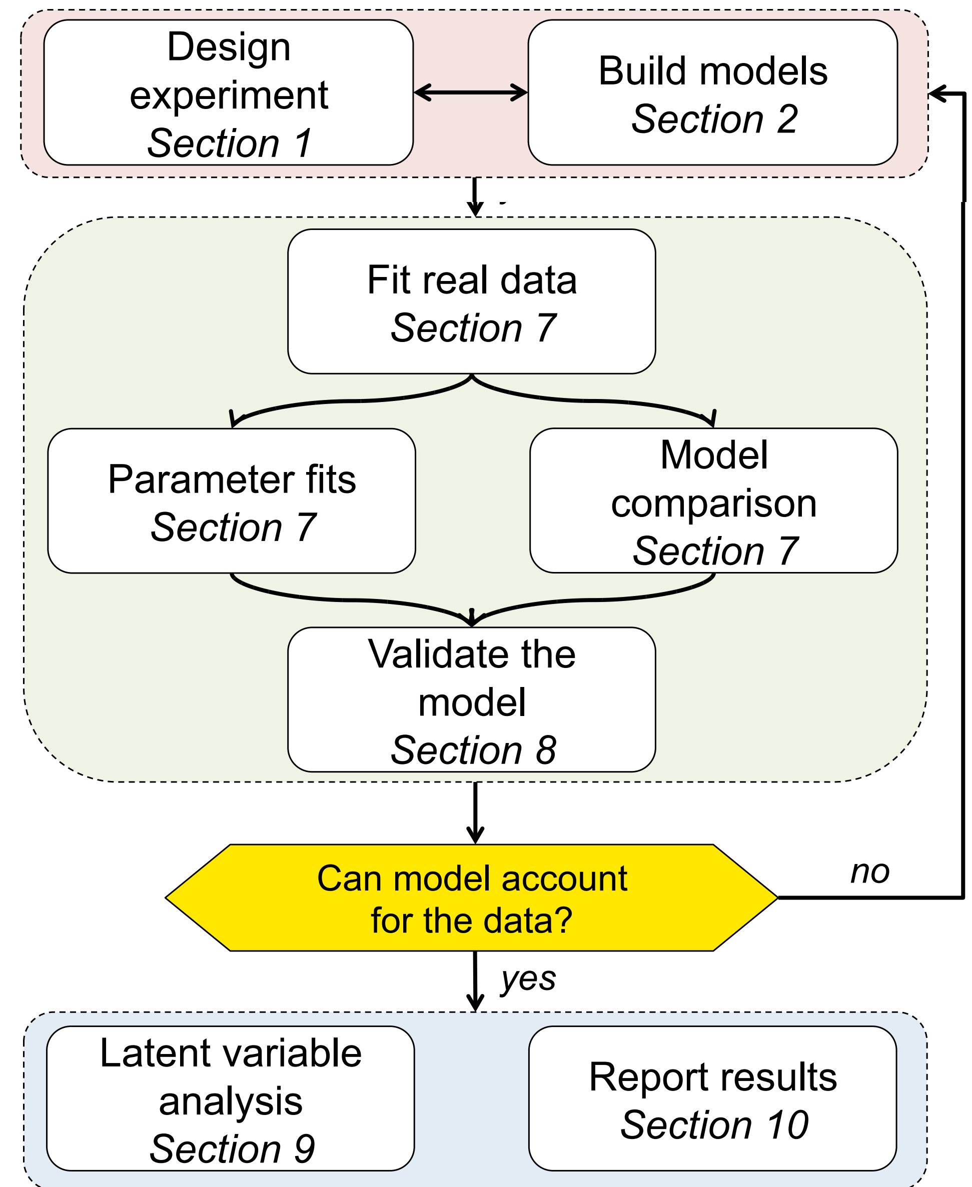
(Design), Parameterization & Selection of Behavioral Models

Benoît Girard

benoit.girard@isir.upmc.fr

Modeling Process

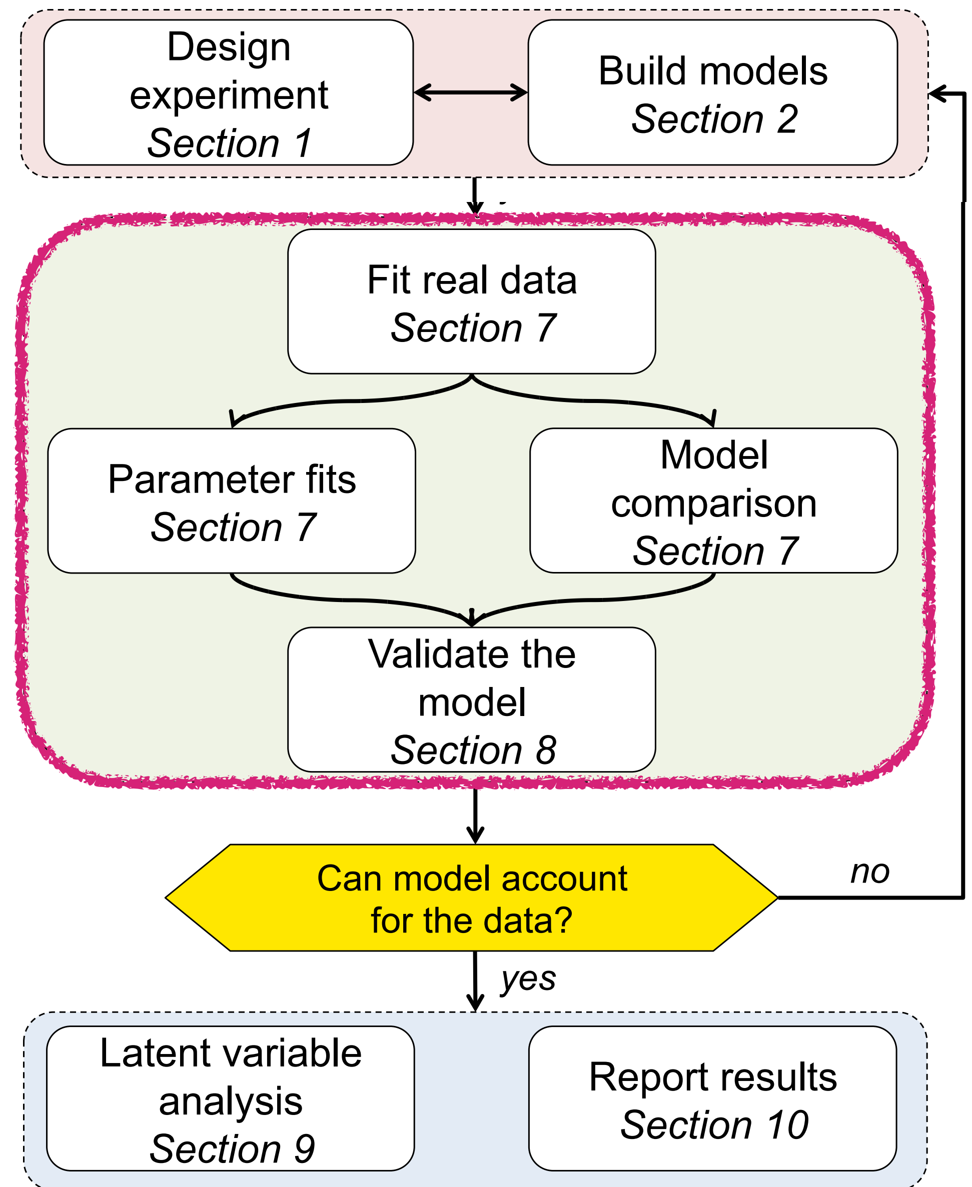
- Wilson, R. C., & Collins, A. G. (2019).
  Ten simple rules for the computational modeling of behavioral data.
  Elife, 8, e49547.



Adapted from (Wilson & Collins, 2019)

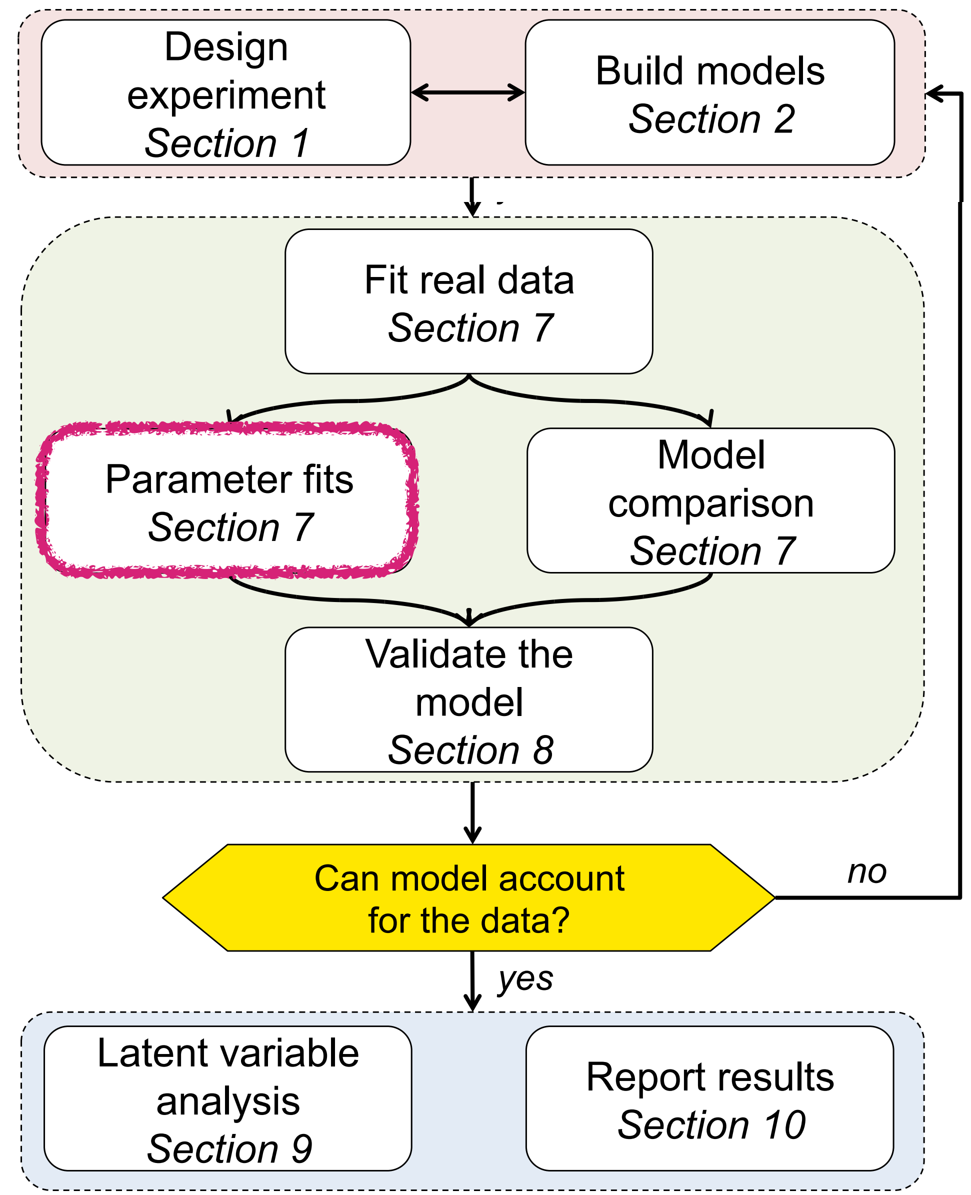
Modeling Process

- Wilson, R. C., & Collins, A. G. (2019).
  Ten simple rules for the computational modeling of behavioral data.
  Elife, 8, e49547.



Adapted from (Wilson & Collins, 2019)

Simple example



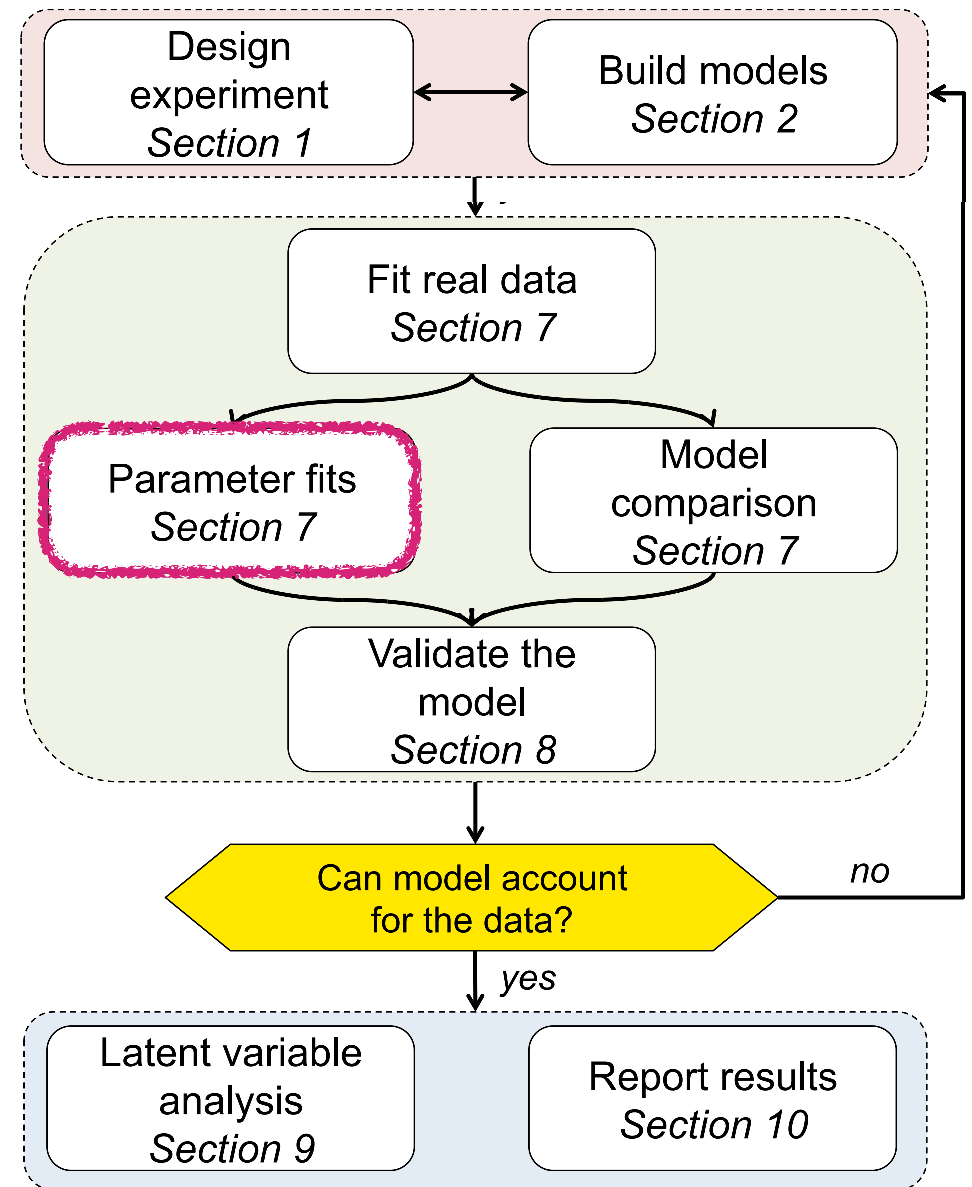
Adapted from (Wilson & Collins, 2019)

Simple example

- I think I measure a phenomenon generated by a normal distribution.

  This is my model!

- I need to find its parameter values: $\theta=(\mu,\sigma)$



Adapted from (Wilson & Collins, 2019)

Model parameter optimization

Log-likelihood maximization

- Model parameter optimization:
  find the parameters θ that best explain the observations $x_{1:N}$
  - Compute for each observation, the probability that the model had to generate that observation:
    $\mathcal{L}(\theta|x_i) = P(x=x_i|\theta)$ *(likelihood of the observation)*
  - Find the θ that maximize this value over all observations:
    - $\mathrm{argmax}_\theta(\mathcal{L}(\theta|x)) = \mathrm{argmax}_\theta(\Pi_{i=1:N} P(x=x_i|\theta))$

Compute log $\mathcal{L}$!

$$\mathcal{L}(\theta) = \prod_{i=1}^{N} p(x = x_i|\theta)$$

$$log\mathcal{L}(\theta) = \sum_{i=1}^{N} \log p(x = x_i|\theta)$$

Simple example

- I think I measure a phenomenon generated by a normal distribution.
- I need to find its parameter values: $\theta=(\mu,\sigma)$
- I test a number of (μ,σ) candidate values.
- For each measurement x_i I compute:

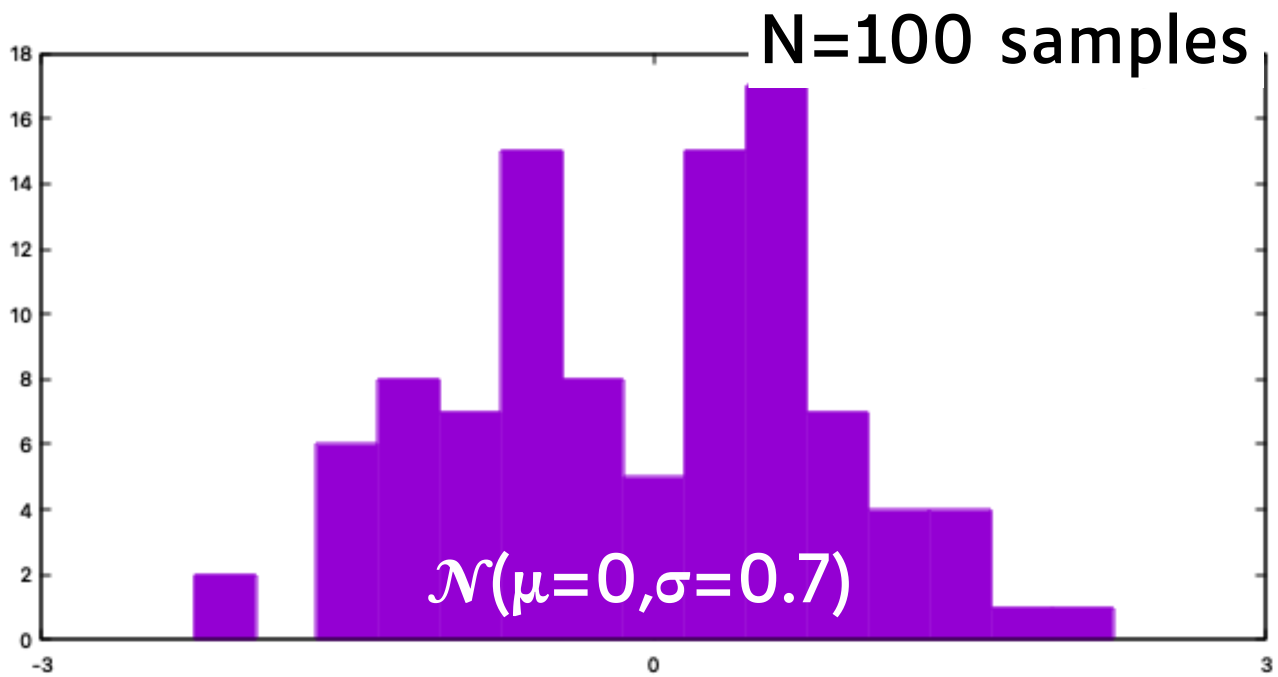
$$log(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x_i-\mu}{\sigma})^2})$$

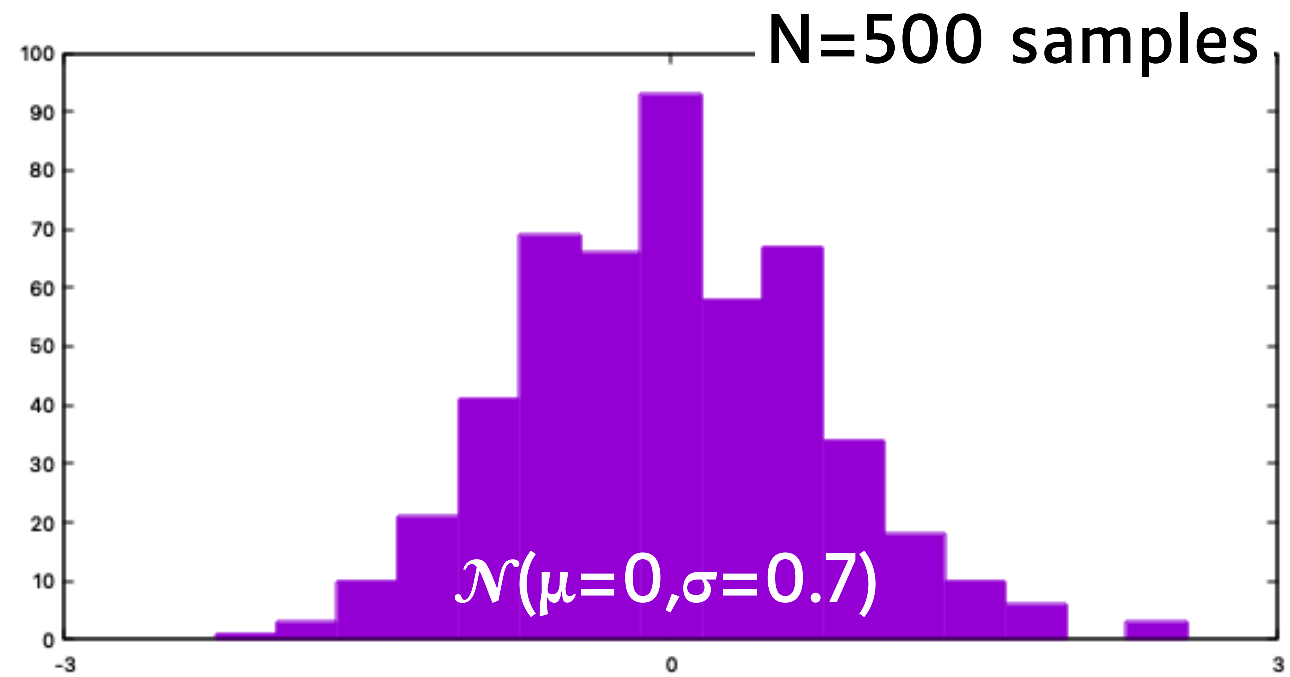
- I sum the values for all x_i

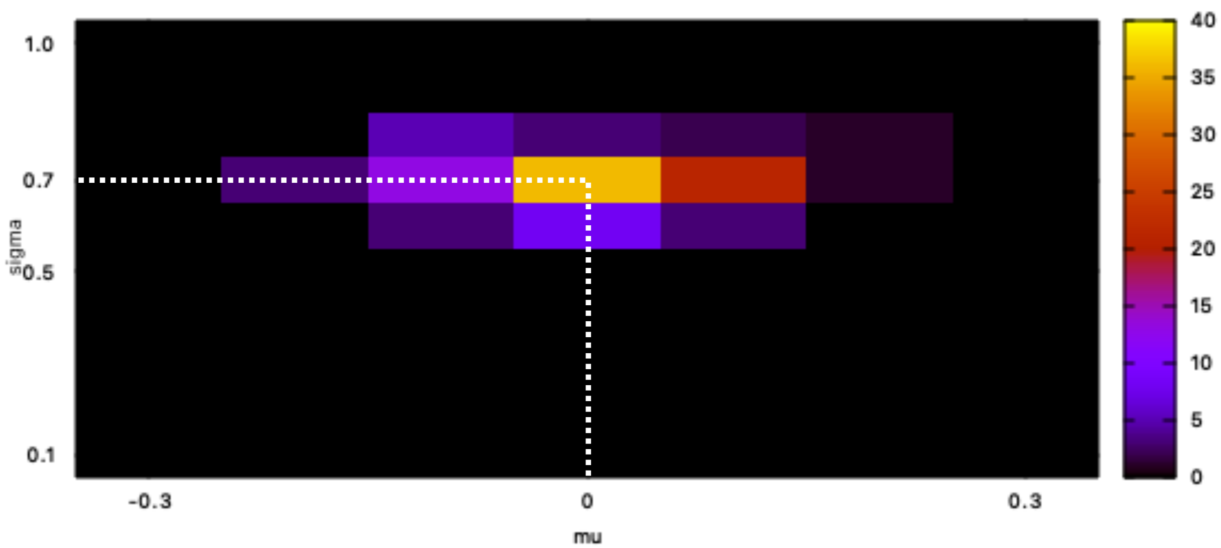
$$log\mathcal{L}(\mu,\sigma)=\sum_{i=1}^{N}log(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}(\frac{x_i-\mu}{\sigma})^2})$$

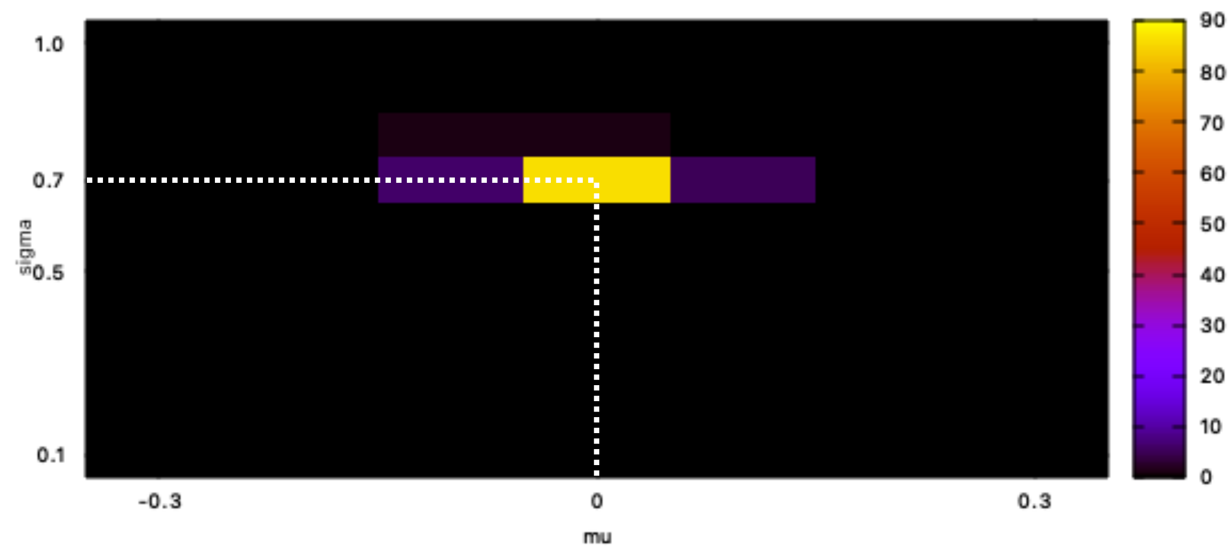
- I select the $\theta=(\mu,\sigma)$ for which this is maximum

Number of samples









Grid search:

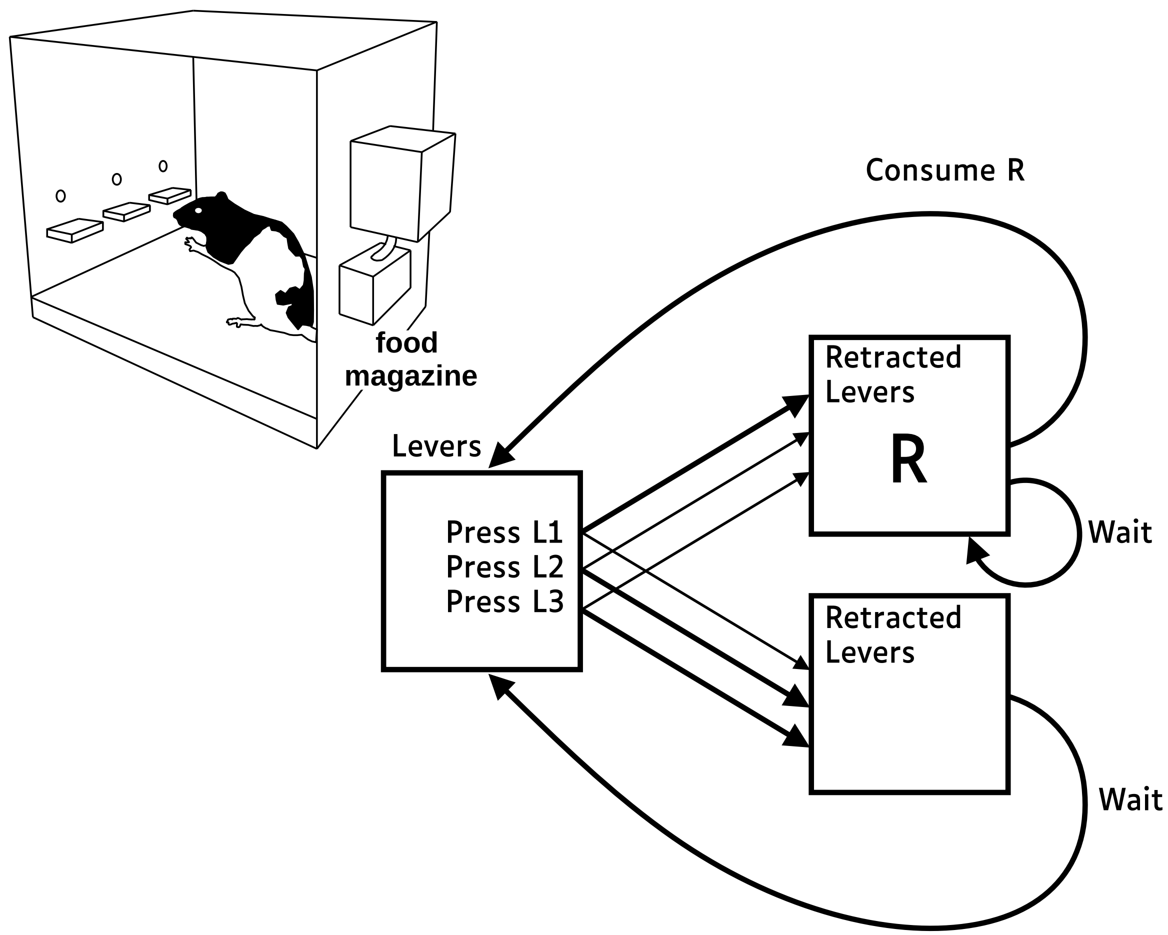
$\mu \in [-0.3,-0.2,-0.1,0.0,0.1,0.2,0.3]$

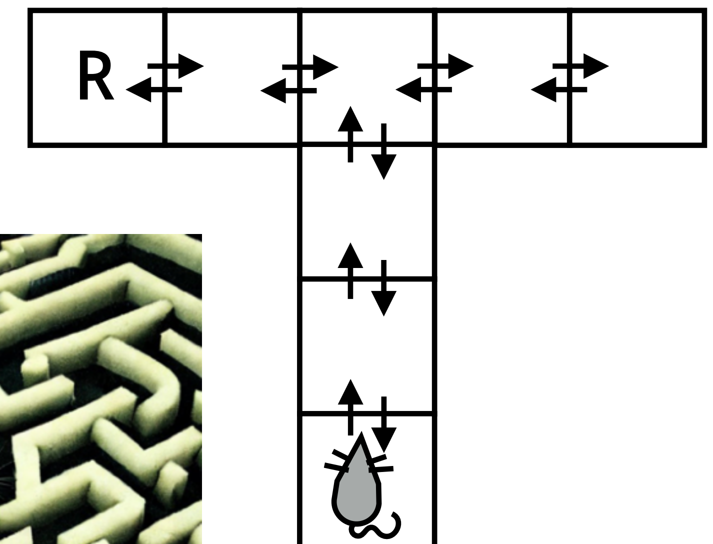
$\sigma \in [0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.]$

Exploring the parameter space

- Grid search:
  becomes expensive in large parameter spaces.
- Gradient descent (typically Matlab *fmincon*):
  can fall into local optima.
- Genetic algorithms (CMA-ES, NSGA-2):
  less prone to local optimization,
  can handle multiple objectives (to optimize).

Behavior unfolding in time





- Sequences of measurements $x_{1:T}$

- Made of states and actions (or RT, or...):
  $(s_1, a_1) \rightarrow (s_2, a_2) \rightarrow (s_3, a_3) \ldots$

With behavior unfolding in time:

- Model parameter optimization:
  find the parameters θ that best explain the sequence of observations ($s_{1:T}$, $a_{1:T}$)

  - Compute at each timestep, the probability that the model had to make the same choice as the subject: $\mathcal{L}(\theta | s_{1:t}, a_t) = P(a=a_t | s_{1:t}, a_{1:t-1}, \theta)$

  - If some learning is involve, learn according to the ($s_{1:t}$, $a_{1:t-1}$) from the actual observations.

  - Find the θ that maximize this value over the whole sequence:

    - $\mathrm{argmax}_\theta (\mathcal{L}(\theta | s_{1:T}, a_{1:T}) = \Pi_{t=1:T} P(a=a_t | s_{1:t}, a_{1:t-1}, \theta))$

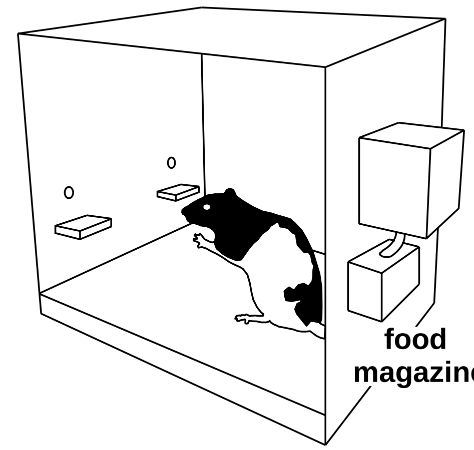
Compute log $\mathcal{L}$!

$$\mathcal{L}(\theta|s_{1:T}, a_{1:T}) = \prod_{i=1}^{T} p(a = a_t|s_{1:t}, a_{1:t}, \theta)$$

$$log\mathcal{L}(\theta|s_{1:T}, a_{1:T}) = \sum_{i=1}^{T} p(a = a_t|s_{1:t}, a_{1:t}, \theta)$$

An example of one model parameter optimization

Log-likelihood maximization



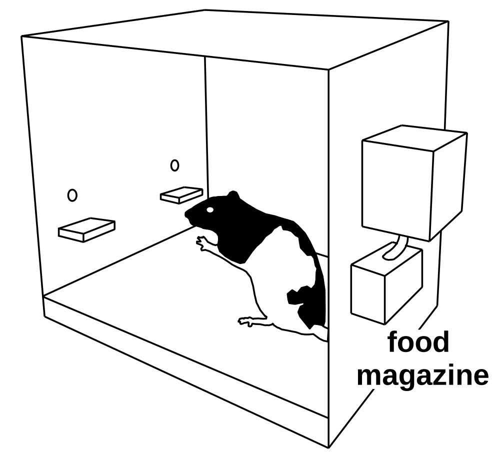
- 2-arm bandit task.
- Biased random response $p_t^1 = b \quad \text{and} \quad p_t^2 = 1 - b$ $\theta_1 = b$

Observations:

$a_0=1 \xrightarrow{R=0} a_1=0 \xrightarrow{R=0} a_2=1 \xrightarrow{R=0} a_3=1 \xrightarrow{R=1} a_4=1 \xrightarrow{R=0} a_5=0 \xrightarrow{R=1} a_6=1 \xrightarrow{R=0} a_7=0$ R=1

Model (θ_1=0.2): p_0=0.2 p_1=0.8

Log-likelihood maximization



- 2-arm bandit task.
- Biased random response $p_t^1 = b \text{ and } p_t^2 = 1 - b$ $\theta_1 = b$

Observations:

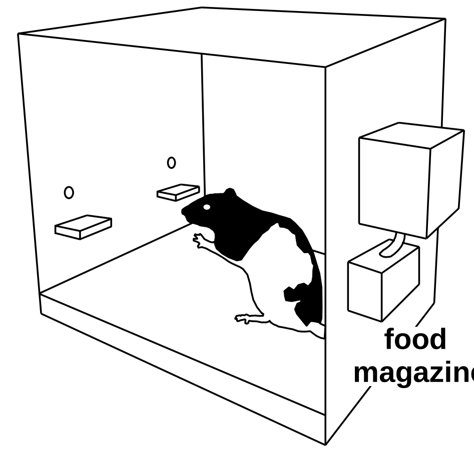
$a_0=1 \xrightarrow{R=0} a_1=0 \xrightarrow{R=0} a_2=1 \xrightarrow{R=0} a_3=1 \xrightarrow{R=1} a_4=1 \xrightarrow{R=0} a_5=0 \xrightarrow{R=1} a_6=1 \xrightarrow{R=0} a_7=0$ R=1

Model (θ_1=0.2): p_0=0.2 p_1=0.8

$\mathcal{L}(\theta|a_i)$= 0.8 0.2 0.8 0.8 0.8 0.2 0.8 0.2

$\mathcal{L}(\theta|a_{0:T})$= 0.8

Log-likelihood maximization



- 2-arm bandit task.
- Biased random response $p_t^1 = b \quad \text{and} \quad p_t^2 = 1 - b$ $\theta_1 = b$

Observations:

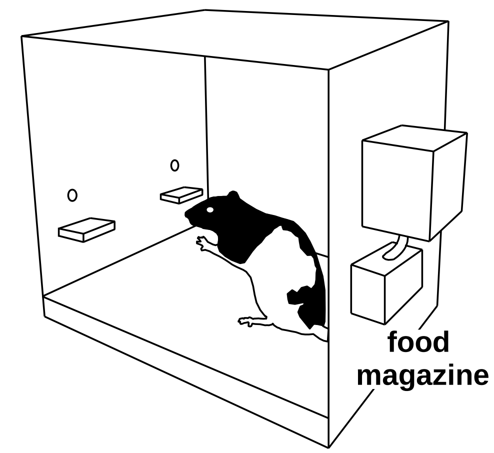
$$a_0=1 \xrightarrow{R=0} a_1=0 \xrightarrow{R=0} a_2=1 \xrightarrow{R=0} a_3=1 \xrightarrow{R=1} a_4=1 \xrightarrow{R=0} a_5=0 \xrightarrow{R=1} a_6=1 \xrightarrow{R=0} a_7=0 \quad R=1$$

Model (θ_1=0.2): p_0=0.2 p_1=0.8

$\mathcal{L}(\theta|a_i)$= 0.8 0.2 0.8 0.8 0.8 0.2 0.8 0.2

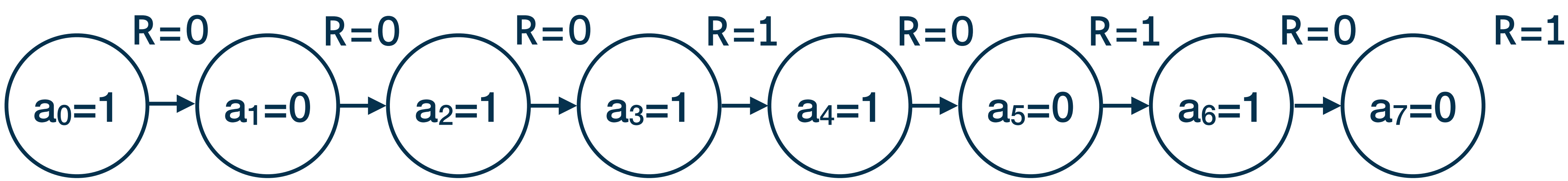
$\mathcal{L}(\theta|a_{0:T})$= 0.8 0.16

Log-likelihood maximization



- 2-arm bandit task.
- Biased random response $p_t^1 = b \text{ and } p_t^2 = 1 - b$ $\theta_1 = b$

Observations:

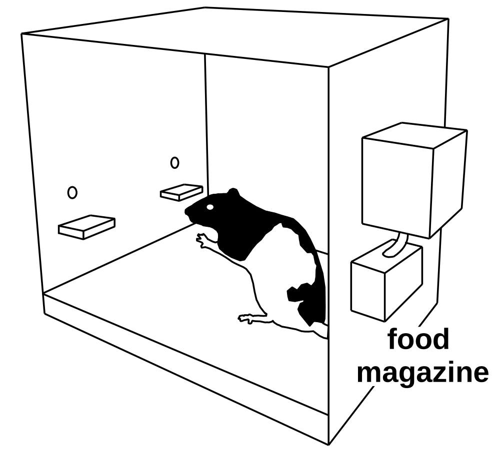


Model (θ_1=0.2): p_0=0.2 p_1=0.8

$\mathcal{L}(\theta|a_i)$= 0.8 0.2 0.8 0.8 0.8 0.2 0.8 0.2

$\mathcal{L}(\theta|a_{0:T})$= 0.8 0.16 0.13

Log-likelihood maximization



- 2-arm bandit task.
- Biased random response $p_t^1 = b \text{ and } p_t^2 = 1 - b$ $\theta_1 = b$

Observations:

R=0 R=0 R=0 R=1 R=0 R=1 R=0 R=1

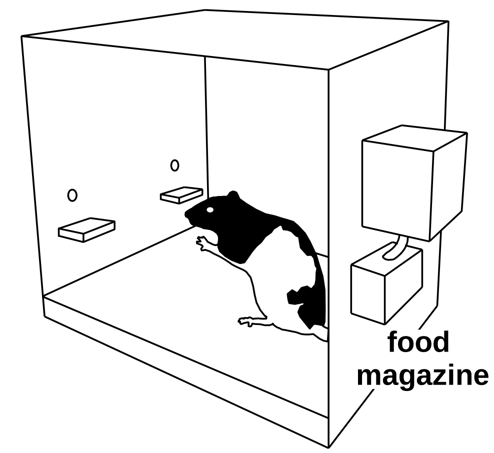
$a_0=1 \rightarrow a_1=0 \rightarrow a_2=1 \rightarrow a_3=1 \rightarrow a_4=1 \rightarrow a_5=0 \rightarrow a_6=1 \rightarrow a_7=0$

Model (θ_1=0.2): p_0=0.2 p_1=0.8

$\mathcal{L}(\theta|a_i)$= 0.8 0.2 0.8 0.8 0.8 0.2 0.8 0.2

$\mathcal{L}(\theta|a_{0:T})$= 0.8 0.16 0.13 0.1 0.08 0.016 0.013 0.0026

Log-likelihood maximization



- 2-arm bandit task.
- Biased random response $p_t^1 = b \text{ and } p_t^2 = 1 - b$ $\theta_1 = b$

Observations:

R=0 R=0 R=0 R=1 R=0 R=1 R=0 R=1

$a_0=1 \rightarrow a_1=0 \rightarrow a_2=1 \rightarrow a_3=1 \rightarrow a_4=1 \rightarrow a_5=0 \rightarrow a_6=1 \rightarrow a_7=0$

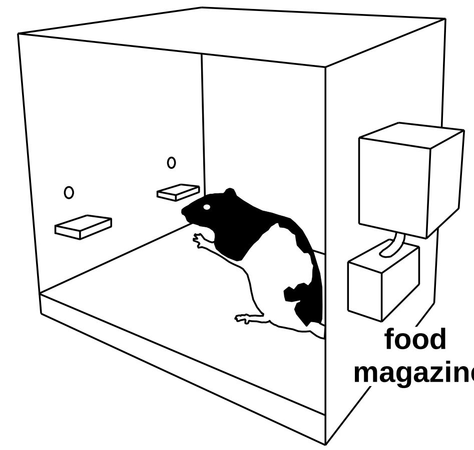
Model (θ_1=0.2): p_0=0.2 p_1=0.8

$\mathcal{L}(\theta|a_i)$= 0.8 0.2 0.8 0.8 0.8 0.2 0.8 0.2

x x x x x x x

$\mathcal{L}(\theta|a_{0:T})$= 0.8 0.16 0.13 0.1 0.08 0.016 0.013 0.0026

Log-likelihood maximization



- 2-arm bandit task.
- Biased random response $p_t^1 = b$ and $p_t^2 = 1 - b$ $\theta_1 = b$

Observations:

R=0	R=0	R=0	R=1	R=0	R=1	R=0	R=1
$a_0=1$	$a_1=0$	$a_2=1$	$a_3=1$	$a_4=1$	$a_5=0$	$a_6=1$	$a_7=0$

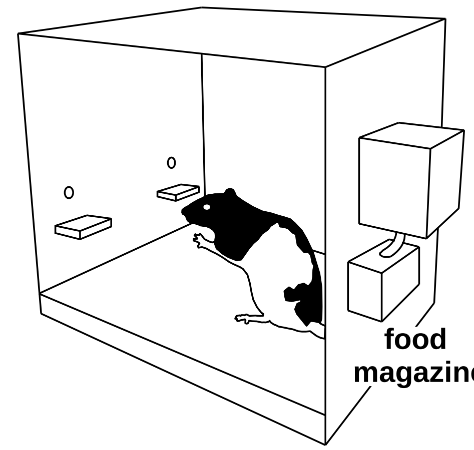
$a_0=1 \rightarrow a_1=0 \rightarrow a_2=1 \rightarrow a_3=1 \rightarrow a_4=1 \rightarrow a_5=0 \rightarrow a_6=1 \rightarrow a_7=0$

Model (θ_1=0.2): p_0=0.2 p_1=0.8

$\mathcal{L}(\theta|a_i)$= 0.8 0.2 0.8 0.8 0.8 0.2 0.8 0.2

Log $\mathcal{L}(\theta|a_{0:T})$=-0.22

Log-likelihood maximization



- 2-arm bandit task.
- Biased random response $p_t^1 = b \;\text{ and }\; p_t^2 = 1 - b$ $\theta_1 = b$

Observations:

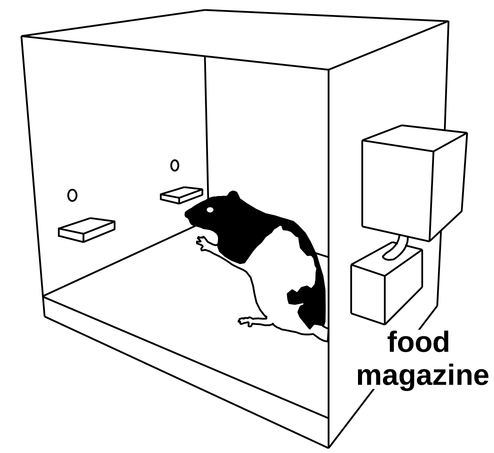
$a_0=1 \xrightarrow{R=0} a_1=0 \xrightarrow{R=0} a_2=1 \xrightarrow{R=0} a_3=1 \xrightarrow{R=1} a_4=1 \xrightarrow{R=0} a_5=0 \xrightarrow{R=1} a_6=1 \xrightarrow{R=0} a_7=0$ R=1

Model (θ_1=0.2): p_0=0.2 p_1=0.8

$\mathcal{L}(\theta|a_i)$= 0.8 0.2 0.8 0.8 0.8 0.2 0.8 0.2

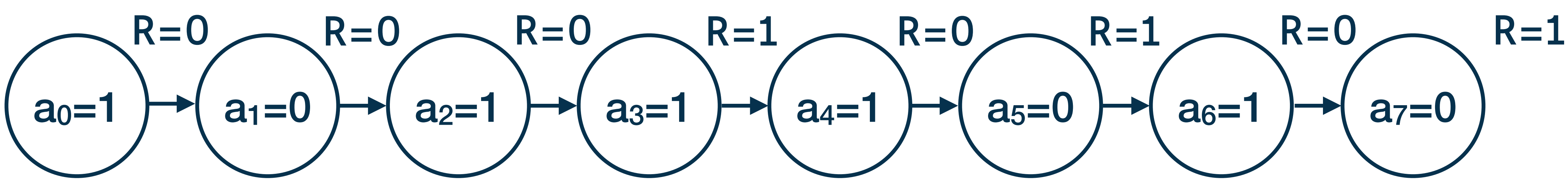
Log $\mathcal{L}(\theta|a_{0:T})$=-0.22 -0.18

Log-likelihood maximization



- 2-arm bandit task.
- Biased random response $p_t^1 = b \text{ and } p_t^2 = 1 - b$ $\theta_1 = b$

Observations:



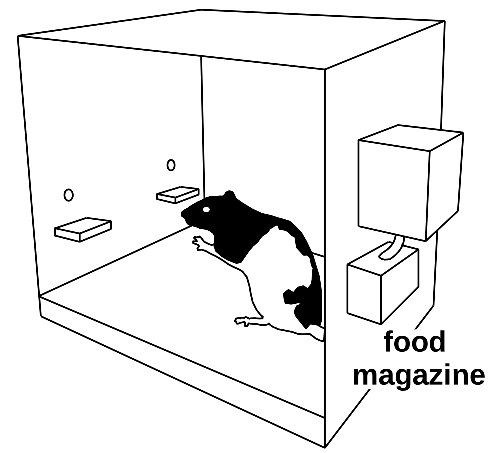
Model (θ_1=0.2): p_0=0.2 p_1=0.8

$\mathcal{L}(\theta|a_i)$= 0.8 0.2 0.8 0.8 0.8 0.2 0.8 0.2

Log $\mathcal{L}(\theta|a_{0:T})$=-0.22 -0.18 -2.0

Log-likelihood maximization

- 2-arm bandit task.



- Biased random response $p_t^1 = b \;\; \text{and} \;\; p_t^2 = 1 - b$ $\quad \theta_1 = b$

Observations:

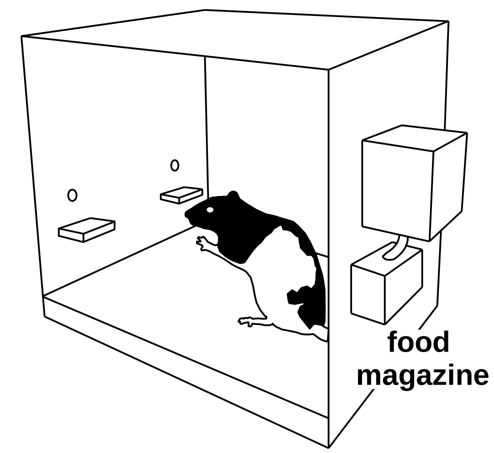
R=0	R=0	R=0	R=1	R=0	R=1	R=0	R=1
$a_0=1$	$a_1=0$	$a_2=1$	$a_3=1$	$a_4=1$	$a_5=0$	$a_6=1$	$a_7=0$

Model (θ_1=0.2): p_0=0.2 p_1=0.8

$\mathcal{L}(\theta \mid a_i)=$	0.8	0.2	0.8	0.8	0.8	0.2	0.8	0.2
Log $\mathcal{L}(\theta \mid a_{0:T})=$	-0.22	-0.18	-2.0	-2.3	-2.5	-4.1	-4.3	-5.9

Log-likelihood maximization

- 2-arm bandit task.



