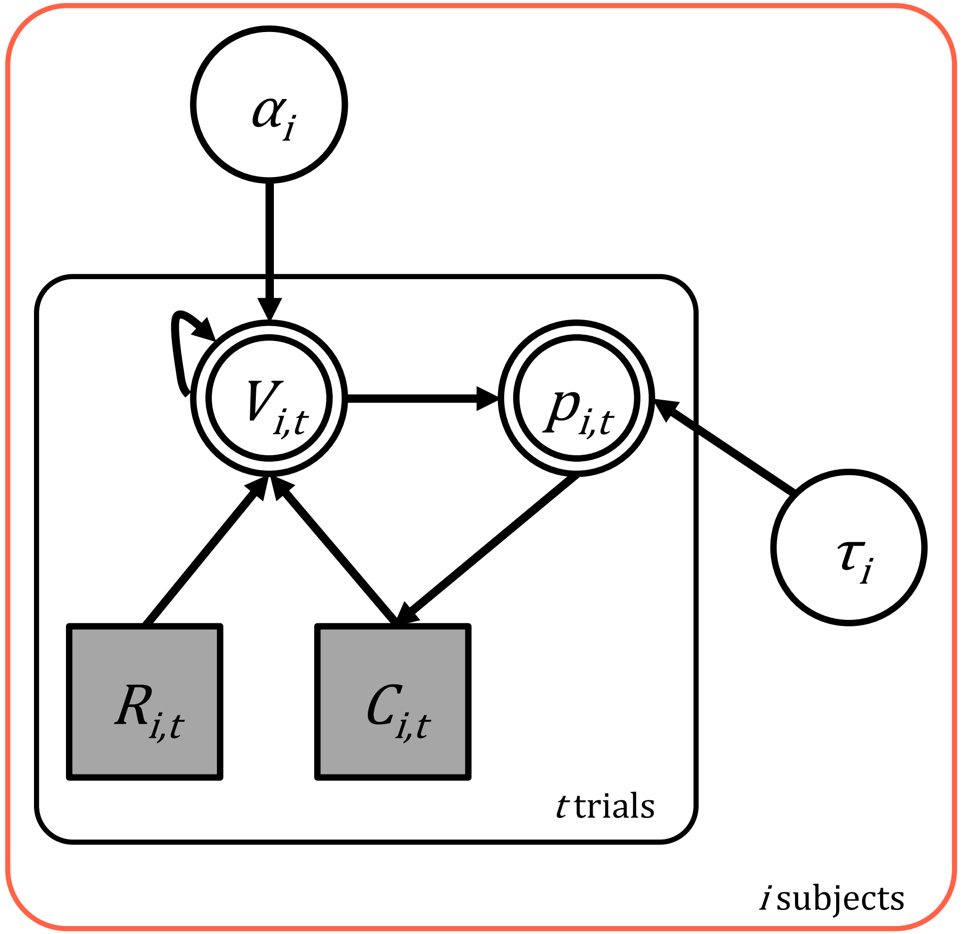
deterministic variable

Fitting Multiple Participants as ONE



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

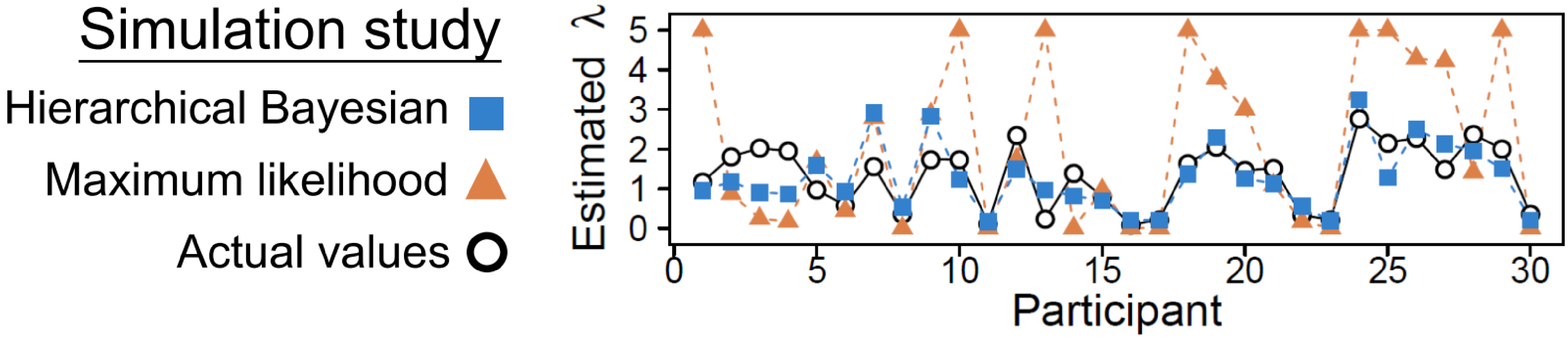
- each subject is fitted independently of the others
- best model fit for each subject
- parameter estimates can be noisy