- Biased random response $p_t^1 = b \text{ and } p_t^2 = 1 - b$ $\theta_1 = b$

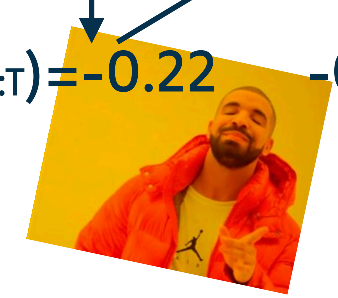
Observations:

$a_0=1 \xrightarrow{R=0} a_1=0 \xrightarrow{R=0} a_2=1 \xrightarrow{R=0} a_3=1 \xrightarrow{R=1} a_4=1 \xrightarrow{R=0} a_5=0 \xrightarrow{R=1} a_6=1 \xrightarrow{R=0} a_7=0$ R=1

Model (θ_1=0.2): p_0=0.2 p_1=0.8

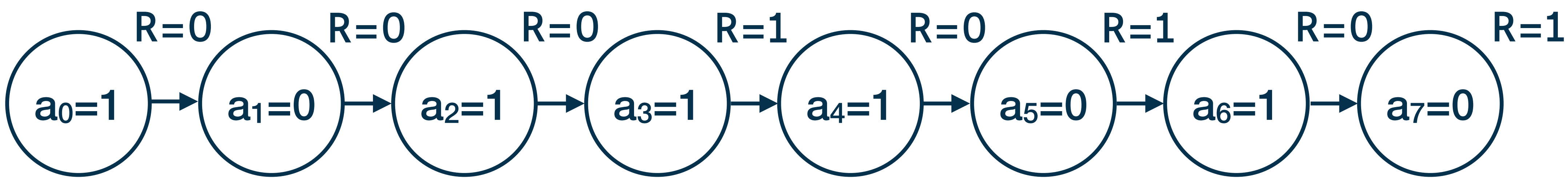
$\mathcal{L}(\theta|a_i)$= 0.8 0.2 0.8 0.8 0.8 0.2 0.8 0.2

Log $\mathcal{L}(\theta|a_{0:T})$=-0.22 -0.18 -2.0 -2.3 -2.5 -4.1 -4.3 -5.9

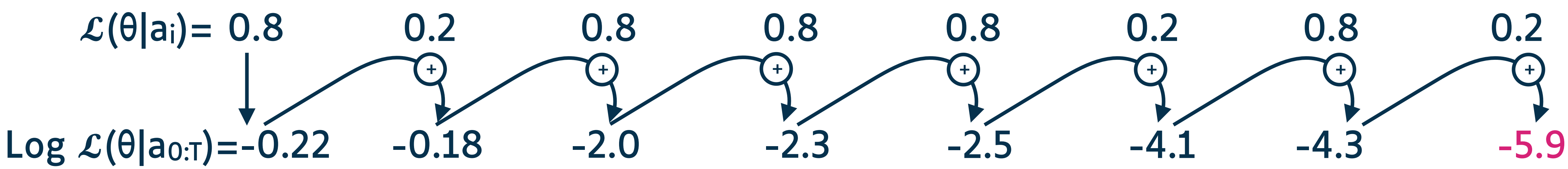


Log-likelihood maximization

Observations:

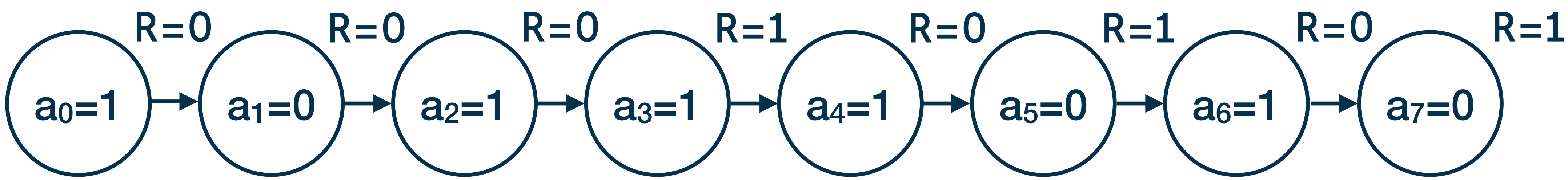


Model (θ_1=0.2): p_1=0.2 p_2=0.8



Log-likelihood maximization

Observations:



Model (θ_1=0.2): p_1=0.2 p_2=0.8

$\mathcal{L}(\theta|a_i)$= 0.8 0.2 0.8 0.8 0.8 0.2 0.8 0.2

Log $\mathcal{L}(\theta|a_{0:T})$=-0.22 -0.18 -2.0 -2.3 -2.5 -4.1 -4.3 -5.9

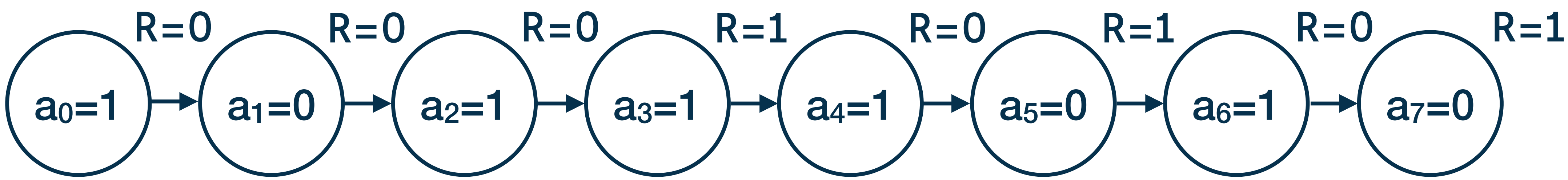
Model (θ_1=0.3): p_1=0.3 p_2=0.7

$\mathcal{L}(\theta|a_i)$= 0.7 0.3 0.7 0.7 0.7 0.3 0.7 0.3

Log $\mathcal{L}(\theta|a_{0:T})$=-0.35 -1.6 -1.9 -2.3 -2.6 -3.8 -4.2 -5.4

Log-likelihood maximization

Observations:



Model (θ_1=0.2): p_1=0.2 p_2=0.8

$\mathcal{L}(\theta \vert a_i)$=	0.8	0.2	0.8	0.8	0.8	0.2	0.8	0.2
Log $\mathcal{L}(\theta \vert a_{0:T})$=	-0.22	-0.18	-2.0	-2.3	-2.5	-4.1	-4.3	-5.9

Model (θ_1=0.3): p_1=0.3 p_2=0.7

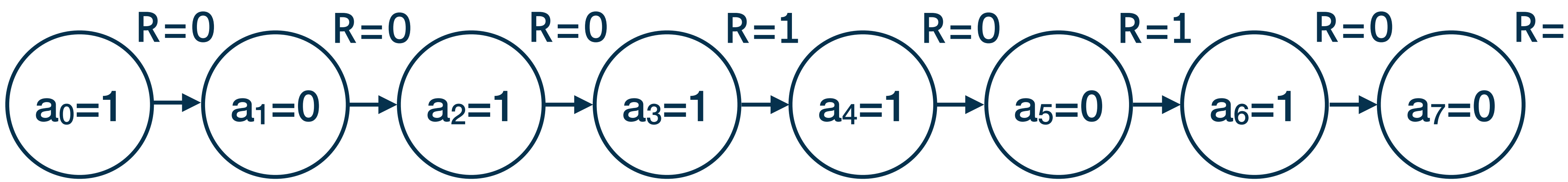
$\mathcal{L}(\theta \vert a_i)$=	0.7	0.3	0.7	0.7	0.7	0.3	0.7	0.3
Log $\mathcal{L}(\theta \vert a_{0:T})$=	-0.35	-1.6	-1.9	-2.3	-2.6	-3.8	-4.2	-5.4

Model (θ_1=0.1): p_1=0.1 p_2=0.9

$\mathcal{L}(\theta \vert a_i)$=	0.9	0.1	0.9	0.9	0.9	0.1	0.9	0.1
Log $\mathcal{L}(\theta \vert a_{0:T})$=	-0.1	-2.4	-2.5	-2.6	-2.7	-5.0	-5.1	-7.4

Log-likelihood maximization

Observations:



Model (θ_1=0.2): p_1=0.2 p_2=0.8

$\mathcal{L}(\theta\|a_i)=$	0.8	0.2	0.8	0.8	0.8	0.2	0.8	0.2
Log $\mathcal{L}(\theta\|a_{0:T})=$	-0.22	-0.18	-2.0	-2.3	-2.5	-4.1	-4.3	-5.9

Model (θ_1=0.3): p_1=0.3 p_2=0.7

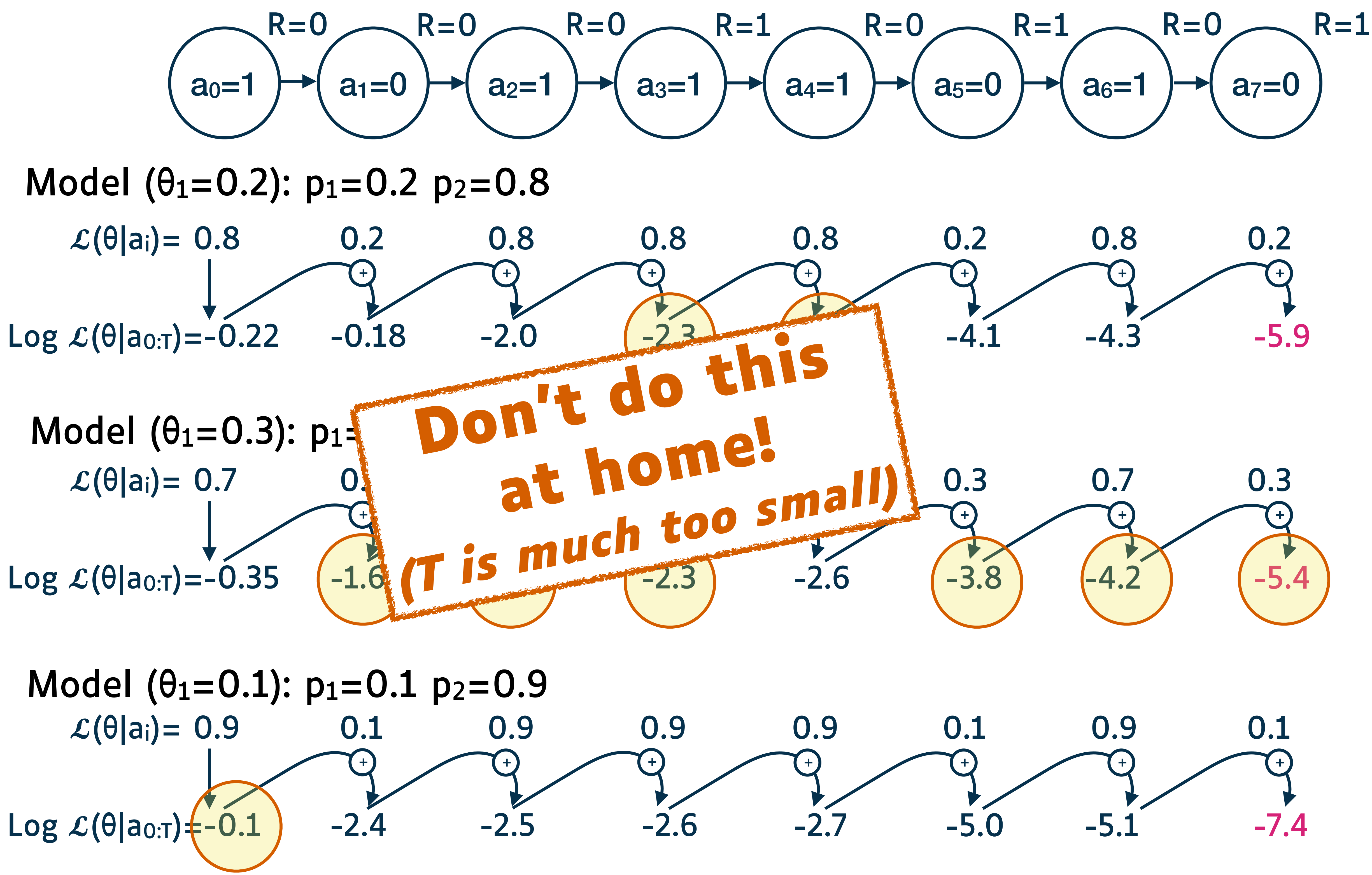
$\mathcal{L}(\theta\|a_i)=$	0.7	0.3	0.7	0.7	0.7	0.3	0.7	0.3
Log $\mathcal{L}(\theta\|a_{0:T})=$	-0.35	-1.6	-1.9	-2.3	-2.6	-3.8	-4.2	-5.4

Model (θ_1=0.1): p_1=0.1 p_2=0.9

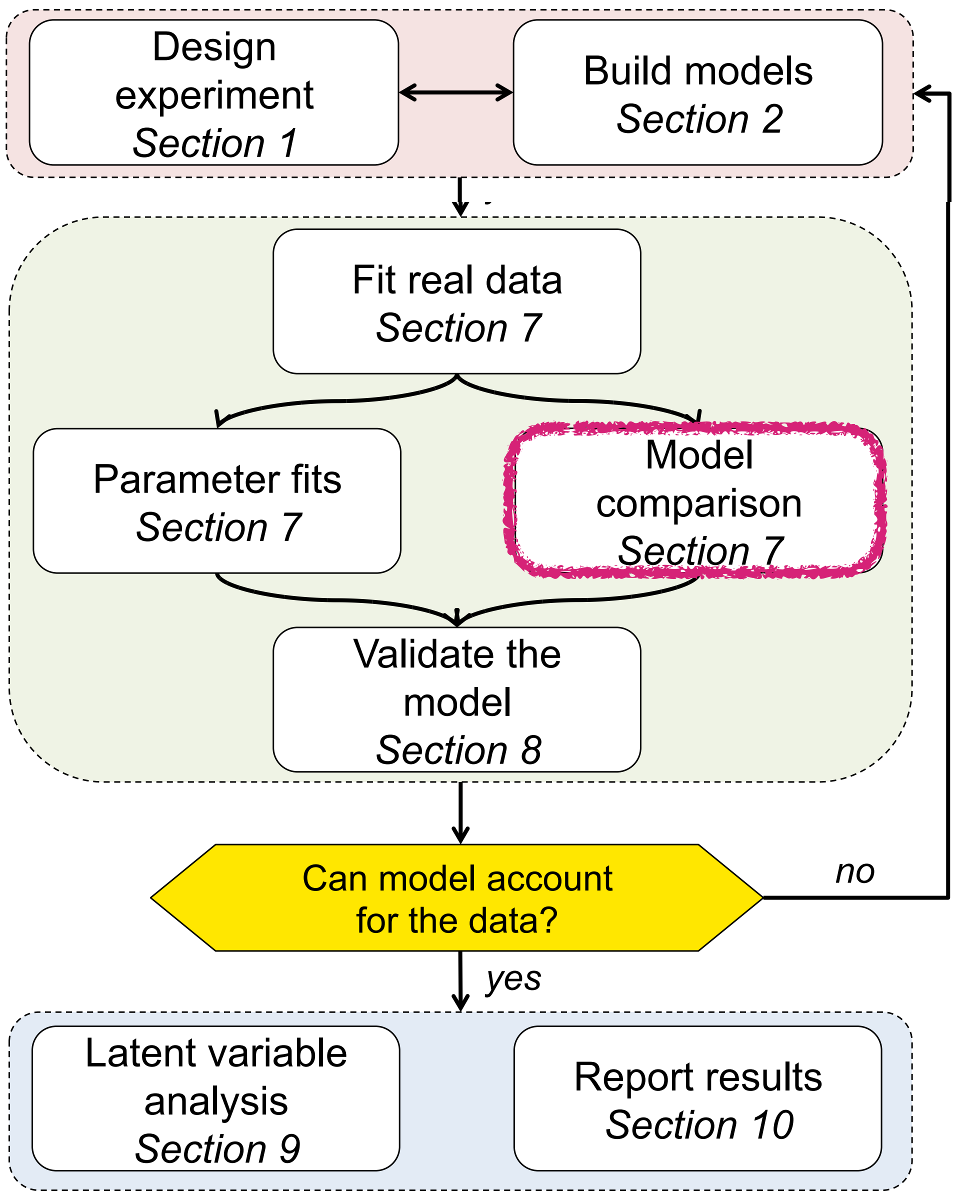
$\mathcal{L}(\theta\|a_i)=$	0.9	0.1	0.9	0.9	0.9	0.1	0.9	0.1
Log $\mathcal{L}(\theta\|a_{0:T})=$	-0.1	-2.4	-2.5	-2.6	-2.7	-5.0	-5.1	-7.4

Log-likelihood maximization

Observations:



Modeling Process



Adapted from (Wilson & Collins, 2019)

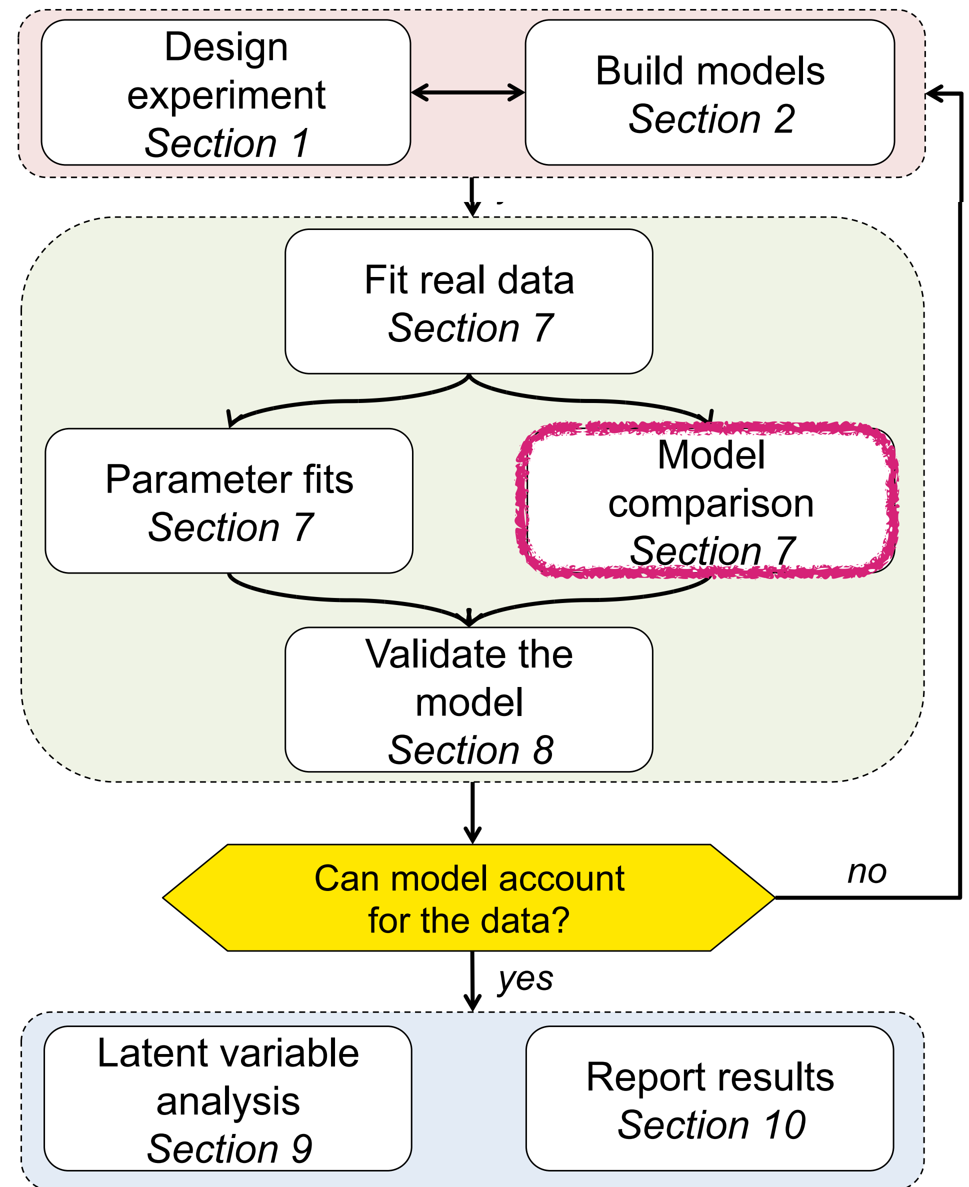
Model comparison & selection

Log-likelihood maximization

- If the models g_i have the same number of parameters:

  After optimizing θ for each model *i*, directly compare the log-likelihoods:

$$\mathrm{argmax}_i \mathcal{L}(\hat{\theta}|x, g_i)$$



Adapted from (Wilson & Collins, 2019)

One more classical decision-making model

« Building the set of candidate models is partially a subjective art »
Burnham & Anderson, 2002

- *2-arm bandit task.*
- Biased random response $p_t^1 = b \quad \text{and} \quad p_t^2 = 1 - b$ $\theta_1 = b$
- Noisy win-stay lose-shift

$$p_t^k = \begin{cases} 1 - \epsilon/2 & \text{if } (c_{t-1} = k \text{ and } r_{t-1} = 1) \text{ OR } (c_{t-1} \neq k \text{ and } r_{t-1} = 0) \\ \epsilon/2 & \text{if } (c_{t-1} \neq k \text{ and } r_{t-1} = 1) \text{ OR } (c_{t-1} = k \text{ and } r_{t-1} = 0) \end{cases}$$

$\theta_2 = \epsilon$

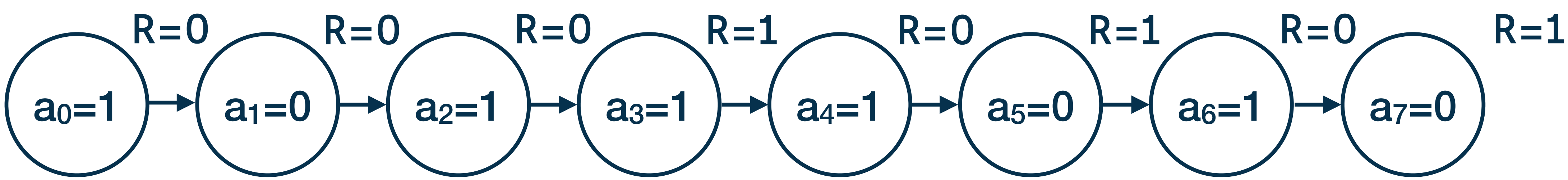
- Rescorla-Wagner $Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k)$ $p_t^k = \frac{\exp(\beta Q_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i)}$ $\theta_3 = (\alpha, \beta)$
- Choice Kernel (behavioral persistence) $CK_{t+1}^k = CK_t^k + \alpha_c(a_t^k - CK_t^k)$ $p_t^k = \frac{\exp(\beta_c CK_t^k)}{\sum_{i=1}^K \exp(\beta_c CK_t^i)}$ $\theta_4 = (\alpha_c, \beta_c)$
- RW+CK

$$Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k)$$
$$CK_{t+1}^k = CK_t^k + \alpha_c(a_t^k - CK_t^k)$$
$$p_t^k = \frac{\exp(\beta Q_t^k + \beta_c CK_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i + \beta_c CK_t^i)}$$
$$\theta_5 = (\alpha, \beta, \alpha_c, \beta_c)$$

From (Wilson & Collins, 2019)

Log-likelihood maximization

Observations:



- Biased random response $p_t^1 = b \quad \text{and} \quad p_t^2 = 1 - b$ $\theta_1 = b$

  Model (θ_1=0.4):

  Log $\mathcal{L}(\theta|a_{0:T},g1)$= -5.3

- Win-stay lose-shift

$$p_t^k = \begin{cases} 1-\epsilon/2 & \text{if } (c_{t-1} = k \text{ and } r_{t-1} = 1) \text{ OR } (c_{t-1} \neq k \text{ and } r_{t-1} = 0) \\ \epsilon/2 & \text{if } (c_{t-1} \neq k \text{ and } r_{t-1} = 1) \text{ OR } (c_{t-1} = k \text{ and } r_{t-1} = 0) \end{cases}$$

$$\theta_2 = \epsilon$$

Log-likelihood maximization

Observations:

R=0, R=0, R=0, R=1, R=0, R=1, R=0, R=1

$a_0=1 \rightarrow a_1=0 \rightarrow a_2=1 \rightarrow a_3=1 \rightarrow a_4=1 \rightarrow a_5=0 \rightarrow a_6=1 \rightarrow a_7=0$

- **Biased random response**

$p_t^1 = b \quad \text{and} \quad p_t^2 = 1 - b$ $\qquad \theta_1 = b$

Model (θ_1=0.4):

Log $\mathcal{L}(\theta|a_{0:T},g1)$= -5.3

- **Win-stay lose-shift**

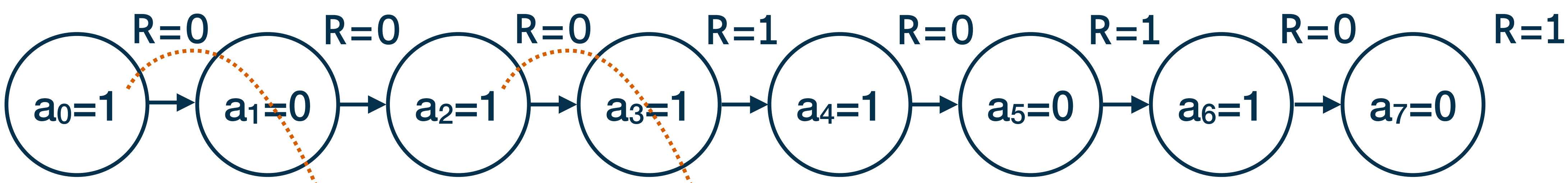
$$p_t^k = \begin{cases} 1-\epsilon/2 & \text{if } (c_{t-1}=k \text{ and } r_{t-1}=1) \text{ OR } (c_{t-1}\neq k \text{ and } r_{t-1}=0) \\ \epsilon/2 & \text{if } (c_{t-1}\neq k \text{ and } r_{t-1}=1) \text{ OR } (c_{t-1}=k \text{ and } r_{t-1}=0) \end{cases}$$

Model (θ_2=0.1): $\qquad \theta_2 = \epsilon$

$\mathcal{L}(\theta\|a_i)$=	0.5	0.95	0.95	0.05	0.95	0.95	0.05	0.95
Log $\mathcal{L}(\theta\|a_{0:T})$=	-0.7	-0.74	-0.8	-3.8	-3.85	-3.9	-6.9	-6.94

Log-likelihood maximization

Observations:



- **Biased random response** $p_t^1 = b \text{ and } p_t^2 = 1-b$ $\theta_1 = b$

  Model (θ_1=0.4):

  Log $\mathcal{L}(\theta|a_{0:T}, g1)$= -5.3

- **Win-stay lose-shift**

$$p_t^k = \begin{cases} 1-\epsilon/2 & \text{if } (c_{t-1}=k \text{ and } r_{t-1}=1) \text{ OR } (c_{t-1}\neq k \text{ and } r_{t-1}=0) \\ \epsilon/2 & \text{if } (c_{t-1}\neq k \text{ and } r_{t-1}=1) \text{ OR } (c_{t-1}=k \text{ and } r_{t-1}=0) \end{cases}$$

  Model (θ_2=0.1): $\theta_2 = \epsilon$

$\mathcal{L}(\theta\|a_i)$=	0.5	0.95	0.95	0.05	0.95	0.95	0.05	0.95
Log $\mathcal{L}(\theta\|a_{0:T})$=	-0.7	-0.74	-0.8	-3.8	-3.85	-3.9	-6.9	-6.94

Log-likelihood maximization

Observations:

$a_0=1 \xrightarrow{R=0} a_1=0 \xrightarrow{R=0} a_2=1 \xrightarrow{R=0} a_3=1 \xrightarrow{R=1} a_4=1 \xrightarrow{R=0} a_5=0 \xrightarrow{R=1} a_6=1 \xrightarrow{R=0} a_7=0$ R=1

- **Biased random response**

  $p_t^1 = b \quad \text{and} \quad p_t^2 = 1-b$ $\qquad \theta_1 = b$

  Model (θ_1=0.4):

  Log $\mathcal{L}(\theta|a_{0:T}, g1)$= -5.3

- **Win-stay lose-shift**

$$p_t^k = \begin{cases} 1-\epsilon/2 & \text{if } (c_{t-1}=k \text{ and } r_{t-1}=1) \text{ OR } (c_{t-1}\neq k \text{ and } r_{t-1}=0) \\ \epsilon/2 & \text{if } (c_{t-1}\neq k \text{ and } r_{t-1}=1) \text{ OR } (c_{t-1}= k \text{ and } r_{t-1}=0) \end{cases}$$

$\theta_2 = \epsilon$

  Model (θ_2=0.1):

$\mathcal{L}(\theta\|a_i)$=	0.5	0.95	0.95	0.05	0.95	0.95	0.05	0.95
Log $\mathcal{L}(\theta\|a_{0:T})$=	-0.7	-0.74	-0.8	-3.8	-3.85	-3.9	-6.9	-6.94

  Model (θ_2=0.2):

$\mathcal{L}(\theta\|a_i)$=	0.5	0.9	0.9	0.1	0.9	0.9	0.1	0.9
Log $\mathcal{L}(\theta\|a_{0:T})$=	-0.7	-0.8	-0.9	-3.2	-3.3	-3.4	-5.7	-5.8

Log-likelihood maximization

We have a winner!

- Biased random response :
  $\theta_1=0.4$
  $\log \mathcal{L}(\theta_1|a_{0:T},g1)=-5.3$

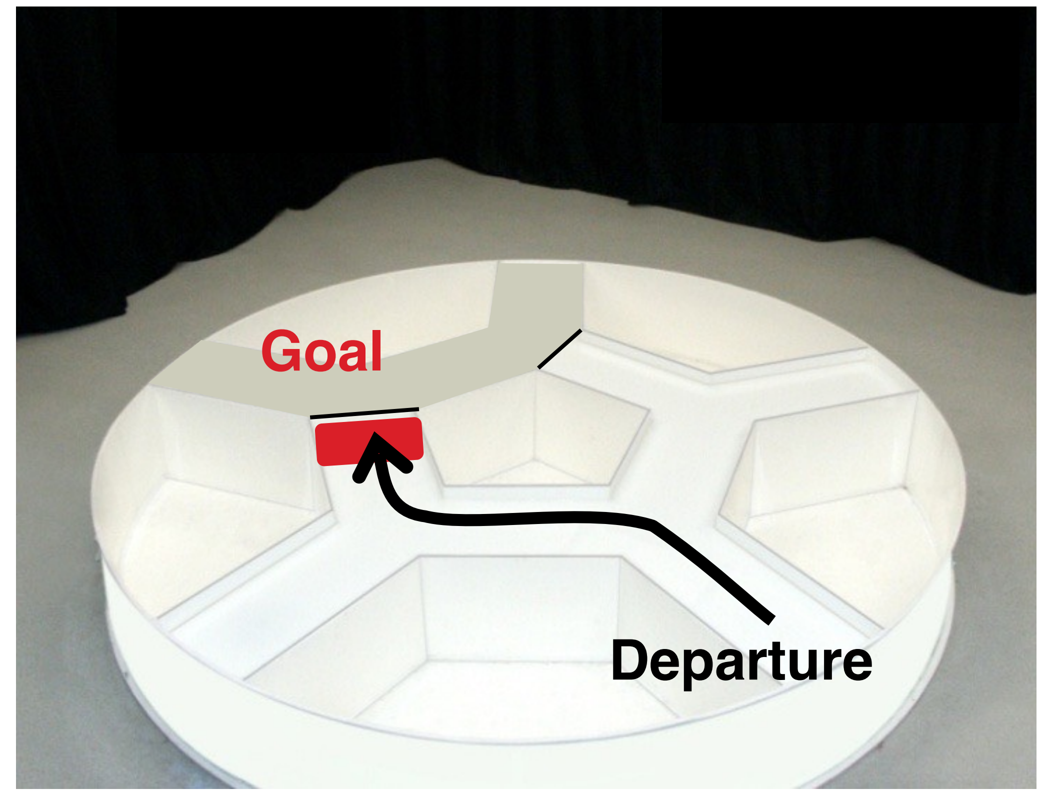
- Win-stay lose-shift:
  $\theta_2=0.6$
  $\log \mathcal{L}(\theta_2|a_{0:T},g2)=-4.88$

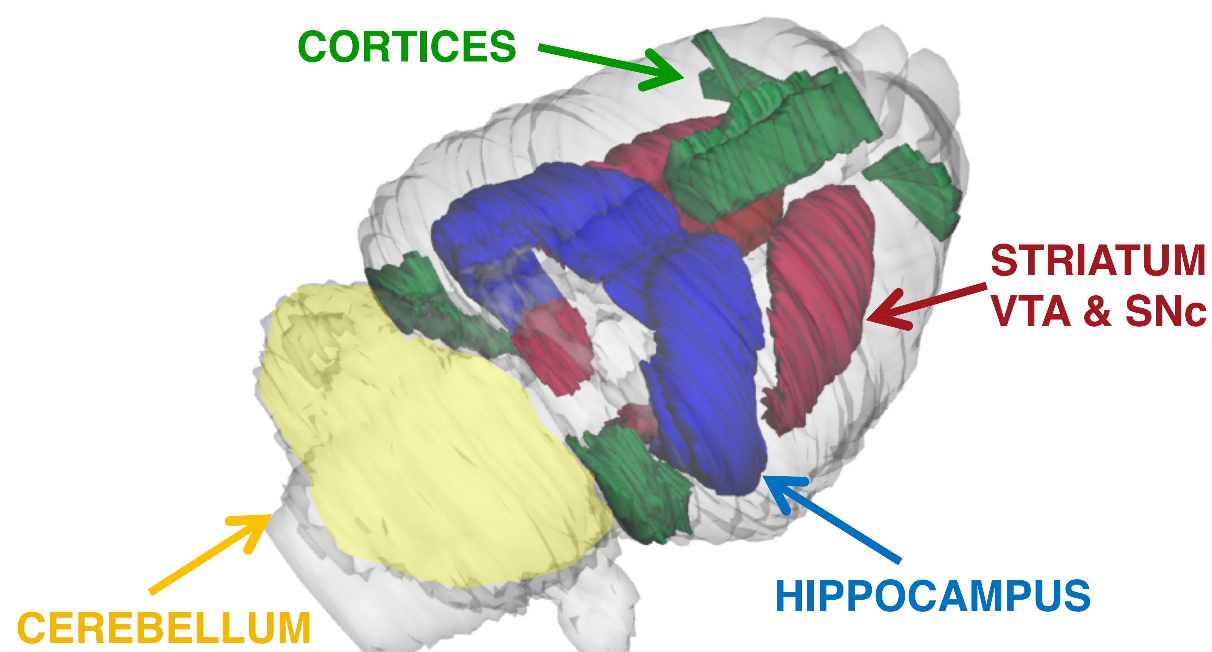
Example of multiple model selection.

Sequential egocentric strategy

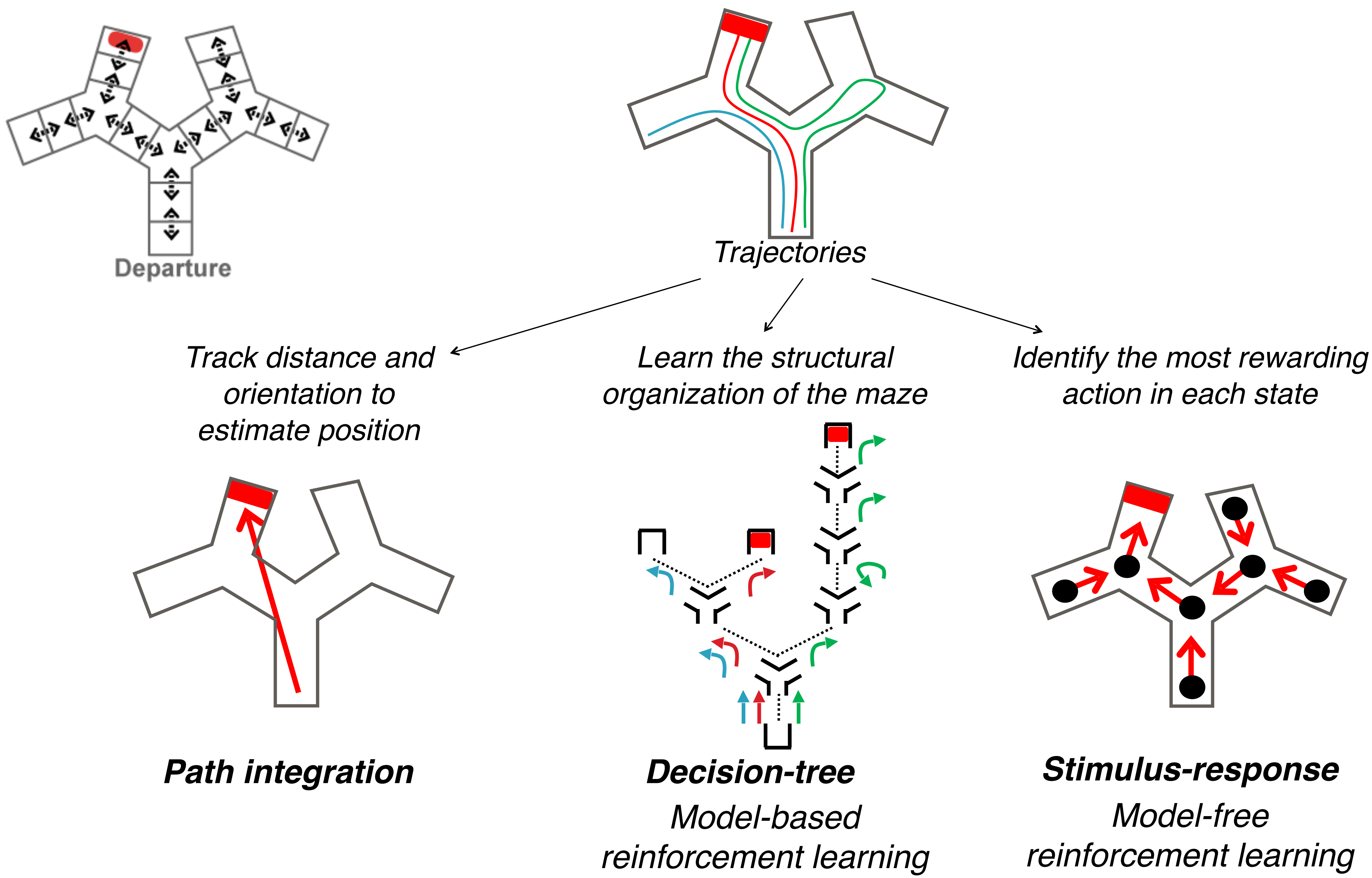
- Triple-Y maze
- Task specificities
  - no allocentric information
  - self-motion-based
- Methodological specificity
  - unbiased exploration of the whole brain

(Babayan et al., 2019)

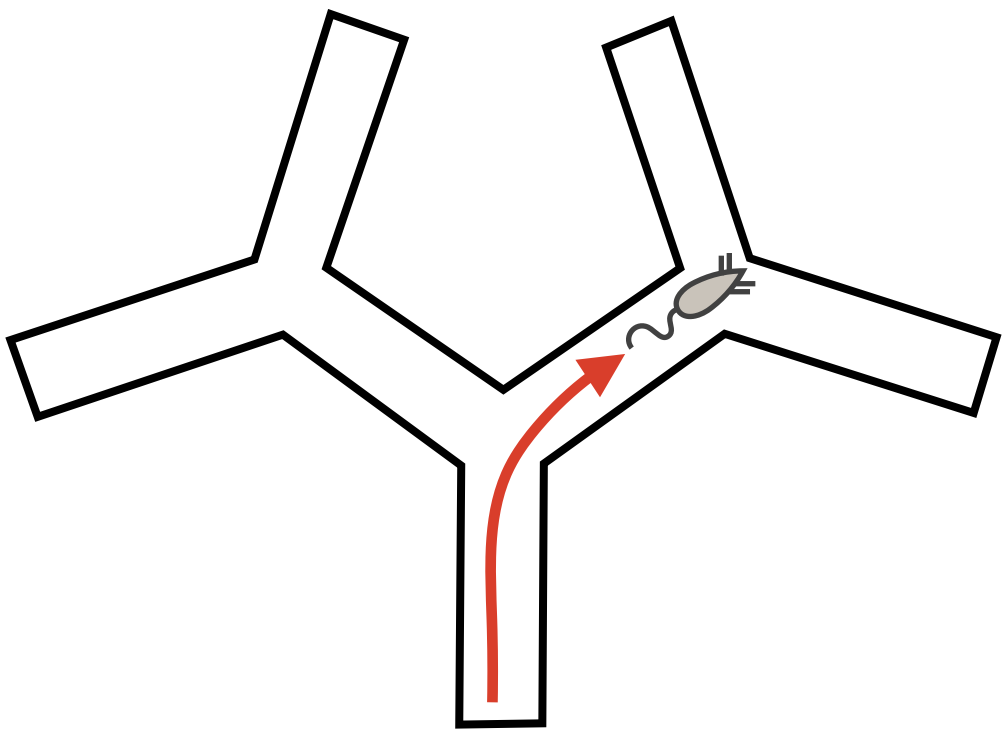




Which models for this task?

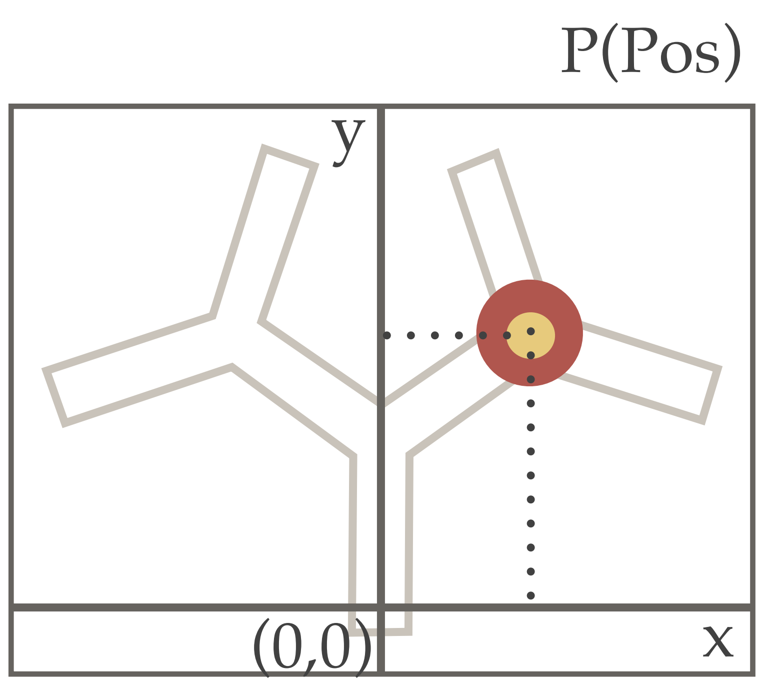


Path Integration

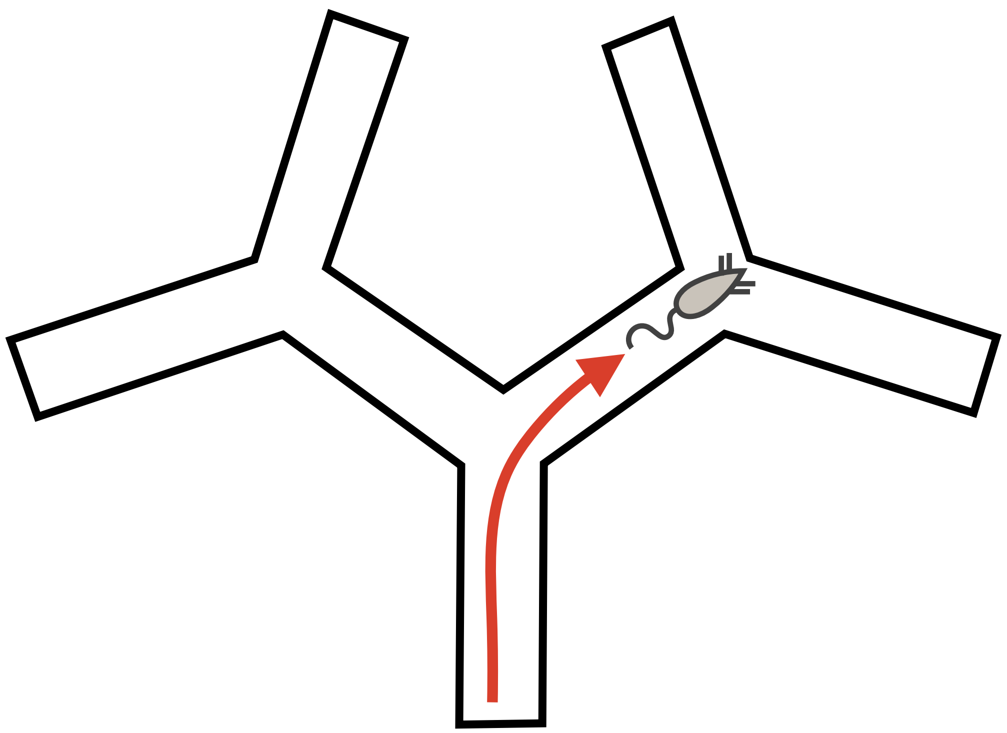


$$P^t(Pos) = \mathcal{N}(\mu = (x(t), y(t)), \sigma = \sigma_0 t)$$

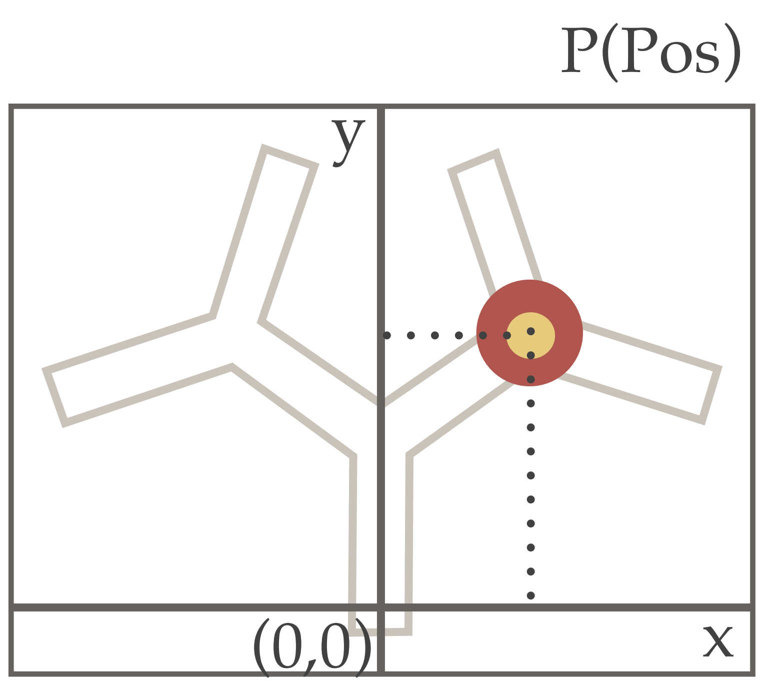
- 4 actions: Fwd, L, R, U-turn
- **Position estimation update**



Path Integration

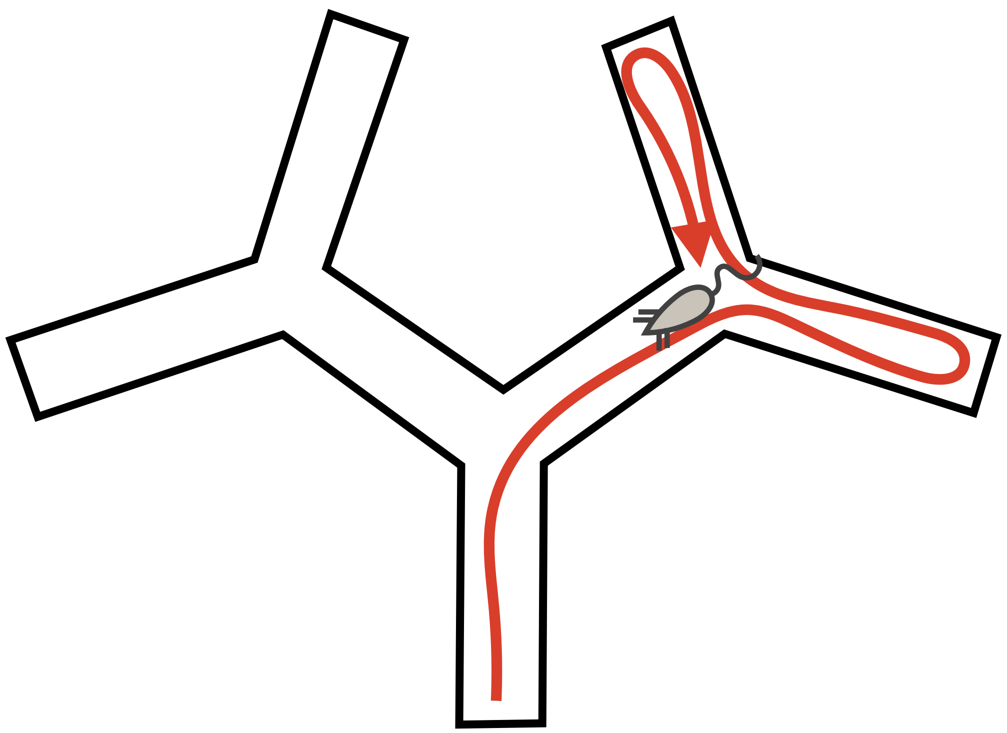


$$P^t(Pos) = \mathcal{N}(\mu = (x(t), y(t)), \sigma = \sigma_0 t)$$



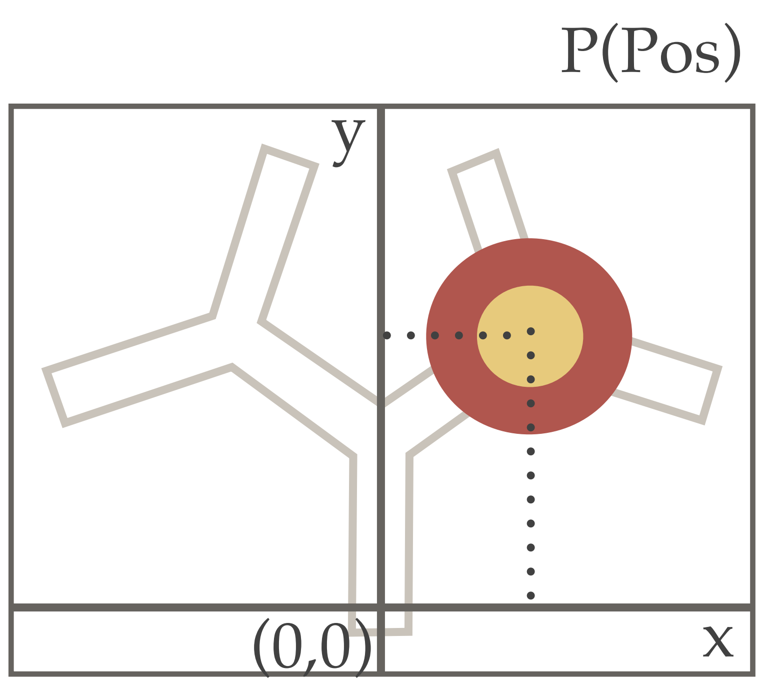
- 4 actions: Fwd, L, R, U-turn
- **Position estimation update**