Why Hierarchical Bayesian Cognitive Modeling?

cognitive model

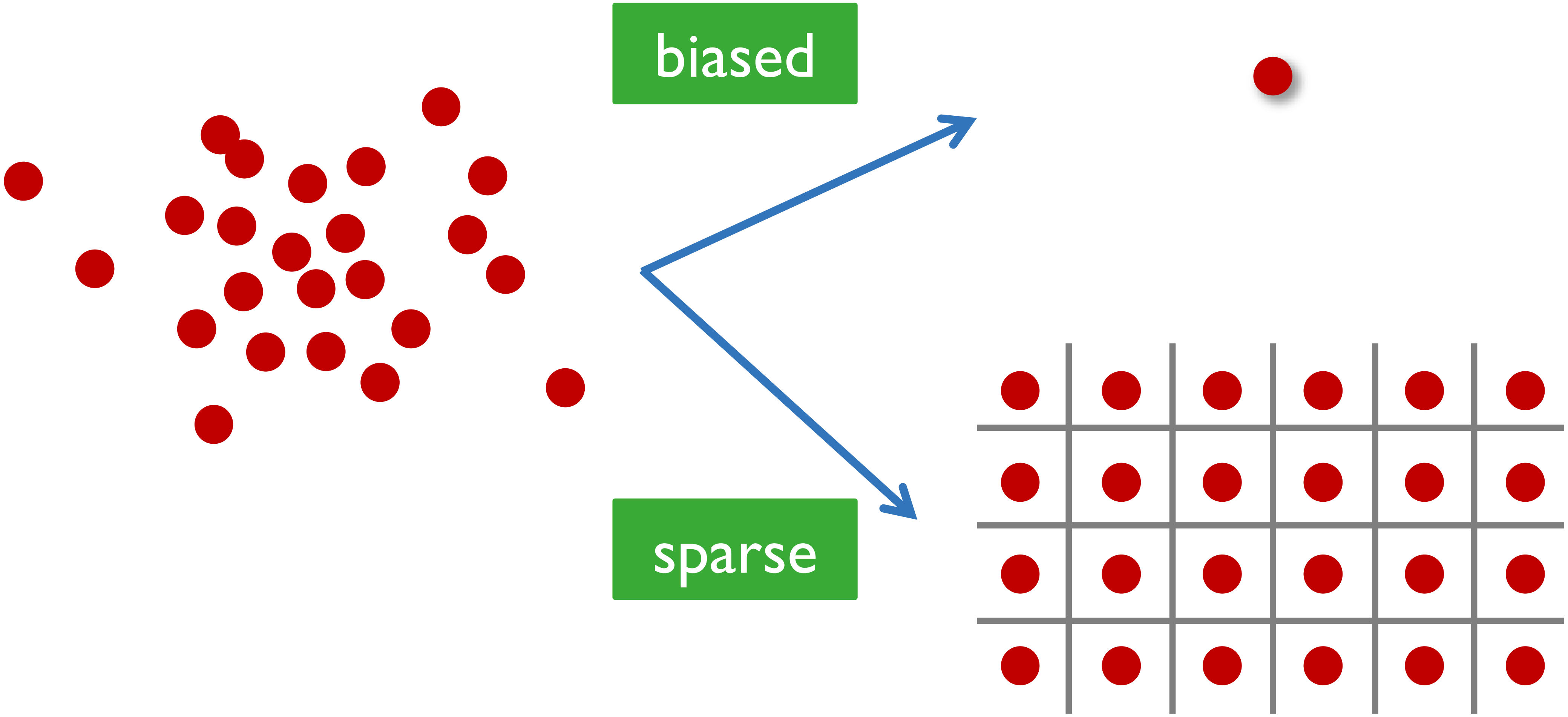
statistics

computing

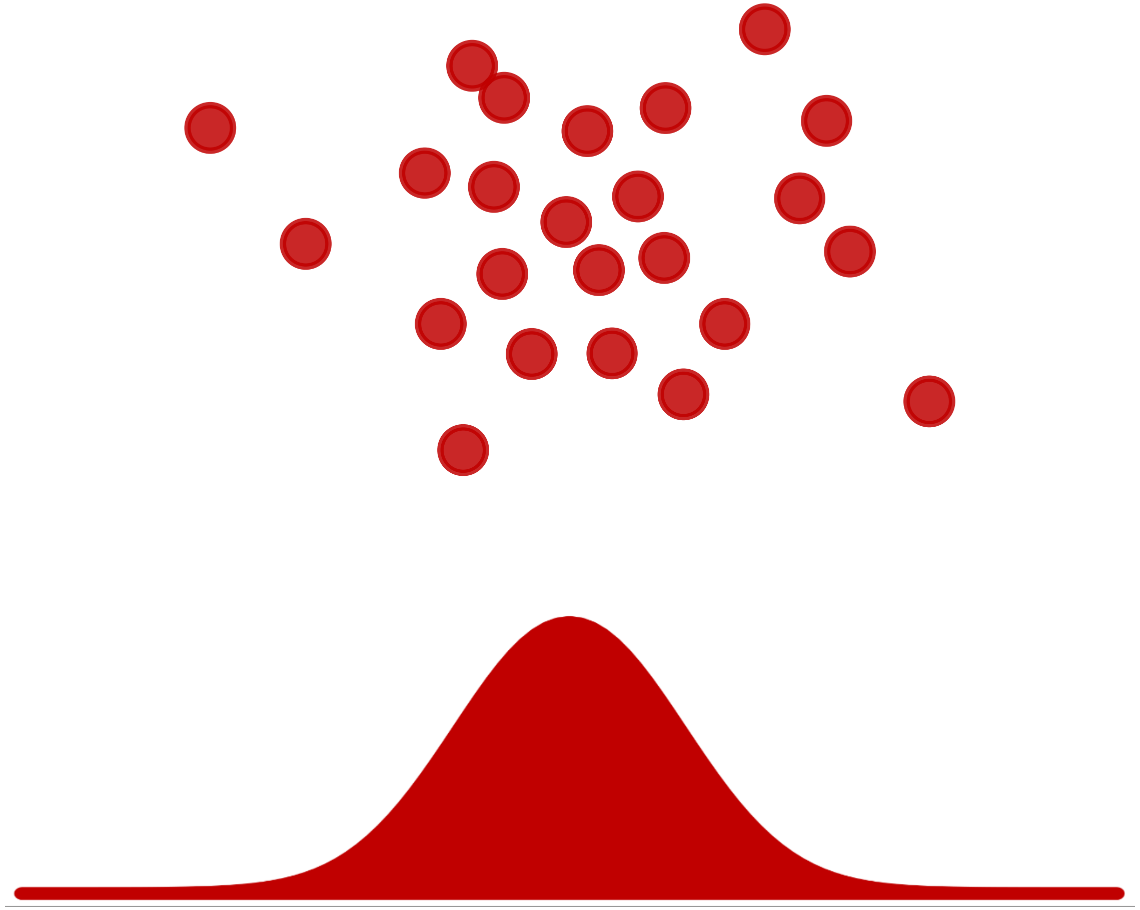


Ahn et al. (2011)