Path Integration

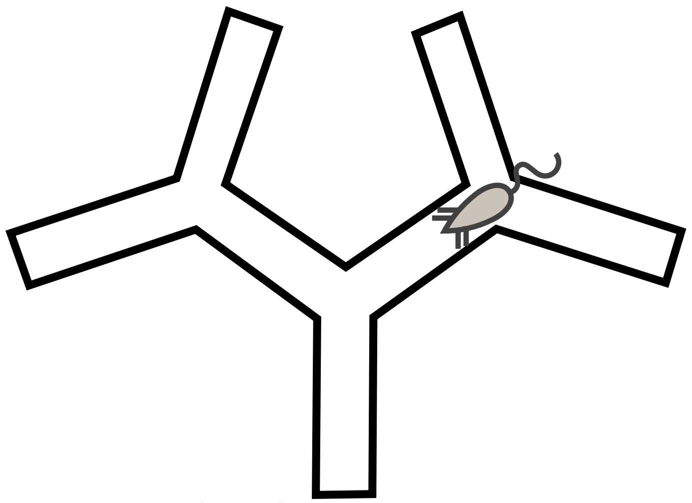


$$P^t(Pos) = \mathcal{N}(\mu = (x(t), y(t)), \sigma = \sigma_0 t)$$

- 4 actions: Fwd, L, R, U-turn
- **Position estimation update**

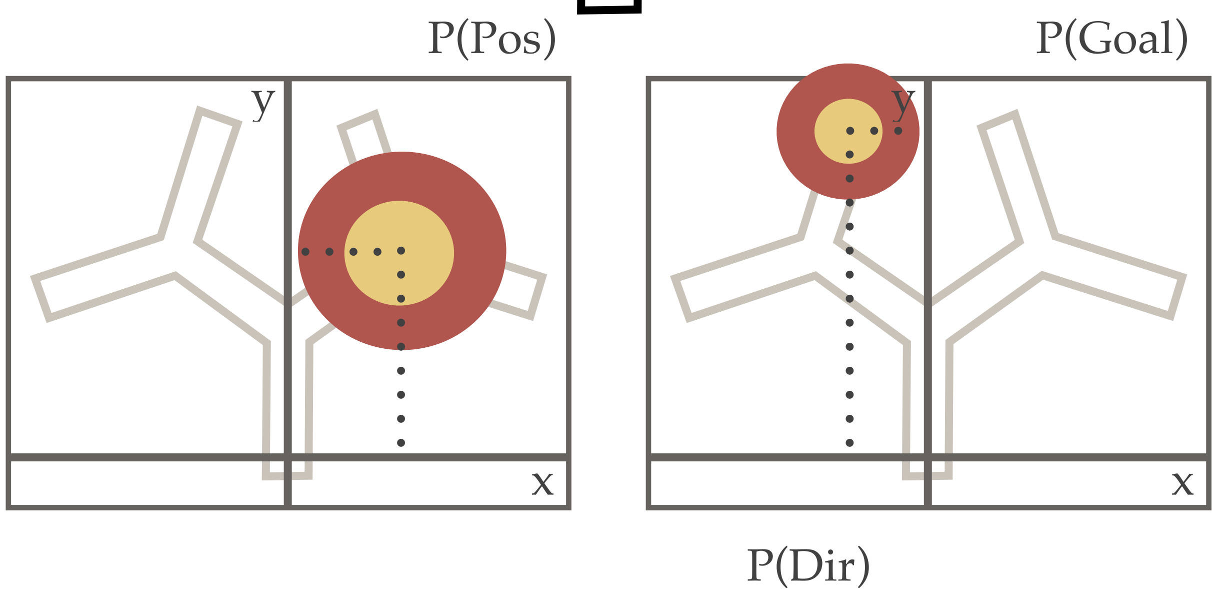


Path Integration



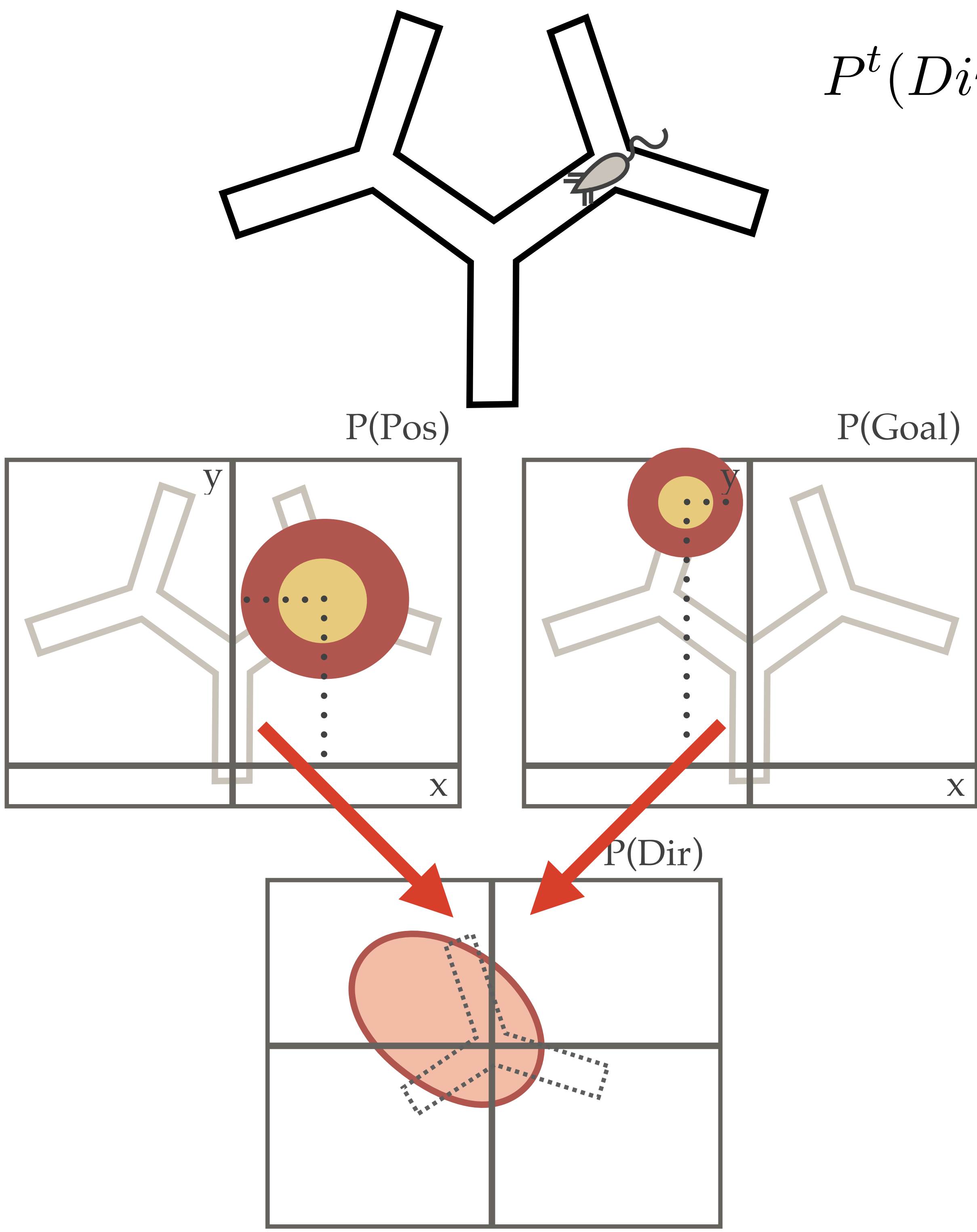
$$P^t(Dir = \alpha) = \frac{\sum_{(Pos,Goal)\in\mathcal{D}^\alpha} P^t(Pos) \times P(Goal)}{\sum_{(Pos,Goal)} P^t(Pos) \times P(Goal)}$$

$$\mathcal{D}^\alpha = \{(Pos, Goal) | atan2(Goal - Pos) = \alpha\}$$



- 4 actions: Fwd, L, R, U-turn
- Position estimation update
- **Inference**

Path Integration

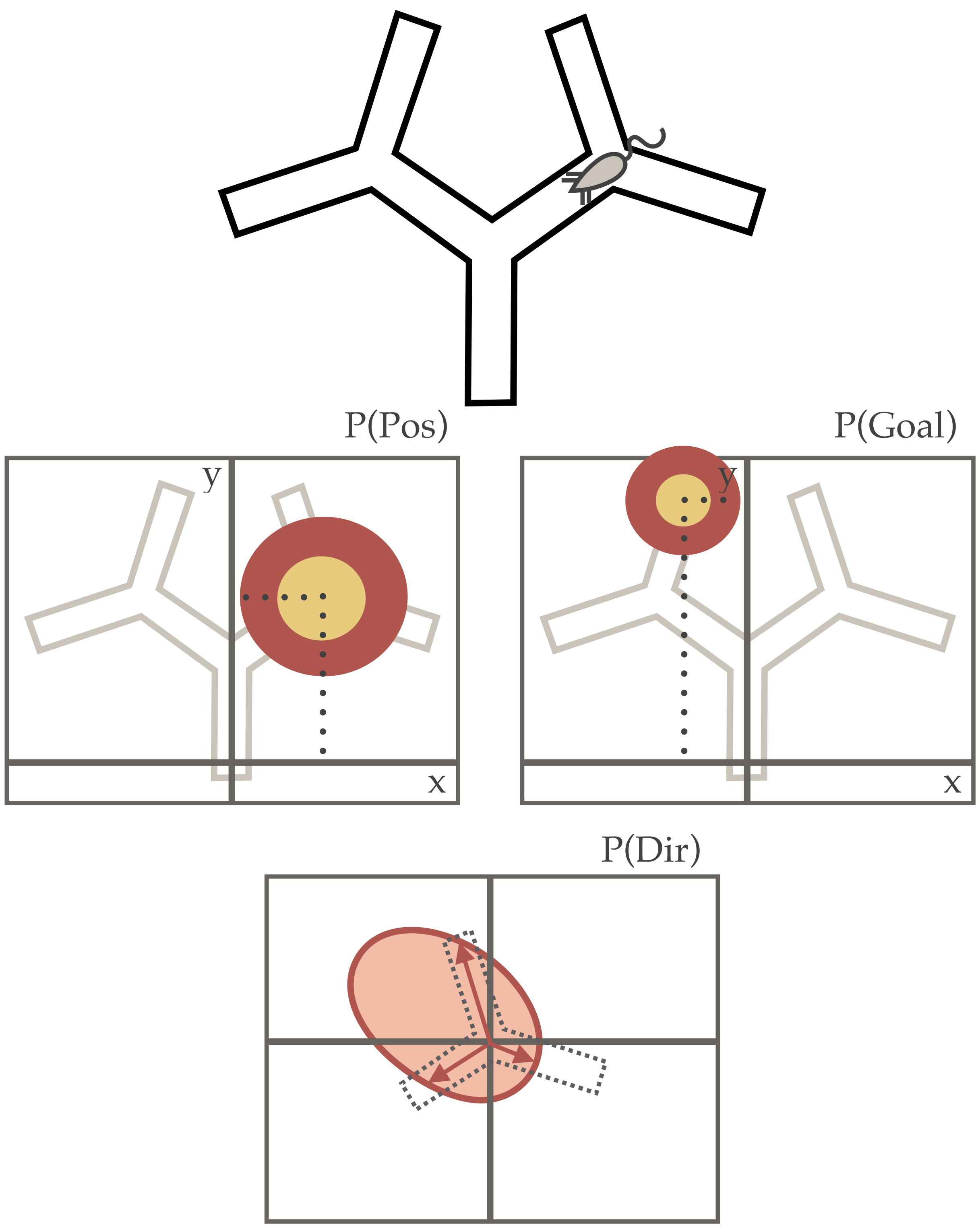


$$P^t(Dir = \alpha) = \frac{\sum_{(Pos,Goal)\in\mathcal{D}^\alpha} P^t(Pos) \times P(Goal)}{\sum_{(Pos,Goal)} P^t(Pos) \times P(Goal)}$$

$$\mathcal{D}^\alpha = \{(Pos, Goal)|atan2(Goal - Pos) = \alpha\}$$

- 4 actions: Fwd, L, R, U-turn
- Position estimation update
- **Inference**

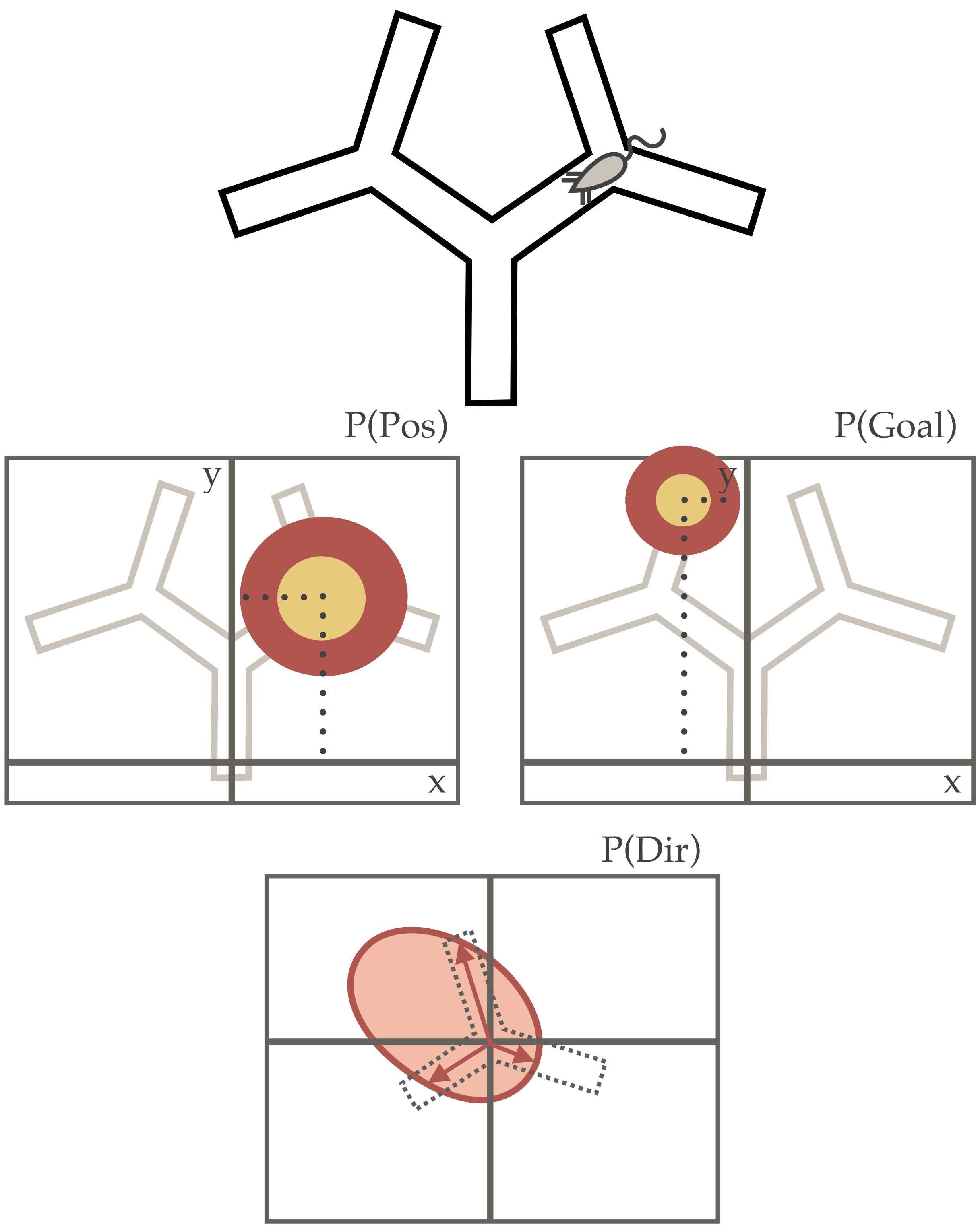
Path Integration



$$P^t(a_i) = \frac{e^{\beta P^t(Dir=a_i)}}{\sum_j e^{\beta P^t(Dir=a_j)}}$$

- 4 actions: Fwd, L, R, U-turn
- Position estimation update
- Inference
- **Action selection (softmax)**

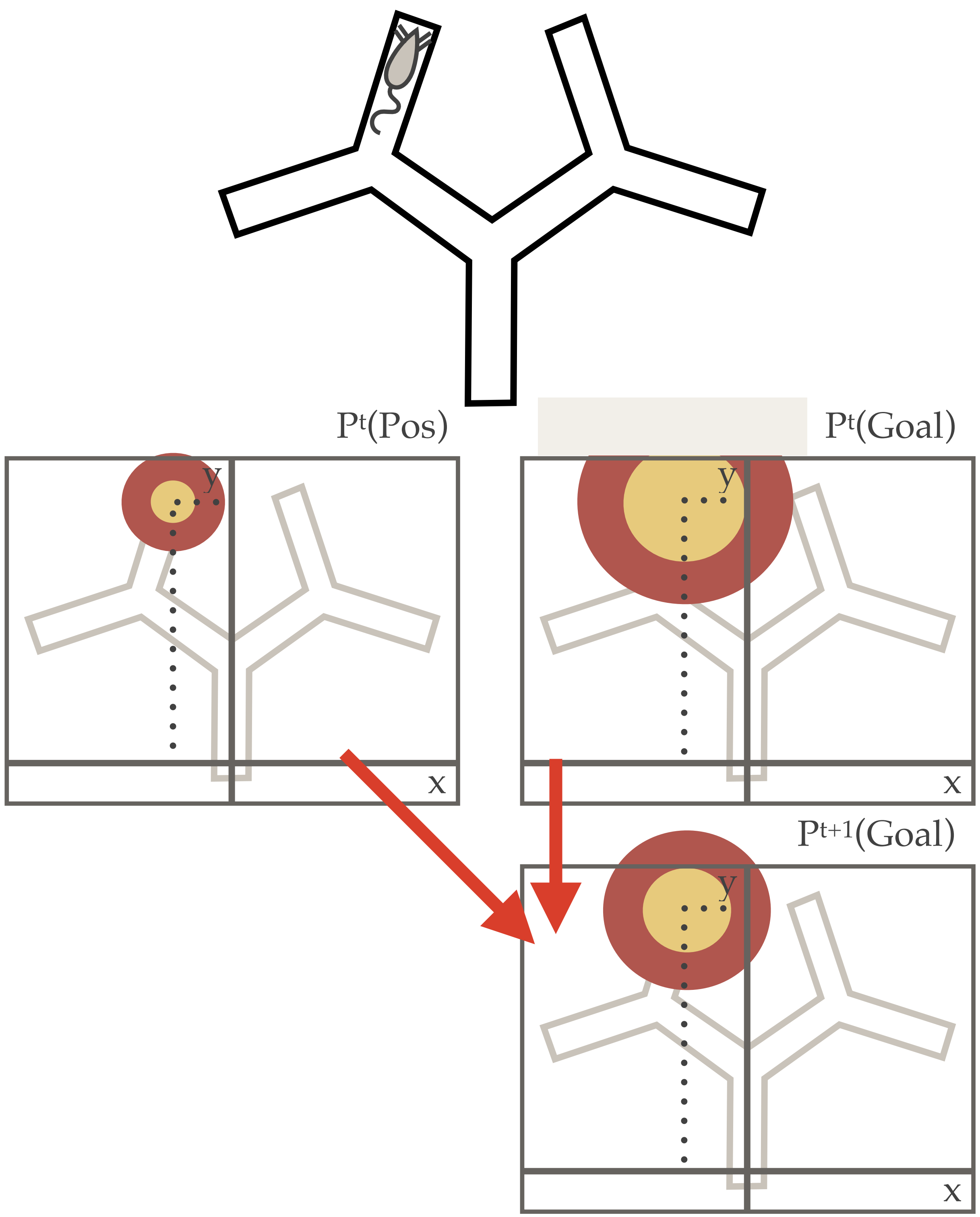
Path Integration



$$P^t(a_i) = \frac{e^{\beta P^t(Dir=a_i)}}{\sum_j e^{\beta P^t(Dir=a_j)}}$$

- 4 actions: Fwd, L, R, U-turn
- Position estimation update
- Inference
- **Action selection (softmax)**

Path Integration

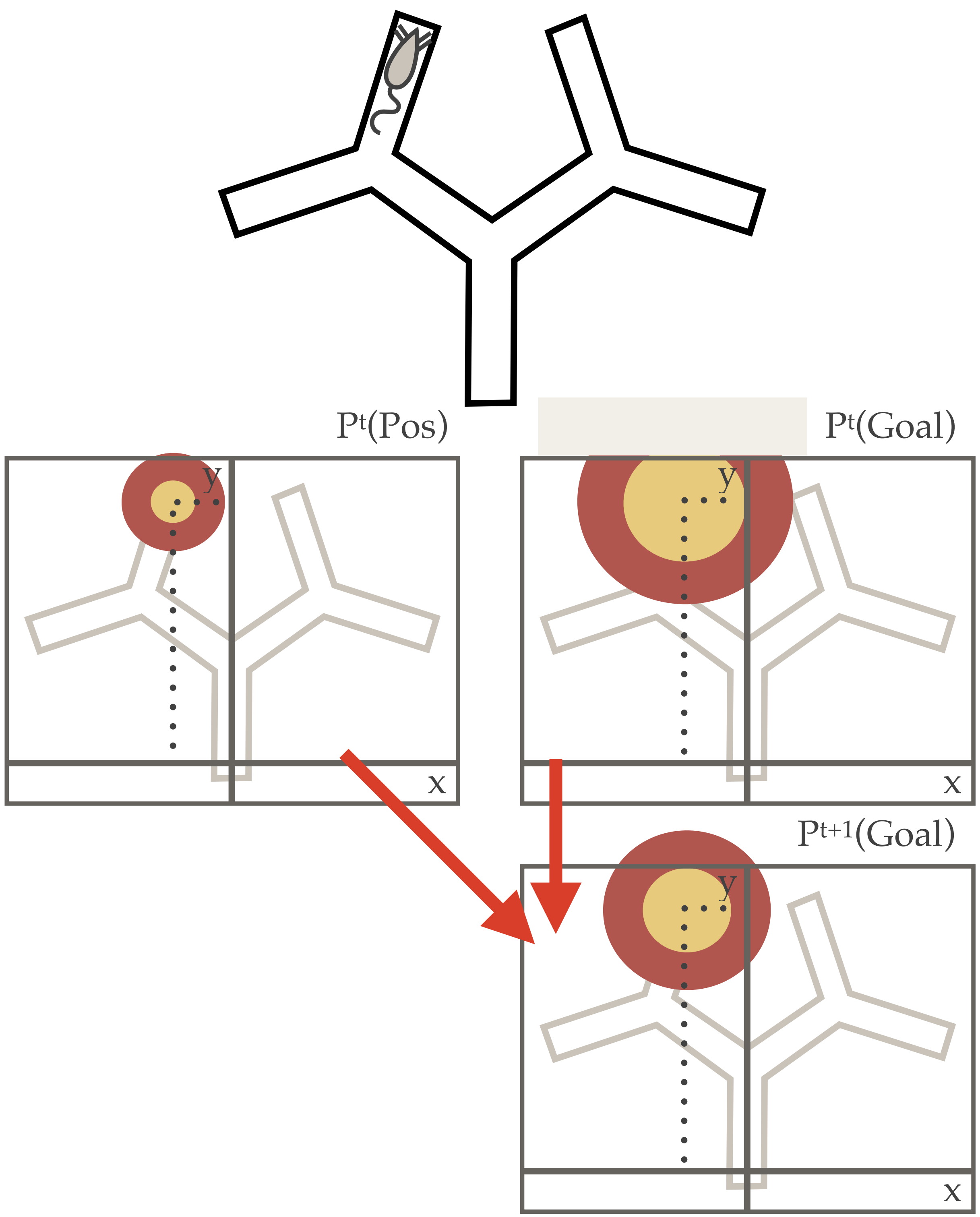


$$P(Goal) \leftarrow (1-\eta)P(Goal) + \eta P^t(Pos)$$

- 4 actions: Fwd, L, R, U-turn
- Position estimation update
- Inference
- Action selection (softmax)
- **Learning**

Path Integration

$$P(Goal) \leftarrow (1 - \eta) P(Goal) + \eta P^t(Pos)$$

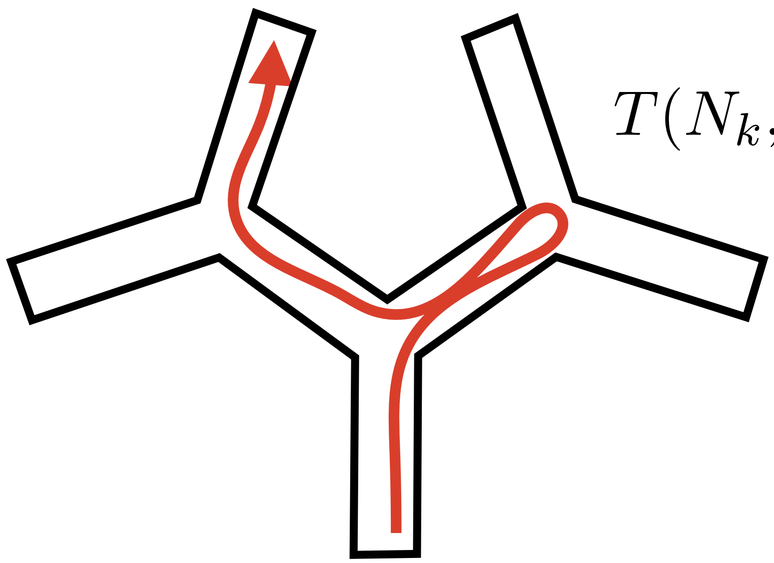


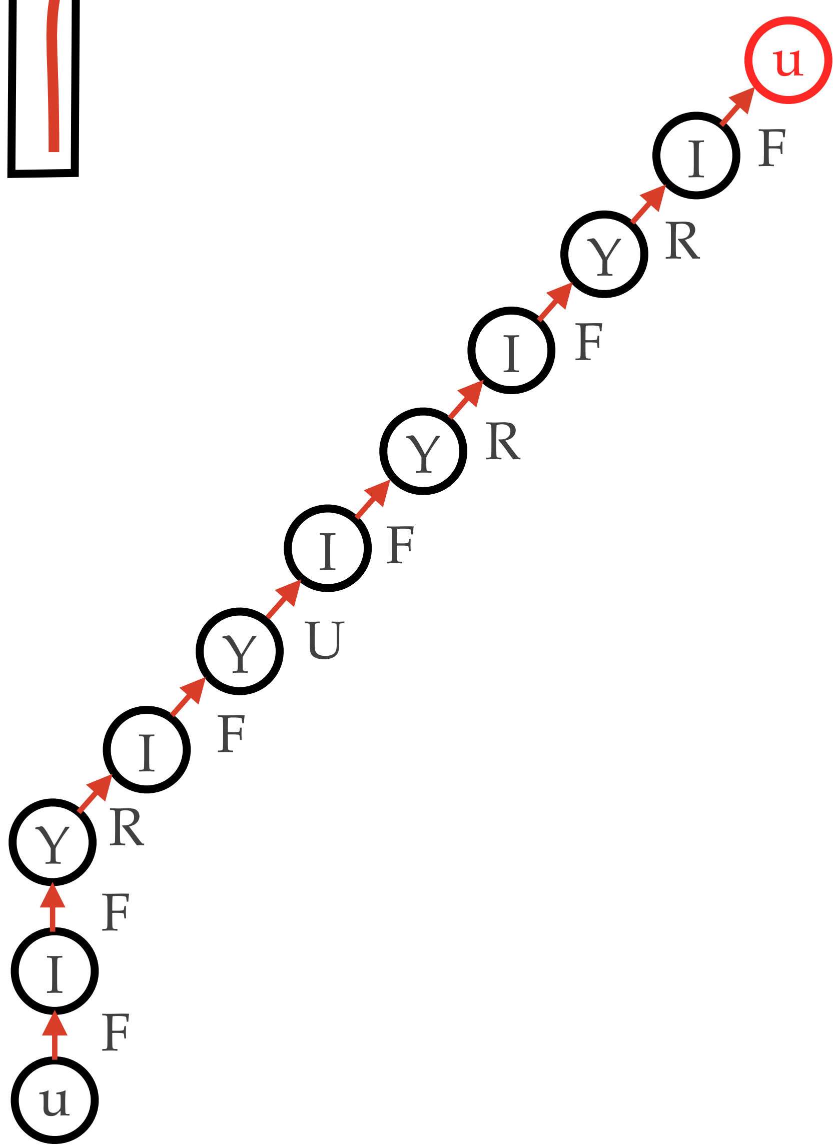
- 4 actions: Fwd, L, R, U-turn
- Position estimation update
- Inference
- Action selection (softmax)
- **Learning**

Model-Based Reinforcement Learning

$$T(N_k, a_i, N_{m+1}) = \eta$$

$$T(N_k, a_i, N_{m+1}) \leftarrow T(N_k, a_i, N_{m+1}) + \eta(1 - T(N_k, a_i, N_{m+1}))$$



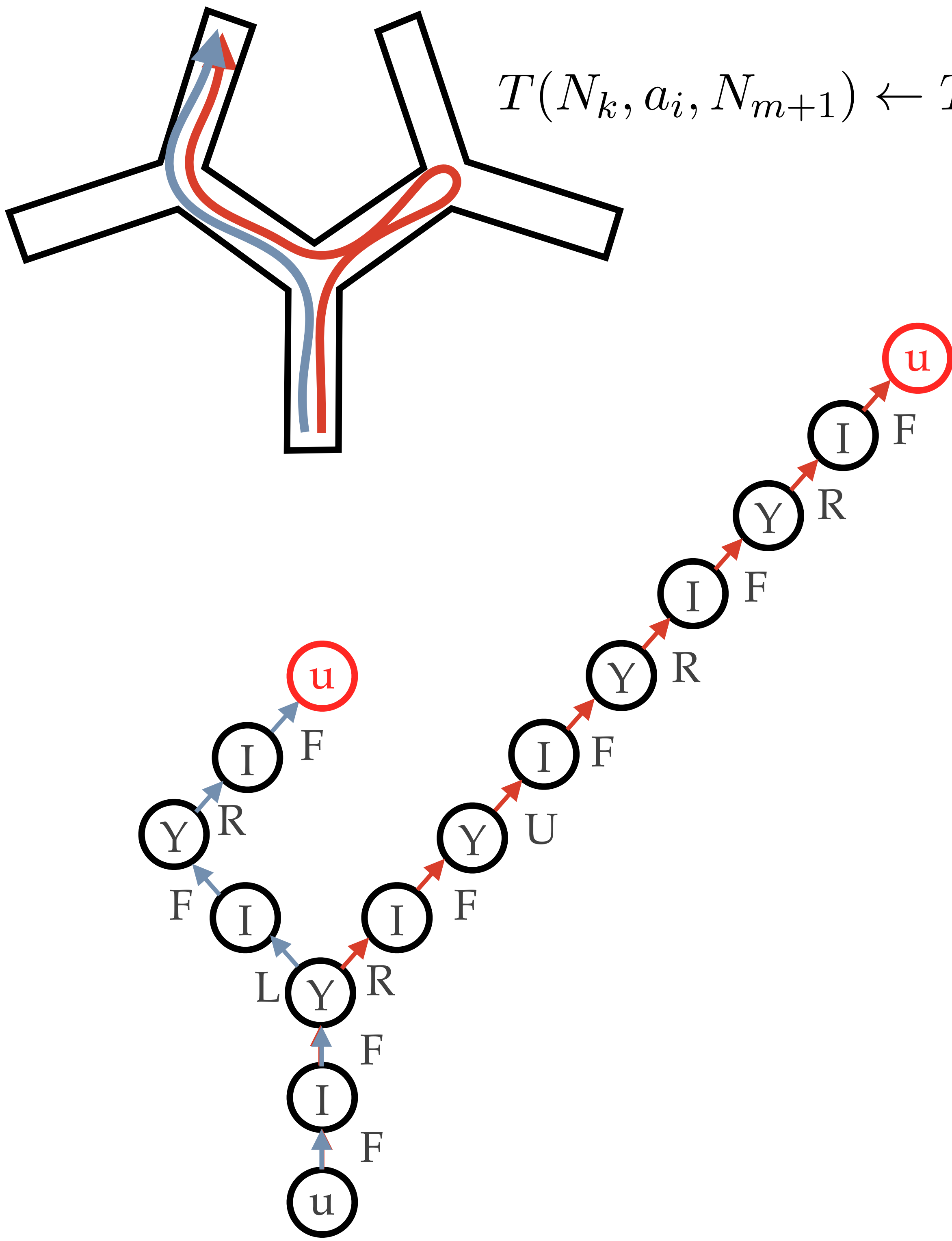


- 3 types of states: u ; I ; Y
- 4 actions: Fwd, L, R, U-turn
- **Transitions learning**

Model-Based Reinforcement Learning

$$T(N_k, a_i, N_{m+1}) = \eta$$

$$T(N_k, a_i, N_{m+1}) \leftarrow T(N_k, a_i, N_{m+1}) + \eta(1 - T(N_k, a_i, N_{m+1}))$$

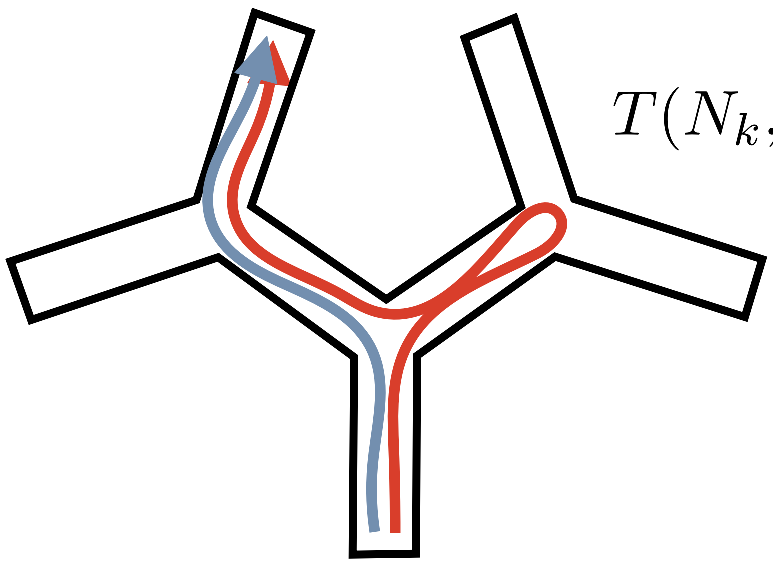


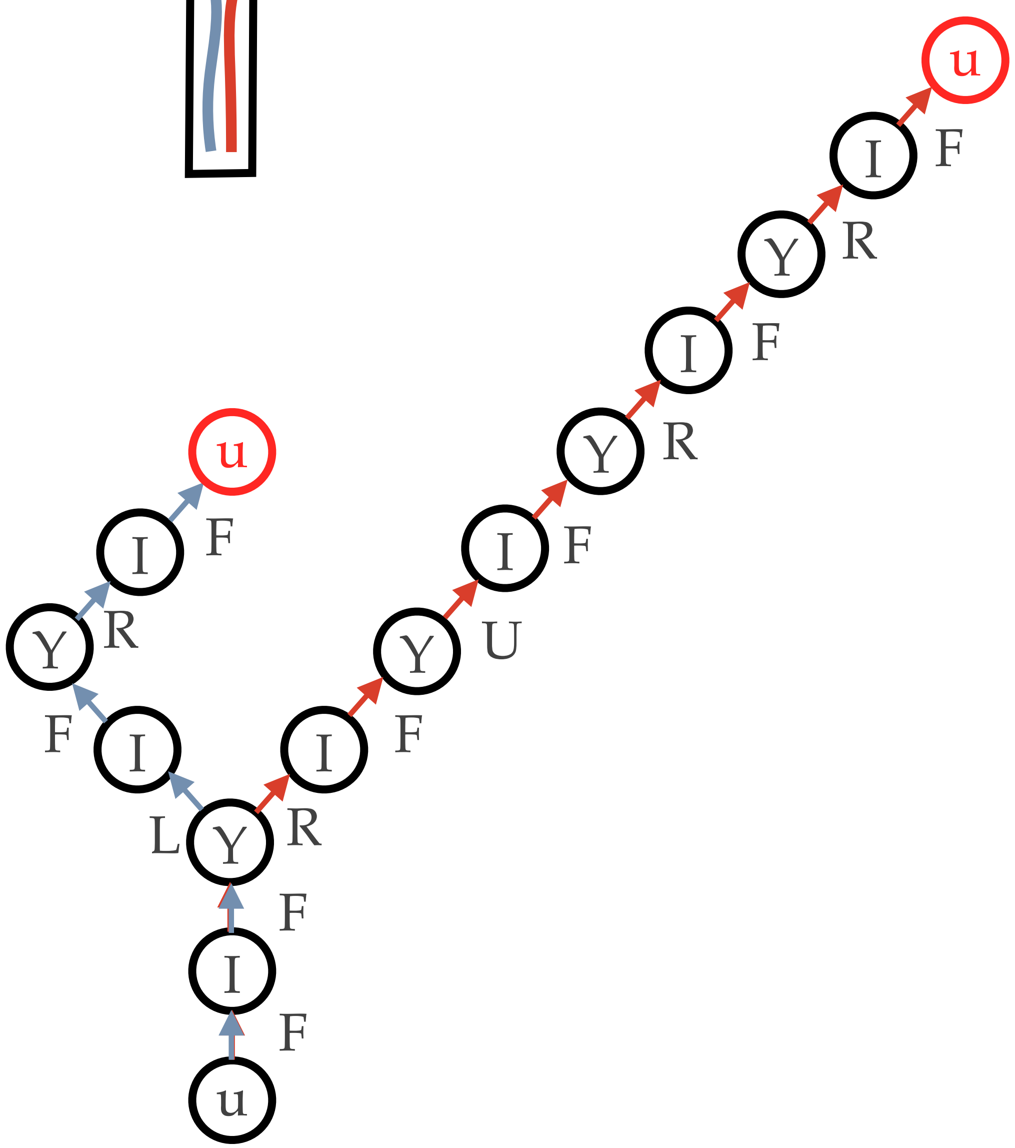
- 3 types of states: u ; I ; Y
- 4 actions: Fwd, L, R, U-turn
- **Transitions learning**

Model-Based Reinforcement Learning

$$T(N_k, a_i, N_{m+1}) = \eta$$

$$T(N_k, a_i, N_{m+1}) \leftarrow T(N_k, a_i, N_{m+1}) + \eta(1 - T(N_k, a_i, N_{m+1}))$$

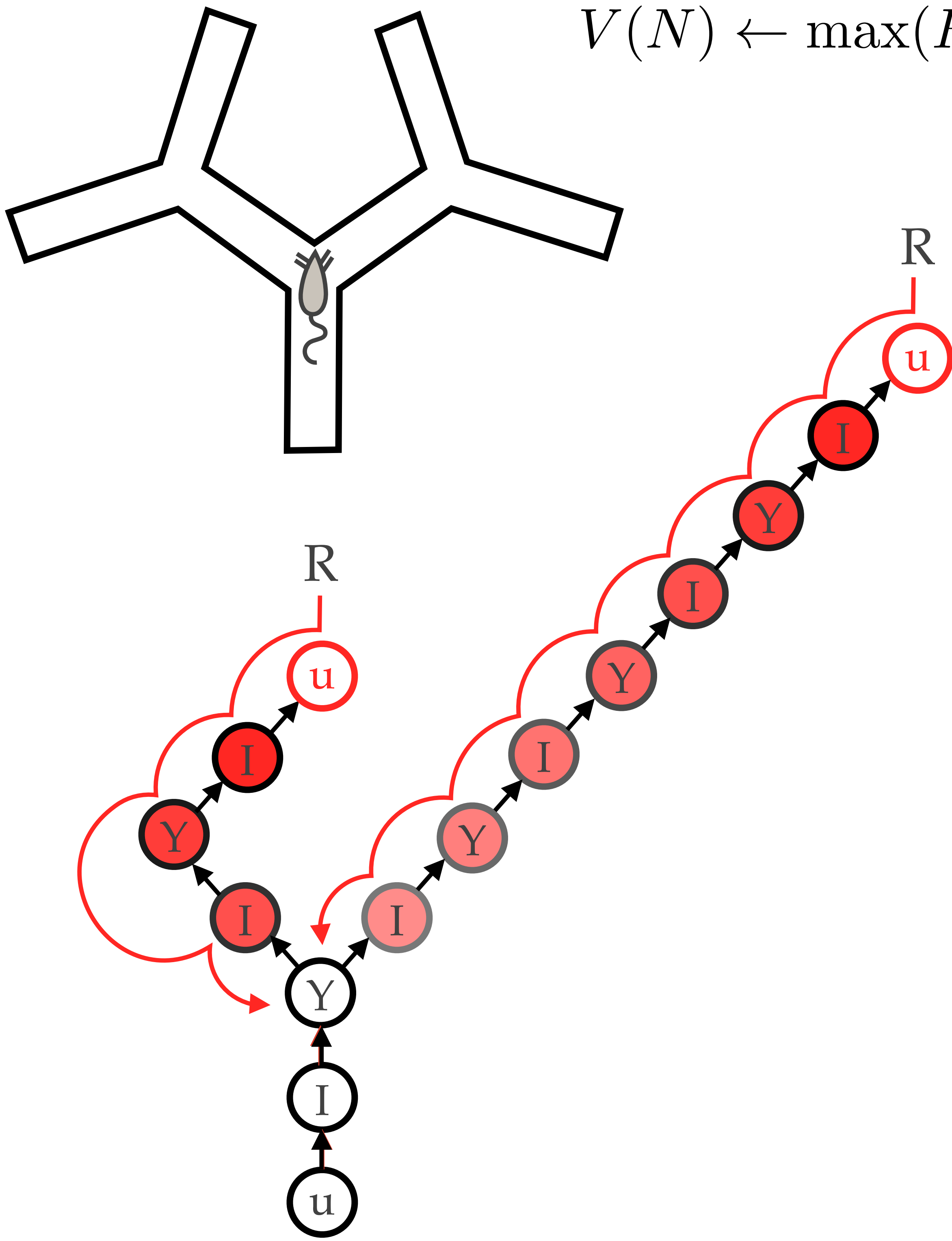




- 3 types of states: u ; I ; Y
- 4 actions: Fwd, L, R, U-turn
- **Transitions learning**

Model-Based Reinforcement Learning

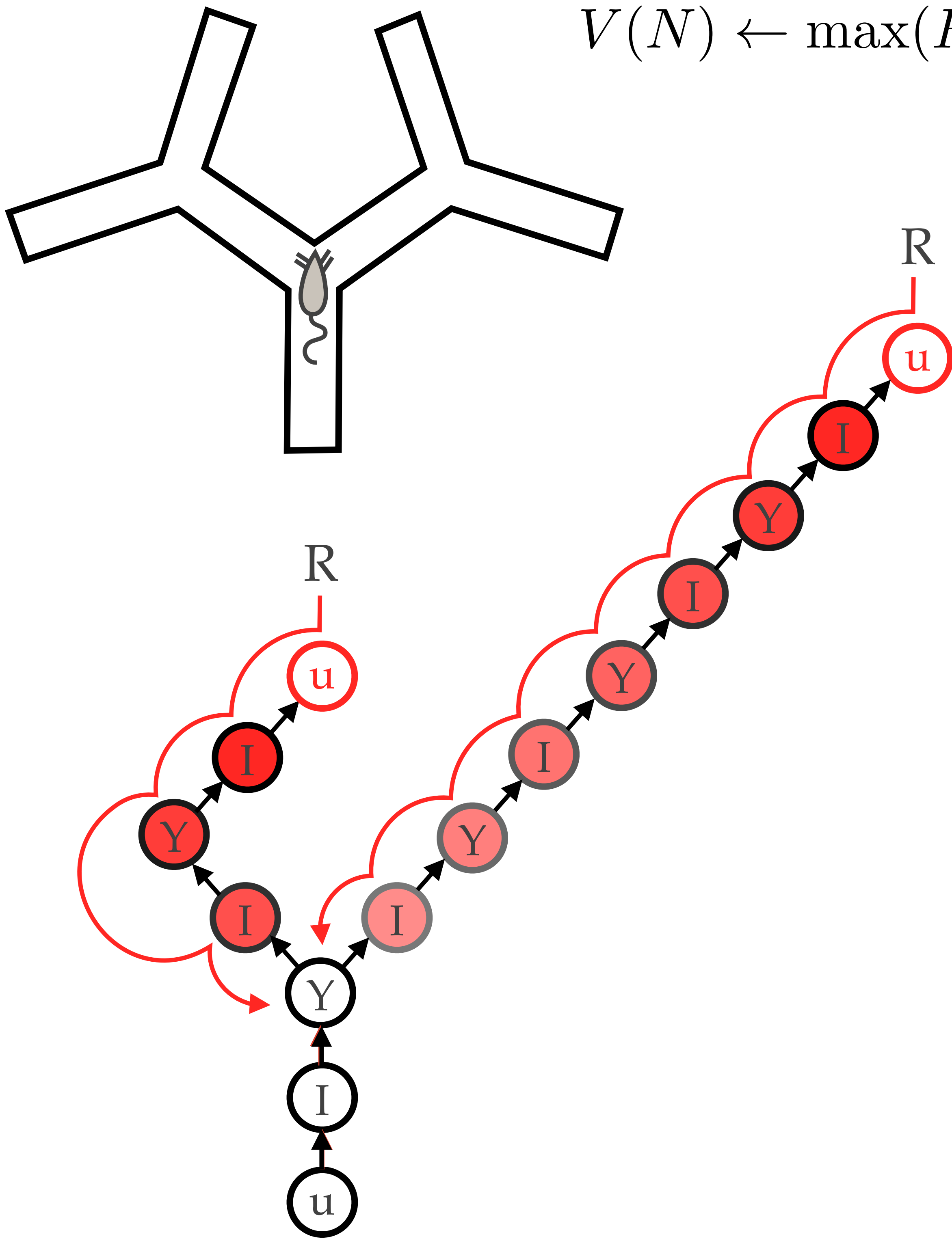
$$V(N) \leftarrow \max(R(N), V(N), \max_i(\gamma T(N, a_i, N')V(N'))$$



- 3 types of states: u ; I ; Y
- 4 actions: Fwd, L, R, U-turn
- Transitions learning
- **Inference (value interation)**

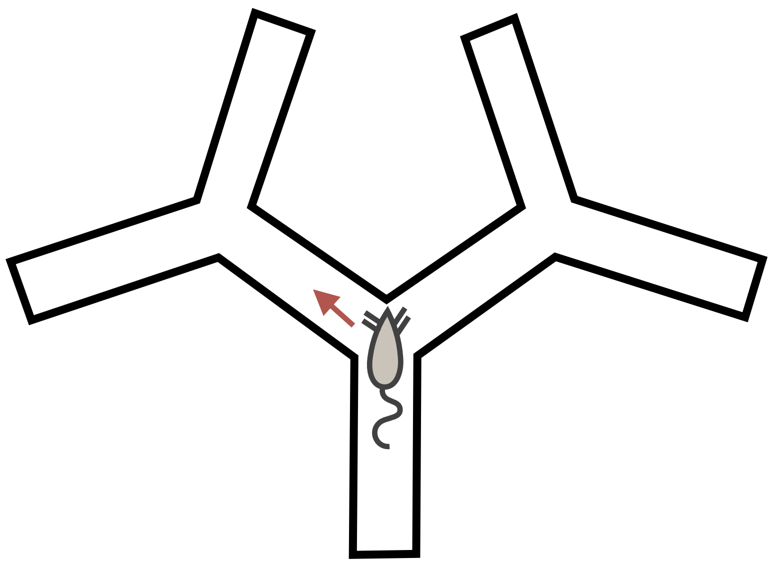
Model-Based Reinforcement Learning

$$V(N) \leftarrow \max(R(N), V(N), \max_i(\gamma T(N, a_i, N')V(N'))$$

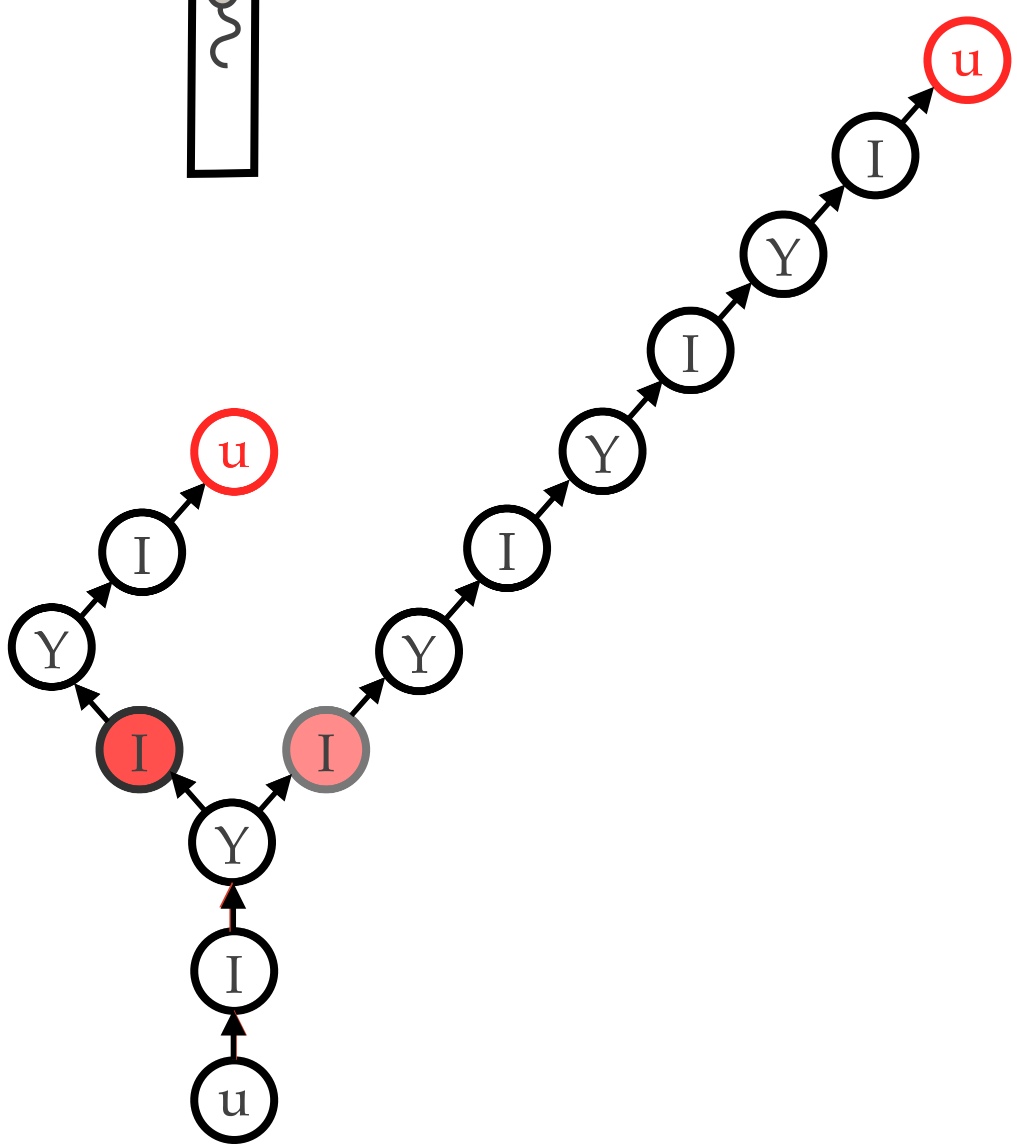


- 3 types of states: u ; I ; Y
- 4 actions: Fwd, L, R, U-turn
- Transitions learning
- **Inference (value interation)**

Model-Based Reinforcement Learning

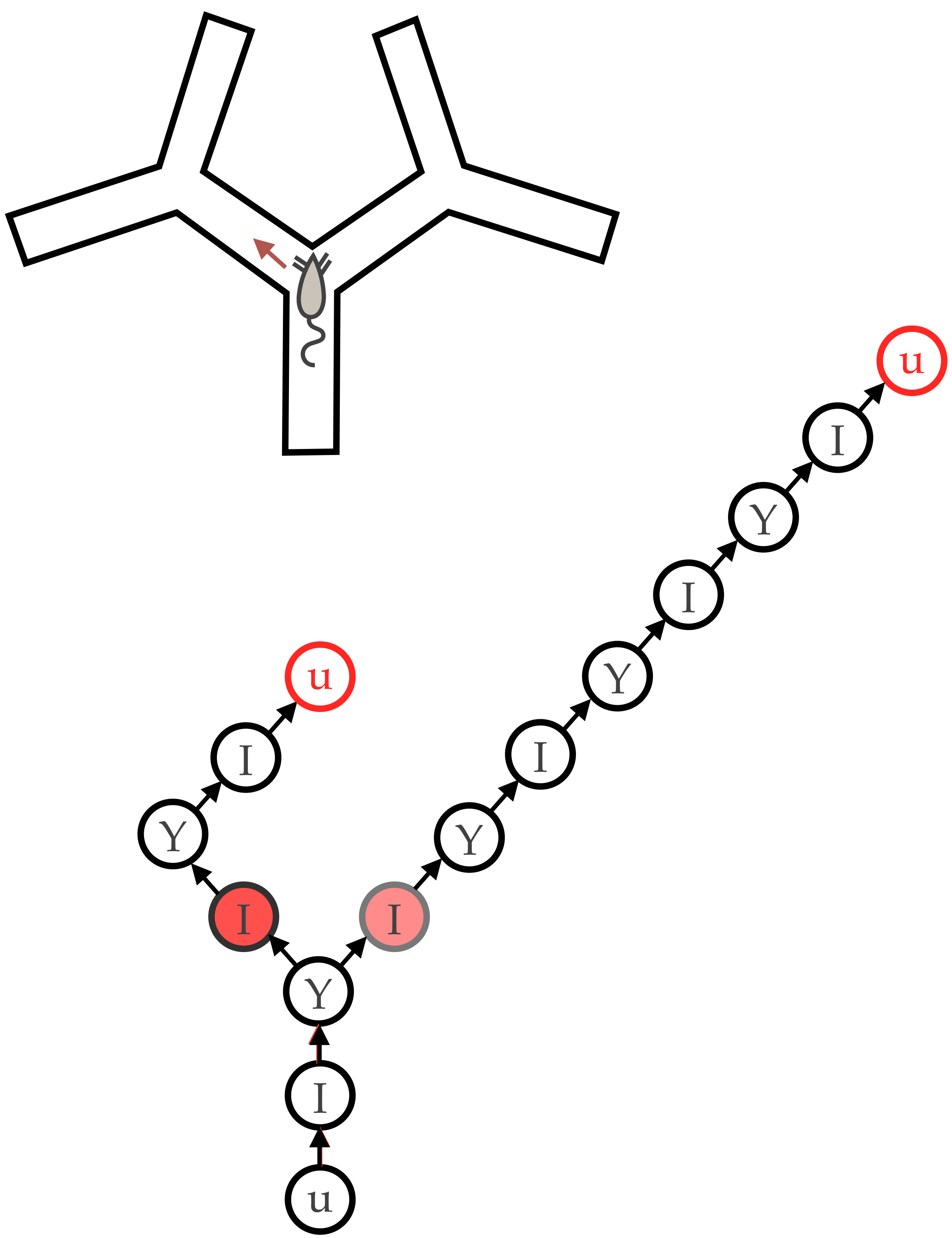


$$P(N_k, a_i) = \frac{e^{\beta Q(N_k, a_i)}}{\sum_j e^{\beta Q(N_k, a_j)}}$$



- 3 types of states: u ; I ; Y
- 4 actions: Fwd, L, R, U-turn
- Transitions learning
- Inference (value interation)
- **Action selection (softmax)**