Fitting Multiple Participants



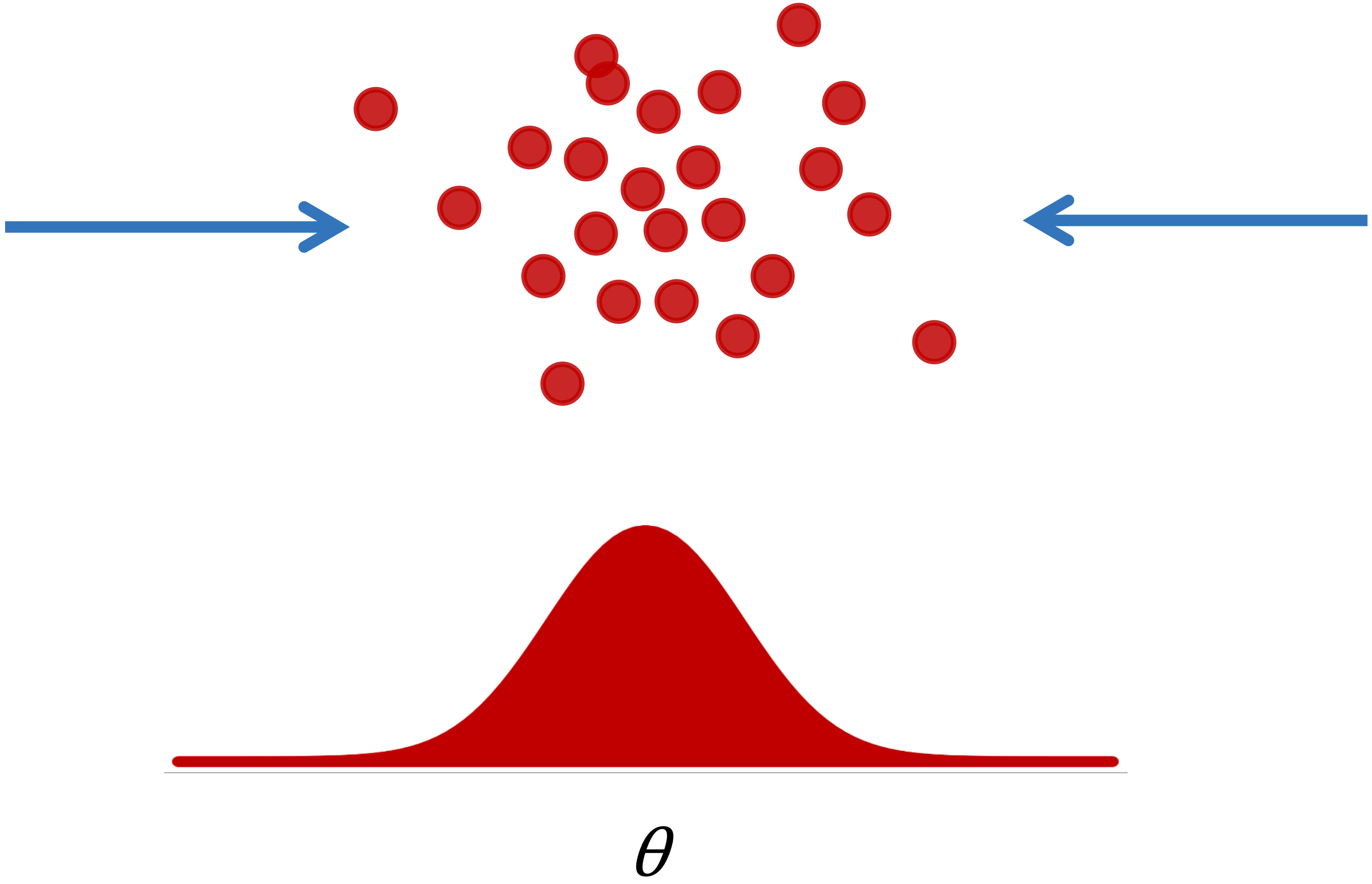
Fitting Multiple Participants



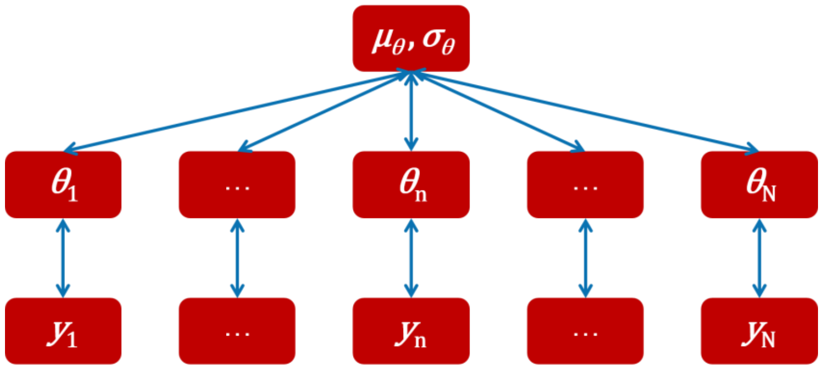
θ

Fitting Multiple Participants

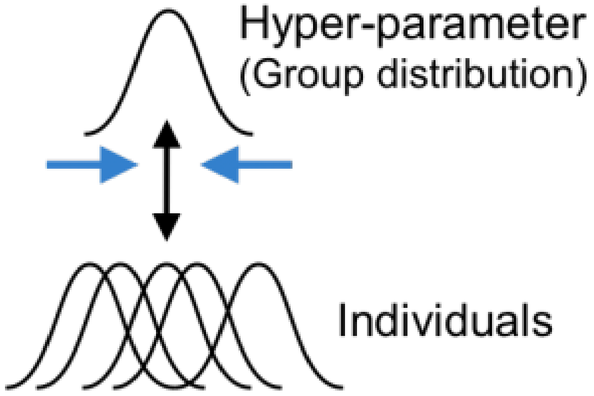
cognitive model
statistics
computing

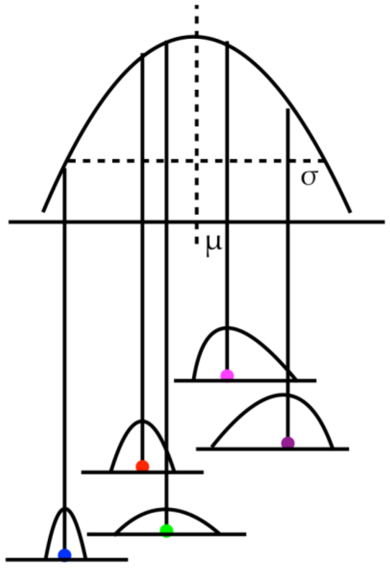


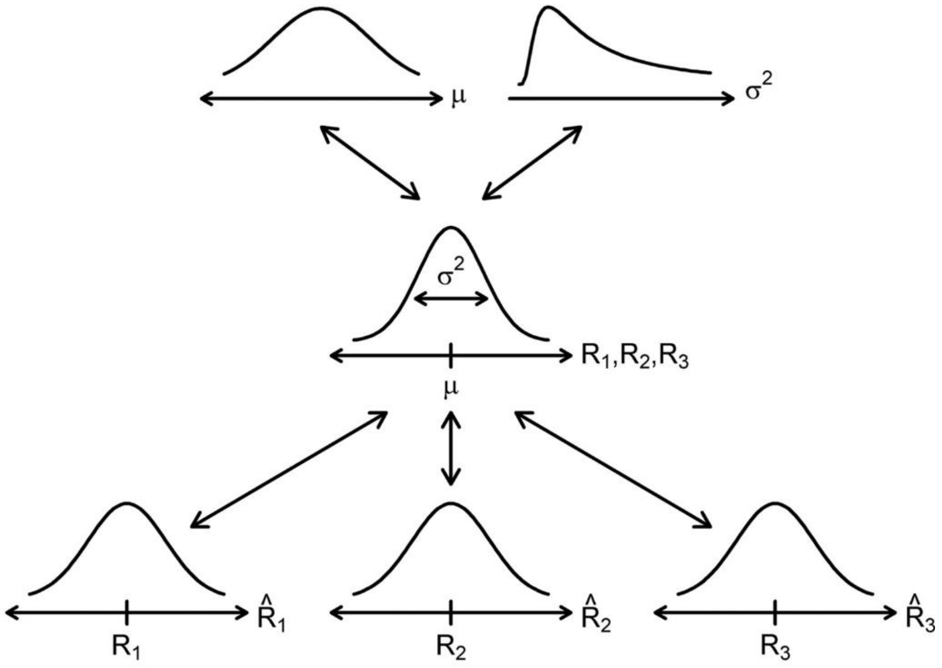
Hierarchical Structure