Model-Based Reinforcement Learning

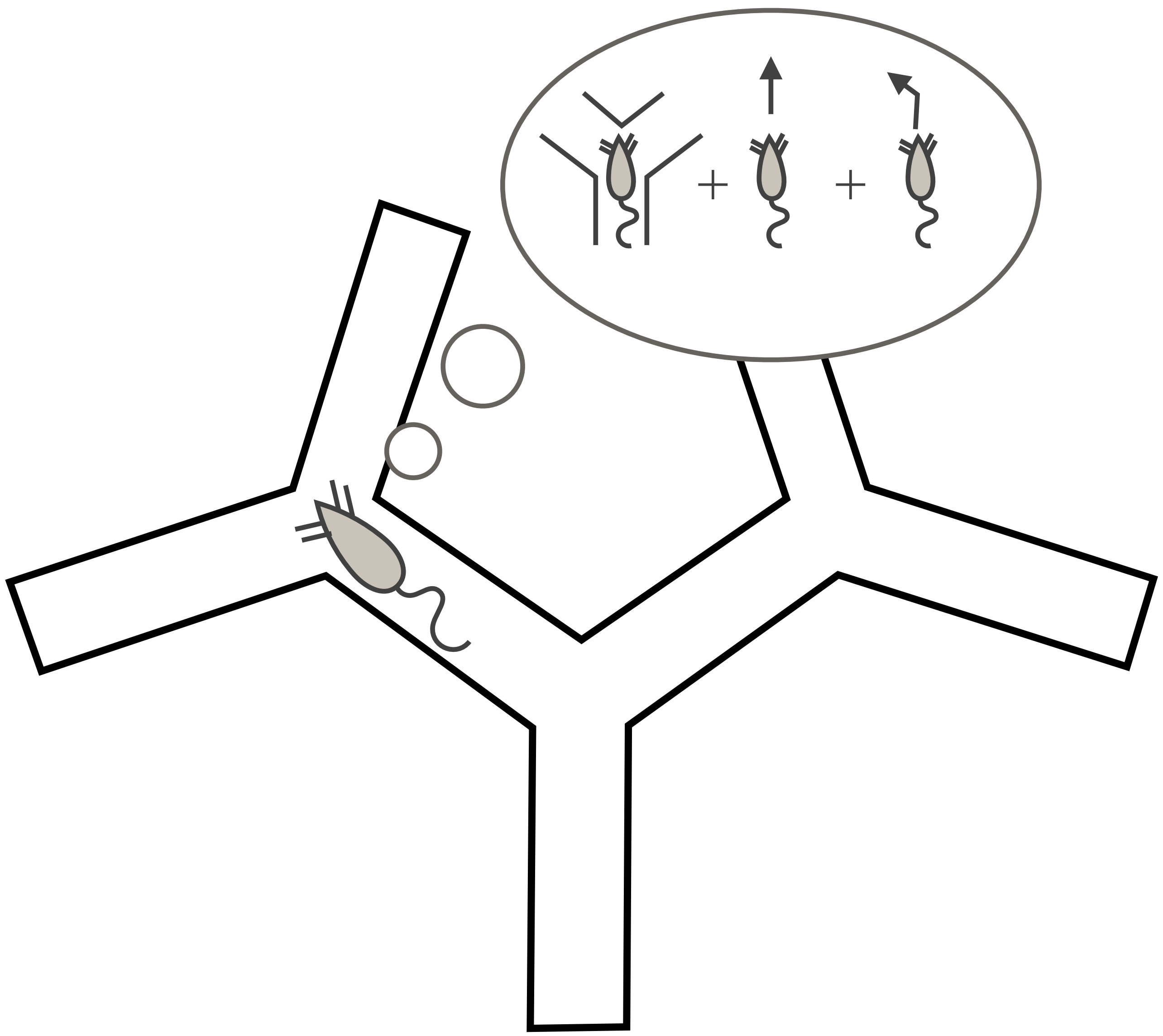


$$P(N_k, a_i) = \frac{e^{\beta Q(N_k, a_i)}}{\sum_j e^{\beta Q(N_k, a_j)}}$$

- 3 types of states: u ; I ; Y
- 4 actions: Fwd, L, R, U-turn
- Transitions learning
- Inference (value interation)
- **Action selection (softmax)**

Model-Free Reinforcement Learning

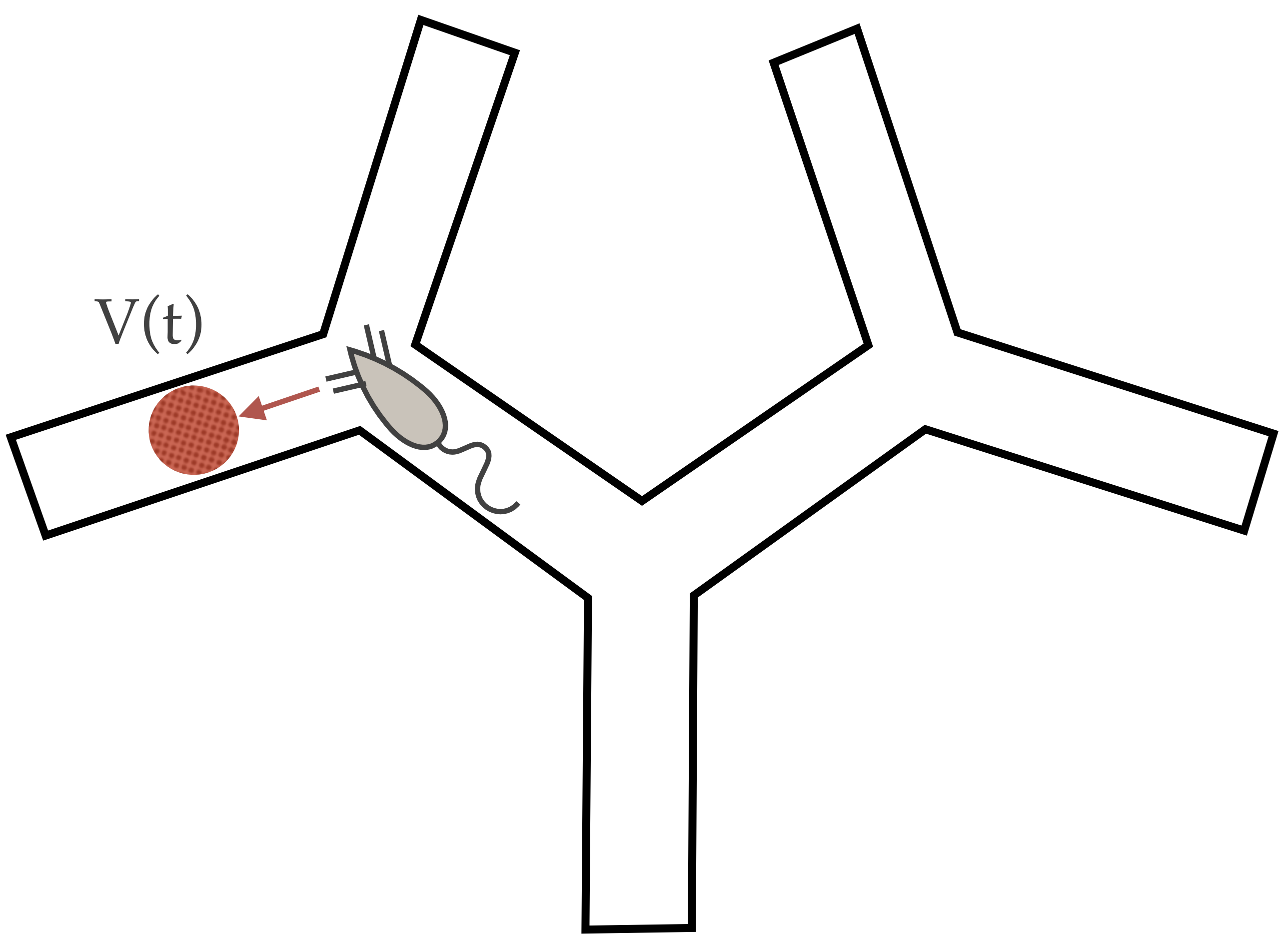
$$S^t = (I^t, a^{t-1}, ..., a^{t-n})$$



- 3 types of states: u ; I ; Y
- 4 actions: Fwd, L, R, U-turn
- **State incl. memory of actions**

Model-Free Reinforcement Learning

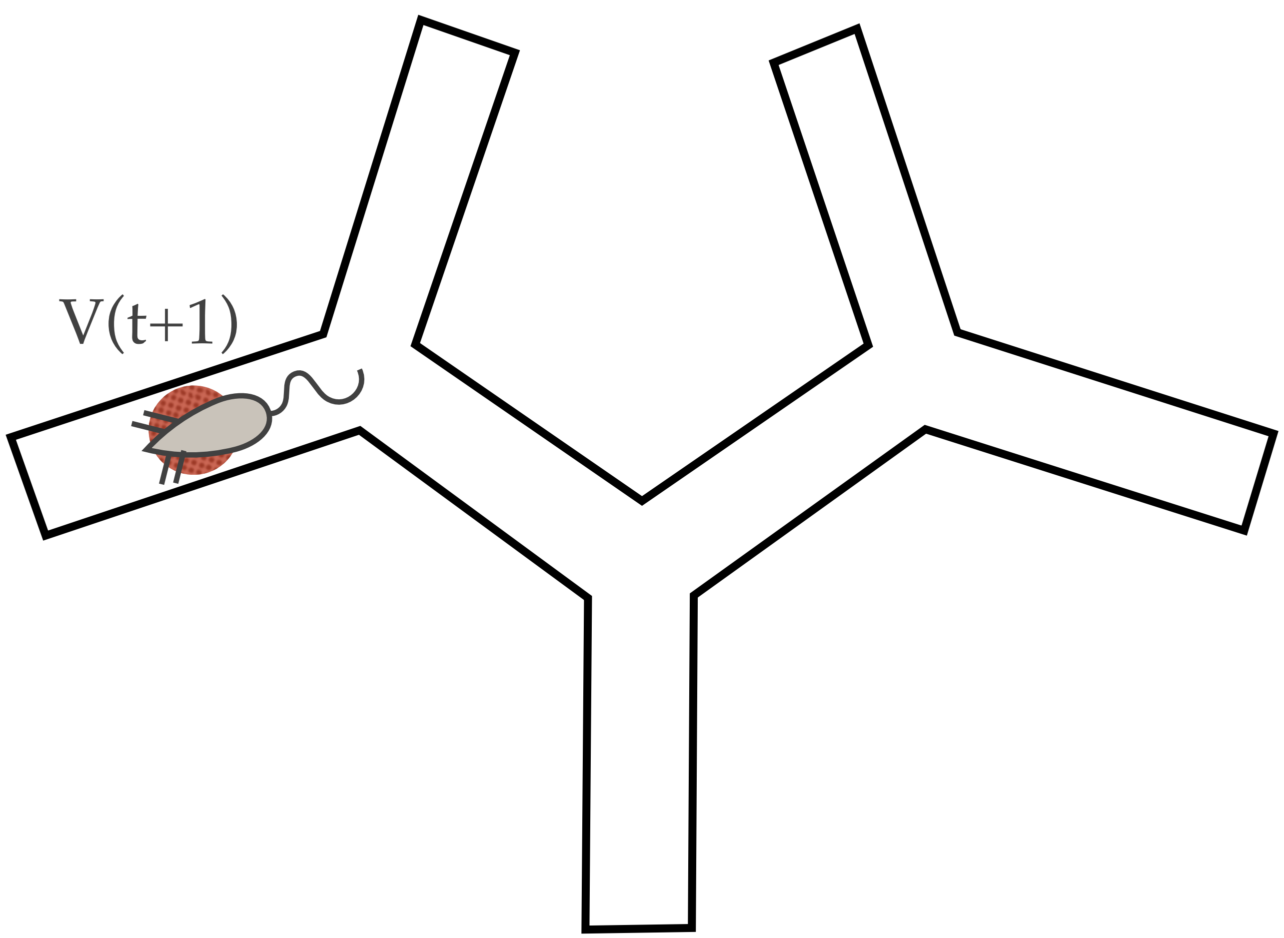
$$\delta^t = r^t + \gamma V(s^{t+1}) - V(s^t)$$
$$V(s^t) \leftarrow V(s^t) + \eta\delta^t$$
$$p(s^t, a^t) \leftarrow p(s^t, a^t) + \eta\delta^t$$



- 3 types of states: u ; I ; Y
- 4 actions: Fwd, L, R, U-turn
- State incl. memory of actions
- **Inference (Actor-critic)**

Model-Free Reinforcement Learning

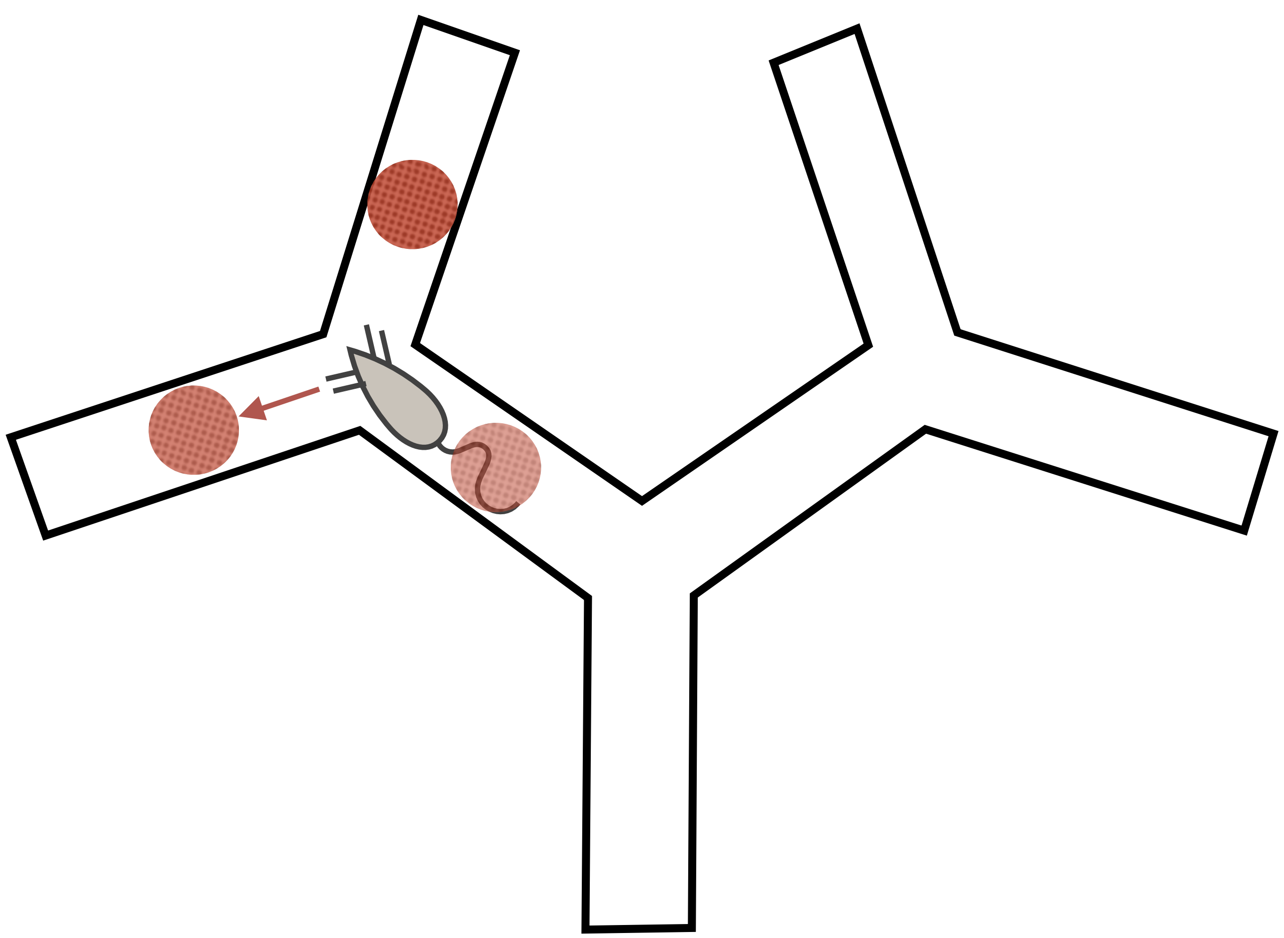
$$\delta^t = r^t + \gamma V(s^{t+1}) - V(s^t)$$
$$V(s^t) \leftarrow V(s^t) + \eta\delta^t$$
$$p(s^t, a^t) \leftarrow p(s^t, a^t) + \eta\delta^t$$



- 3 types of states: u ; I ; Y
- 4 actions: Fwd, L, R, U-turn
- State incl. memory of actions
- **Inference (Actor-critic)**

Model-Free Reinforcement Learning

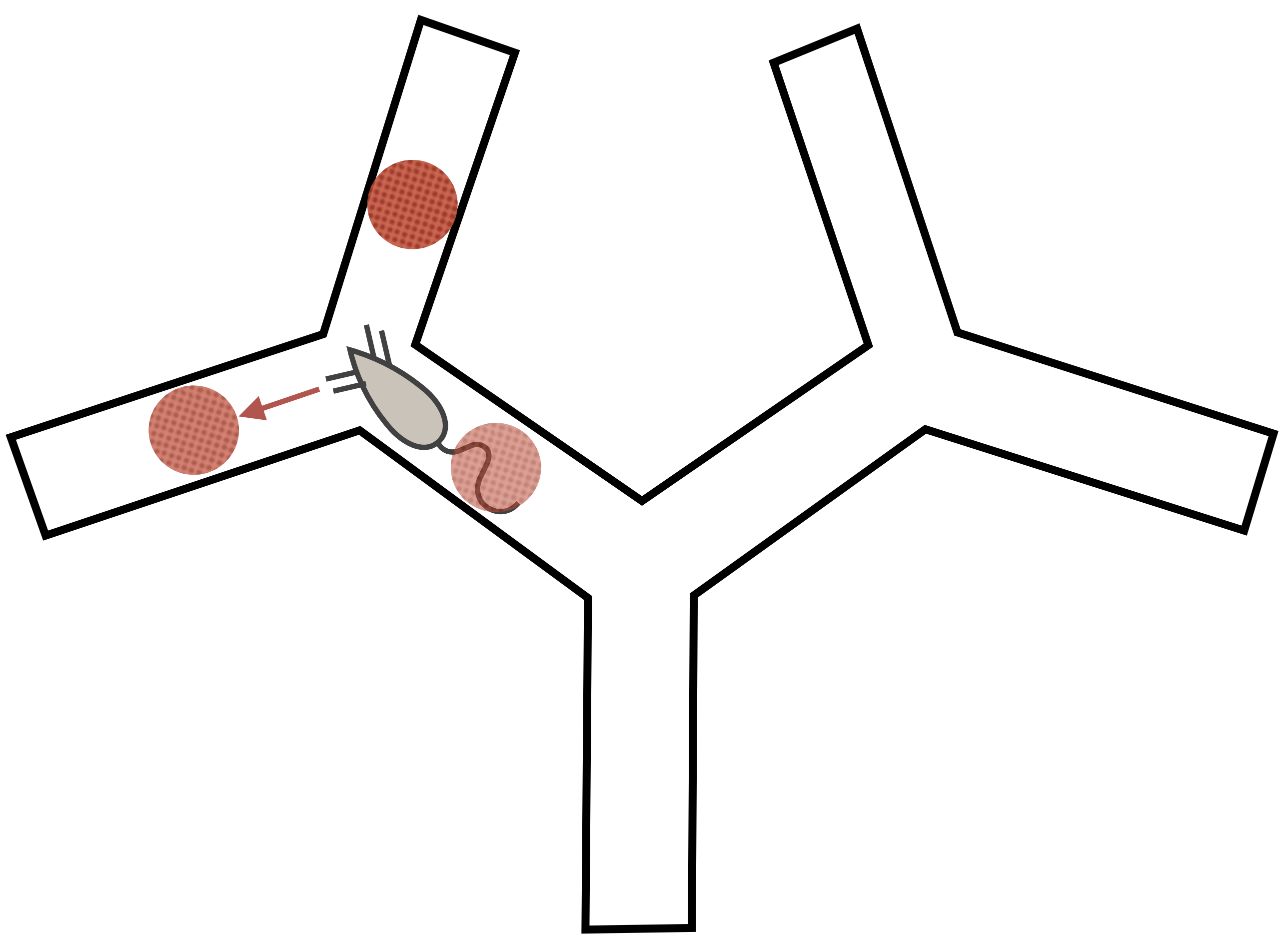
$$P(N_k, a_i) = \frac{e^{\beta Q(N_k, a_i)}}{\sum_j e^{\beta Q(N_k, a_j)}}$$



- 3 types of states: u ; I ; Y
- 4 actions: Fwd, L, R, U-turn
- State incl. memory of actions
- Inference (Actor-critic)
- **Action selection (softmax)**

Model-Free Reinforcement Learning

$$P(N_k, a_i) = \frac{e^{\beta Q(N_k, a_i)}}{\sum_j e^{\beta Q(N_k, a_j)}}$$



- 3 types of states: u ; I ; Y
- 4 actions: Fwd, L, R, U-turn
- State incl. memory of actions
- Inference (Actor-critic)
- **Action selection (softmax)**

Individual Parameter Optimization

RL models:

- learning rate
- exploration / exploitation
- discount factor

Path integration model:

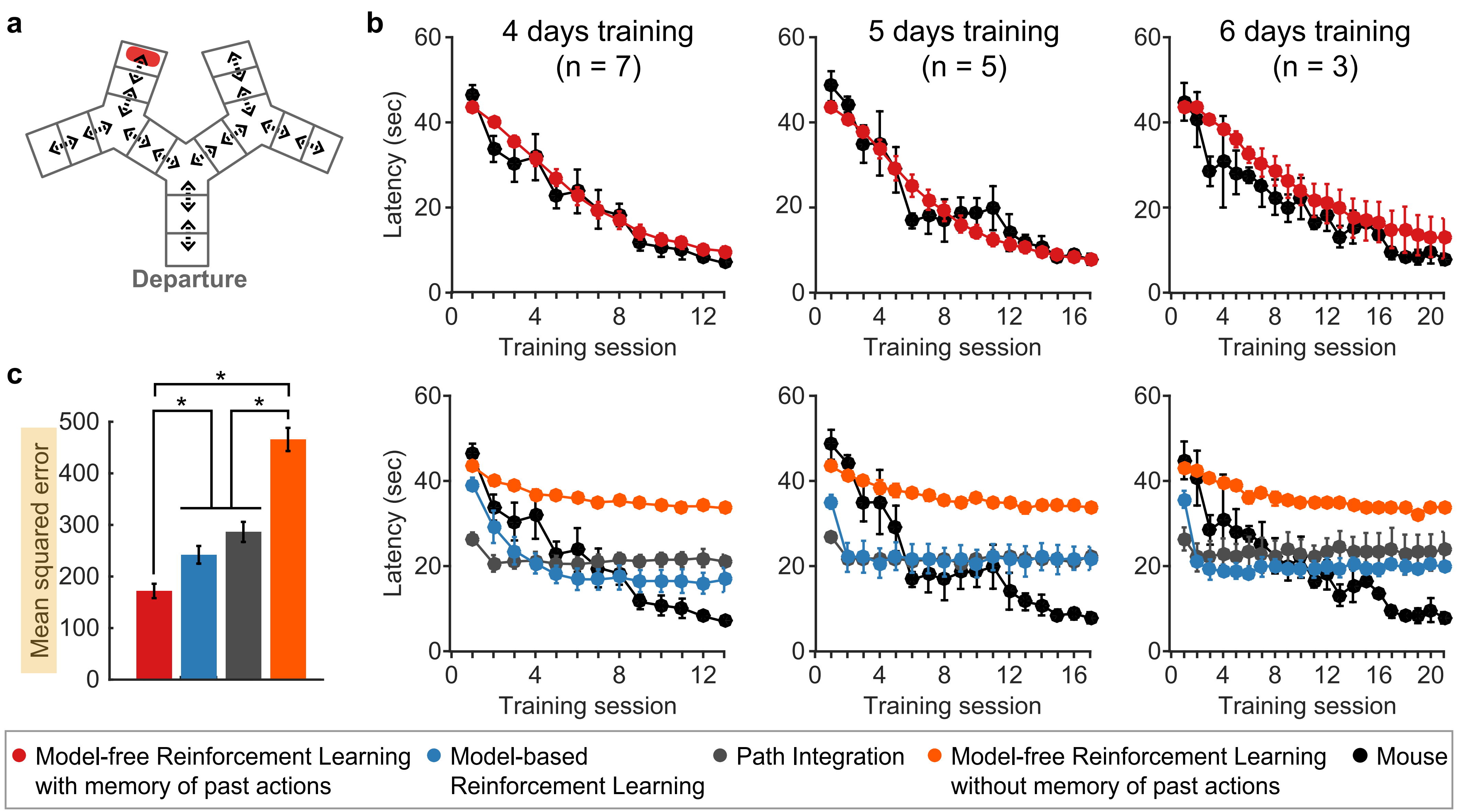
- learning rate
- exploration / exploitation
- position error accumulation

NSGA-2 optimization

fitnesses: log-likelihood, diversity

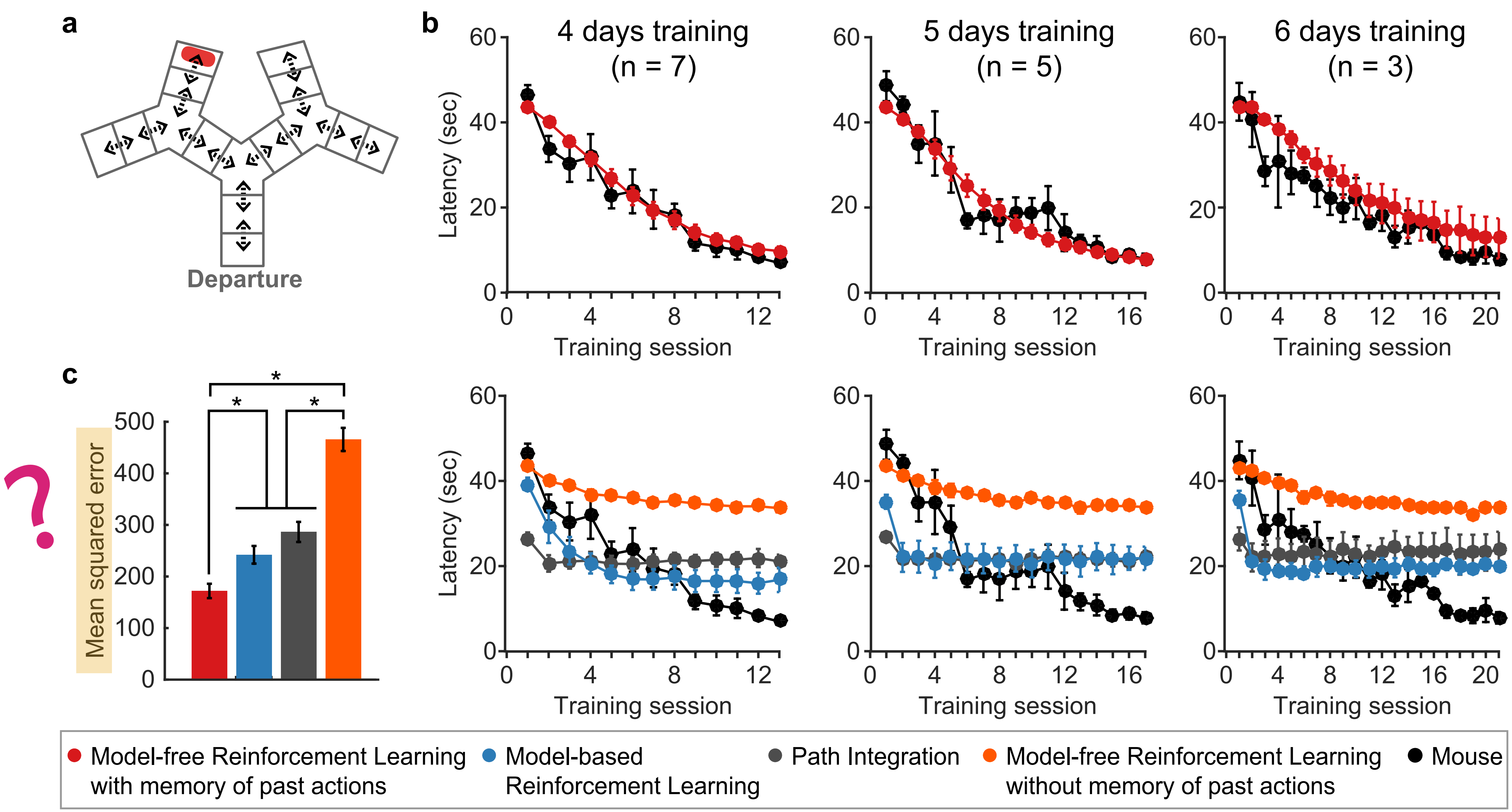
Best model (main paper)

Free wheel simulations (100 agent/mouse)



Best model (main paper)

Free wheel simulations (100 agent/mouse)



Best models (suppl. Information)

a

Mouse	Path Integration: Learning rate (η)	Path Integration: Error accumulation (σ)	Path Integration: Exploration/ exploitation trade-off (β)	Path Integration: Max (log likelihood)	Model-based Reinforcement Learning: Learning rate (η)	Model-based Reinforcement Learning: Discount factor (γ)	Model-based Reinforcement Learning: Exploration/ exploitation trade-off (β)	Model-based Reinforcement Learning: Max (log likelihood)	Model-free Reinforcement Learning (With memory of past actions): Learning rate (η)	Model-free Reinforcement Learning (With memory of past actions): Discount factor (γ)	Model-free Reinforcement Learning (With memory of past actions): Exploration/ exploitation trade-off (β)	Model-free Reinforcement Learning (With memory of past actions): Max (log likelihood)
1	0.001	0.0008	1.14	**-376.09**	0.47	0.93	4.43	**-353.48**	0.87	1.00	0.76	**-345.85**
2	0.001	1.0000	6.47	**-490.20**	0.13	1.00	6.58	**-472.41**	0.85	1.00	1.05	**-421.82**
3	0.002	0.0003	0.98	**-408.16**	0.39	0.78	6.93	**-379.45**	0.16	1.00	5.92	**-340.73**
4	0.001	0.0015	1.28	**-348.19**	0.89	0.87	4.83	**-328.25**	0.20	1.00	6.05	**-302.24**
5	0.001	0.0003	0.83	**-444.17**	0.80	0.97	3.03	**-423.04**	0.17	0.84	4.92	**-379.34**
6	0.996	0.0009	0.70	**-538.02**	0.76	0.29	200.00	**-503.53**	0.08	1.00	9.41	**-482.90**
7	0.001	1.0000	7.83	**-514.61**	0.89	0.88	4.39	**-491.09**	0.15	1.00	3.68	**-484.85**
8	0.982	0.0031	1.29	**-611.51**	0.89	0.90	4.25	**-576.73**	0.36	1.00	1.50	**-552.72**
9	1.000	0.0019	1.29	**-518.58**	0.76	0.96	3.53	**-466.68**	0.31	1.00	2.86	**-417.21**
10	0.001	0.0003	0.89	**-563.74**	0.75	0.28	200.00	**-519.95**	0.08	1.00	6.55	**-502.52**
11	0.628	0.0013	0.93	**-646.86**	0.88	0.56	11.69	**-639.53**	0.19	1.00	2.56	**-589.24**
12	0.001	0.9568	5.38	**-812.72**	0.80	0.94	3.50	**-750.07**	0.12	1.00	3.24	**-743.17**
13	0.936	0.0024	1.47	**-599.09**	0.97	0.68	8.93	**-579.27**	0.19	1.00	2.13	**-576.56**
14	0.001	0.0017	0.93	**-495.97**	0.32	0.85	4.96	**-469.88**	0.11	1.00	4.89	**-473.38**
15	0.576	0.0041	1.34	**-756.02**	0.71	0.99	3.01	**-691.71**	1.00	1.00	0.20	**-745.84**

b

Mouse	Model-free Reinforcement Learning (without memory of past actions): Learning rate (η)	Discount factor (γ)	Exploration/ exploitation trade-off (β)	Max (log likelihood)
1	0.45	1.00	0.34	**-332.66**
2	0.02	1.00	18.40	**-344.68**
3	0.07	1.00	2.04	**-355.87**
4	0.05	1.00	4.07	**-318.48**
5	0.03	1.00	3.15	**-426.17**
6	0.03	1.00	1.80	**-523.92**
7	0.64	1.00	0.47	**-395.15**
8	0.58	1.00	0.28	**-524.22**
9	0.68	1.00	0.31	**-407.27**
10	0.64	1.00	0.07	**-579.21**
11	0.83	1.00	0.12	**-604.87**
12	0.66	1.00	0.23	**-653.92**
13	0.97	1.00	0.15	**-534.31**
14	0.03	1.00	4.15	**-444.91**
15	0.90	1.00	0.14	**-671.09**

Supplementary Table 5: Parameter optimization results for the three models capable of learning the sequence (a) and for the model-free algorithm without memory (b). The parameters were optimized for each mouse and each model by identifying the set giving the maximum log likelihood. In red are highlighted the highest maximum log likelihood values across the four learning models and orange values highlight the highest values amongst the three model capable of learning the sequence (a). These optimized parameters were used for the learning simulations (Figure 3, Supplementary Figures 7) and the correlation analysis between the model-free reinforcement learning parameters and c-Fos densities (Figures 3, Supplementary Figure 8 and Supplementary Table 4).

Best models (suppl. Information)

a

	Path Integration				Model-based Reinforcement Learning				Model-free Reinforcement Learning (With memory of past actions)			
Mouse	**Learning rate (η)**	**Error accumulation (σ)**	**Exploration/ exploitation trade-off (β)**	**Max (log likelihood)**	**Learning rate (η)**	**Discount factor (γ)**	**Exploration/ exploitation trade-off (β)**	**Max (log likelihood)**	**Learning rate (η)**	**Discount factor (γ)**	**Exploration/ exploitation trade-off (β)**	**Max (log likelihood)**
1	0.001	0.0008	1.14	**-376.09**	0.47	0.93	4.43	**-353.48**	0.87	1.00	0.76	**-345.85**
2	0.001	1.0000	6.47	**-490.20**	0.13	1.00	6.58	**-472.41**	0.85	1.00	1.05	**-421.82**
3	0.002	0.0003	0.98	**-408.16**	0.39	0.78	6.93	**-379.45**	0.16	1.00	5.92	**-340.73**
4	0.001	0.0015	1.28	**-348.19**	0.89	0.87	4.83	**-328.25**	0.20	1.00	6.05	**-302.24**
5	0.001	0.0003	0.83	**-444.17**	0.80	0.97	3.03	**-423.04**	0.17	0.84	4.92	**-379.34**
6	0.996	0.0009	0.70	**-538.02**	0.76	0.29	200.00	**-503.53**	0.08	1.00	9.41	**-482.90**
7	0.001	1.0000	7.83	**-514.61**	0.89	0.88	4.39	**-491.09**	0.15	1.00	3.68	**-484.85**
8	0.982	0.0031	1.29	**-611.51**	0.89	0.90	4.25	**-576.73**	0.36	1.00	1.50	**-552.72**
9	1.000	0.0019	1.29	**-518.58**	0.76	0.96	3.53	**-466.68**	0.31	1.00	2.86	**-417.21**
10	0.001	0.0003	0.89	**-563.74**	0.75	0.28	200.00	**-519.95**	0.08	1.00	6.55	**-502.52**
11	0.628	0.0013	0.93	**-646.86**	0.88	0.56	11.69	**-639.53**	0.19	1.00	2.56	**-589.24**
12	0.001	0.9568	5.38	**-812.72**	0.80	0.94	3.50	**-750.07**	0.12	1.00	3.24	**-743.17**
13	0.936	0.0024	1.47	**-599.09**	0.97	0.68	8.93	**-579.27**	0.19	1.00	2.13	**-576.56**
14	0.001	0.0017	0.93	**-495.97**	0.32	0.85	4.96	**-469.88**	0.11	1.00	4.89	**-473.38**
15	0.576	0.0041	1.34	**-756.02**	0.71	0.99	3.01	**-691.71**	1.00	1.00	0.20	**-745.84**

b

	Model-free Reinforcement Learning (without memory of past actions)			
Mouse	**Learning rate (η)**	**Discount factor (γ)**	**Exploration/ exploitation trade-off (β)**	**Max (log likelihood)**
1	0.45	1.00	0.34	**-332.66**
2	0.02	1.00	18.40	**-344.68**
3	0.07	1.00	2.04	**-355.87**
4	0.05	1.00	4.07	**-318.48**
5	0.03	1.00	3.15	**-426.17**
6	0.03	1.00	1.80	**-523.92**
7	0.64	1.00	0.47	**-395.15**
8	0.58	1.00	0.28	**-524.22**
9	0.68	1.00	0.31	**-407.27**
10	0.64	1.00	0.07	**-579.21**
11	0.83	1.00	0.12	**-604.87**
12	0.66	1.00	0.23	**-653.92**
13	0.97	1.00	0.15	**-534.31**
14	0.03	1.00	4.15	**-444.91**
15	0.90	1.00	0.14	**-671.09**

?

Supplementary Table 5: Parameter optimization results for the three models capable of learning the sequence (a) and for the model-free algorithm without memory (b). The parameters were optimized for each mouse and each model by identifying the set giving the maximum log likelihood. In red are highlighted the highest maximum log likelihood values across the four learning models and orange values highlight the highest values amongst the three model capable of learning the sequence (a). These optimized parameters were used for the learning simulations (Figure 3, Supplementary Figures 7) and the correlation analysis between the model-free reinforcement learning parameters and c-Fos densities (Figures 3, Supplementary Figure 8 and Supplementary Table 4).

Best models (suppl. Information)

a

	Path Integration				Model-based Reinforcement Learning				Model-free Reinforcement Learning (With memory of past actions)			
Mouse	**Learning rate (η)**	**Error accumulation (σ)**	**Exploration/ exploitation trade-off (β)**	**Max (log likelihood)**	**Learning rate (η)**	**Discount factor (γ)**	**Exploration/ exploitation trade-off (β)**	**Max (log likelihood)**	**Learning rate (η)**	**Discount factor (γ)**	**Exploration/ exploitation trade-off (β)**	**Max (log likelihood)**
1	0.001	0.0008	1.14	**-376.09**	0.47	0.93	4.43	**-353.48**	0.87	1.00	0.76	**-345.85**
2	0.001	1.0000	6.47	**-490.20**	0.13	1.00	6.58	**-472.41**	0.85	1.00	1.05	**-421.82**
3	0.002	0.0003	0.98	**-408.16**	0.39	0.78	6.93	**-379.45**	0.16	1.00	5.92	**-340.73**
4	0.001	0.0015	1.28	**-348.19**	0.89	0.87	4.83	**-328.25**	0.20	1.00	6.05	**-302.24**
5	0.001	0.0003	0.83	**-444.17**	0.80	0.97	3.03	**-423.04**	0.17	0.84	4.92	**-379.34**
6	0.996	0.0009	0.70	**-538.02**	0.76	0.29	200.00	**-503.53**	0.08	1.00	9.41	**-482.90**
7	0.001	1.0000	7.83	**-514.61**	0.89	0.88	4.39	**-491.09**	0.15	1.00	3.68	**-484.85**
8	0.982	0.0031	1.29	**-611.51**	0.89	0.90	4.25	**-576.73**	0.36	1.00	1.50	**-552.72**
9	1.000	0.0019	1.29	**-518.58**	0.76	0.96	3.53	**-466.68**	0.31	1.00	2.86	**-417.21**
10	0.001	0.0003	0.89	**-563.74**	0.75	0.28	200.00	**-519.95**	0.08	1.00	6.55	**-502.52**
11	0.628	0.0013	0.93	**-646.86**	0.88	0.56	11.69	**-639.53**	0.19	1.00	2.56	**-589.24**
12	0.001	0.9568	5.38	**-812.72**	0.80	0.94	3.50	**-750.07**	0.12	1.00	3.24	**-743.17**
13	0.936	0.0024	1.47	**-599.09**	0.97	0.68	8.93	**-579.27**	0.19	1.00	2.13	**-576.56**
14	0.001	0.0017	0.93	**-495.97**	0.32	0.85	4.96	**-469.88**	0.11	1.00	4.89	**-473.38**
15	0.576	0.0041	1.34	**-756.02**	0.71	0.99	3.01	**-691.71**	1.00	1.00	0.20	**-745.84**

b

	Model-free Reinforcement Learning (without memory of past actions)			
Mouse	**Learning rate (η)**	**Discount factor (γ)**	**Exploration/ exploitation trade-off (β)**	**Max (log likelihood)**
1	0.45	1.00	0.34	**-332.66**
2	0.02	1.00	18.40	**-344.68**
3	0.07	1.00	2.04	**-355.87**
4	0.05	1.00	4.07	**-318.48**
5	0.03	1.00	3.15	**-426.17**
6	0.03	1.00	1.80	**-523.92**
7	0.64	1.00	0.47	**-395.15**
8	0.58	1.00	0.28	**-524.22**
9	0.68	1.00	0.31	**-407.27**
10	0.64	1.00	0.07	**-579.21**
11	0.83	1.00	0.12	**-604.87**
12	0.66	1.00	0.23	**-653.92**
13	0.97	1.00	0.15	**-534.31**
14	0.03	1.00	4.15	**-444.91**
15	0.90	1.00	0.14	**-671.09**

Supplementary Table 5: Parameter optimization results for the three models capable of learning the sequence (a) and for the model-free algorithm without memory (b). The parameters were optimized for each mouse and each model by identifying the set giving the maximum log likelihood. In red are highlighted the highest maximum log likelihood values across the four learning models and orange values highlight the highest values amongst the three model capable of learning the sequence (a). These optimized parameters were used for the learning simulations (Figure 3, Supplementary Figures 7) and the correlation analysis between the model-free reinforcement learning parameters and c-Fos densities (Figures 3, Supplementary Figure 8 and Supplementary Table 4).

Best models (suppl. Information)

a

Mouse	Path Integration				Model-based Reinforcement Learning				Model-free Reinforcement Learning (With memory of past actions)			
	Learning rate (η)	Error accumulation (σ)	Exploration/ exploitation trade-off (β)	Max (log likelihood)	Learning rate (η)	Discount factor (γ)	Exploration/ exploitation trade-off (β)	Max (log likelihood)	Learning rate (η)	Discount factor (γ)	Exploration/ exploitation trade-off (β)	Max (log likelihood)
1	0.001	0.0008	1.14	**-376.09**	0.47	0.93	4.43	**-353.48**	0.87	1.00	0.76	**-345.85**
2	0.001	1.0000	6.47	**-490.20**	0.13	1.00	6.58	**-472.41**	0.85	1.00	1.05	**-421.82**
3	0.002	0.0003	0.98	**-408.16**	0.39	0.78	6.93	**-379.45**	0.16	1.00	5.92	**-340.73**
4	0.001	0.0015	1.28	**-348.19**	0.89	0.87	4.83	**-328.25**	0.20	1.00	6.05	**-302.24**
5	0.001	0.0003	0.83	**-444.17**	0.80	0.97	3.03	**-423.04**	0.17	0.84	4.92	**-379.34**
6	0.996	0.0009	0.70	**-538.02**	0.76	0.29	200.00	**-503.53**	0.08	1.00	9.41	**-482.90**
7	0.001	1.0000	7.83	**-514.61**	0.89	0.88	4.39	**-491.09**	0.15	1.00	3.68	**-484.85**
8	0.982	0.0031	1.29	**-611.51**	0.89	0.90	4.25	**-576.73**	0.36	1.00	1.50	**-552.72**
9	1.000	0.0019	1.29	**-518.58**	0.76	0.96	3.53	**-466.68**	0.31	1.00	2.86	**-417.21**
10	0.001	0.0003	0.89	**-563.74**	0.75	0.28	200.00	**-519.95**	0.08	1.00	6.55	**-502.52**
11	0.628	0.0013	0.93	**-646.86**	0.88	0.56	11.69	**-639.53**	0.19	1.00	2.56	**-589.24**
12	0.001	0.9568	5.38	**-812.72**	0.80	0.94	3.50	**-750.07**	0.12	1.00	3.24	**-743.17**
13	0.936	0.0024	1.47	**-599.09**	0.97	0.68	8.93	**-579.27**	0.19	1.00	2.13	**-576.56**
14	0.001	0.0017	0.93	**-495.97**	0.32	0.85	4.96	**-469.88**	0.11	1.00	4.89	**-473.38**
15	0.576	0.0041	1.34	**-756.02**	0.71	0.99	3.01	**-691.71**	1.00	1.00	0.20	**-745.84**

b

Mouse	Model-free Reinforcement Learning (without memory of past actions)			
	Learning rate (η)	Discount factor (γ)	Exploration/ exploitation trade-off (β)	Max (log likelihood)
1	0.45	1.00	0.34	**-332.66**
2	0.02	1.00	18.40	**-344.68**
3	0.07	1.00	2.04	**-355.87**
4	0.05	1.00	4.07	**-318.48**
5	0.03	1.00	3.15	**-426.17**
6	0.03	1.00	1.80	**-523.92**
7	0.64	1.00	0.47	**-395.15**
8	0.58	1.00	0.28	**-524.22**
9	0.68	1.00	0.31	**-407.27**
10	0.64	1.00	0.07	**-579.21**
11	0.83	1.00	0.12	**-604.87**
12	0.66	1.00	0.23	**-653.92**
13	0.97	1.00	0.15	**-534.31**
14	0.03	1.00	4.15	**-444.91**
15	0.90	1.00	0.14	**-671.09**

Supplementary Table 5: Parameter optimization results for the three models capable of learning the sequence (a) and for the model-free algorithm without memory (b). The parameters were optimized for each mouse and each model by identifying the set giving the maximum log likelihood. In red are highlighted the highest maximum log likelihood values across the four learning models and orange values highlight the highest values amongst the three model capable of learning the sequence (a). These optimized parameters were used for the learning simulations (Figure 3, Supplementary Figures 7) and the correlation analysis between the model-free reinforcement learning parameters and c-Fos densities (Figures 3, Supplementary Figure 8 and Supplementary Table 4).

Best models (suppl. Information)

a

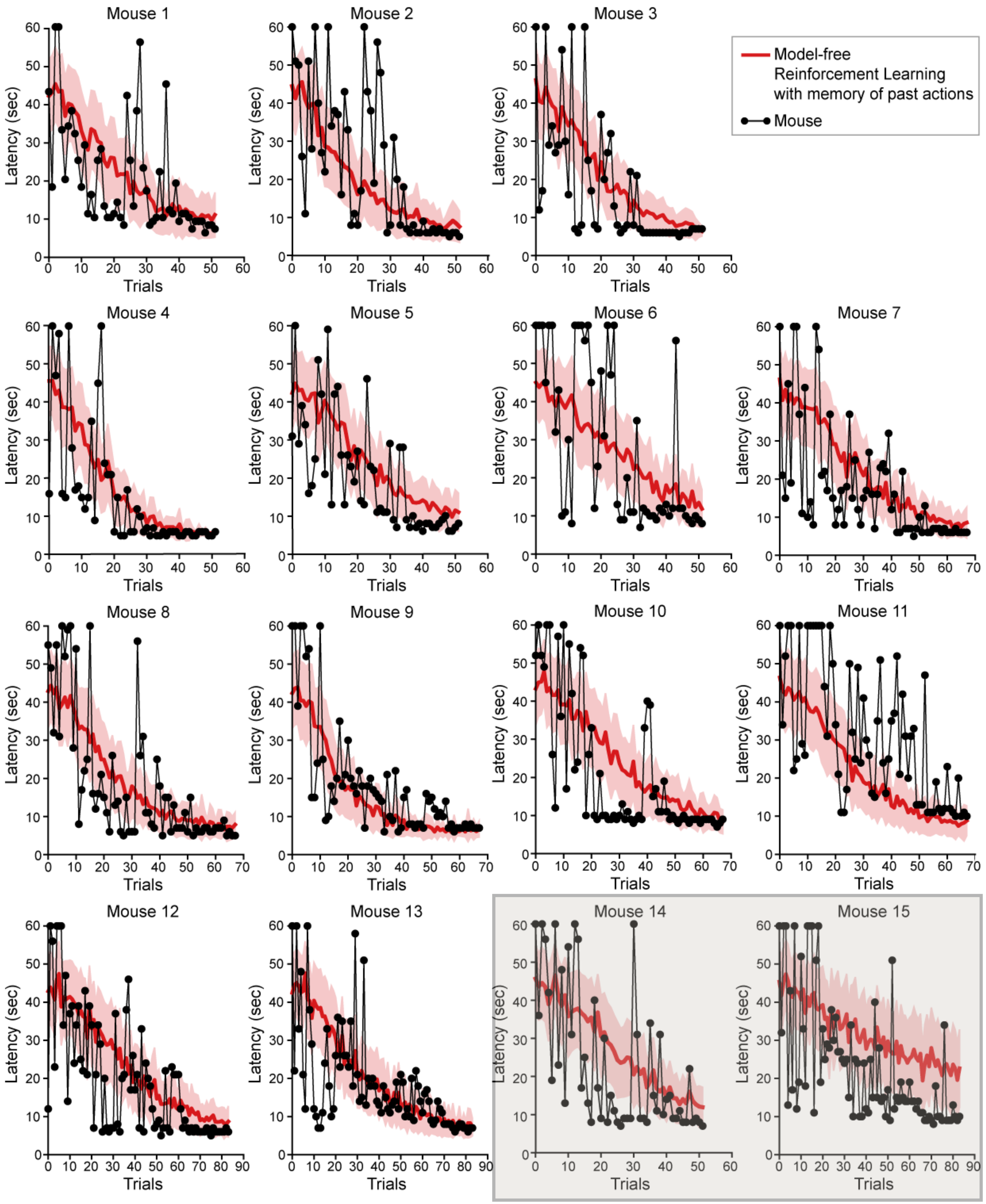
Mouse	Path Integration: Learning rate (η)	Path Integration: Error accumulation (σ)	Path Integration: Exploration/ exploitation trade-off (β)	Path Integration: Max (log likelihood)	Model-based Reinforcement Learning: Learning rate (η)	Model-based Reinforcement Learning: Discount factor (γ)	Model-based Reinforcement Learning: Exploration/ exploitation trade-off (β)	Model-based Reinforcement Learning: Max (log likelihood)	Model-free Reinforcement Learning (With memory of past actions): Learning rate (η)	Model-free Reinforcement Learning (With memory of past actions): Discount factor (γ)	Model-free Reinforcement Learning (With memory of past actions): Exploration/ exploitation trade-off (β)	Model-free Reinforcement Learning (With memory of past actions): Max (log likelihood)
1	0.001	0.0008	1.14	**-376.09**	0.47	0.93	4.43	**-353.48**	0.87	1.00	0.76	**-345.85**
2	0.001	1.0000	6.47	**-490.20**	0.13	1.00	6.58	**-472.41**	0.85	1.00	1.05	**-421.82**
3	0.002	0.0003	0.98	**-408.16**	0.39	0.78	6.93	**-379.45**	0.16	1.00	5.92	**-340.73**
4	0.001	0.0015	1.28	**-348.19**	0.89	0.87	4.83	**-328.25**	0.20	1.00	6.05	**-302.24**
5	0.001	0.0003	0.83	**-444.17**	0.80	0.97	3.03	**-423.04**	0.17	0.84	4.92	**-379.34**
6	0.996	0.0009	0.70	**-538.02**	0.76	0.29	200.00	**-503.53**	0.08	1.00	9.41	**-482.90**
7	0.001	1.0000	7.83	**-514.61**	0.89	0.88	4.39	**-491.09**	0.15	1.00	3.68	**-484.85**
8	0.982	0.0031	1.29	**-611.51**	0.89	0.90	4.25	**-576.73**	0.36	1.00	1.50	**-552.72**
9	1.000	0.0019	1.29	**-518.58**	0.76	0.96	3.53	**-466.68**	0.31	1.00	2.86	**-417.21**
10	0.001	0.0003	0.89	**-563.74**	0.75	0.28	200.00	**-519.95**	0.08	1.00	6.55	**-502.52**
11	0.628	0.0013	0.93	**-646.86**	0.88	0.56	11.69	**-639.53**	0.19	1.00	2.56	**-589.24**
12	0.001	0.9568	5.38	**-812.72**	0.80	0.94	3.50	**-750.07**	0.12	1.00	3.24	**-743.17**
13	0.936	0.0024	1.47	**-599.09**	0.97	0.68	8.93	**-579.27**	0.19	1.00	2.13	**-576.56**
14	0.001	0.0017	0.93	**-495.97**	0.32	0.85	4.96	**-469.88**	0.11	1.00	4.89	**-473.38**
15	0.576	0.0041	1.34	**-756.02**	0.71	0.99	3.01	**-691.71**	1.00	1.00	0.20	**-745.84**

b

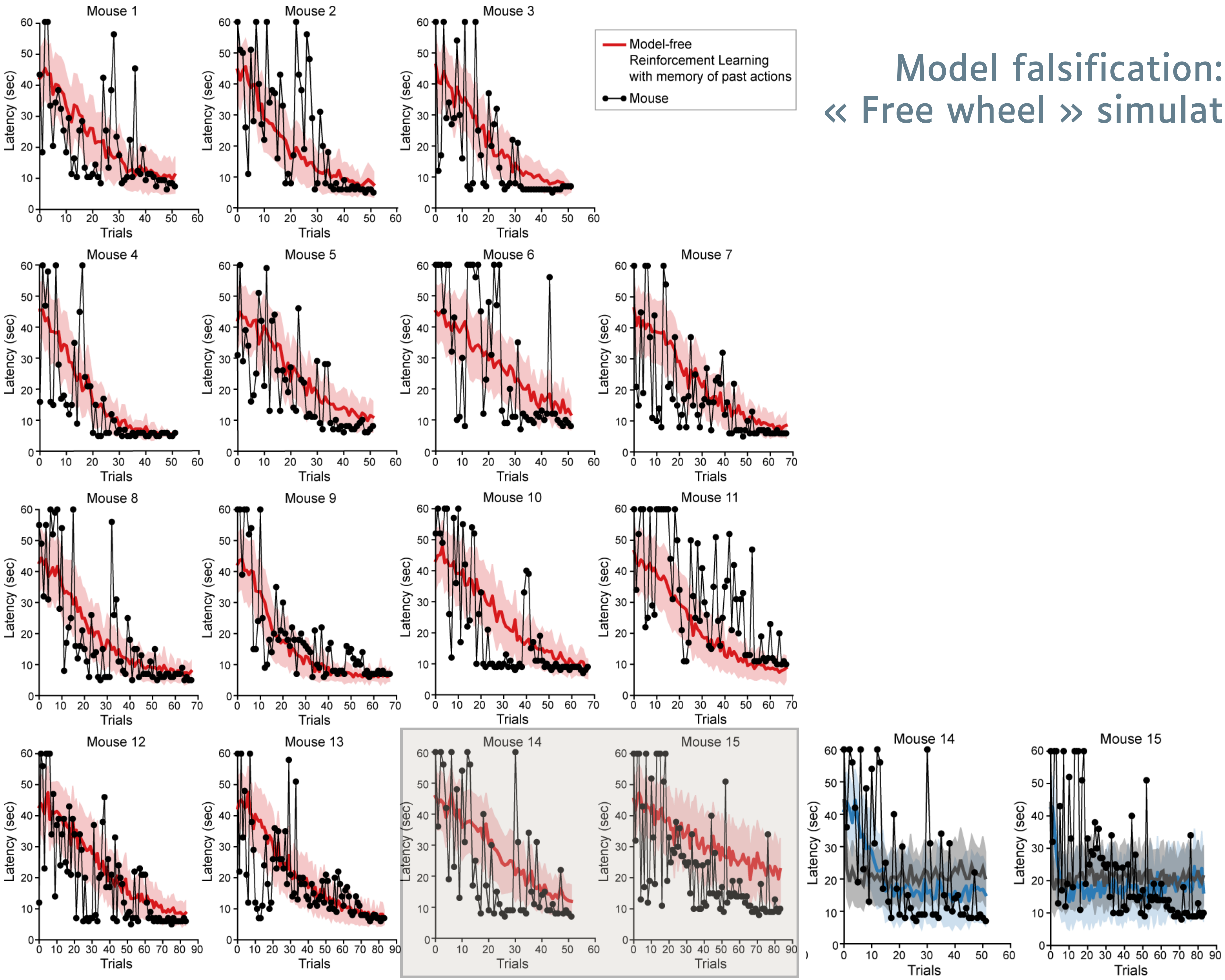
Mouse	Model-free Reinforcement Learning (without memory of past actions): Learning rate (η)	Discount factor (γ)	Exploration/ exploitation trade-off (β)	Max (log likelihood)
1	0.45	1.00	0.34	**-332.66**
2	0.02	1.00	18.40	**-344.68**
3	0.07	1.00	2.04	**-355.87**
4	0.05	1.00	4.07	**-318.48**
5	0.03	1.00	3.15	**-426.17**
6	0.03	1.00	1.80	**-523.92**
7	0.64	1.00	0.47	**-395.15**
8	0.58	1.00	0.28	**-524.22**
9	0.68	1.00	0.31	**-407.27**
10	0.64	1.00	0.07	**-579.21**
11	0.83	1.00	0.12	**-604.87**
12	0.66	1.00	0.23	**-653.92**
13	0.97	1.00	0.15	**-534.31**
14	0.03	1.00	4.15	**-444.91**
15	0.90	1.00	0.14	**-671.09**

Supplementary Table 5: Parameter optimization results for the three models capable of learning the sequence (a) and for the model-free algorithm without memory (b). The parameters were optimized for each mouse and each model by identifying the set giving the maximum log likelihood. In red are highlighted the highest maximum log likelihood values across the four learning models and orange values highlight the highest values amongst the three model capable of learning the sequence (a). These optimized parameters were used for the learning simulations (Figure 3, Supplementary Figures 7) and the correlation analysis between the model-free reinforcement learning parameters and c-Fos densities (Figures 3, Supplementary Figure 8 and Supplementary Table 4).