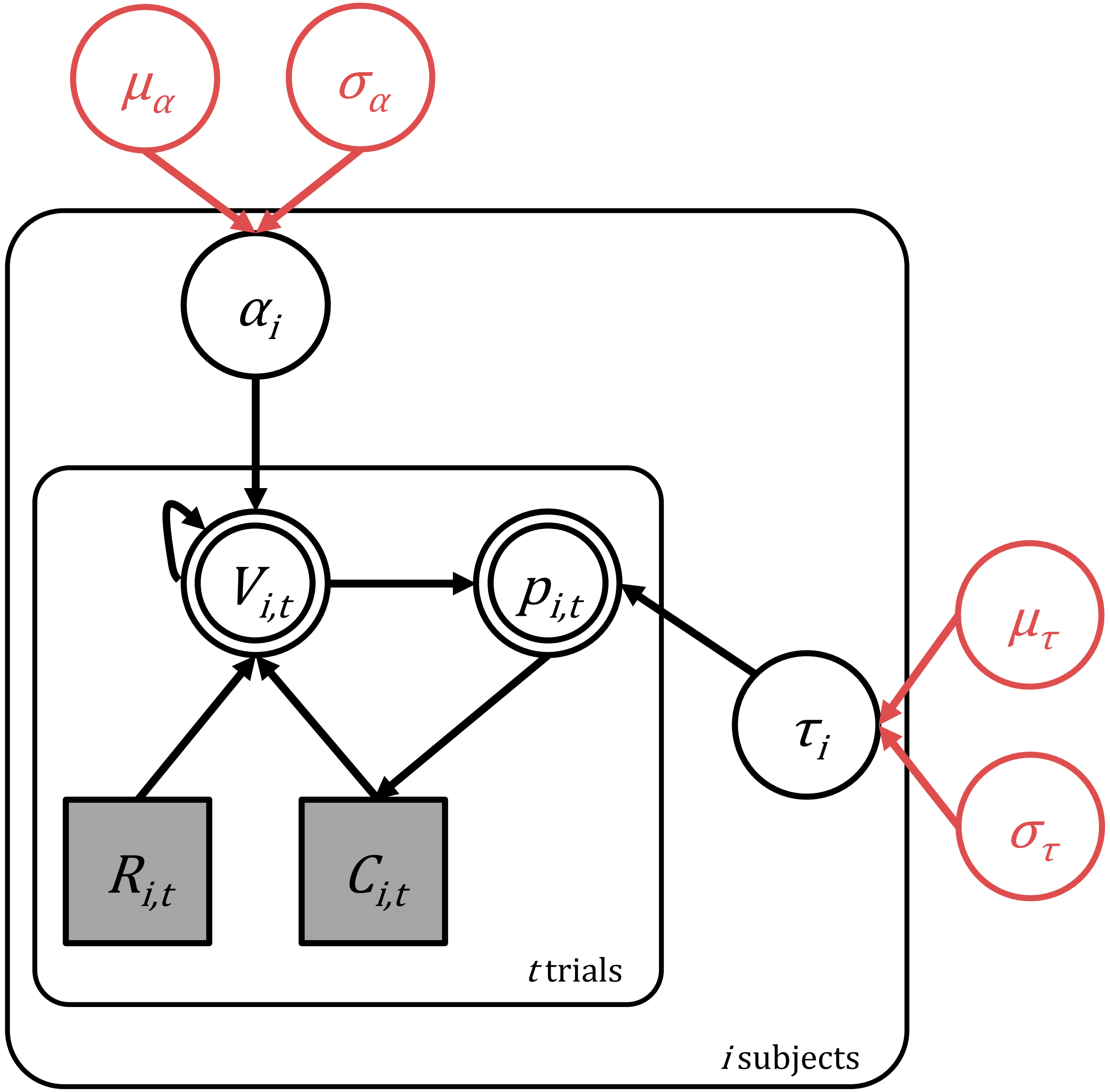
$$P(\Theta,\Phi \mid D) = \frac{P(D \mid \Theta,\Phi)P(\Theta,\Phi)}{P(D)} \propto P(D \mid \Theta)P(\Theta \mid \Phi)P(\Phi)$$





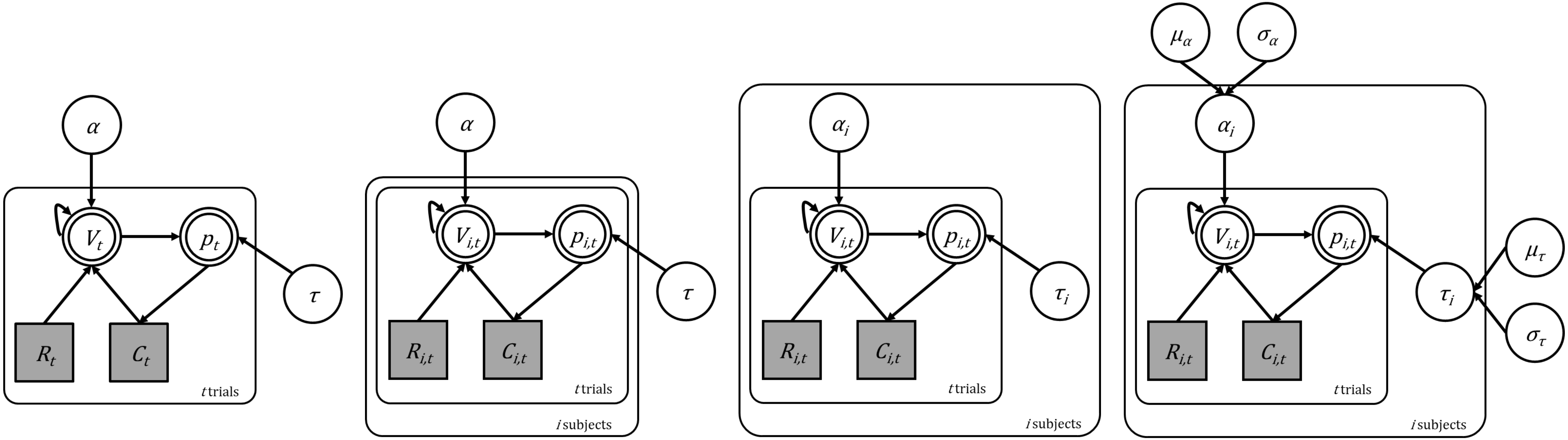


Hierarchical RL Model



HOW DID WE GET HERE?

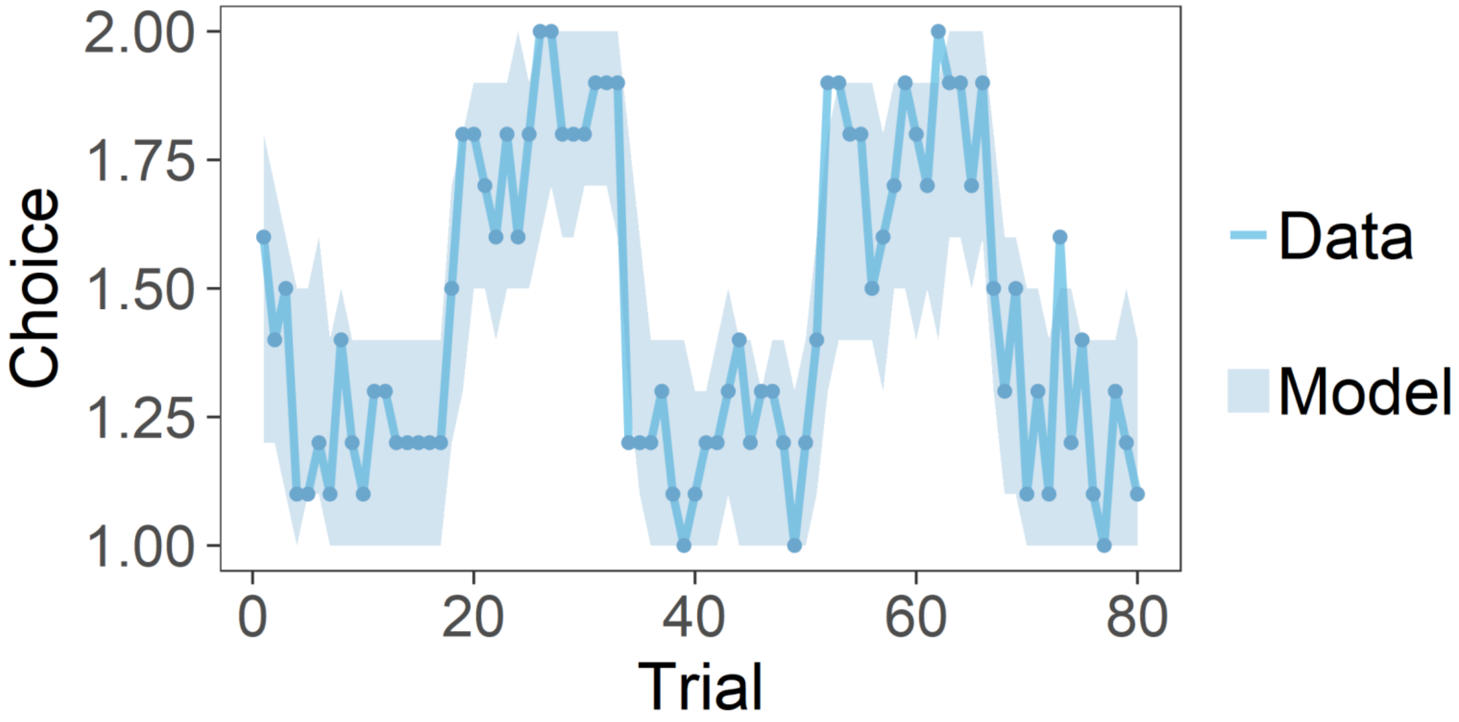
cognitive model
statistics
computing



The cognitive model *per se* is the same!