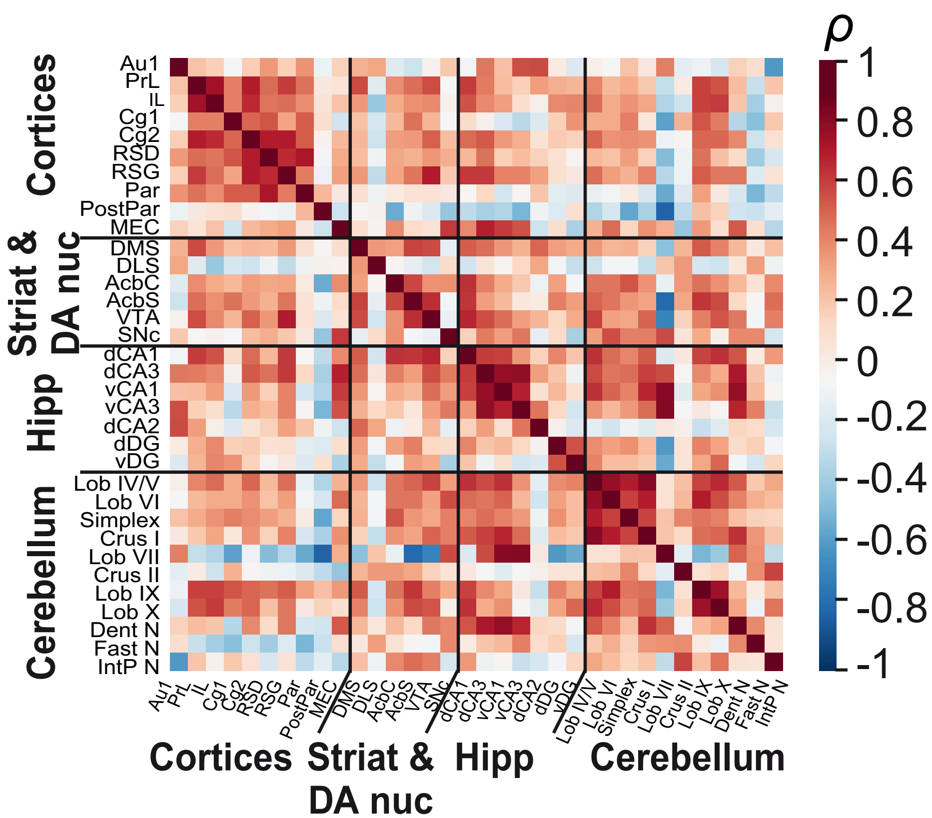
Model falsification: « Free wheel » simulations

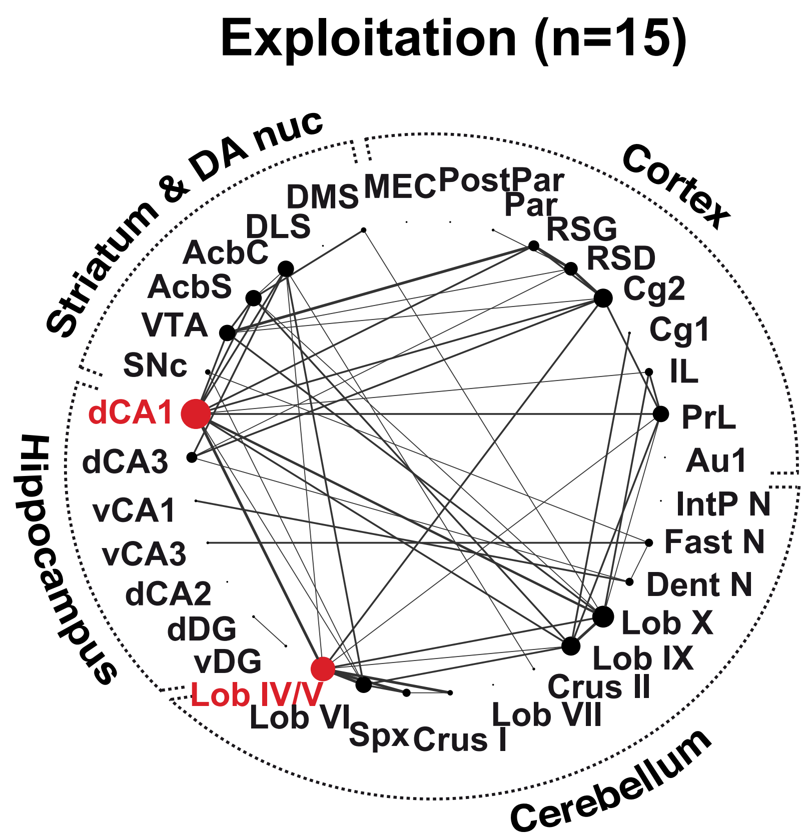


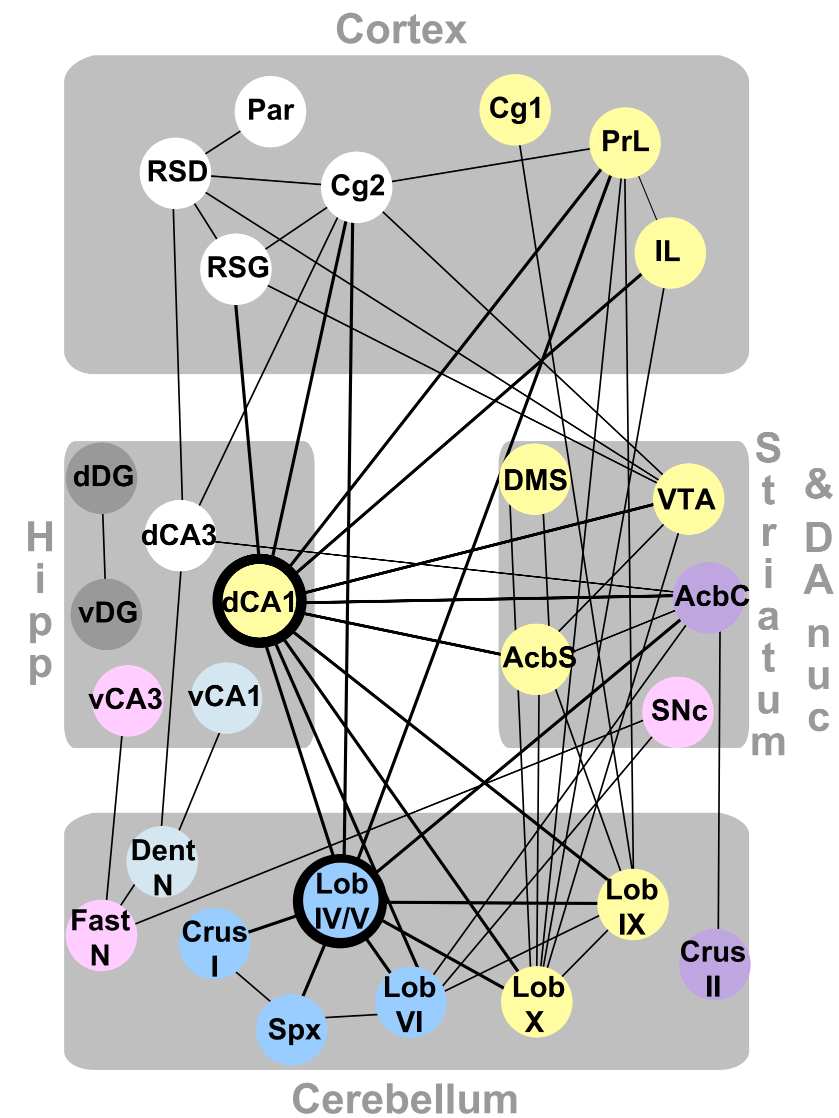
Model falsification:
« Free wheel » simulations



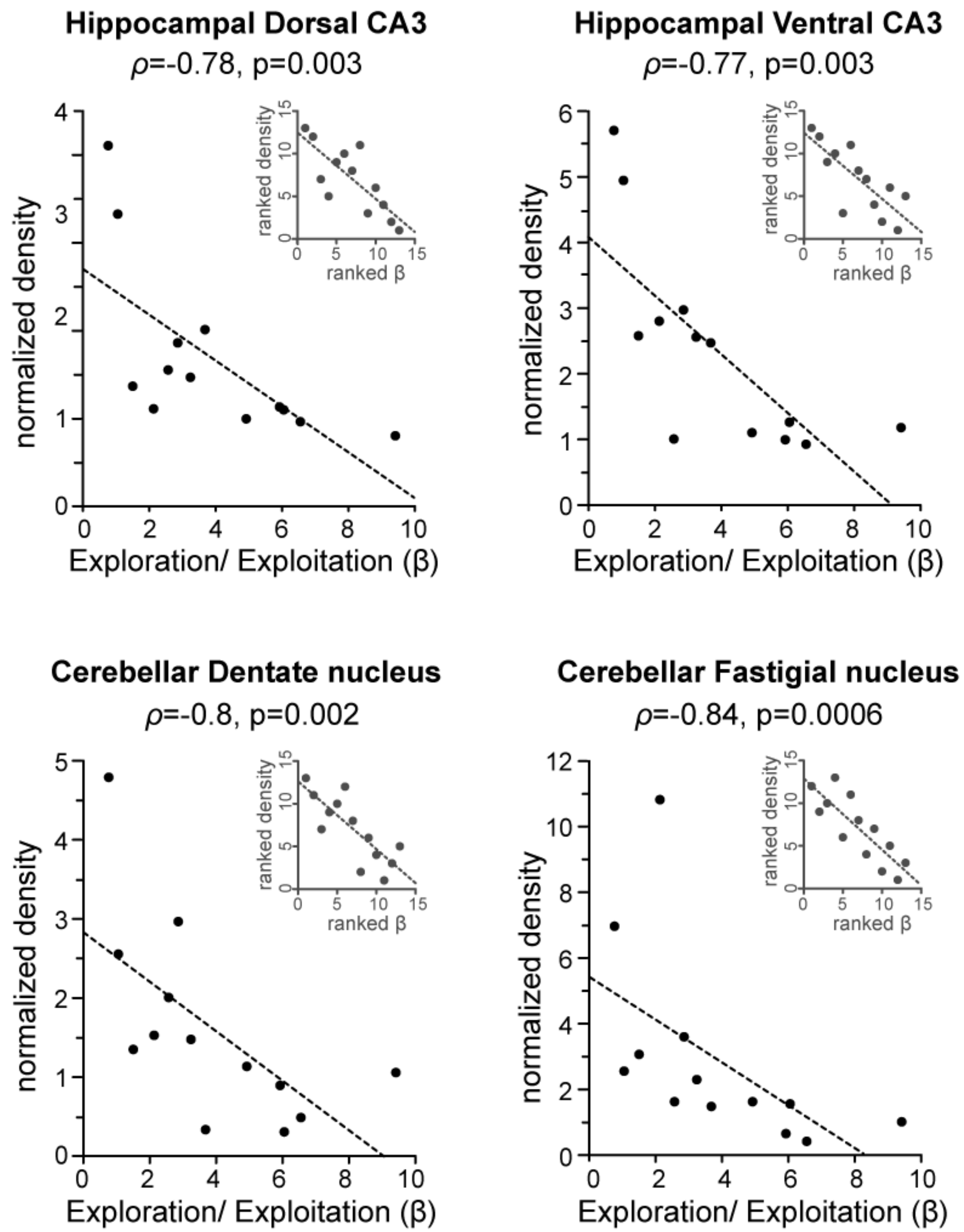
Back to the data:

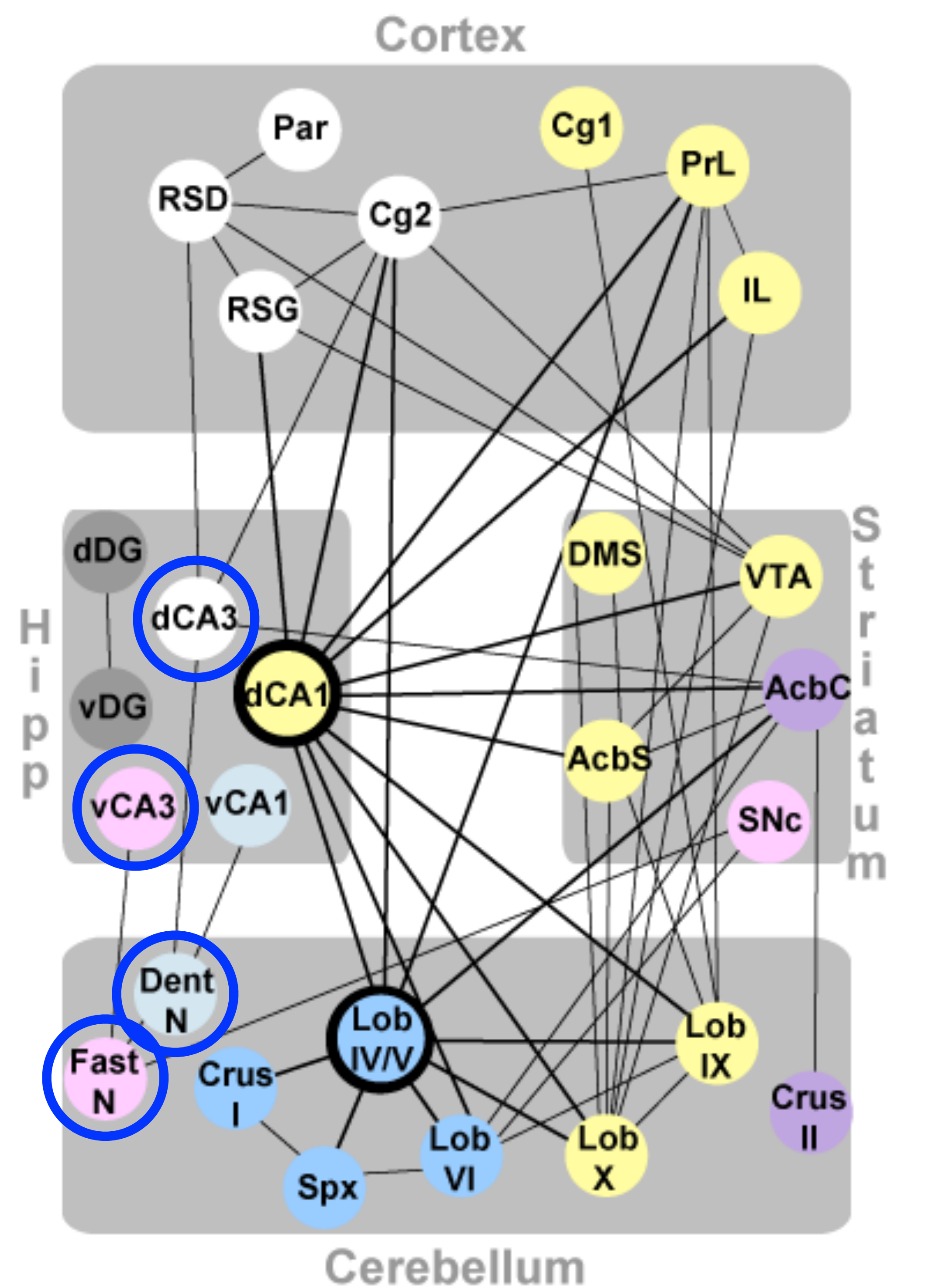




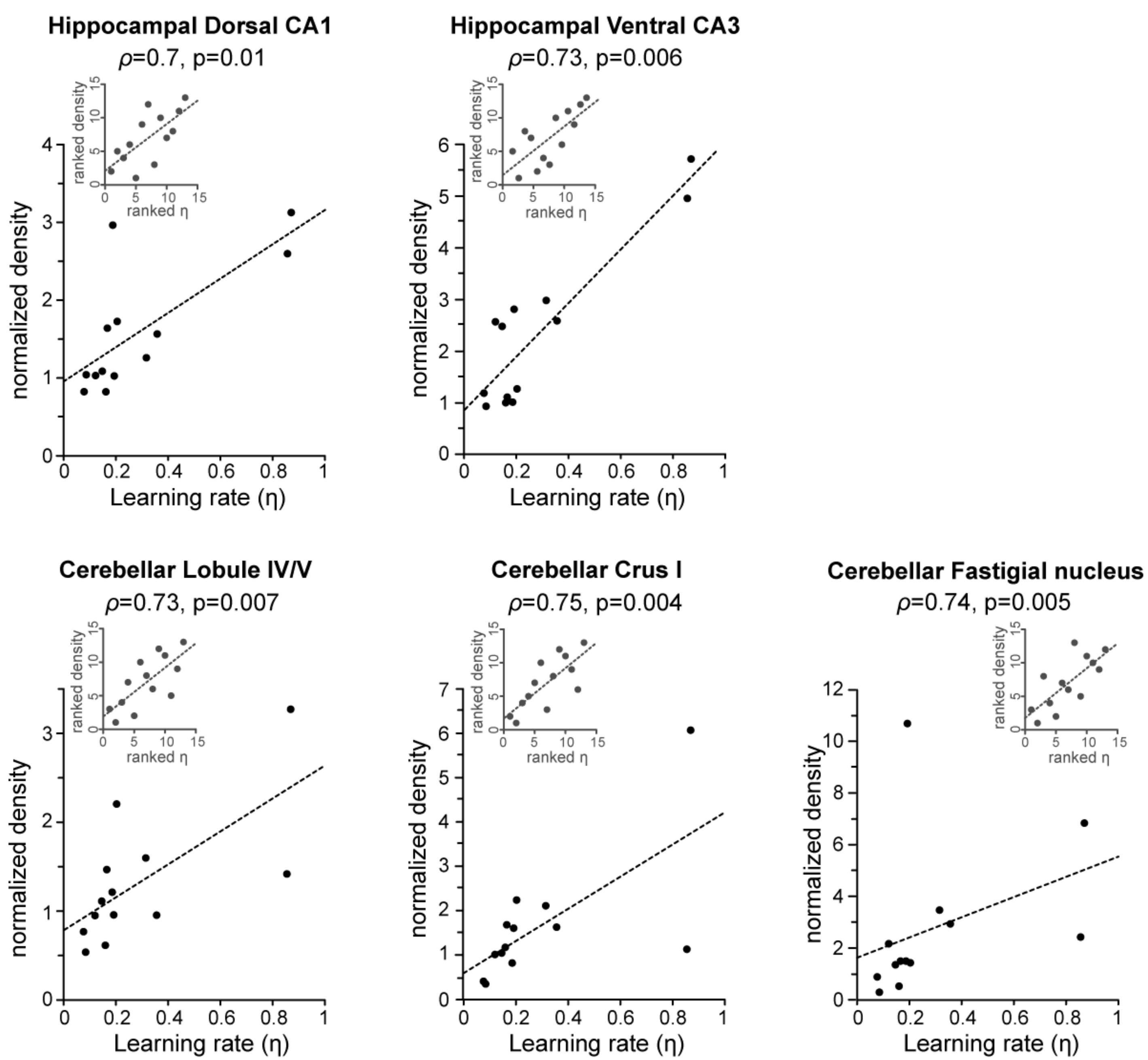


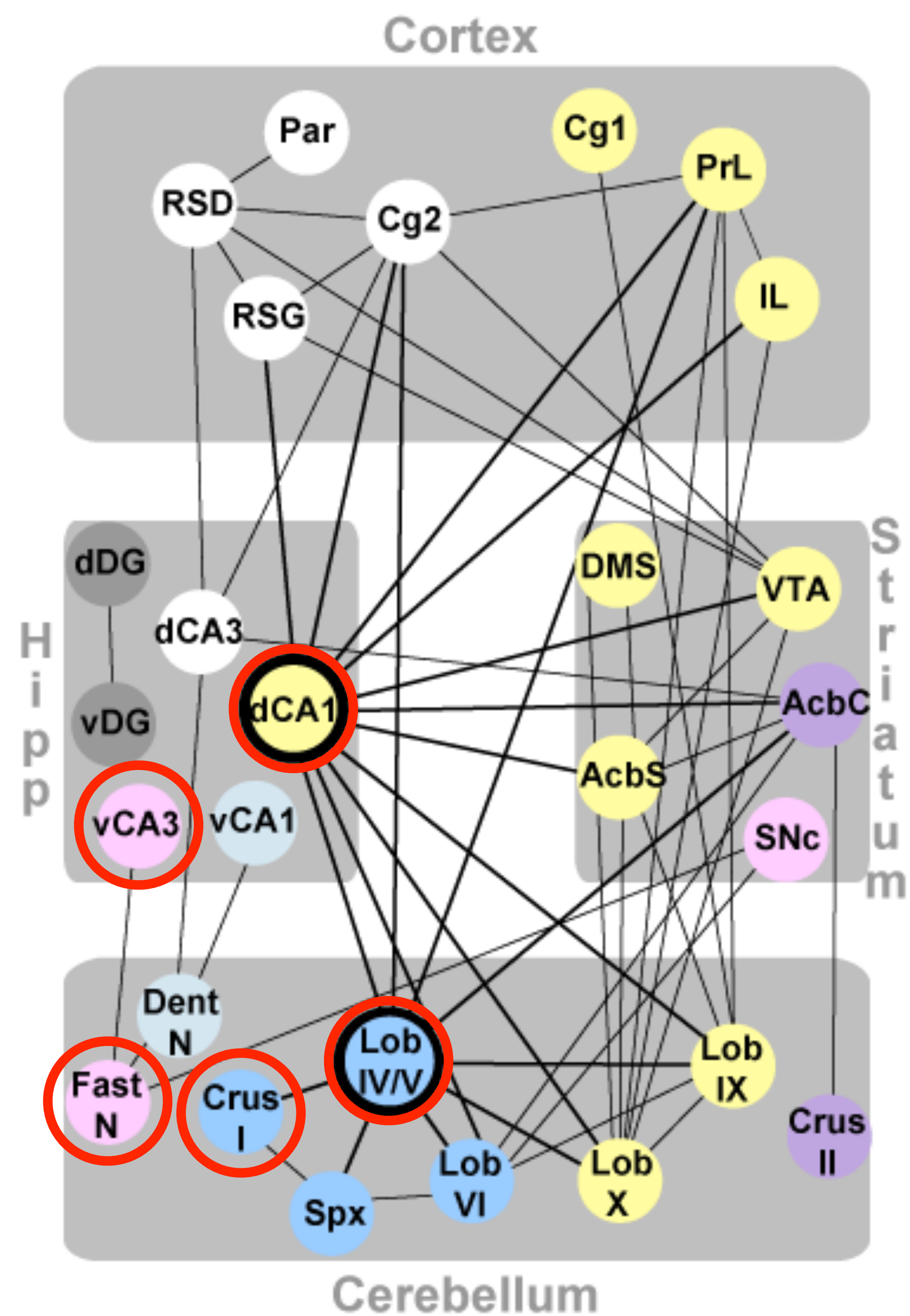
Correlations with cFos: exploration/exploitation





Correlations with cFos: learning rate





Model comparison & selection (continued)

Models with different numbers of parameters

- Akaike (1973) Information Criterion (AIC)
  - Derived from Kullback-Leibler divergence
  - Minimize:

$$AIC = -2 \log \mathcal{L}(\hat{\theta}|y) + 2K$$

  « 2 » is arbitrary, K: #param

  - Small #sample *T* (Hurvich & Tsai, 1989):

$$AIC_c = -2 \log \mathcal{L}(\hat{\theta}|y) + 2K \frac{T}{T-K-1}$$

Models with different numbers of parameters

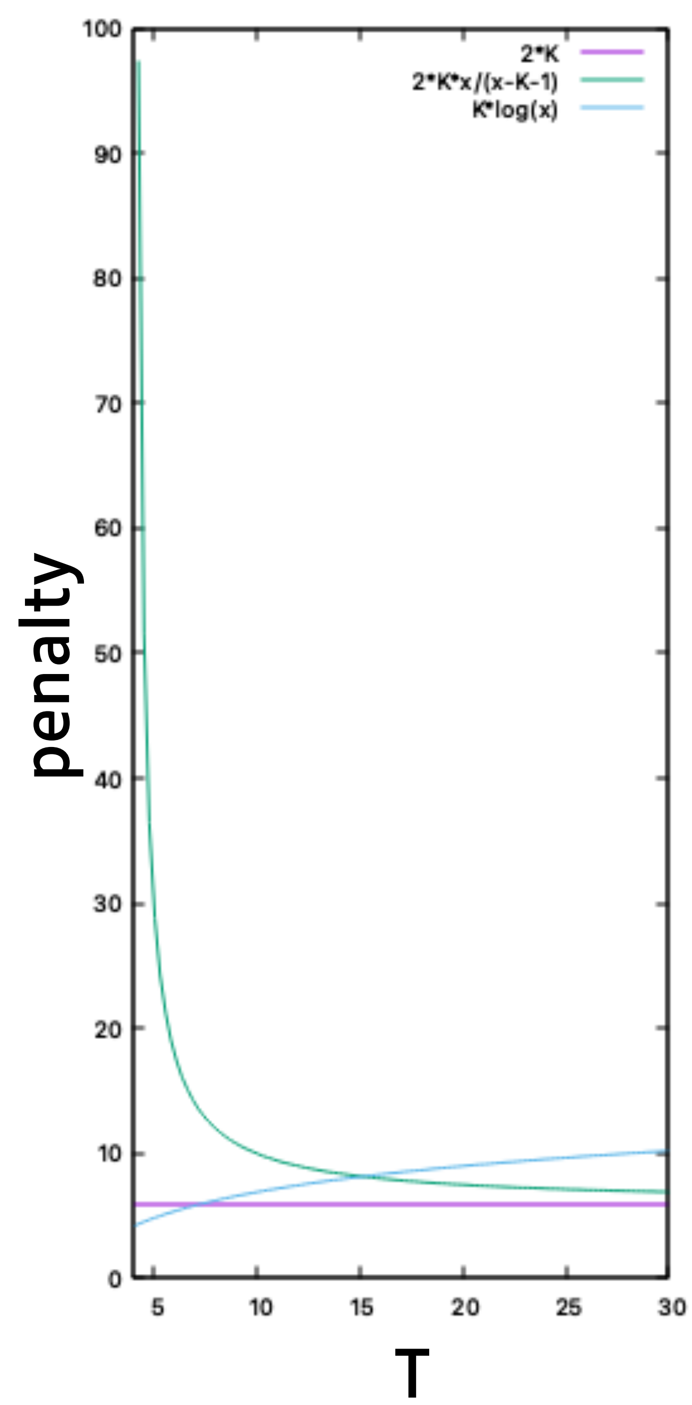
- **Akaike (1973) Information Criterion (AIC)**

$$AIC = -2\log\mathcal{L}(\hat{\theta}|y) + 2K$$

$$AIC_c = -2\log\mathcal{L}(\hat{\theta}|y) + 2K\frac{T}{T-K-1}$$

- **Bayesian Information Criterion (BIC)**

$$BIC = -2\log\mathcal{L}(\hat{\theta}|y) + K\log(T)$$



Even more classical decision-making models

« Building the set of candidate models is partially a subjective art »
Burnham & Anderson, 2002

- *2-arm bandit task.*
- Biased random response $p_t^1 = b \quad \text{and} \quad p_t^2 = 1 - b$ $\quad \theta_1 = b$
- Noisy win-stay lose-shift
$$p_t^k = \begin{cases} 1 - \epsilon/2 & \text{if } (c_{t-1} = k \text{ and } r_{t-1} = 1) \text{ OR } (c_{t-1} \neq k \text{ and } r_{t-1} = 0) \\ \epsilon/2 & \text{if } (c_{t-1} \neq k \text{ and } r_{t-1} = 1) \text{ OR } (c_{t-1} = k \text{ and } r_{t-1} = 0) \end{cases}$$
$$\theta_2 = \epsilon$$
- Rescorla-Wagner $Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k)$ $\quad p_t^k = \frac{\exp(\beta Q_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i)}$ $\quad \theta_3 = (\alpha, \beta)$
- Choice Kernel (behavioral persistence) $CK_{t+1}^k = CK_t^k + \alpha_c(a_t^k - CK_t^k)$ $\quad p_t^k = \frac{\exp(\beta_c CK_t^k)}{\sum_{i=1}^K \exp(\beta_c CK_t^i)}$ $\quad \theta_4 = (\alpha_c, \beta_c)$
- RW+CK
$$Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k)$$
$$CK_{t+1}^k = CK_t^k + \alpha_c(a_t^k - CK_t^k)$$
$$p_t^k = \frac{\exp(\beta Q_t^k + \beta_c CK_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i + \beta_c CK_t^i)}$$
$$\theta_5 = (\alpha, \beta, \alpha_c, \beta_c)$$

From (Wilson & Collins, 2019)

Even more classical decision-making models

« Building the set of candidate models is partially a subjective art »
Burnham & Anderson, 2002

- *2-arm bandit task.*

- Biased random response $p_t^1 = b \quad \text{and} \quad p_t^2 = 1 - b$ $\theta_1 = b$

- Noisy win-stay lose-shift
$$p_t^k = \begin{cases} 1 - \epsilon/2 & \text{if } (c_{t-1} = k \text{ and } r_{t-1} = 1) \text{ OR } (c_{t-1} \neq k \text{ and } r_{t-1} = 0) \\ \epsilon/2 & \text{if } (c_{t-1} \neq k \text{ and } r_{t-1} = 1) \text{ OR } (c_{t-1} = k \text{ and } r_{t-1} = 0) \end{cases}$$
$\theta_2 = \epsilon$

- Rescorla-Wagner $Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k)$ $p_t^k = \frac{\exp(\beta Q_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i)}$ $\theta_3 = (\alpha, \beta)$

- Choice Kernel (behavioral persistence) $CK_{t+1}^k = CK_t^k + \alpha_c(a_t^k - CK_t^k)$ $p_t^k = \frac{\exp(\beta_c CK_t^k)}{\sum_{i=1}^K \exp(\beta_c CK_t^i)}$ $\theta_4 = (\alpha_c, \beta_c)$

- RW+CK
$$Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k)$$
$$CK_{t+1}^k = CK_t^k + \alpha_c(a_t^k - CK_t^k)$$
$$p_t^k = \frac{\exp(\beta Q_t^k + \beta_c CK_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i + \beta_c CK_t^i)}$$
$\theta_5 = (\alpha, \beta, \alpha_c, \beta_c)$

From (Wilson & Collins, 2019)

Even more classical decision-making models

« Building the set of candidate models is partially a subjective art »
Burnham & Anderson, 2002

- *2-arm bandit task.*

- Biased random response
  $$p_t^1 = b \quad \text{and} \quad p_t^2 = 1 - b \qquad \theta_1 = b$$

- Noisy win-stay lose-shift
  $$p_t^k = \begin{cases} 1 - \epsilon/2 & \text{if } (c_{t-1} = k \text{ and } r_{t-1} = 1) \text{ OR } (c_{t-1} \neq k \text{ and } r_{t-1} = 0) \\ \epsilon/2 & \text{if } (c_{t-1} \neq k \text{ and } r_{t-1} = 1) \text{ OR } (c_{t-1} = k \text{ and } r_{t-1} = 0) \end{cases}$$
  $$\theta_2 = \epsilon$$

- Rescorla-Wagner
  $$Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k) \qquad p_t^k = \frac{\exp(\beta Q_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i)} \qquad \theta_3 = (\alpha, \beta)$$

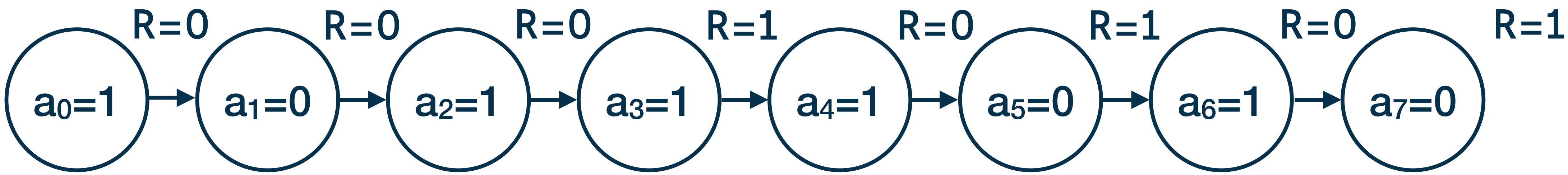
- Choice Kernel
  (behavioral persistence)
  $$CK_{t+1}^k = CK_t^k + \alpha_c(a_t^k - CK_t^k) \qquad p_t^k = \frac{\exp(\beta_c CK_t^k)}{\sum_{i=1}^K \exp(\beta_c CK_t^i)} \qquad \theta_4 = (\alpha_c, \beta_c)$$

- RW+CK
  $$Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k)$$
  $$CK_{t+1}^k = CK_t^k + \alpha_c(a_t^k - CK_t^k)$$
  $$p_t^k = \frac{\exp(\beta Q_t^k + \beta_c CK_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i + \beta_c CK_t^i)} \qquad \theta_5 = (\alpha, \beta, \alpha_c, \beta_c)$$

From (Wilson & Collins, 2019)

AIC maximization

Observations:



- **Rescorla-Wagner**

$$Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k)$$

$$p_t^k = \frac{\exp(\beta Q_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i)}$$

$$\theta_3 = (\alpha, \beta)$$

Model (θ_3=(α=0.3; β=1)):

AIC maximization

Observations:

$a_0=1 \xrightarrow{R=0} a_1=0 \xrightarrow{R=0} a_2=1 \xrightarrow{R=0} a_3=1 \xrightarrow{R=1} a_4=1 \xrightarrow{R=0} a_5=0 \xrightarrow{R=1} a_6=1 \xrightarrow{R=0} a_7=0 \quad R=1$

$$Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k) \qquad p_t^k = \frac{\exp(\beta Q_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i)} \qquad \theta_3 = (\alpha, \beta)$$

- ## Rescorla-Wagner

Model ($\theta_3=(\alpha=0.3;\ \beta=1)$):

$Q^0= 0$

$Q^1= 0$

$p^0= 0.5$

$p^1= 0.5$

$\mathcal{L}(\theta|a_i)= 0.5$

$\downarrow$

Log $\mathcal{L}(\theta|a_{0:T})=-0.7$

AIC maximization

Observations:

$a_0=1 \xrightarrow{R=0} a_1=0 \xrightarrow{R=0} a_2=1 \xrightarrow{R=0} a_3=1 \xrightarrow{R=1} a_4=1 \xrightarrow{R=0} a_5=0 \xrightarrow{R=1} a_6=1 \xrightarrow{R=0} a_7=0 \quad R=1$

$Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k)$ $p_t^k = \frac{\exp(\beta Q_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i)}$ $\theta_3 = (\alpha, \beta)$

- **Rescorla-Wagner**

Model ($\theta_3=(\alpha=0.3;\ \beta=1)$):

$Q^0 = 0$ 0

$Q^1 = 0$ **0**

$p^0 = 0.5$

$p^1 = 0.5$

$\mathcal{L}(\theta|a_i) = 0.5$

$\downarrow$

Log $\mathcal{L}(\theta|a_{0:T}) = -0.7$

AIC maximization

Observations:

$a_0=1 \xrightarrow{R=0} a_1=0 \xrightarrow{R=0} a_2=1 \xrightarrow{R=0} a_3=1 \xrightarrow{R=1} a_4=1 \xrightarrow{R=0} a_5=0 \xrightarrow{R=1} a_6=1 \xrightarrow{R=0} a_7=0$ R=1

$$Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k)$$

$$p_t^k = \frac{\exp(\beta Q_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i)}$$

$$\theta_3 = (\alpha, \beta)$$

- **Rescorla-Wagner**

  Model ($\theta_3=(\alpha=0.3;\ \beta=1)$):

$Q^0=$	0	0
$Q^1=$	0	**0**
$p^0=$	0.5	0.5
$p^1=$	0.5	0.5
$\mathcal{L}(\theta \mid a_i)=$	0.5	0.5

$\operatorname{Log} \mathcal{L}(\theta \mid a_{0:T})=-0.7$

AIC maximization

Observations:

R=0 R=0 R=0 R=1 R=0 R=1 R=0 R=1

$a_0=1 \rightarrow a_1=0 \rightarrow a_2=1 \rightarrow a_3=1 \rightarrow a_4=1 \rightarrow a_5=0 \rightarrow a_6=1 \rightarrow a_7=0$

$$Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k) \qquad p_t^k = \frac{\exp(\beta Q_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i)} \qquad \theta_3 = (\alpha, \beta)$$

- Rescorla-Wagner

Model ($\theta_3=(\alpha=0.3;\ \beta=1)$):

$Q^0=$	0	0
$Q^1=$	0	**0**
$p^0=$	0.5	0.5
$p^1=$	0.5	0.5
$\mathcal{L}(\theta \mid a_i)=$	0.5	0.5
Log $\mathcal{L}(\theta \mid a_{0:T})=$	-0.7	-1.4

AIC maximization

Observations:

$$a_0=1 \xrightarrow{R=0} a_1=0 \xrightarrow{R=0} a_2=1 \xrightarrow{R=0} a_3=1 \xrightarrow{R=1} a_4=1 \xrightarrow{R=0} a_5=0 \xrightarrow{R=1} a_6=1 \xrightarrow{R=0} a_7=0 \quad R=1$$

$$Q^k_{t+1} = Q^k_t + \alpha(r_t - Q^k_t) \qquad p^k_t = \frac{\exp(\beta Q^k_t)}{\sum_{i=1}^K \exp(\beta Q^i_t)} \qquad \theta_3 = (\alpha, \beta)$$

- **Rescorla-Wagner**

  Model ($\theta_3=(\alpha=0.3;\ \beta=1)$):

$Q^0=$	0	0	**0**
$Q^1=$	0	**0**	0
$p^0=$	0.5	0.5	0.5
$p^1=$	0.5	0.5	0.5
$\mathcal{L}(\theta\|a_i)=$	0.5	0.5	0.5
$\text{Log}\ \mathcal{L}(\theta\|a_{0:T})=$	-0.7	-1.4	-2.1

AIC maximization

Observations:

$$a_0=1 \xrightarrow{R=0} a_1=0 \xrightarrow{R=0} a_2=1 \xrightarrow{R=0} a_3=1 \xrightarrow{R=1} a_4=1 \xrightarrow{R=0} a_5=0 \xrightarrow{R=1} a_6=1 \xrightarrow{R=0} a_7=0 \quad R=1$$

- **Rescorla-Wagner**

$$Q^k_{t+1} = Q^k_t + \alpha(r_t - Q^k_t) \qquad p^k_t = \frac{\exp(\beta Q^k_t)}{\sum_{i=1}^{K} \exp(\beta Q^i_t)} \qquad \theta_3 = (\alpha, \beta)$$

Model ($\theta_3=(\alpha=0.3;\ \beta=1)$):

$Q^0=$	0	0	**0**	0	0	0	**0.3**	0.3
$Q^1=$	0	**0**	0	**0**	**0.3**	**0.21**	0.21	**0.15**
$p^0=$	0.5	0.5	0.5	0.5	0.43	0.45	0.52	0.54
$p^1=$	0.5	0.5	0.5	0.5	0.57	0.55	0.48	0.46
$\mathcal{L}(\theta\mid a_i)=$	0.5	0.5	0.5	0.5	0.57	0.45	0.48	0.54
$\log \mathcal{L}(\theta\mid a_{0:T})=$	-0.7	-1.4	-2.1	-2.8	-3.3	-4.1	-4.9	-5.5

AIC maximization

Observations:

$a_0=1$	$a_1=0$	$a_2=1$	$a_3=1$	$a_4=1$	$a_5=0$	$a_6=1$	$a_7=0$
R=0	R=0	R=0	R=1	R=0	R=1	R=0	R=1

- Rescorla-Wagner

$$Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k)$$

$$p_t^k = \frac{\exp(\beta Q_t^k)}{\sum_{i=1}^{K} \exp(\beta Q_t^i)}$$

$$\theta_3 = (\alpha, \beta)$$

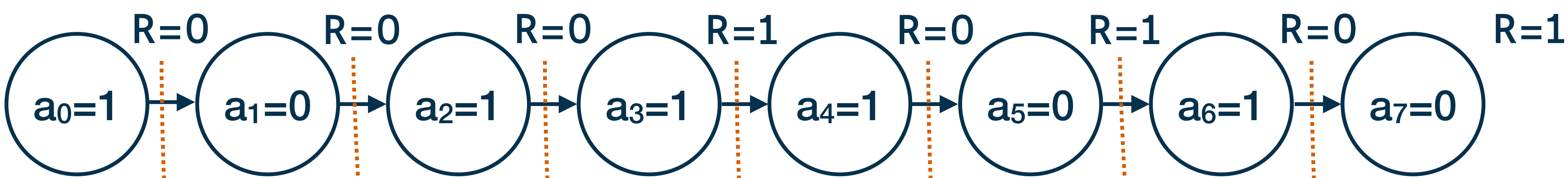
Model (θ_3=(α=0.3; β=1)):

Q^0=	0	0	**0**	0	0	0	**0.3**	0.3
Q^1=	0	**0**	0	**0**	**0.3**	**0.21**	0.21	**0.15**
p^0=	0.5	0.5	0.5	0.5	0.43	0.45	0.52	0.54
p^1=	0.5	0.5	0.5	0.5	0.57	0.55	0.48	0.46
$\mathcal{L}(\theta\|a_i)$=	0.5	0.5	0.5	0.5	0.57	0.45	0.48	0.54
Log $\mathcal{L}(\theta\|a_{0:T})$=	-0.7	-1.4	-2.1	-2.8	-3.3	-4.1	-4.9	-5.5

AIC= 15.0 ; AICc = 17.4 ; BIC = 15.1

AIC maximization

Observations:



- **Rescorla-Wagner**

$$Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k) \qquad p_t^k = \frac{\exp(\beta Q_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i)} \qquad \theta_3 = (\alpha, \beta)$$

Model (θ_3=(α=0.5; β=4)):

Q^0=	0	0	**0**	0	0	0	**0.5**	0.5
Q^1=	0	**0**	0	**0**	**0.5**	**0.25**	0.25	**0.13**
p^0=	0.5	0.5	0.5	0.5	0.18	0.27	0.73	0.81
p^1=	0.5	0.5	0.5	0.5	0.88	0.73	0.27	0.19
$\mathcal{L}(\theta\|a_i)$=	0.5	0.5	0.5	0.5	0.88	0.27	0.27	0.81
Log $\mathcal{L}(\theta\|a_{0:T})$=	-0.7	-1.4	-2.1	-2.8	-2.9	-4.2	-5.5	-5.7

AIC maximization

Observations:

	$a_0=1$	$a_1=0$	$a_2=1$	$a_3=1$	$a_4=1$	$a_5=0$	$a_6=1$	$a_7=0$
	R=0	R=0	R=0	R=1	R=0	R=1	R=0	R=1

- Rescorla-Wagner

$$Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k)$$

$$p_t^k = \frac{\exp(\beta Q_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i)}$$

$$\theta_3 = (\alpha, \beta)$$

Model (θ_3=(α=0.5; β=4)):

Q^0=	0	0	**0**	0	0	0	**0.5**	0.5
Q^1=	0	**0**	0	**0**	**0.5**	**0.25**	0.25	**0.13**
p^0=	0.5	0.5	0.5	0.5	0.18	0.27	0.73	0.81
p^1=	0.5	0.5	0.5	0.5	0.88	0.73	0.27	0.19
$\mathcal{L}(\theta\|a_i)$=	0.5	0.5	0.5	0.5	0.88	0.27	0.27	0.81
Log $\mathcal{L}(\theta\|a_{0:T})$=	-0.7	-1.4	-2.1	-2.8	-2.9	-4.2	-5.5	-5.7

AIC= 15.5 ; AICc = 17.9 ; BIC = 15.6

AIC maximization

Observations:

$$\underset{R=0}{a_0=1} \rightarrow \underset{R=0}{a_1=0} \rightarrow \underset{R=0}{a_2=1} \rightarrow \underset{R=1}{a_3=1} \rightarrow \underset{R=0}{a_4=1} \rightarrow \underset{R=1}{a_5=0} \rightarrow \underset{R=0}{a_6=1} \rightarrow \underset{R=1}{a_7=0}$$

- **Choice Kernel**

$$CK_{t+1}^k = CK_t^k + \alpha_c(a_t^k - CK_t^k)$$

$$p_t^k = \frac{\exp(\beta_c CK_t^k)}{\sum_{i=1}^K \exp(\beta_c CK_t^i)}$$

$$\theta_4 = (\alpha_c, \beta_c)$$

Model ($\theta_4=(\alpha_c=0.3;\ \beta_c=2)$):

$CK^0 = 0$

$CK^1 = 0$

$p^0 = 0.5$

$p^1 = 0.5$

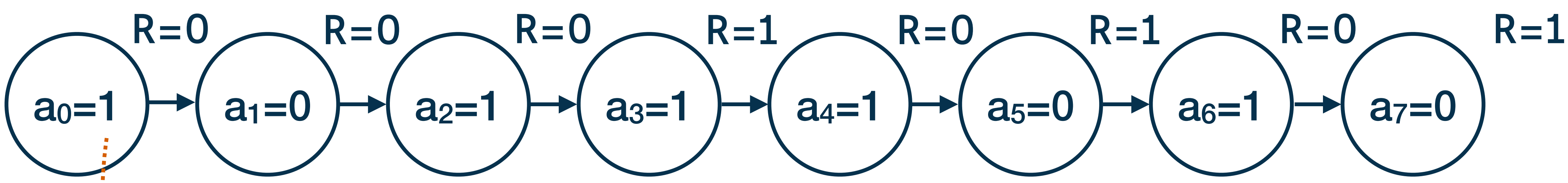
$\mathcal{L}(\theta|a_i) = 0.5$

$\downarrow$

$\text{Log } \mathcal{L}(\theta|a_{0:T}) = -0.7$

AIC maximization

Observations:



- ## Choice Kernel

$CK_{t+1}^{k} = CK_{t}^{k} + \alpha_c (a_t^k - CK_t^k)$ $\quad p_t^k = \frac{\exp(\beta_c CK_t^k)}{\sum_{i=1}^{K} \exp(\beta_c CK_t^i)}$ $\quad \theta_4 = (\alpha_c, \beta_c)$

Model (θ_4=(α_c=0.3; β_c=2)):

CK^0= 0   0

CK^1= 0   **0.3**

p^0= 0.5

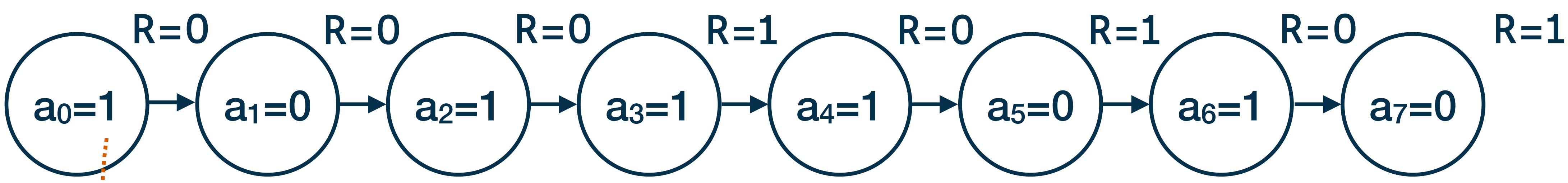
p^1= 0.5

$\mathcal{L}(\theta|a_i)$= 0.5

Log $\mathcal{L}(\theta|a_{0:T})$=-0.7

AIC maximization

Observations:



- **Choice Kernel**

$CK_{t+1}^k = CK_t^k + \alpha_c(a_t^k - CK_t^k)$ $\quad p_t^k = \frac{\exp(\beta_c CK_t^k)}{\sum_{i=1}^K \exp(\beta_c CK_t^i)}$ $\quad \theta_4 = (\alpha_c, \beta_c)$

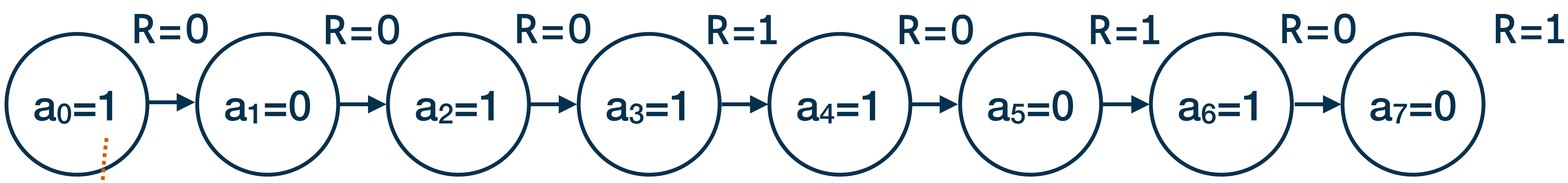
Model (θ_4=(α_c=0.3; β_c=2)):

CK^0=	0	0
CK^1=	0	**0.3**
p^0=	0.5	0.35
p^1=	0.5	0.65
$\mathcal{L}(\theta\|a_i)$=	0.5	0.45

Log $\mathcal{L}(\theta|a_{0:T})$=-0.7

AIC maximization

Observations:



- Choice Kernel

$$CK^k_{t+1} = CK^k_t + \alpha_c(a^k_t - CK^k_t)$$

$$p^k_t = \frac{\exp(\beta_c CK^k_t)}{\sum_{i=1}^K \exp(\beta_c CK^i_t)}$$

$$\theta_4 = (\alpha_c, \beta_c)$$

Model (θ_4=(α_c=0.3; β_c=2)):

CK^0=	0	0
CK^1=	0	**0.3**
p^0=	0.5	0.35
p^1=	0.5	0.65
$\mathcal{L}(\theta\|a_i)$=	0.5	0.45
Log $\mathcal{L}(\theta\|a_{0:T})$=	-0.7	-1.7

AIC maximization

Observations:

$a_0=1 \xrightarrow{R=0} a_1=0 \xrightarrow{R=0} a_2=1 \xrightarrow{R=0} a_3=1 \xrightarrow{R=1} a_4=1 \xrightarrow{R=0} a_5=0 \xrightarrow{R=1} a_6=1 \xrightarrow{R=0} a_7=0$ R=1

- Choice Kernel

$$CK_{t+1}^k = CK_t^k + \alpha_c(a_t^k - CK_t^k)$$

$$p_t^k = \frac{\exp(\beta_c CK_t^k)}{\sum_{i=1}^K \exp(\beta_c CK_t^i)}$$

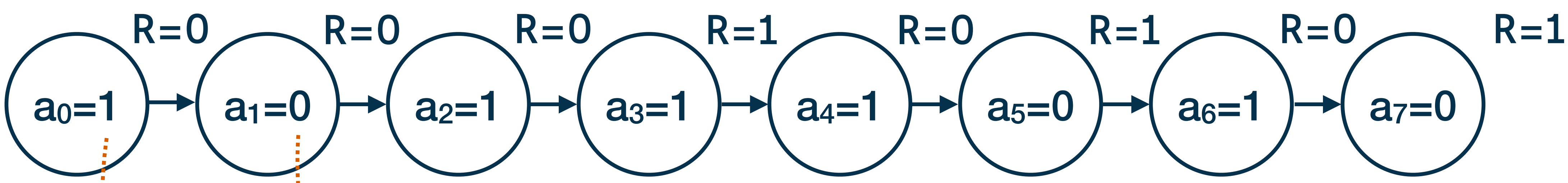
$$\theta_4 = (\alpha_c, \beta_c)$$

Model ($\theta_4=(\alpha_c=0.3;\ \beta_c=2)$):

$CK^0=$	0	0	**0.3**
$CK^1=$	0	**0.3**	**0.21**
$p^0=$	0.5	0.35	
$p^1=$	0.5	0.65	
$\mathcal{L}(\theta\|a_i)=$	0.5	0.45	
Log $\mathcal{L}(\theta\|a_{0:T})=$	-0.7	-1.7	

AIC maximization

Observations:



- **Choice Kernel**

$$CK^k_{t+1} = CK^k_t + \alpha_c(a^k_t - CK^k_t)$$

$$p^k_t = \frac{\exp(\beta_c CK^k_t)}{\sum_{i=1}^{K} \exp(\beta_c CK^i_t)}$$

$$\theta_4 = (\alpha_c, \beta_c)$$

Model ($\theta_4 = (\alpha_c = 0.3;\ \beta_c = 2)$):

$CK^0 =$	0	0	**0.3**
$CK^1 =$	0	**0.3**	**0.21**
$p^0 =$	0.5	0.35	0.54
$p^1 =$	0.5	0.65	0.46
$\mathcal{L}(\theta \mid a_i) =$	0.5	0.45	0.46
$\text{Log } \mathcal{L}(\theta \mid a_{0:T}) =$	-0.7	-1.7	

AIC maximization

Observations:

$a_0=1$	$a_1=0$	$a_2=1$	$a_3=1$	$a_4=1$	$a_5=0$	$a_6=1$	$a_7=0$
R=0	R=0	R=0	R=1	R=0	R=1	R=0	R=1

- Choice Kernel

$$CK_{t+1}^k = CK_t^k + \alpha_c(a_t^k - CK_t^k)$$

$$p_t^k = \frac{\exp(\beta_c CK_t^k)}{\sum_{i=1}^K \exp(\beta_c CK_t^i)}$$

$$\theta_4 = (\alpha_c, \beta_c)$$

Model ($\theta_4=(\alpha_c=0.3;\ \beta_c=2)$):

$CK^0=$	0	0	**0.3**
$CK^1=$	0	**0.3**	**0.21**
$p^0=$	0.5	0.35	0.54
$p^1=$	0.5	0.65	0.46
$\mathcal{L}(\theta\|a_i)=$	0.5	0.45	0.46
Log $\mathcal{L}(\theta\|a_{0:T})=$	-0.7	-1.7	-2.5

AIC maximization

Observations:

$a_0=1$	$a_1=0$	$a_2=1$	$a_3=1$	$a_4=1$	$a_5=0$	$a_6=1$	$a_7=0$
R=0	R=0	R=0	R=1	R=0	R=1	R=0	R=1

- **Choice Kernel**

$CK_{t+1}^{k} = CK_{t}^{k} + \alpha_c (a_t^k - CK_t^k)$

$p_t^k = \frac{\exp(\beta_c CK_t^k)}{\sum_{i=1}^{K} \exp(\beta_c CK_t^i)}$

$\theta_4 = (\alpha_c, \beta_c)$

Model ($\theta_4=(\alpha_c=0.3;\ \beta_c=2)$):

$CK^0=$ 0	0	**0.3**	**0.21**	**0.15**	**0.10**	**0.37**	**0.26**
$CK^1=$ 0	**0.3**	**0.21**	**0.45**	**0.61**	**0.73**	**0.51**	**0.66**
$p^0=$ 0.5	0.35	0.54	0.38	0.28	0.22	0.43	0.31
$p^1=$ 0.5	0.65	0.46	0.62	0.72	0.78	0.57	0.69
$\mathcal{L}(\theta \mid a_i)=$ 0.5	0.45	0.46	0.62	0.72	0.22	0.57	0.31
Log $\mathcal{L}(\theta \mid a_{0:T})=$-0.7	-1.7	-2.5	-3.0	-3.3	-4.8	-5.4	-6.6

AIC= 17.1 ; AICc = 19.5 ; BIC = 17.3

Log-likelihood maximization

- Biased random response :
  $\theta_1=0.4$
  Log $\mathcal{L}(\theta|a_{0:T},g_1)=-5.3$
  AIC = 12.6
  AICc= 13.3
  BIC = 12.7

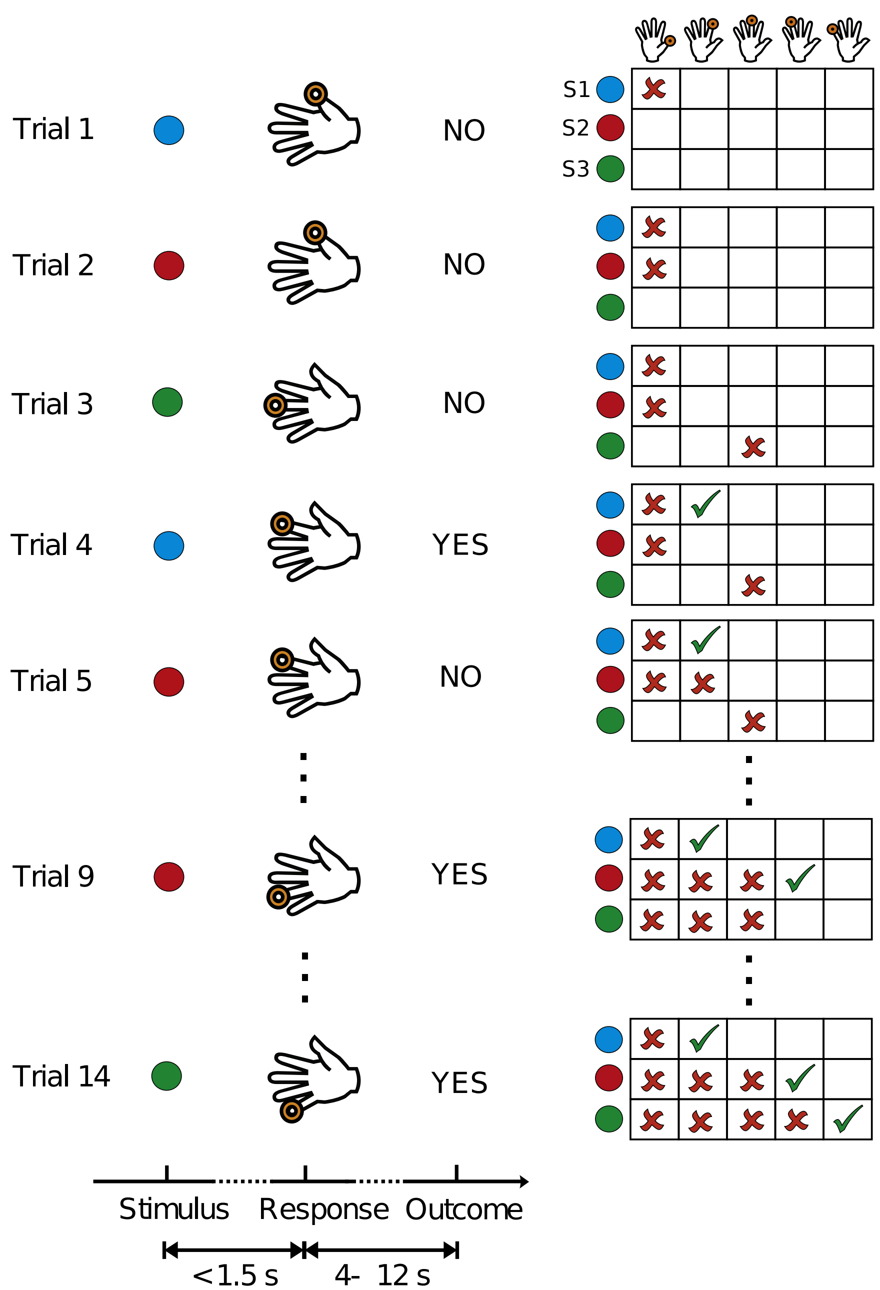
- Win-stay lose-shift:
  $\theta_2=0.6$
  Log $\mathcal{L}(\theta2|a_{0:T},g_2)=-4.88$
  AIC = 11.8
  AICc= 12.4
  BIC = 11.8

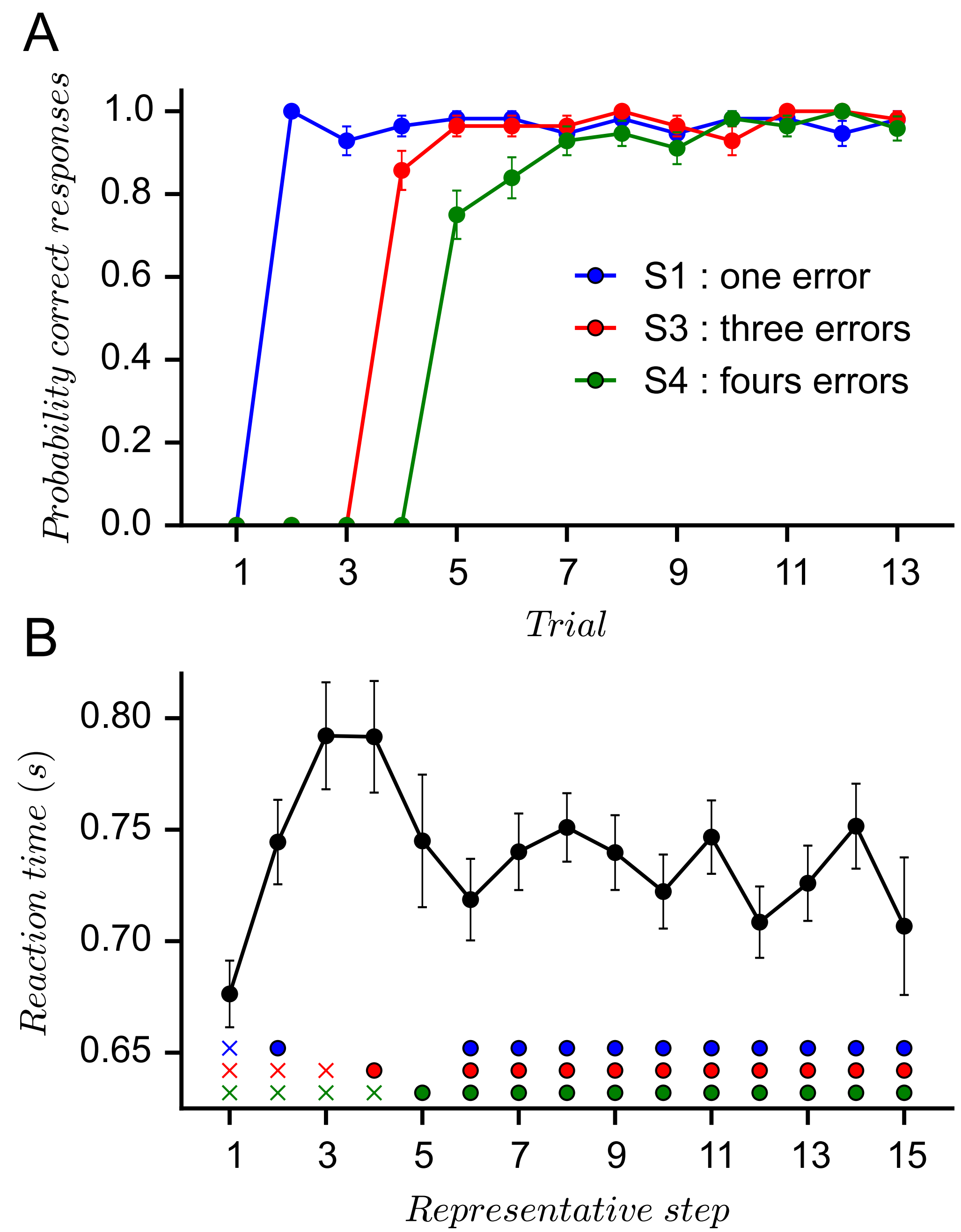
- Rescorla-Wagner:
  $\theta_3=(\alpha=0.4;\ \beta=2)$
  Log $\mathcal{L}(\theta|a_{0:T},g_3)=-5.44$
  AIC = 14.88
  AICc= 17.28
  BIC = 15.0

- Kernel Choice:
  $\theta_4=(\alpha=10^{-6};\ \beta=10^{-3})$
  Log $\mathcal{L}(\theta|a_{0:T},g_4)=-5.54$
  AIC = 15
  AICc= 17.5
  BIC = 15.2

Example of multiple model selection with multiple fitting objectives.

Task (Brovelli et al., 2011)





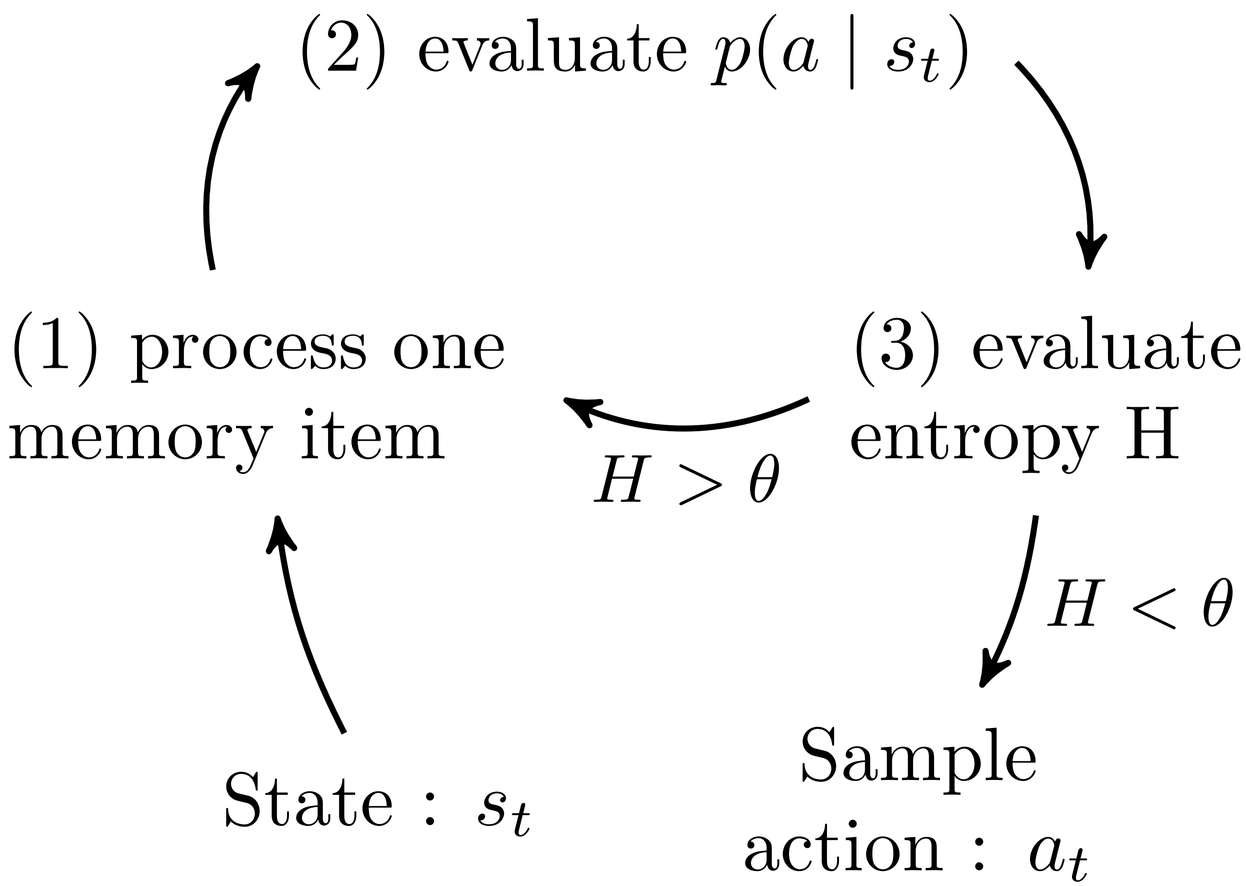
Formal description

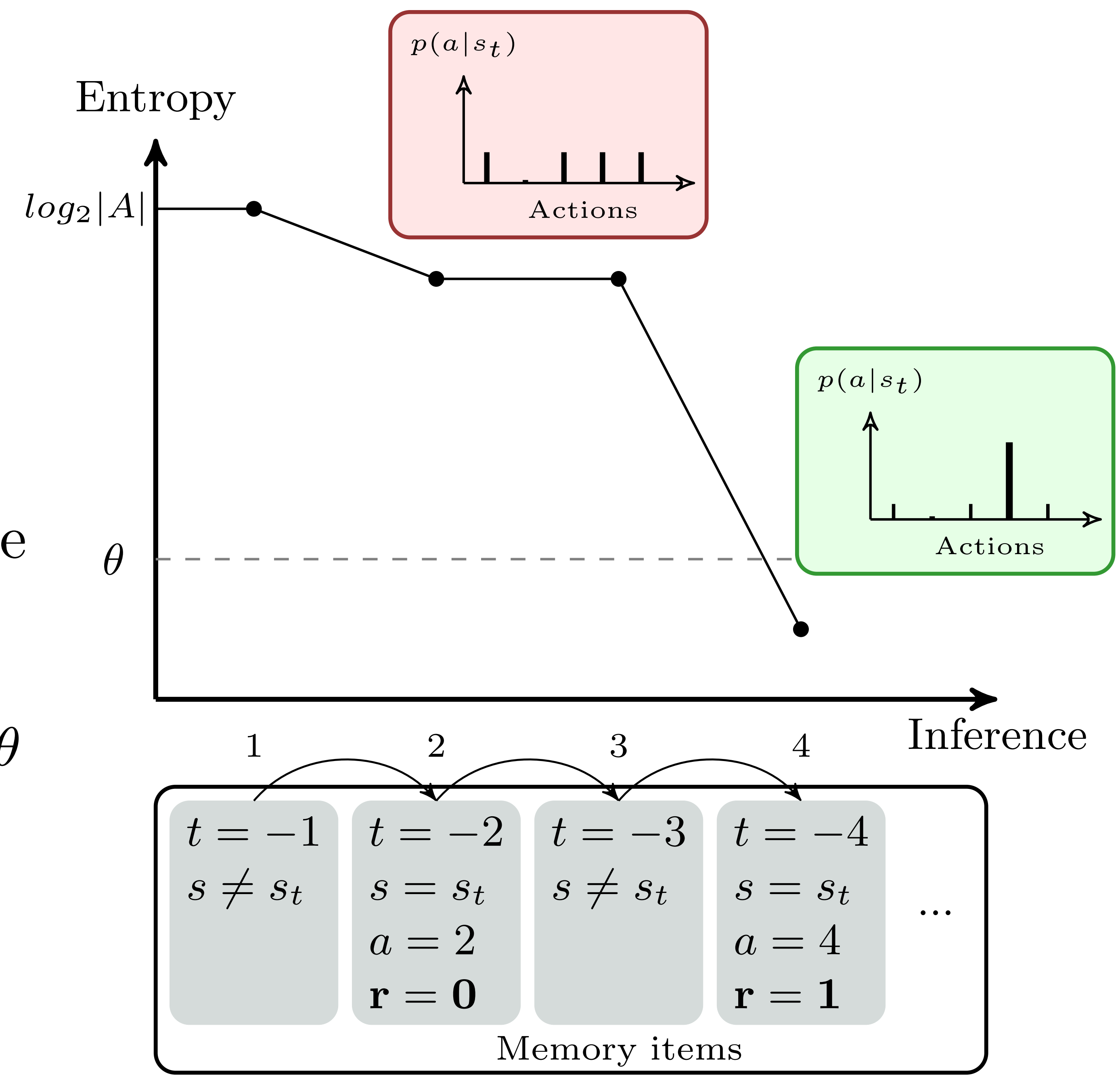
- States = {blue, red, green}
- Actions = {thumb, index, middle, ring, pinky}
- Reward = {0,1}
- Association learning: Q-Learning
- Working memory:
  proposal of a Bayesian Working Memory model

(Viejo et al, 2015, *Front. Behav. Neurosci.*)

Bayesian Working memory

Decision process of BWM:





Simulated reaction times

$$sRT(trial) = (log_2(i + 1))^{\sigma} + H(p(a|s_t))$$

Number of BWM iterations

Think stochastic race models

Adjustable parameter

Coordination methods

- (Keramati et al., 2011): speed accuracy tradeoff

$$Q(s_t, a) = \begin{cases} Q(s_t, a)^{MTB} & \text{if } VPI(s_t, a) > \tau \bar{R}(s_t) \\ Q(s_t, a)^{QL} & \text{if } VPI(s_t, a) < \tau \bar{R}(s_t) \end{cases}$$

- (Collins & Franck, 2012): weighted sum

$$p(a|s_t) = (1 - w(t, s_t))p(a|s_t)^{QL} + w(t, s_t)p(a|s_t)^{MTB}$$

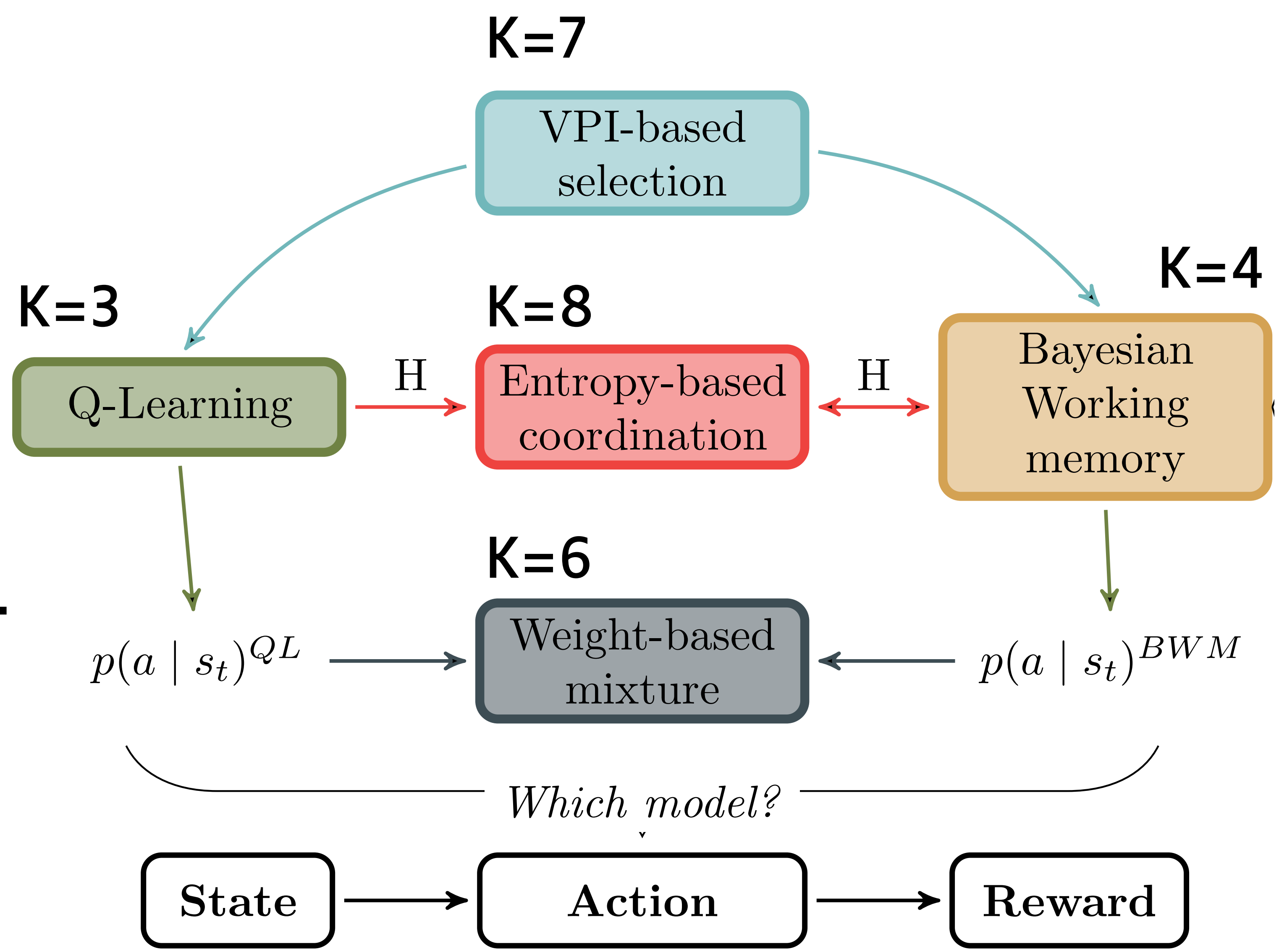
- Entropy coordination:

$$p(Decider|t_{0\rightarrow i}, H^{MTB}, H^{QL}) = \frac{1}{1+\lambda(n-i)\exp^{-\lambda(2H^{max} - H^{MTB}_{0\rightarrow i} - H^{QL})}}$$

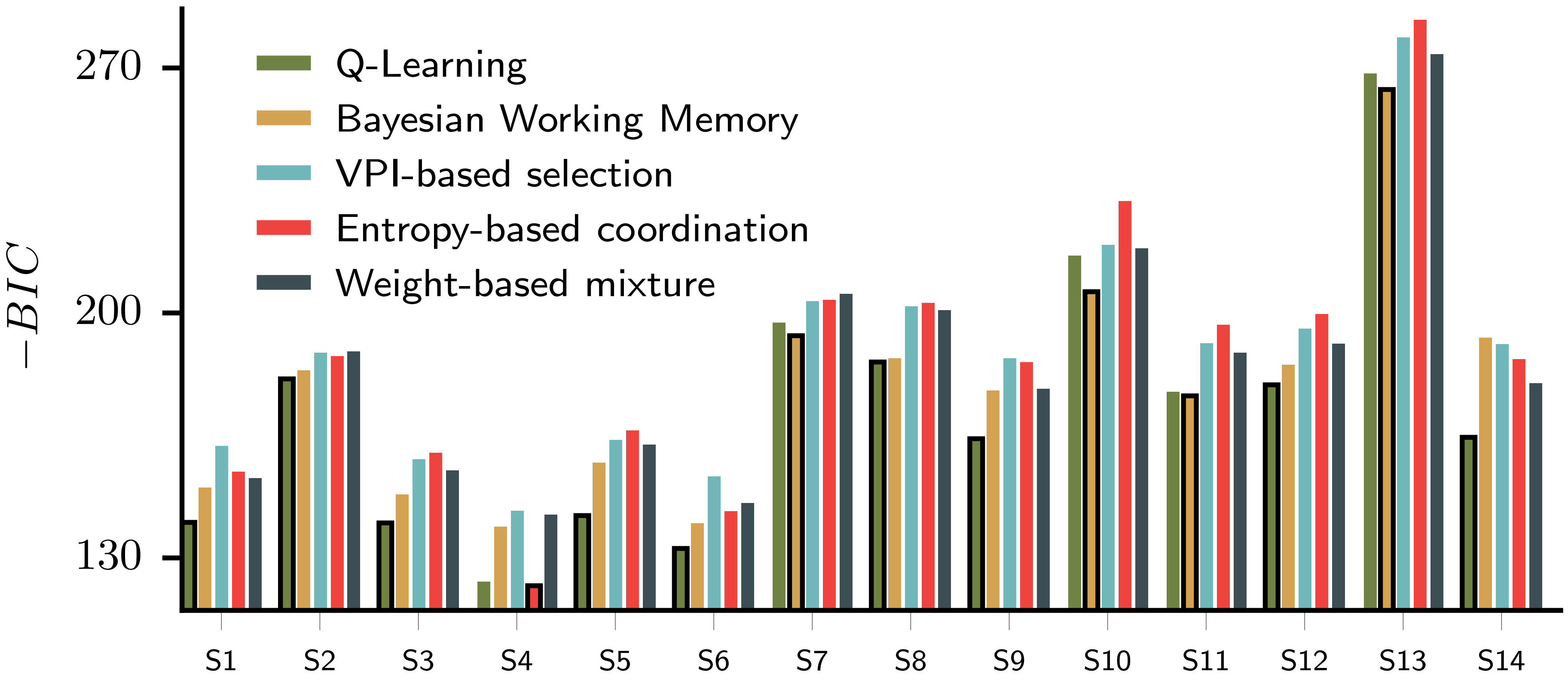
$$Q(s_t, a) = Q(s_t, a)^{MTB}_{0\rightarrow i} + Q(s_t, a)^{QL}$$

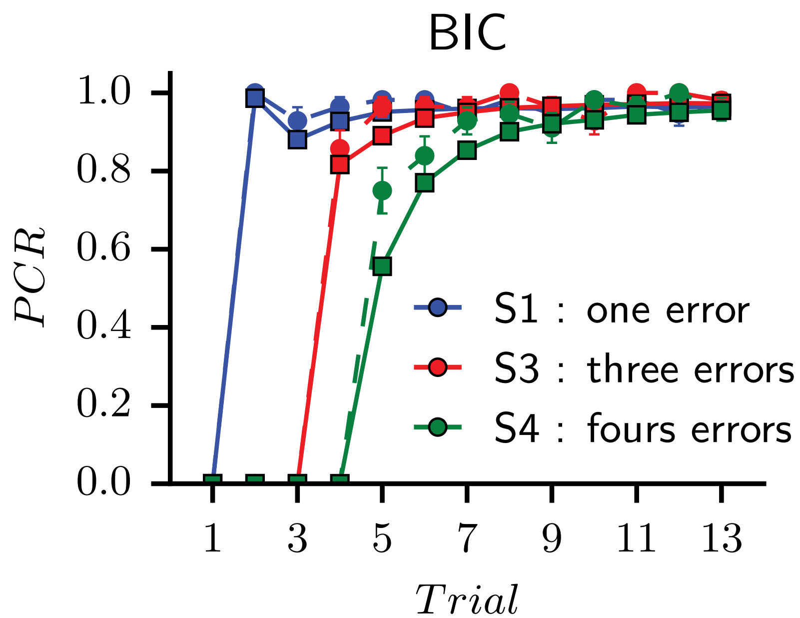
Individual optimizations

- Criteria:
  - LL of the choices
  - MSE of sRT vs. RT



Model selection: decision only

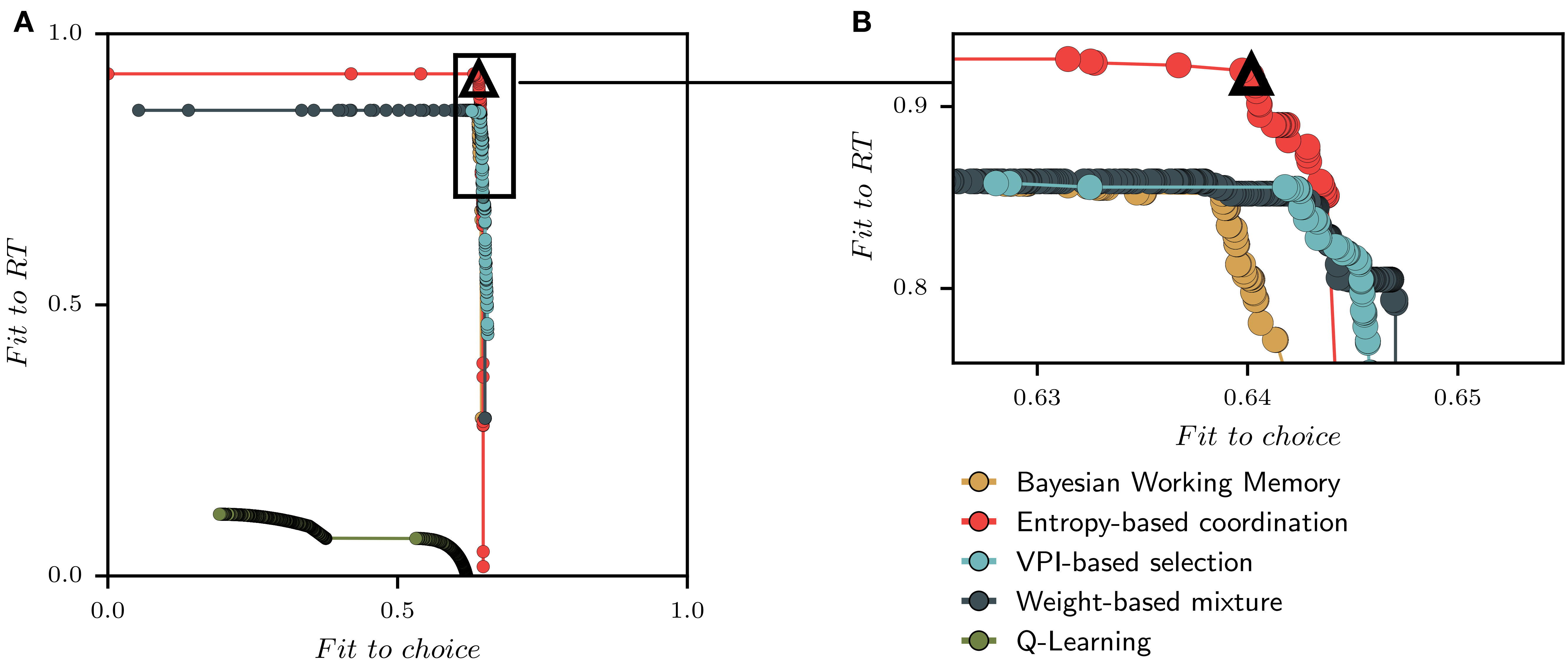




9 QL ; 4 BWM ; 1 Hcoord

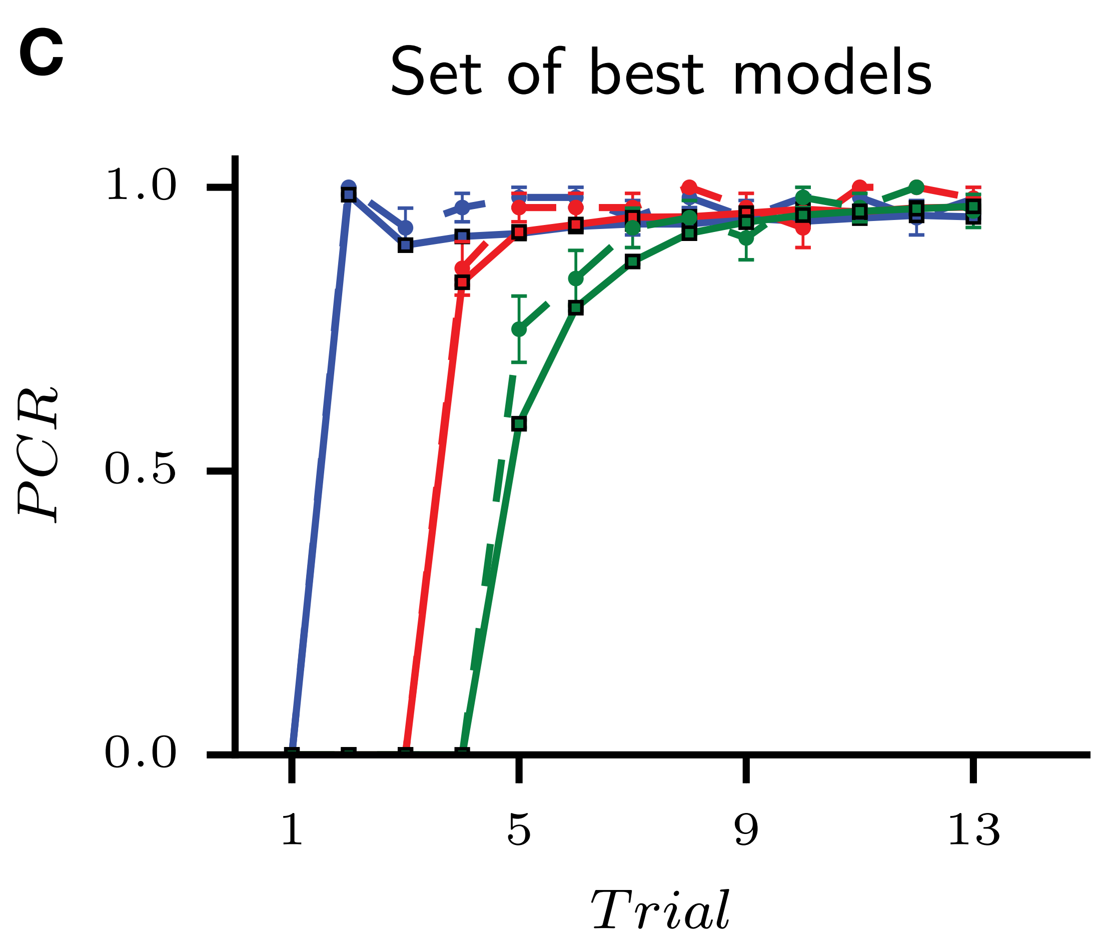
Free wheel simulations

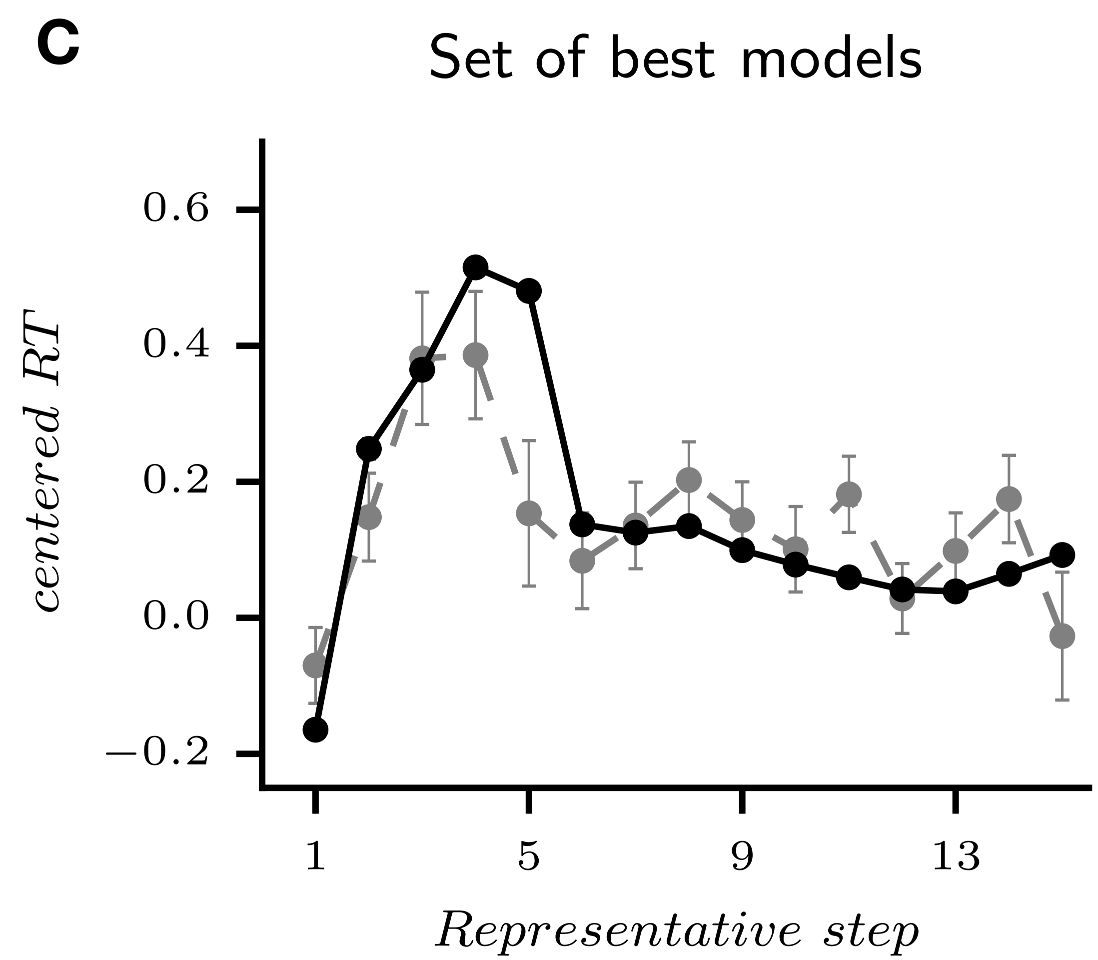
Model selection: decision & RT



9 Hcoord ; 3 wMix ; 2 VPI

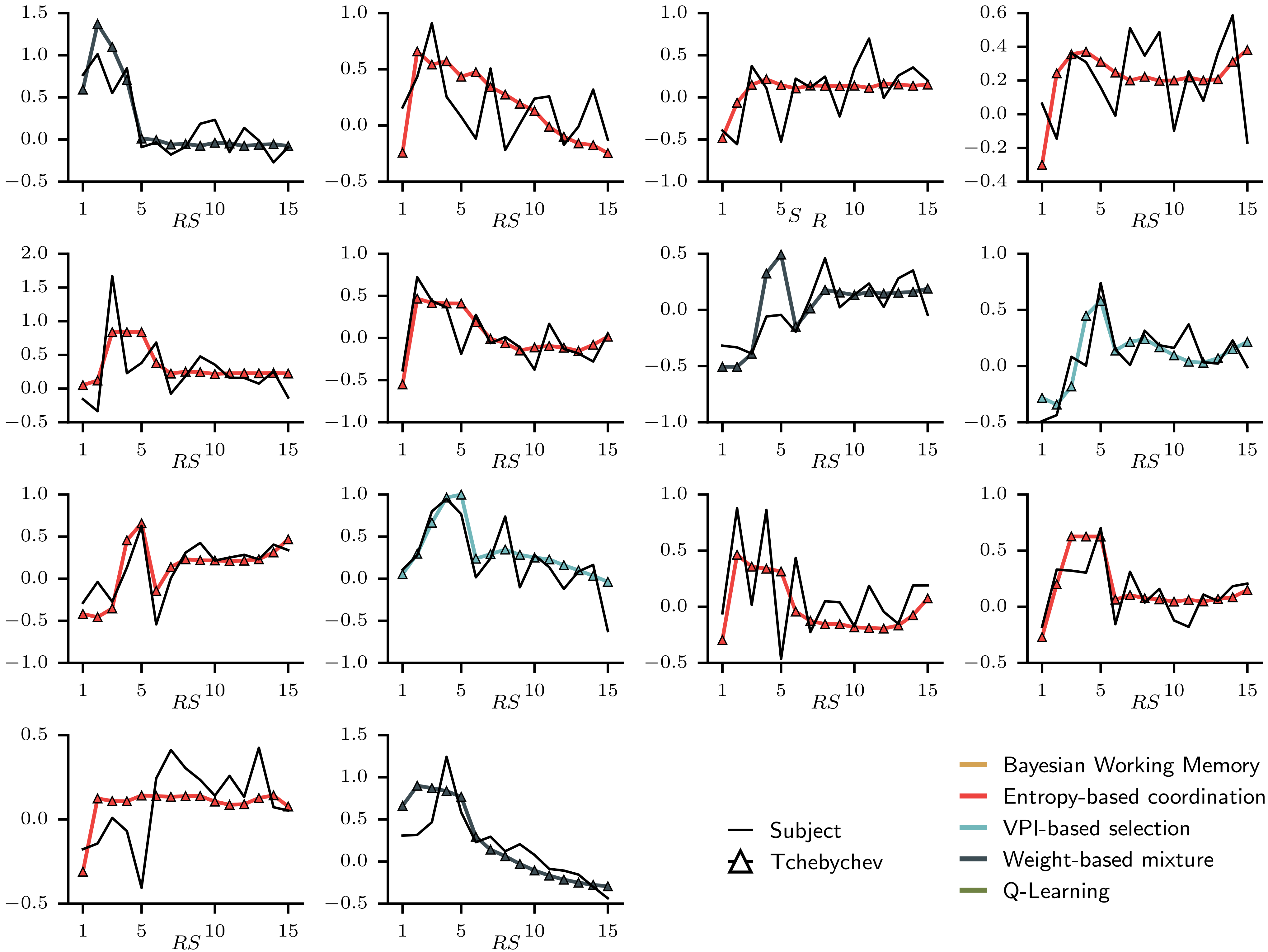
Model falsification: « Free wheel » simulations