Posterior Predictive Check

cognitive model
statistics
computing



Exercise 2

cognitive model
statistics
computing

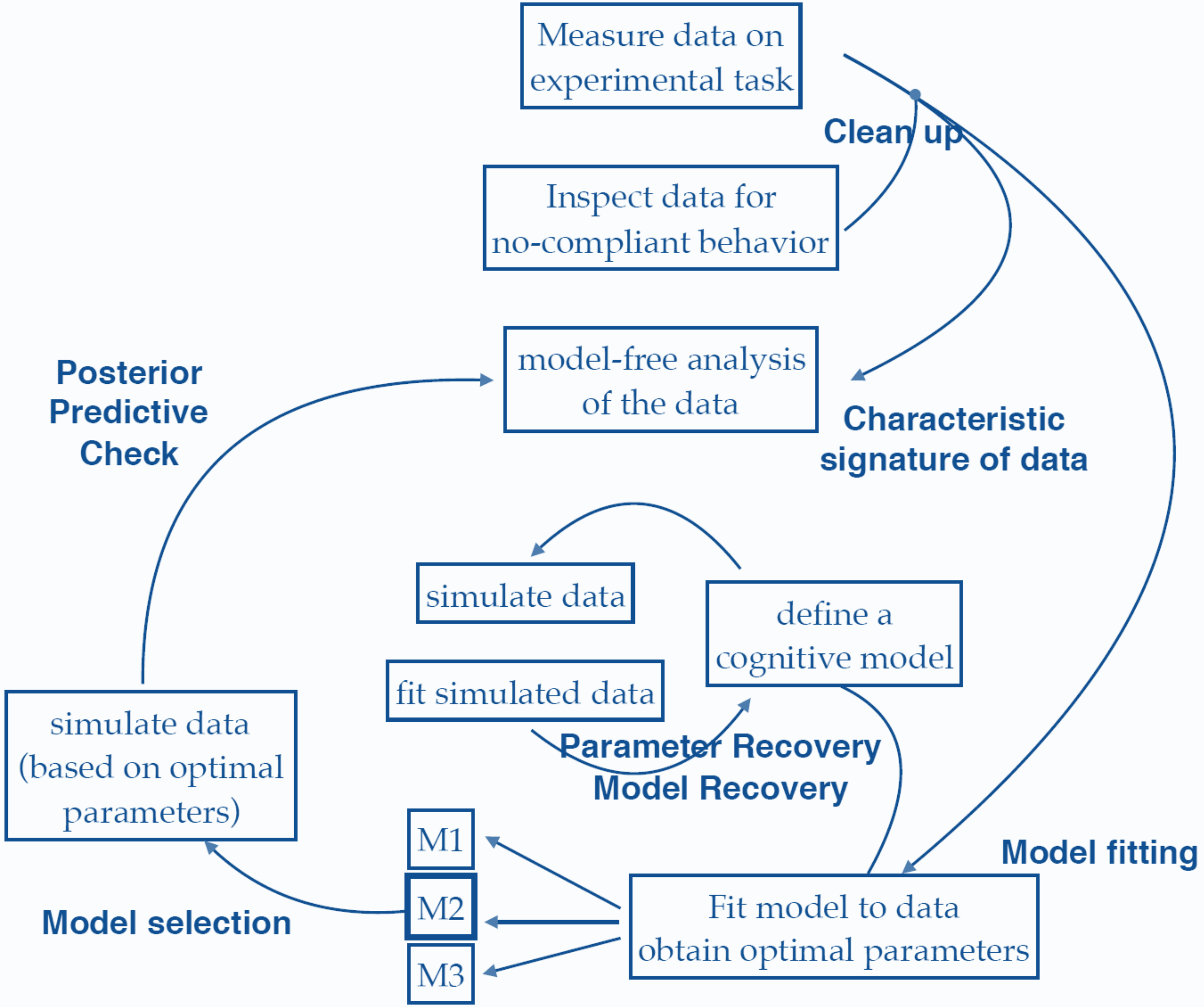
```
…/RL_tutorial/_scripts/MB_analysis.R
```

TASK:

fit the simple RL model

… and plot the model predictions on top of data

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



Measure data on experimental task
Clean up
Inspect data for no-compliant behavior
model-free analysis of the data
Characteristic signature of data
Posterior Predictive Check
simulate data
define a cognitive model
fit simulated data
Parameter Recovery
Model Recovery
simulate data (based on optimal parameters)
M1
M2
M3
Model selection
Model fitting
Fit model to data obtain optimal parameters

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!