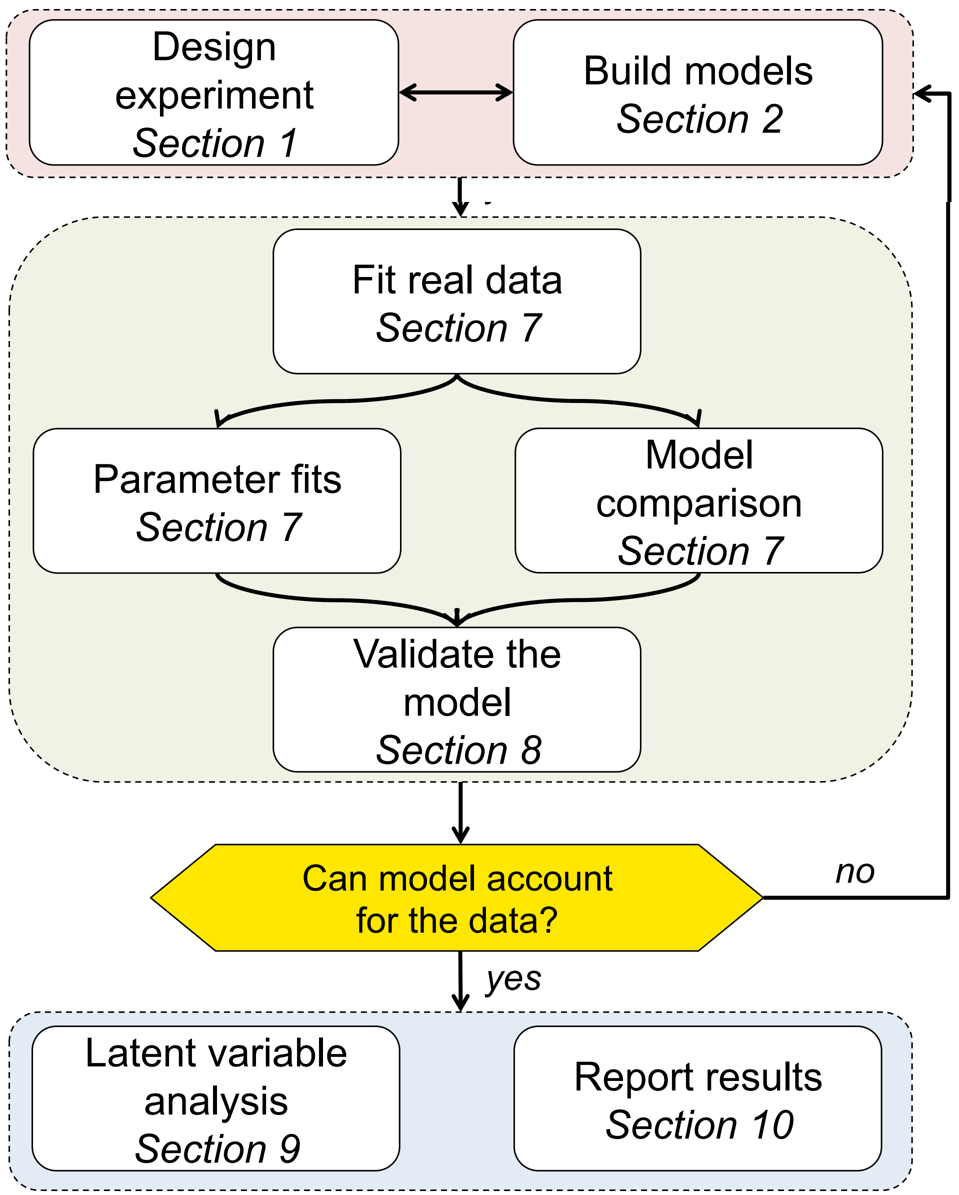


- **14 subjects: 9 Entropy, 3 weighted sums, 2 VPI**

Model falsification:
« Free wheel » simulations

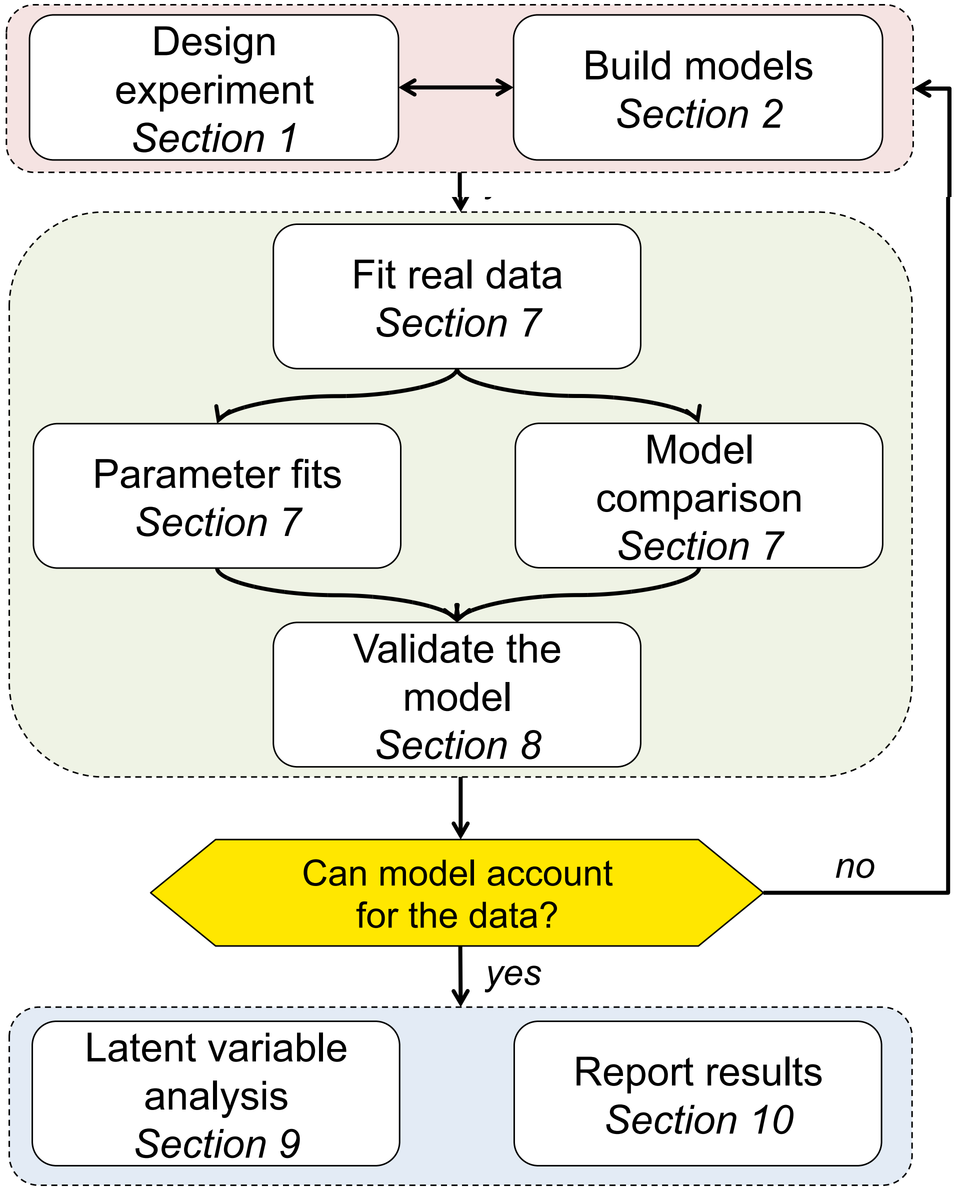


Modeling Process



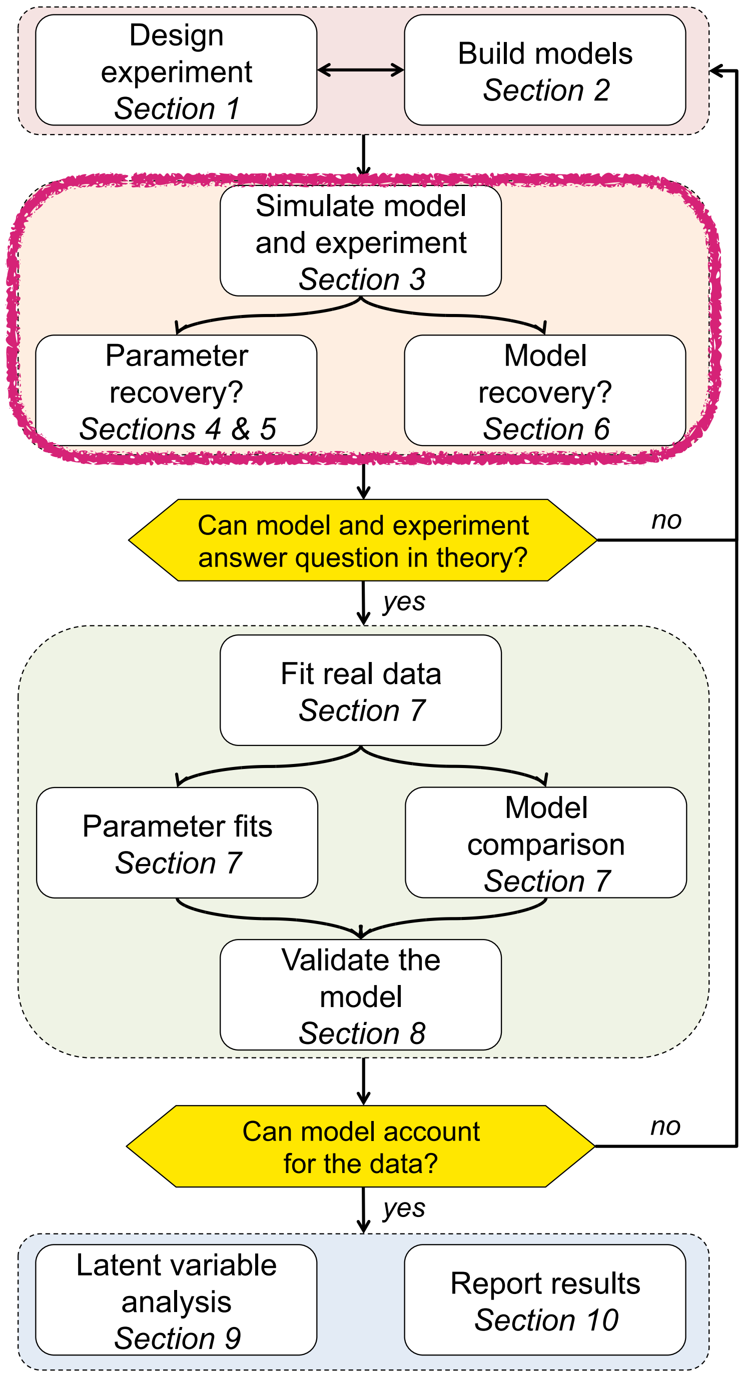
Adapted from (Wilson & Collins, 2019)

Modeling Process



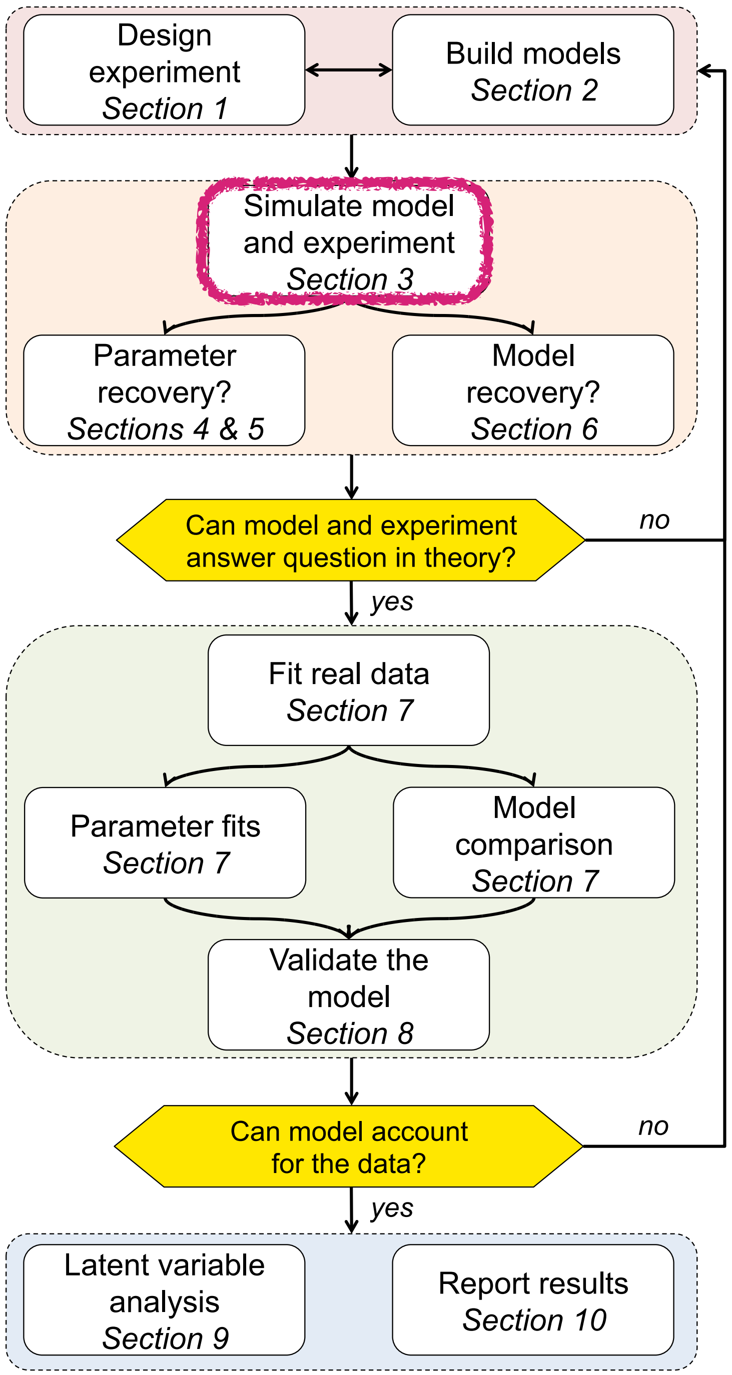
Adapted from (Wilson & Collins, 2019)

Modeling Process



~~Adapted~~ from (Wilson & Collins, 2019)

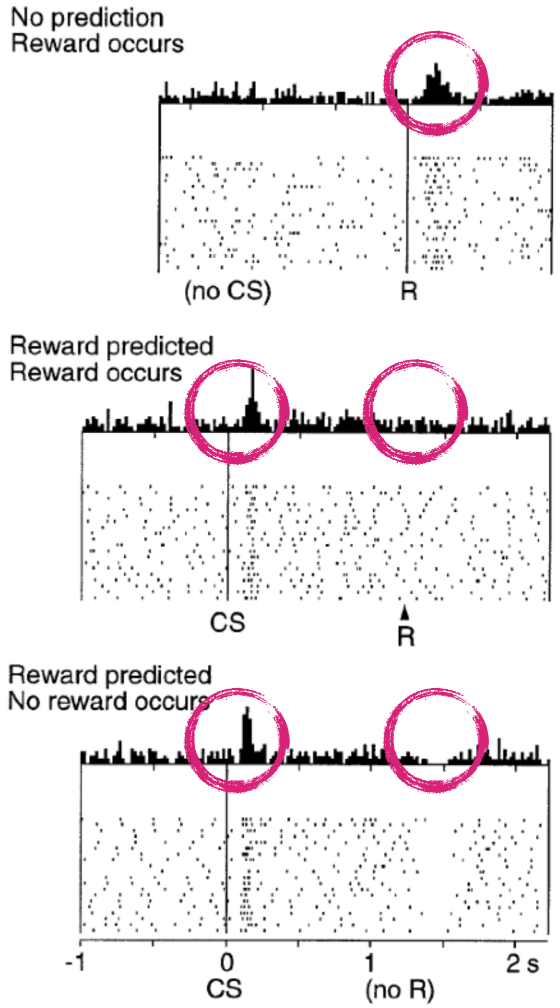
Modeling Process



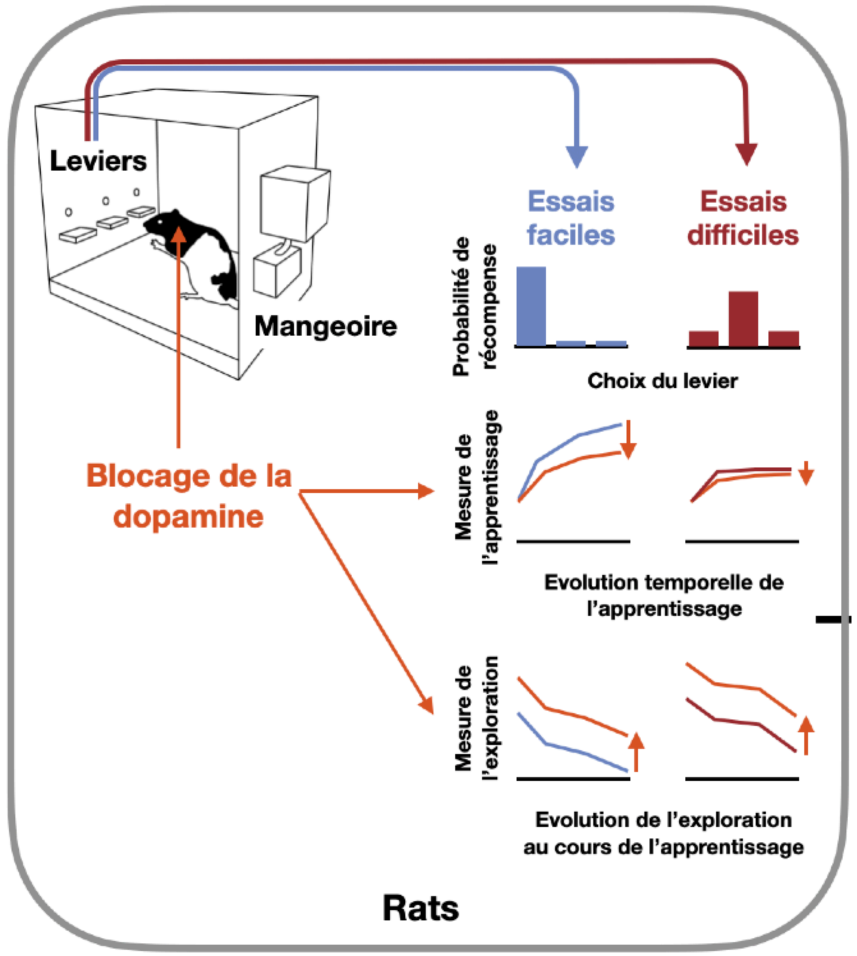
~~Adapted~~ from (Wilson & Collins, 2019)

Model simulations

Dopamine & Exploration



(Schultz et al., 1997)

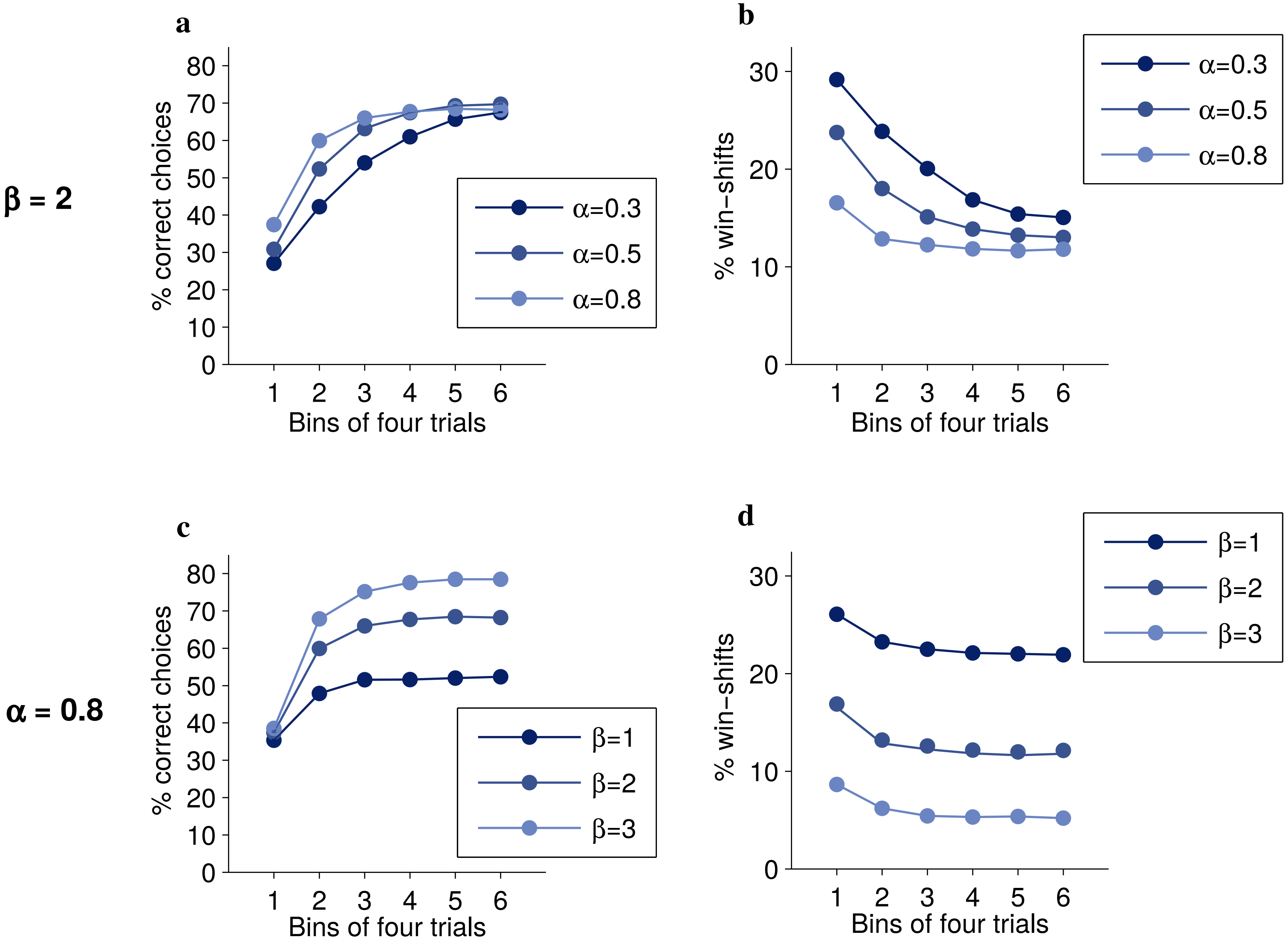


(Cinotti et al., 2019)

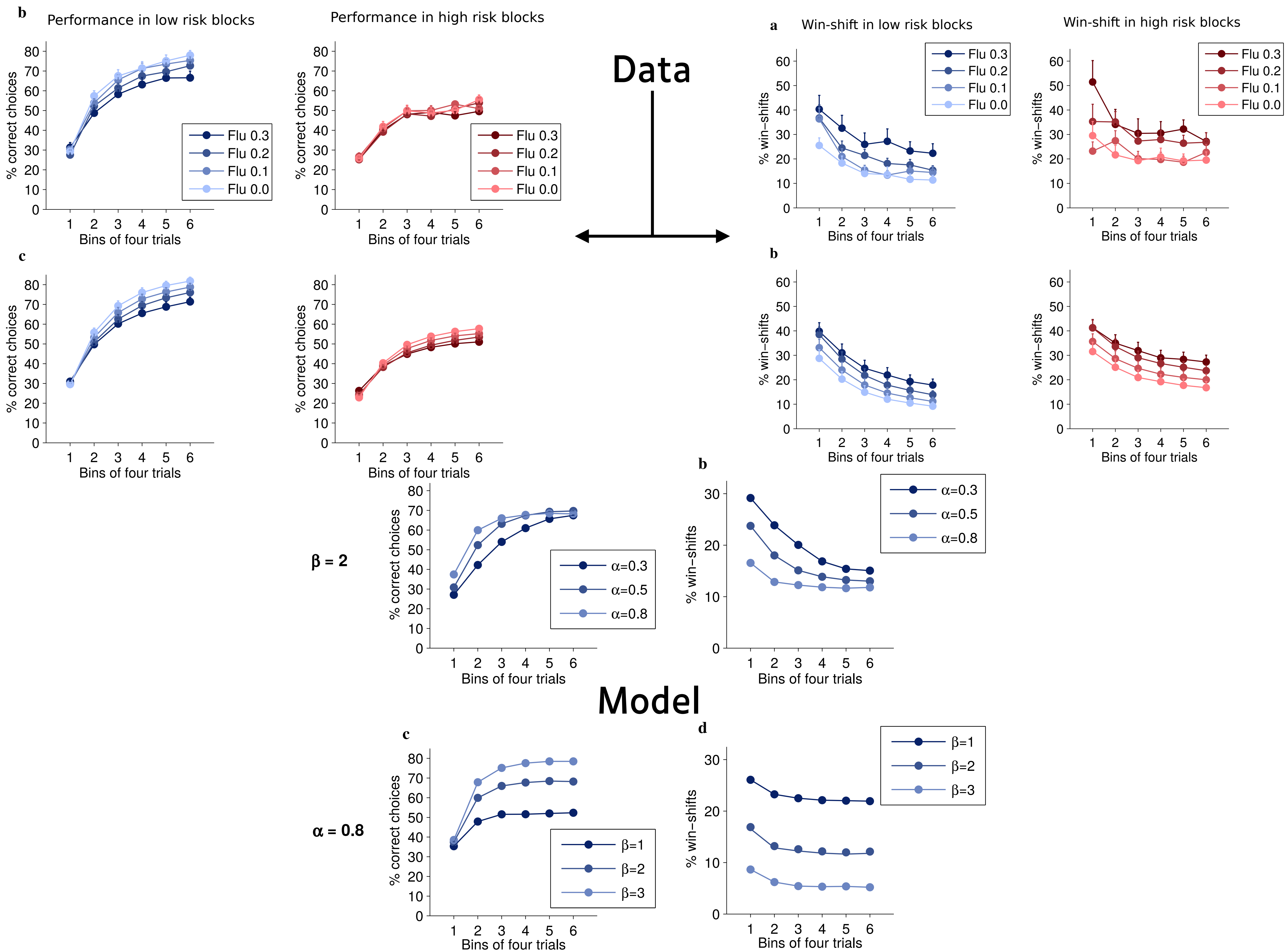
- Well known role of dopamine in RL as the teaching signal (RPE) -> learning rate α
- But also involved in regulating the exploration/exploitation tradeoff?
  -> exploration/exploitation tradeoff β

Hypothesis-informed simulations

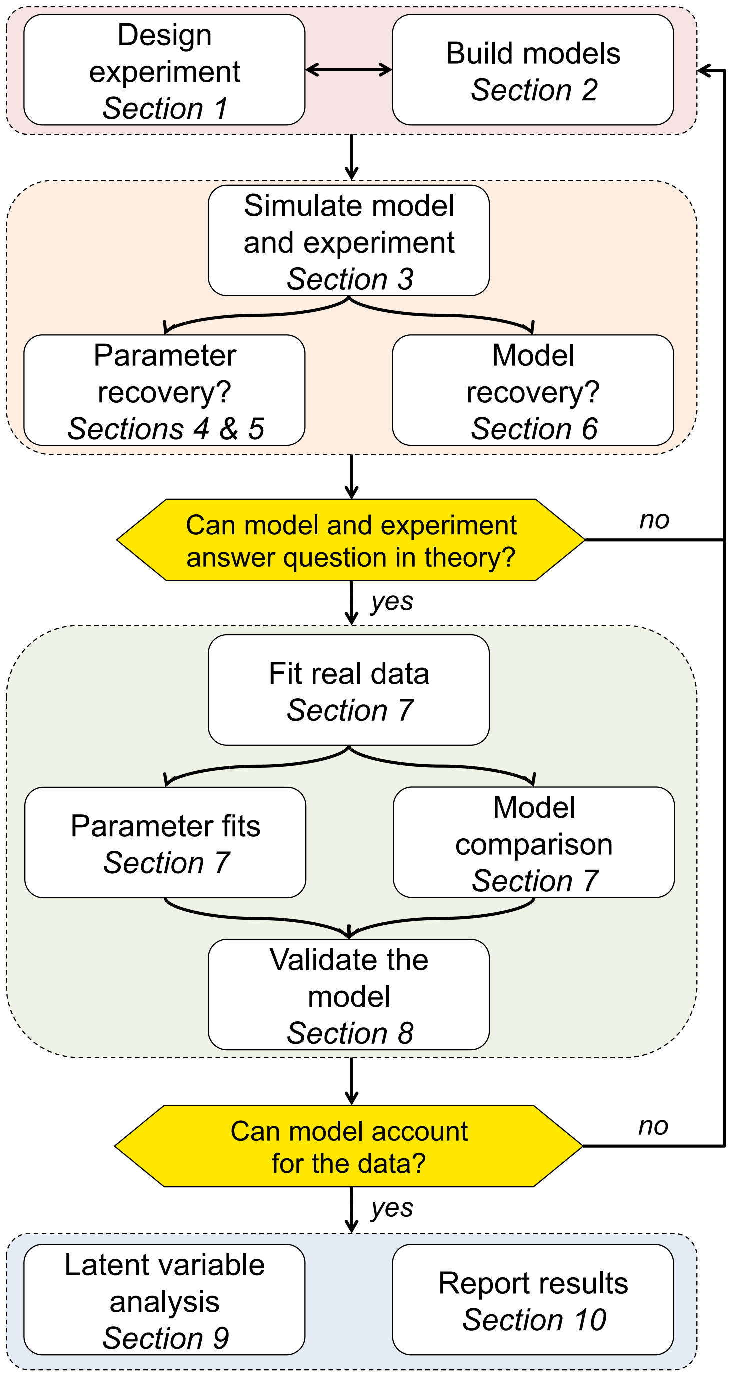
- DA modulation could affect learning rate and/or exploration/exploitation tradeoff.
- What should learning & win-shift measurements look like in these cases?



Hypothesis-informed simulations

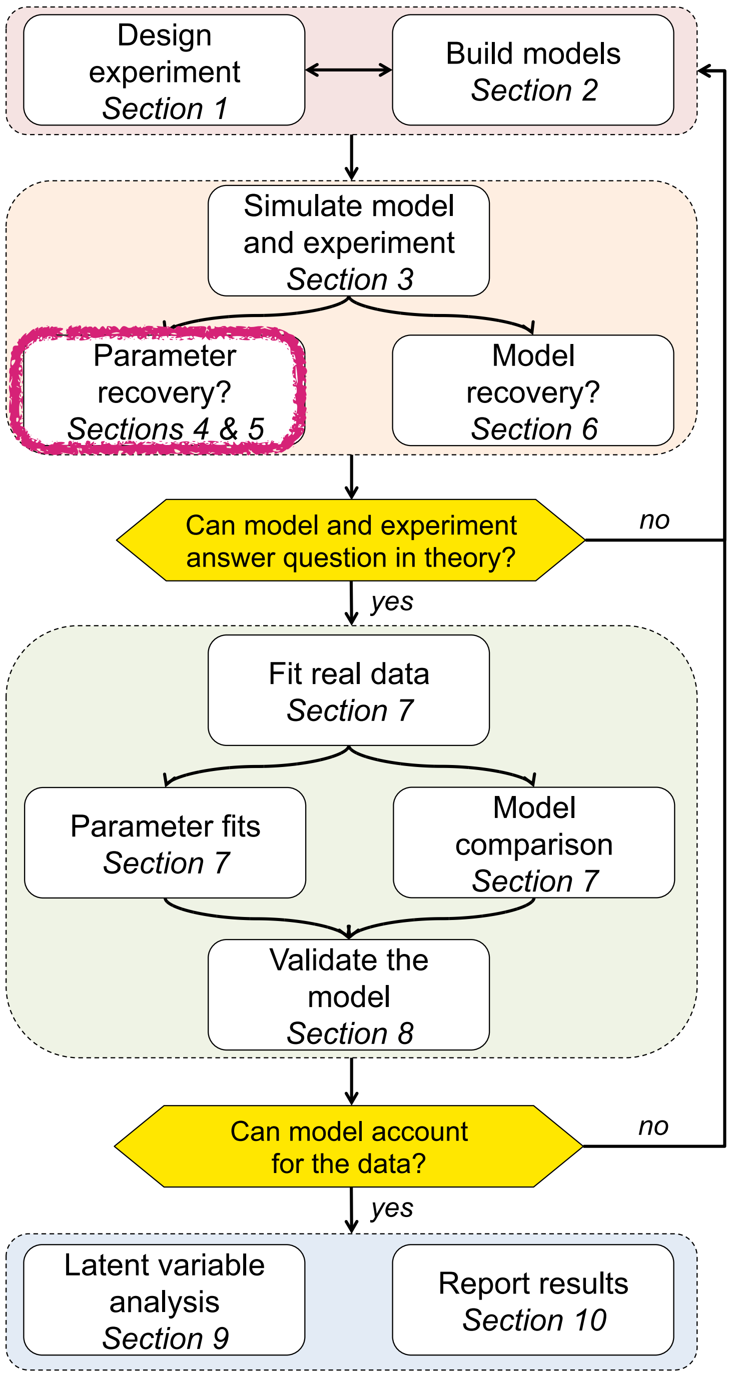


Modeling Process



~~Adapted~~ from (Wilson & Collins, 2019)

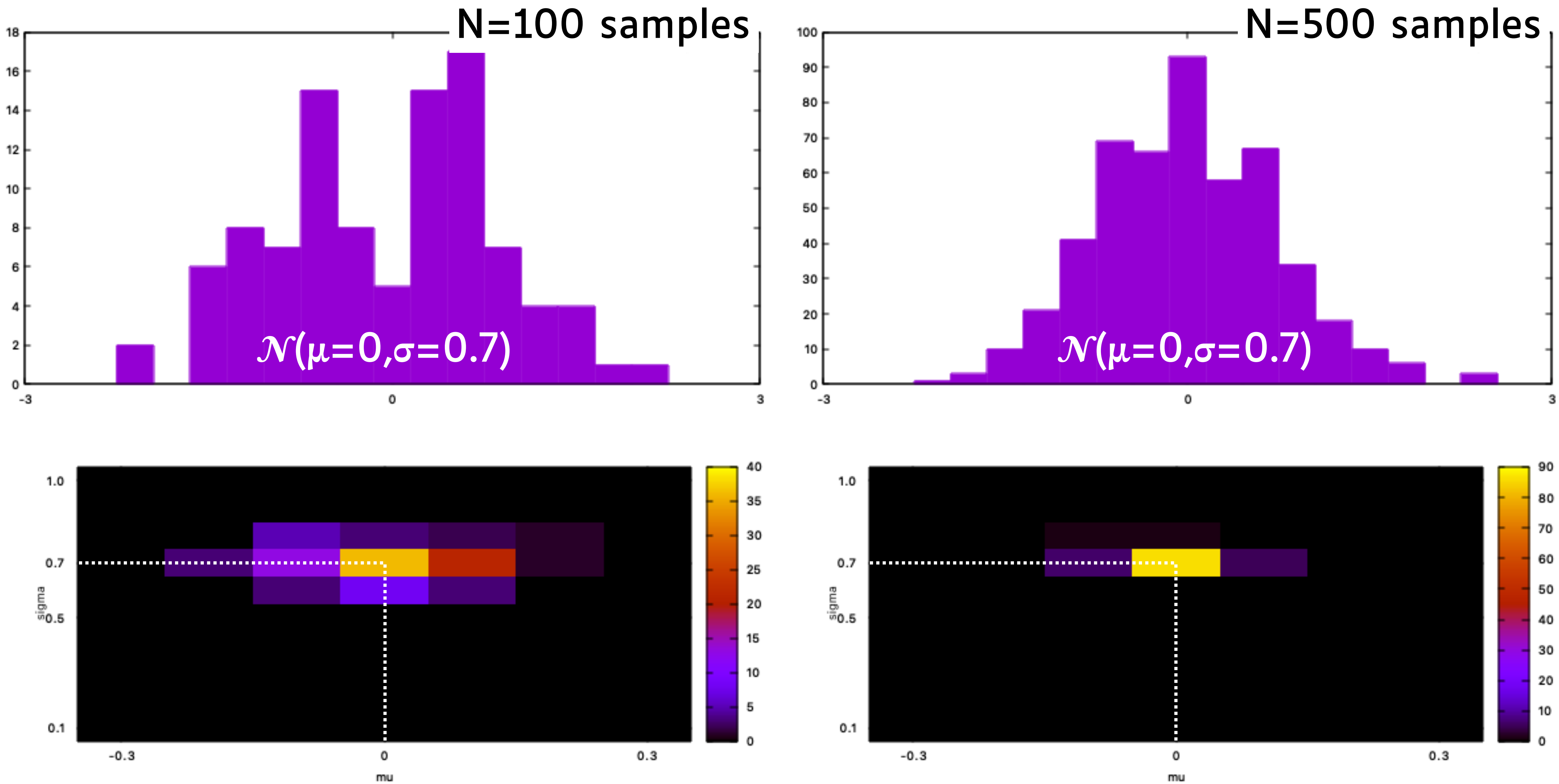
Modeling Process



~~Adapted~~ from (Wilson & Collins, 2019)

Parameter recovery

~~Number of samples~~ Parameter recovery

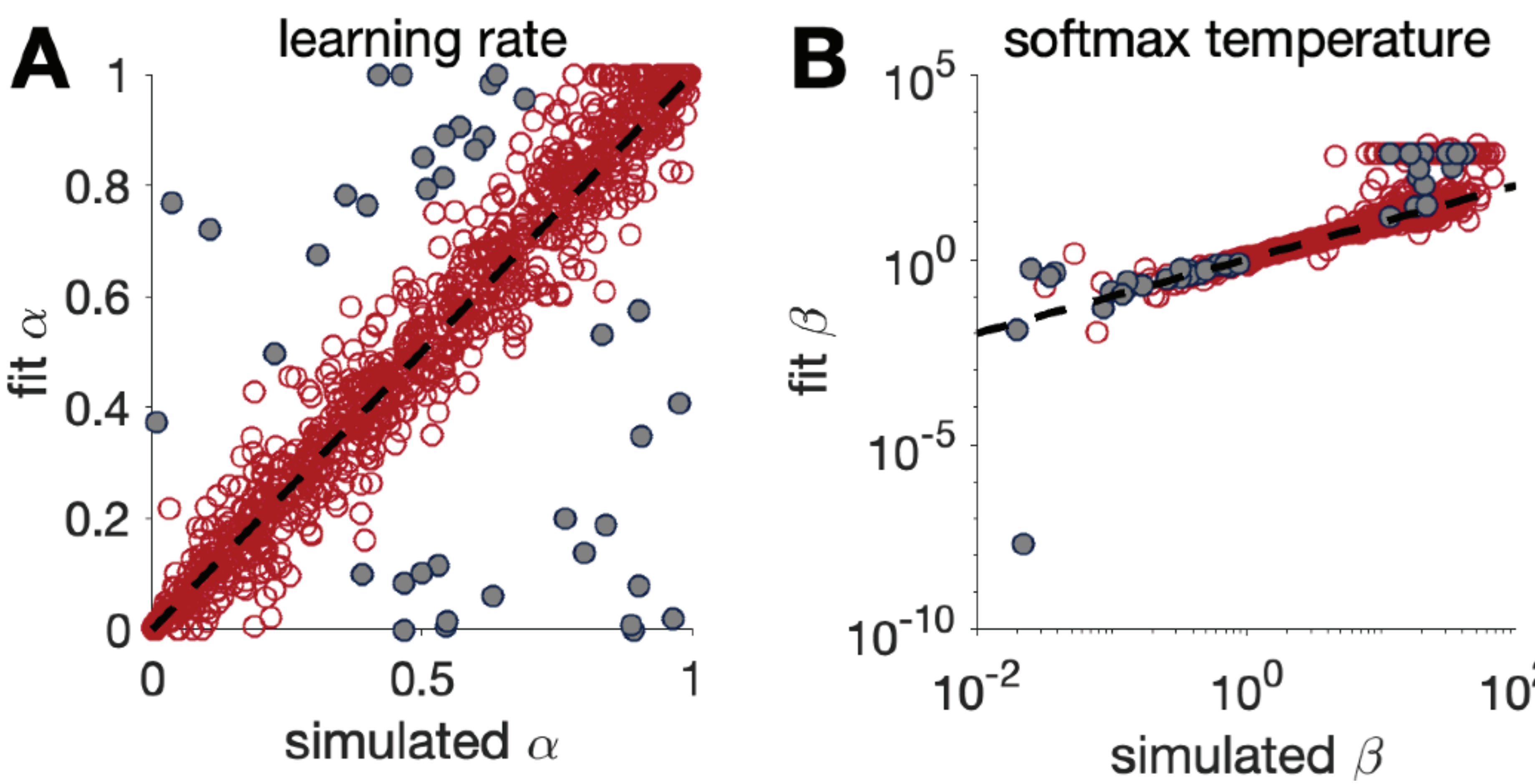


Grid search:

$\mu \in [-0.3,-0.2,-0.1,0.0,0.1,0.2,0.3]$

$\sigma \in [0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.]$

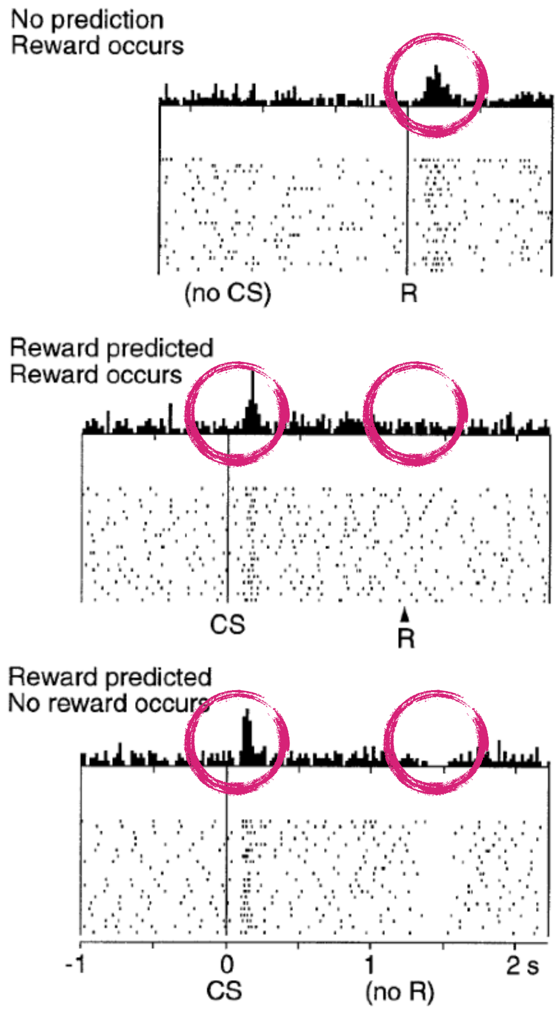
Rescorla-Wagner & 2-arm bandit



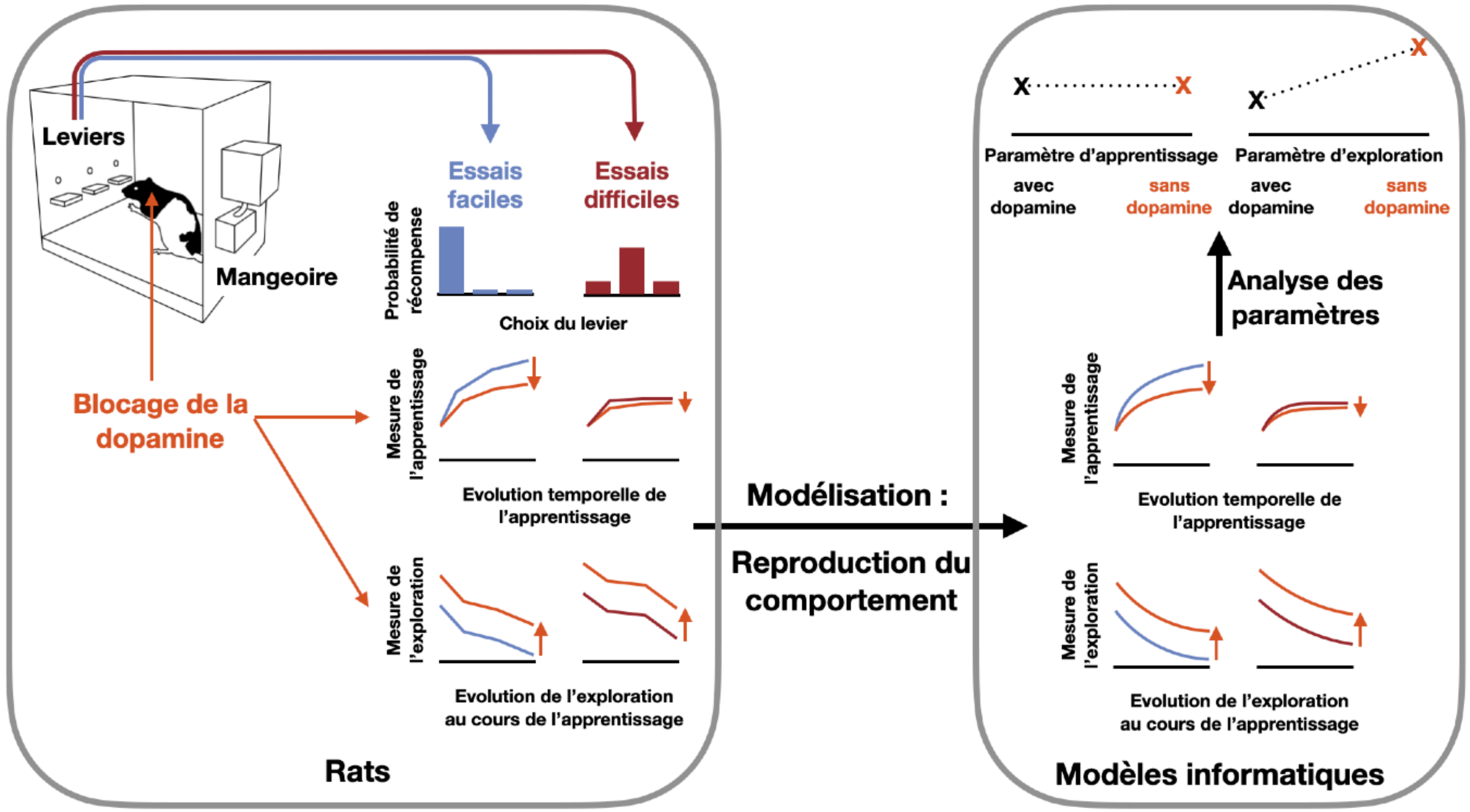
- **Rescorla-Wagner** $Q_{t+1}^k = Q_t^k + \alpha(r_t - Q_t^k)$ $p_t^k = \dfrac{\exp(\beta Q_t^k)}{\sum_{i=1}^K \exp(\beta Q_t^i)}$ $\theta_3 = (\alpha, \beta)$

The role(s) of Dopamine

Dopamine & Exploration



(Schultz et al., 1997)



(Cinotti et al., 2019)

- Well known role of dopamine in RL as the teaching signal (RPE)
  -> learning rate α
- But also involved in regulating the exploration/exploitation tradeoff?
  -> exploration/exploitation tradeoff β

RW with forgetting

- **State-Action values update:**

$$\delta_t = r_t - Q_t(a_t)$$

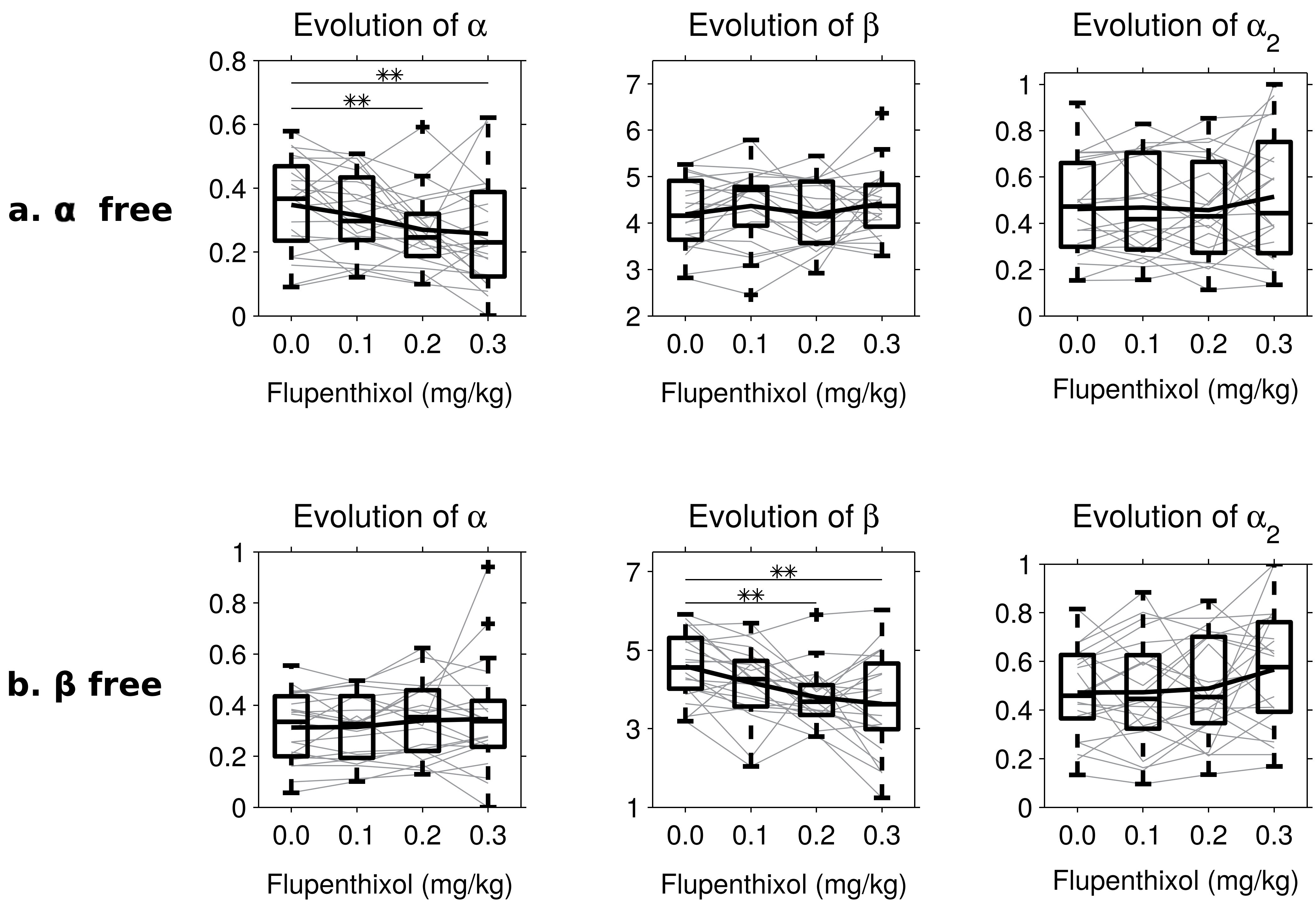
$$Q_{t+1}(a_t) = Q_t(a_t) + \alpha\delta_t$$

$$Q_{t+1}(a \neq a_t) = (1 - \alpha_2) Q_t(a \neq a_t)$$

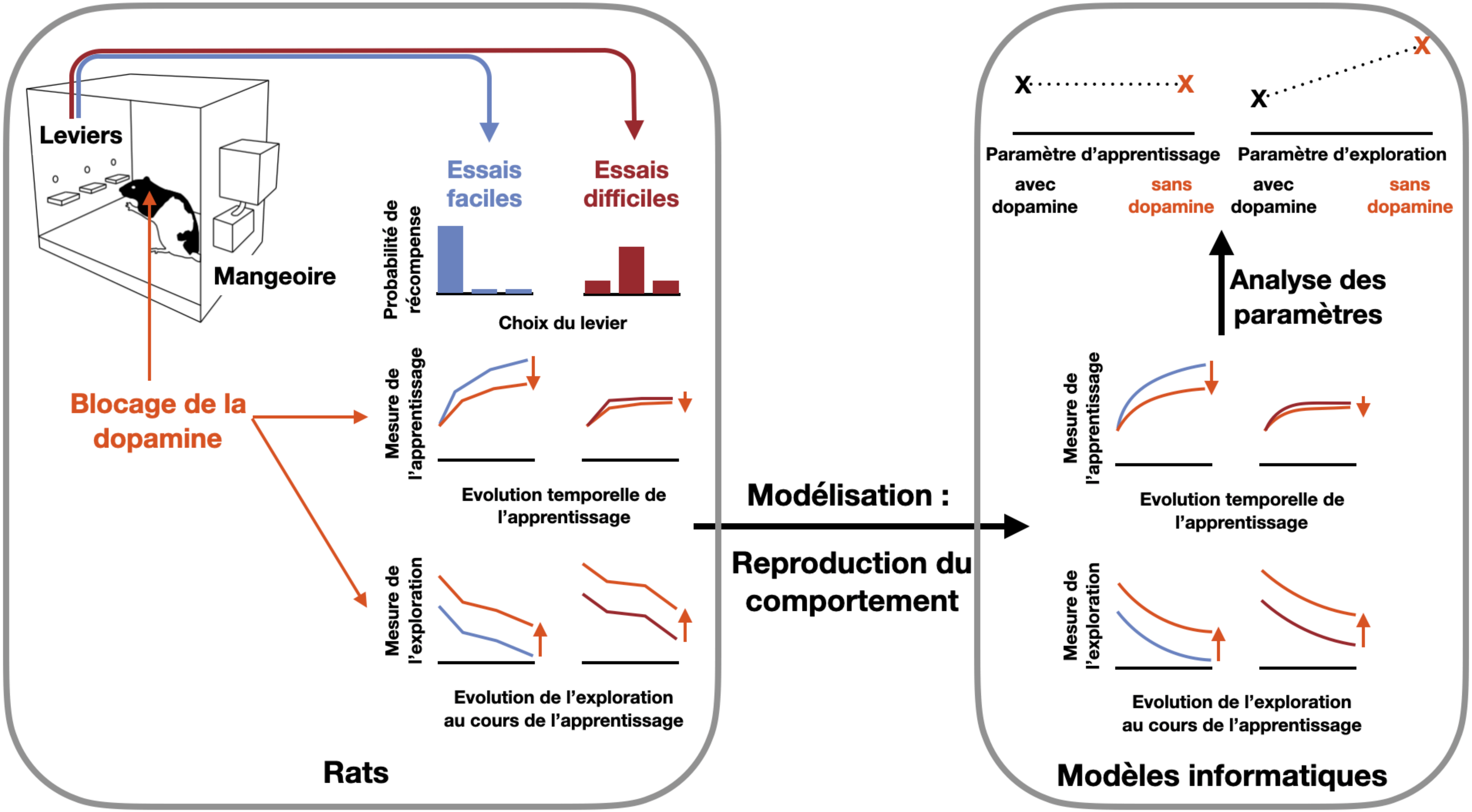
- **Action selection:**

$$P(a_{t+1} = a_i) = \frac{e^{\beta Q_t(a_i)}}{\sum_j e^{\beta Q_t(a_j)}}$$

« Parameter » Recovery

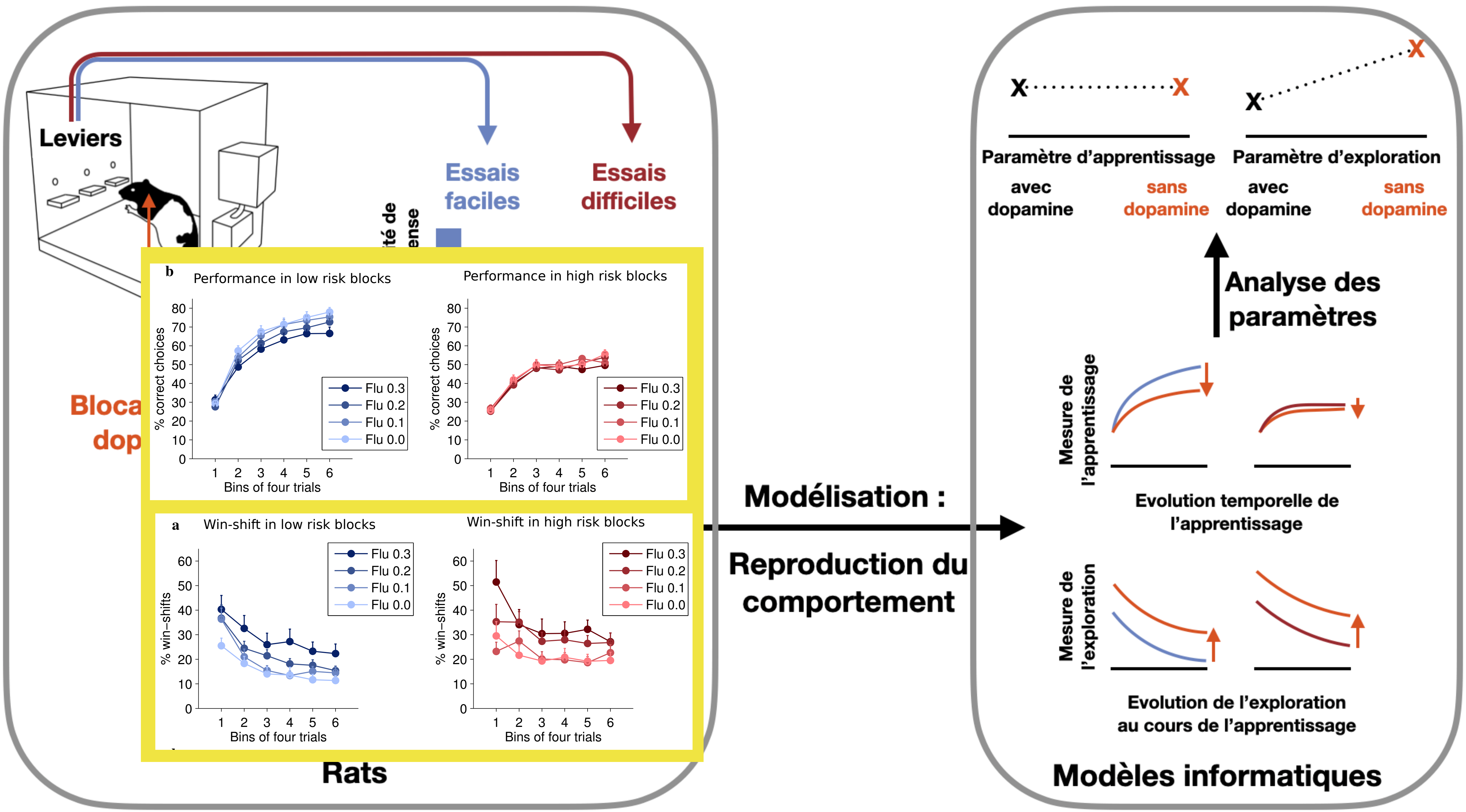


Dopamine & Exploration



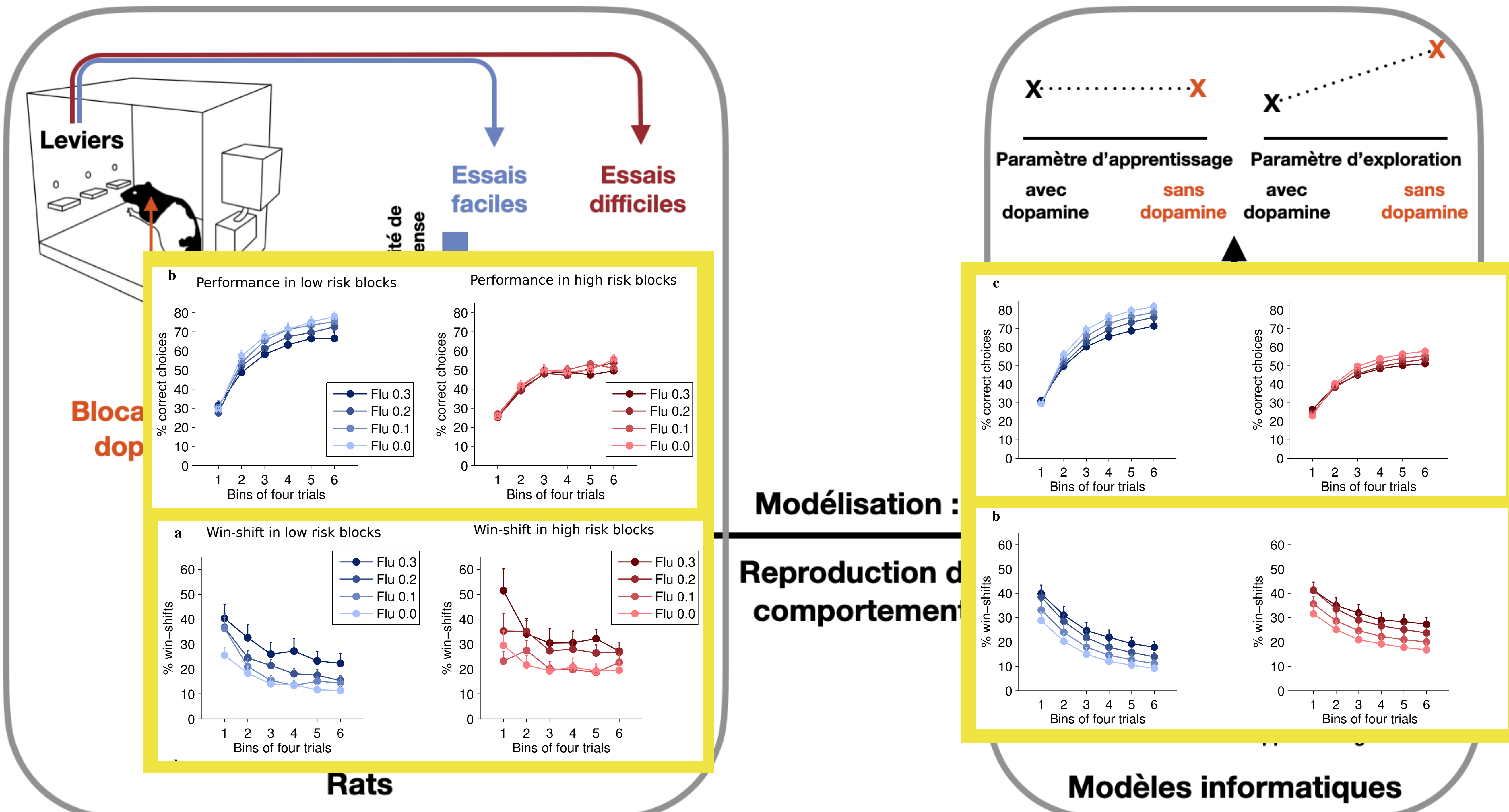
(Cinotti et al., 2019)

Dopamine & Exploration



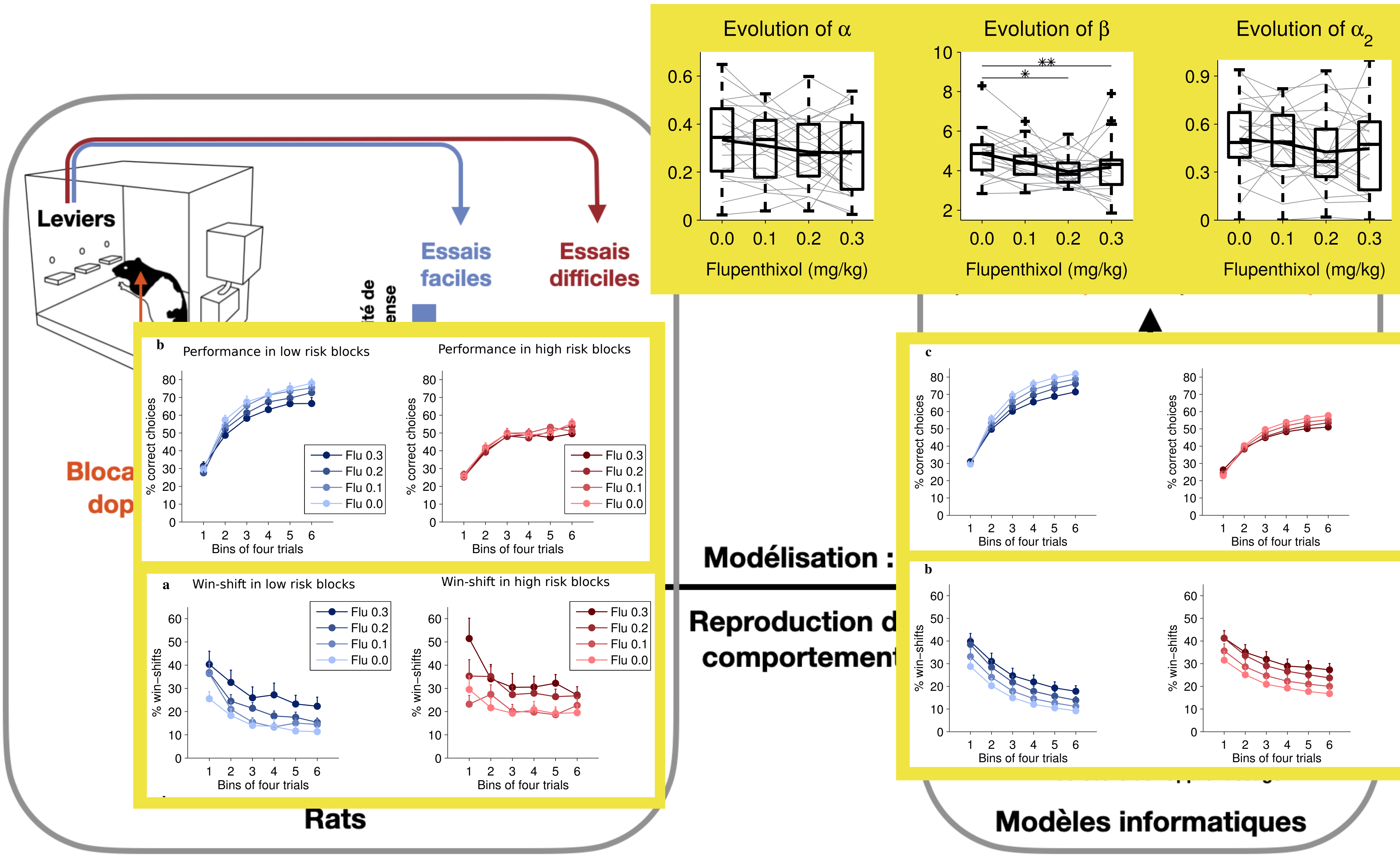
(Cinotti et al., 2019)

Dopamine & Exploration



(Cinotti et al., 2019)

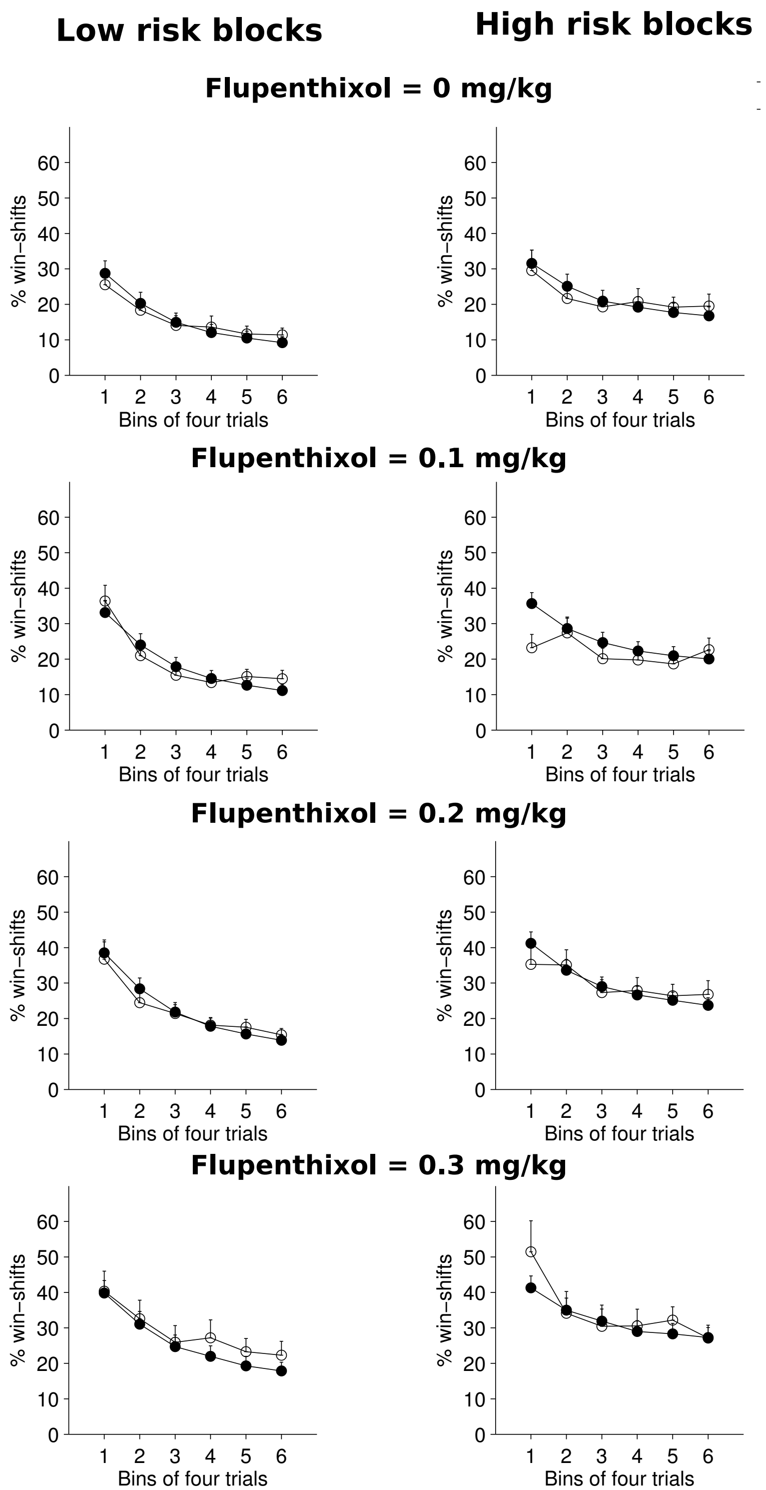
Dopamine & Exploration



(Cinotti et al., 2019)

Model selection

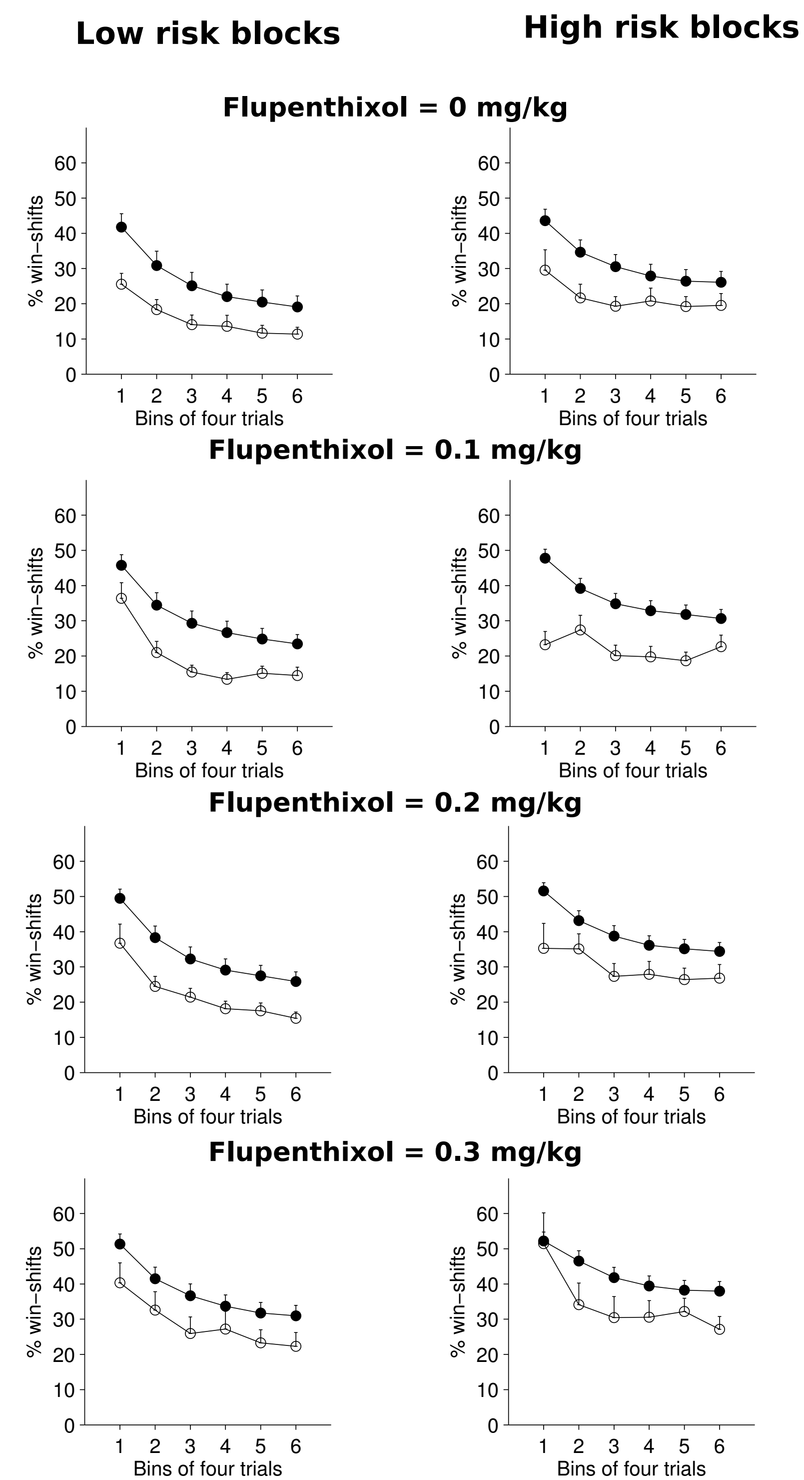
RW+forgetting



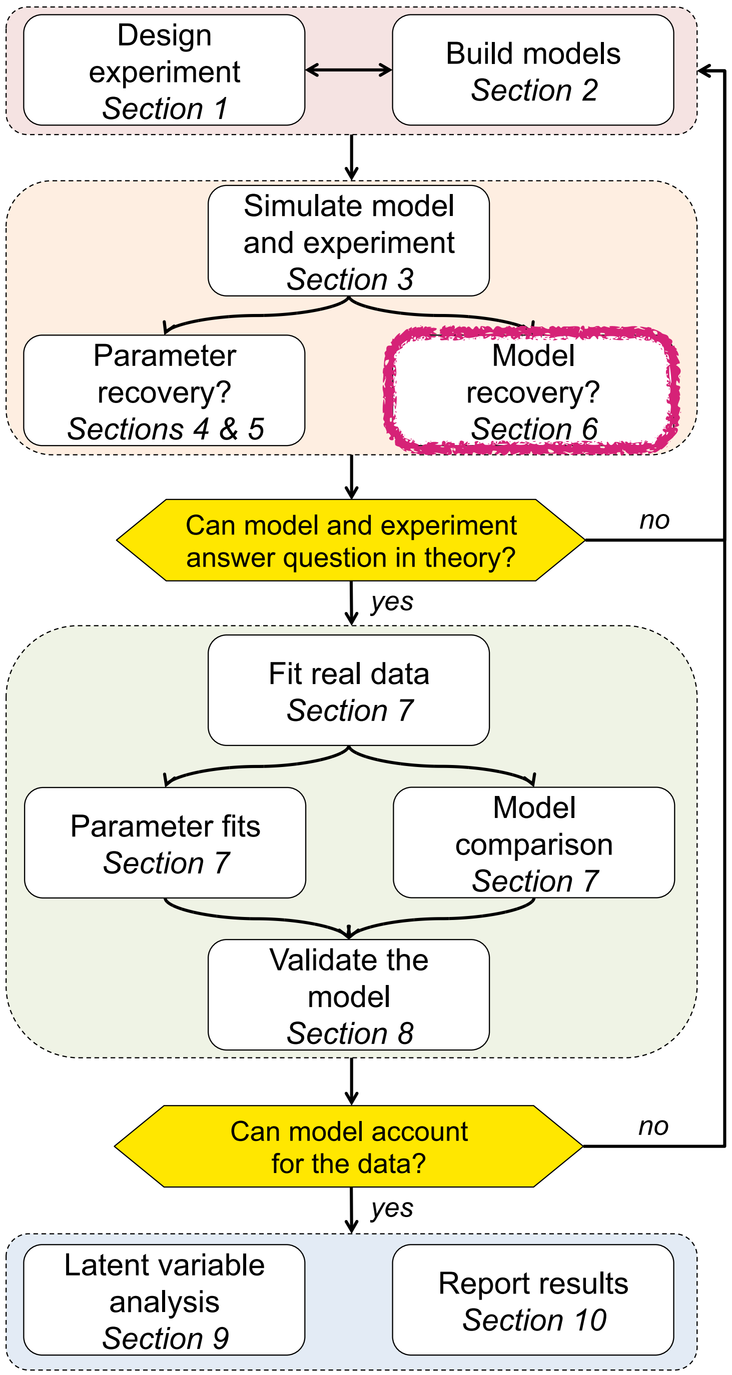
AIC: 74534 *vs.* 82772

BIC: 76149 *vs.* 84386

RW



Modeling Process

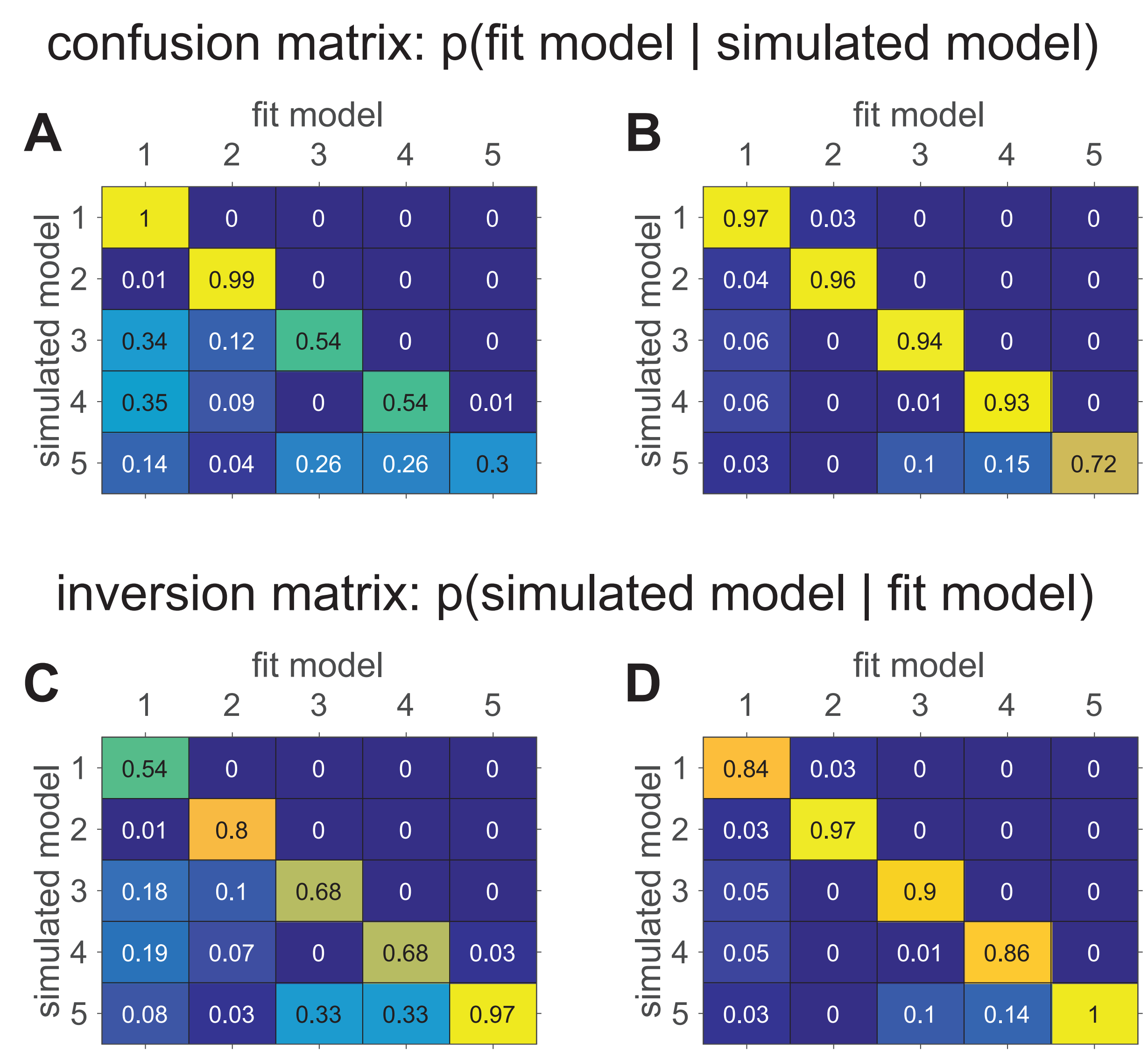


~~Adapted~~ from (Wilson & Collins, 2019)

Model recovery

Model recovery with 2-arm bandit

- $p(R=1|a=0)=0.2$; $p(R=1|a=1)=0.8$
- T=1000 trials
- 100 synthetized data sets per model



Priors
$b \sim U(0,1)$
$\epsilon \sim U(0,1)$
$\alpha \sim U(0,1)$, $\beta \sim \mathrm{Exp}(1)$
$\alpha_c \sim U(0,1)$, $\beta_c \sim \mathrm{Exp}(1)$
$\alpha \sim U(0,1)$, $\beta \sim \mathrm{Exp}(1)$, $\alpha_c \sim U(0,1)$, $\beta_c \sim \mathrm{Exp}(1$

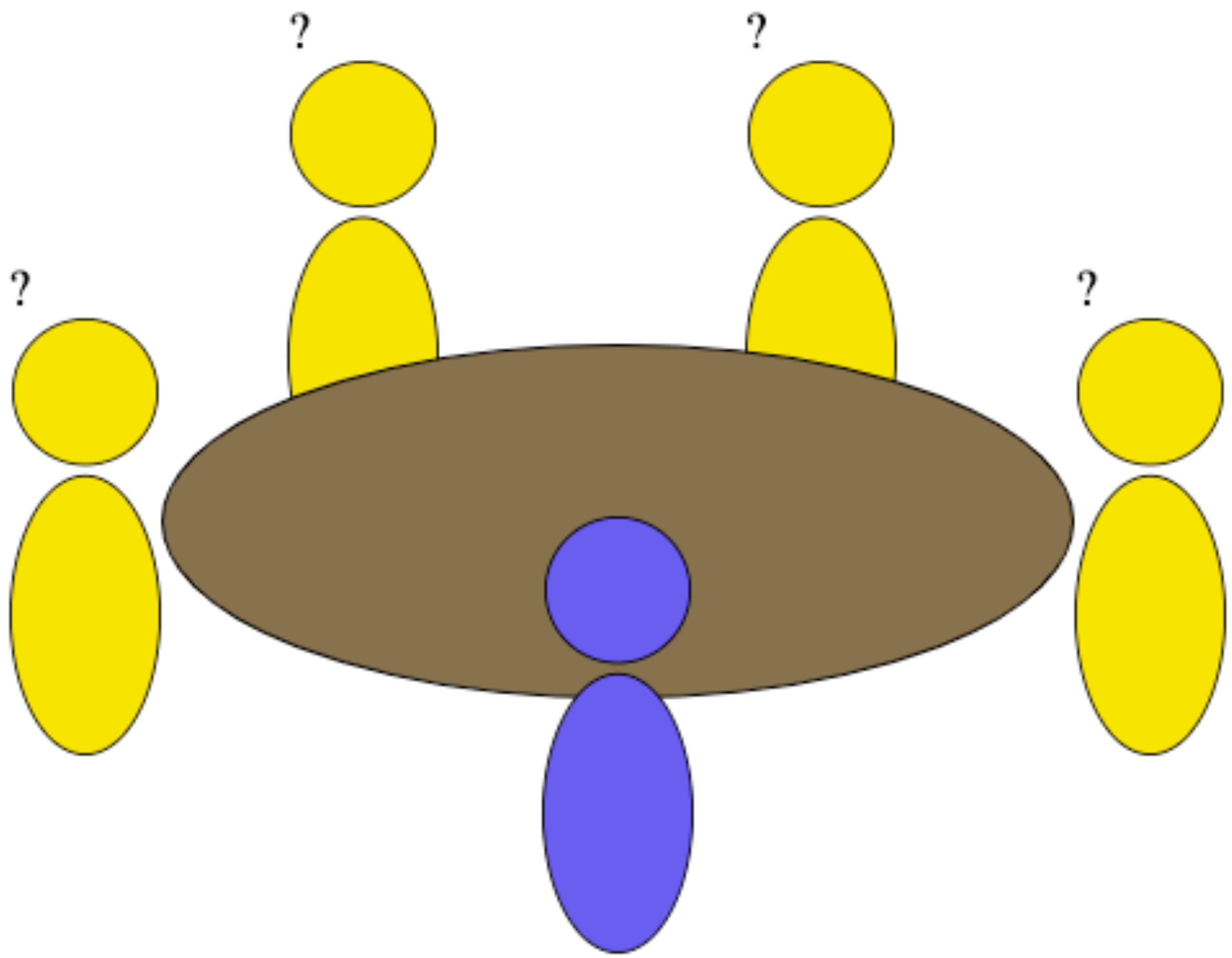
From (Wilson & Collins, 2019)

An example where model selection won't work...

Social influence on dessert choice

- Is there such influence?
  - Social influence has been documented:
    we tend to adjust intake when comparing to commensals (social modeling) (Vartanian et al., 2015)
  - What about choice rather than quantity?
    Not that clear (Cruwys et al., 2015; Robinson, Thomas, et al., 2013)
- Does it depend on the healthiness/unhealthiness of the desserts?
  - Possible differences for low vs. high energy density food (Pliner & Mann, 2004).
- Mainly tested with dyades: what about groups?

Task: a trial

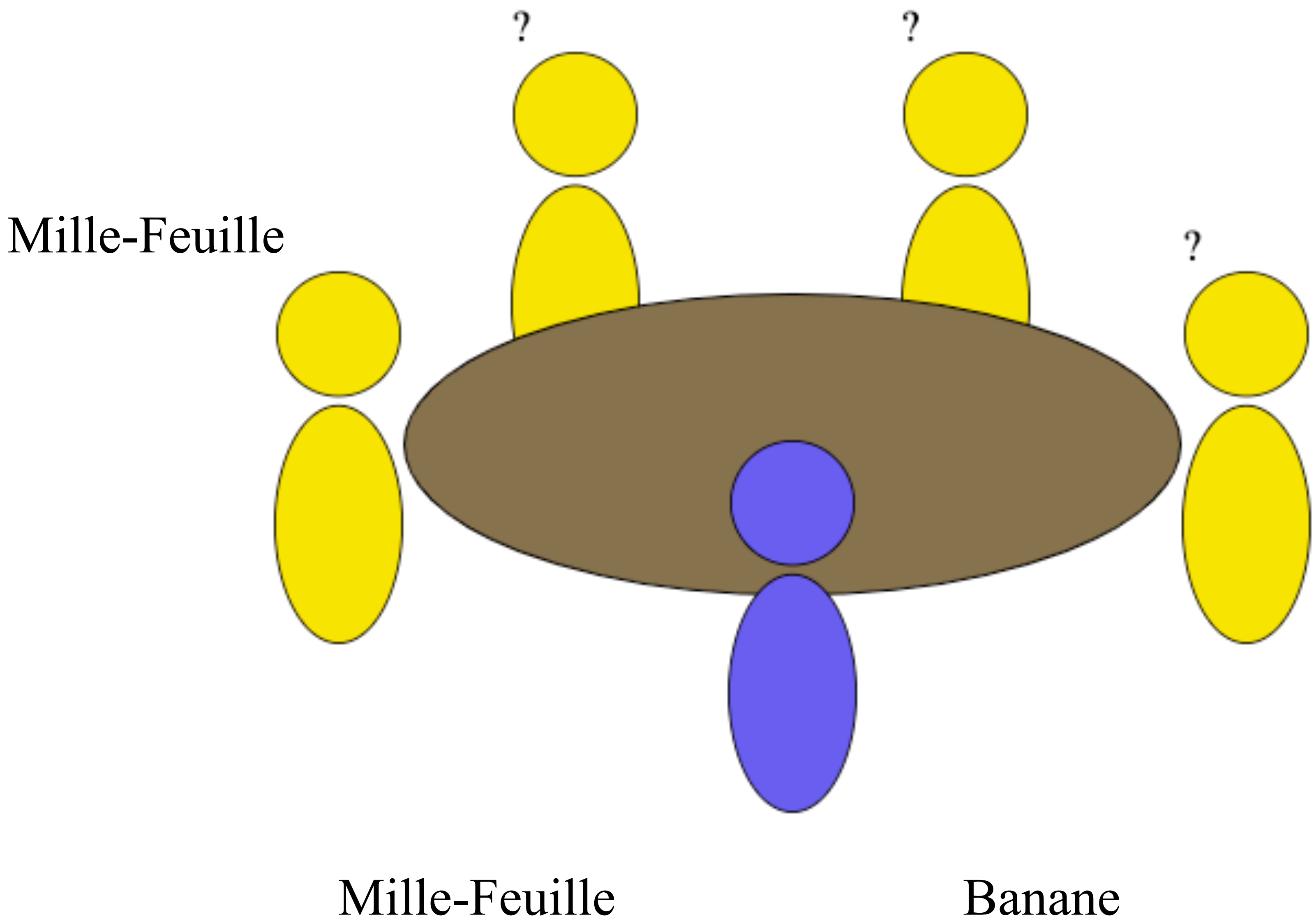


Mille-Feuille Banane

[3 s]

Task: a trial

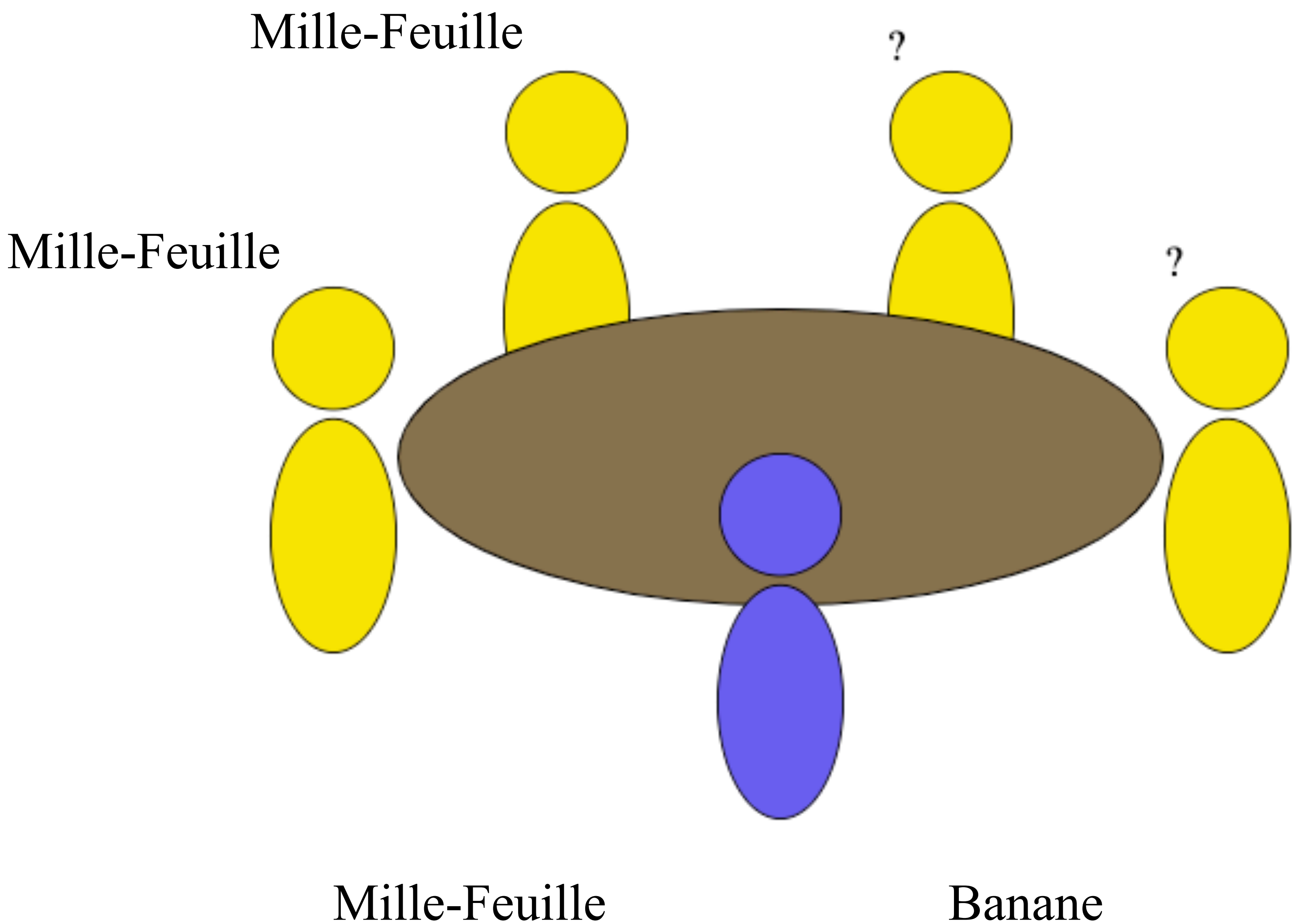
[1.2-3 s]



[3 s]

Task: a trial

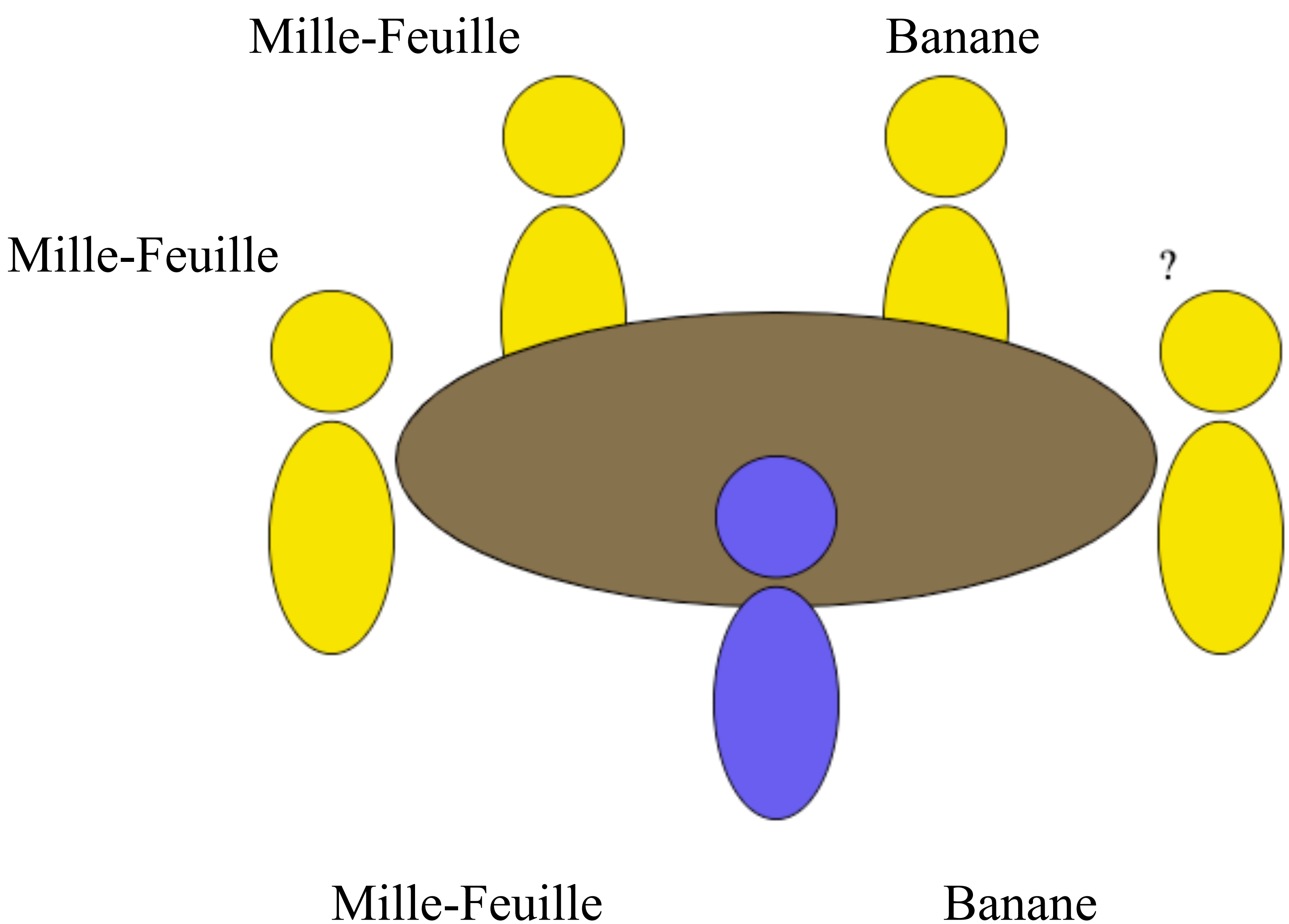
[1.2-3 s]



[3 s]

Task: a trial

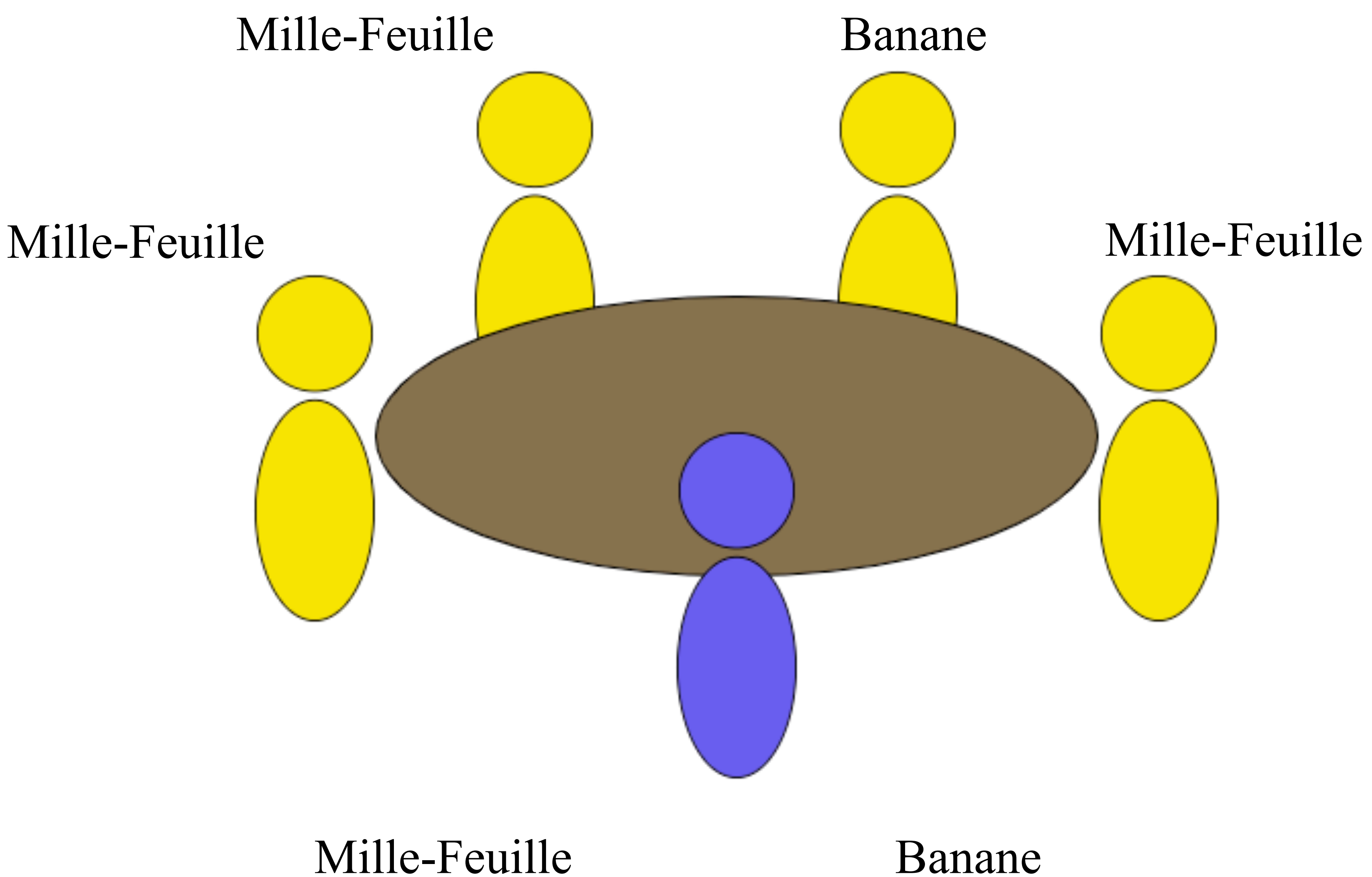
[1.2-3 s]



[3 s]

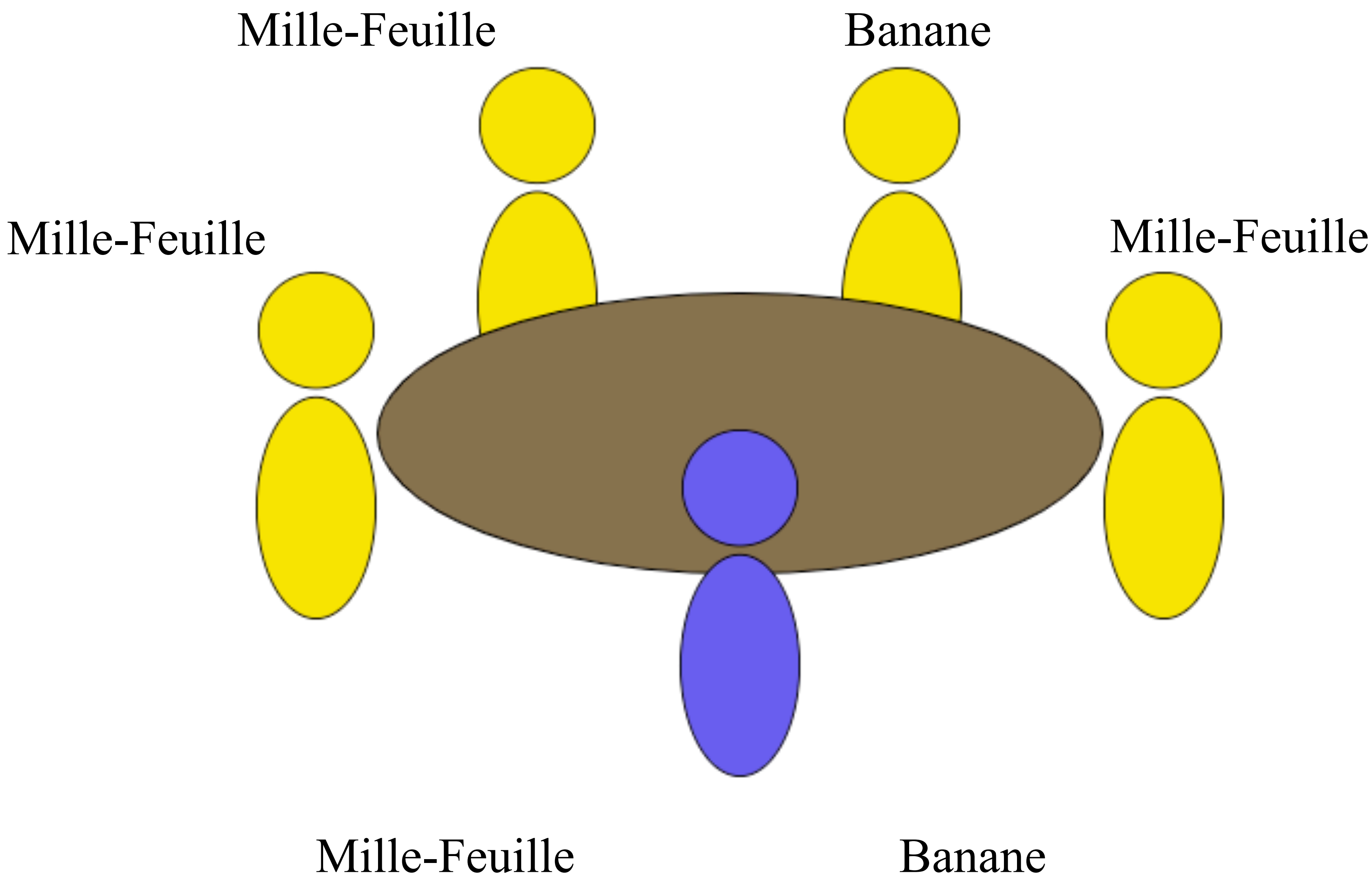
Task: a trial

[1.2-3 s]



[3 s]

Task: a trial



[1.2-3 s]

[3 s]

[choice: 0-7 s]

Task design

- 82 trials

- For each subject:
  subset of 14 desserts
  (7 H, 7 UH),

- Excluding as much as possible those rated 1 & 7,

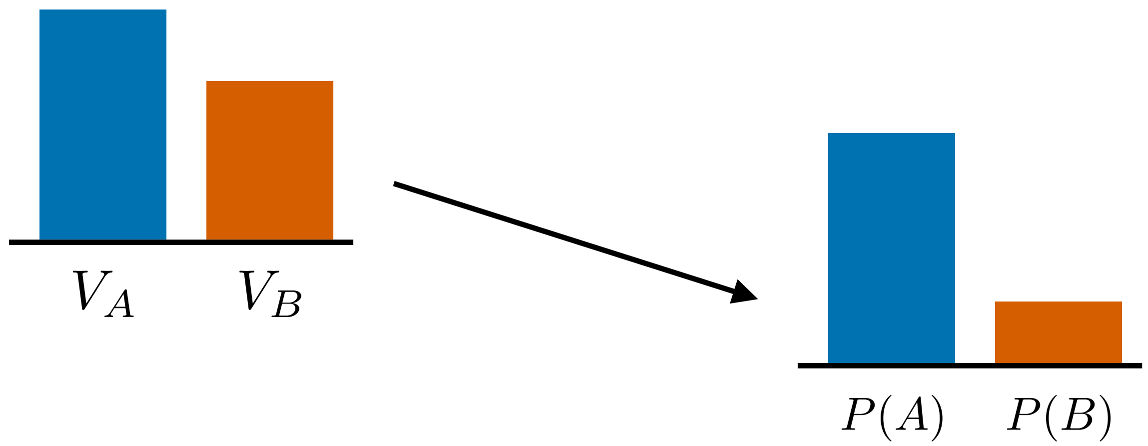
- As uniformly as possible.

Others' Choices	H vs. H	UH vs. UH	H vs. UH
4/0	7	7	9
3/1	7	7	7
2/2	7	7	8
1/3	0	0	7
0/4	0	0	9

Simple models:

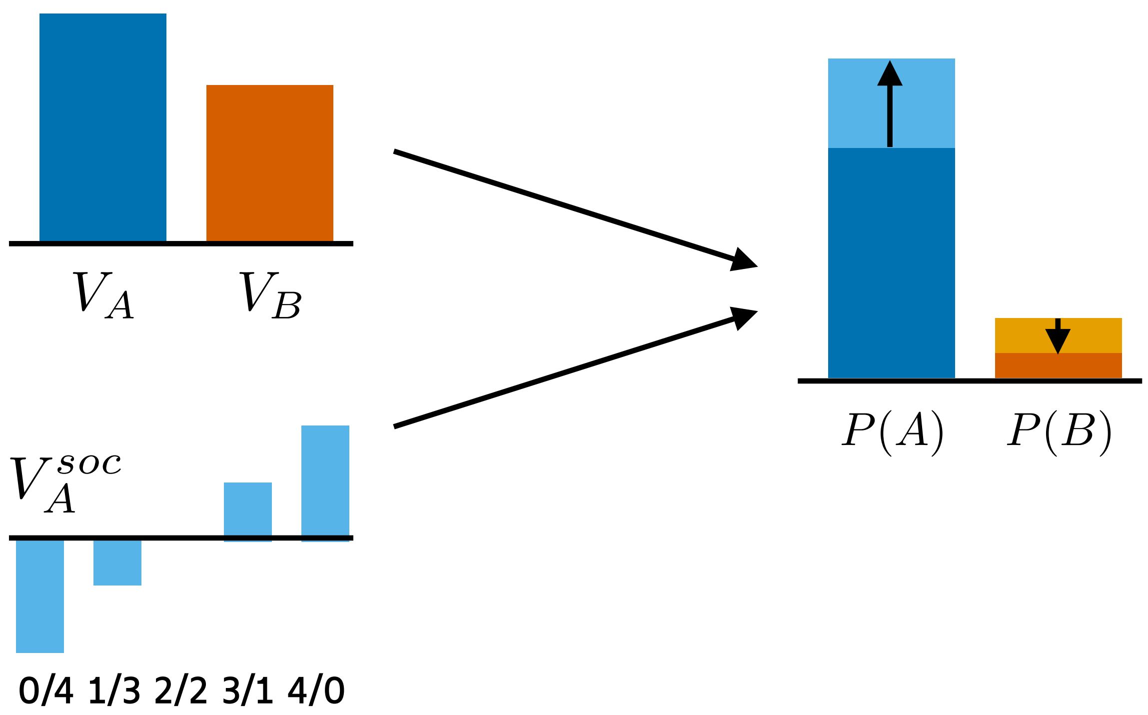
- **Value-based decision-making:**

$$P(A) = \frac{exp(\beta V_A)}{exp(\beta V_A) + exp(\beta V_B)}$$



- **Linear or threshold social modulation:**

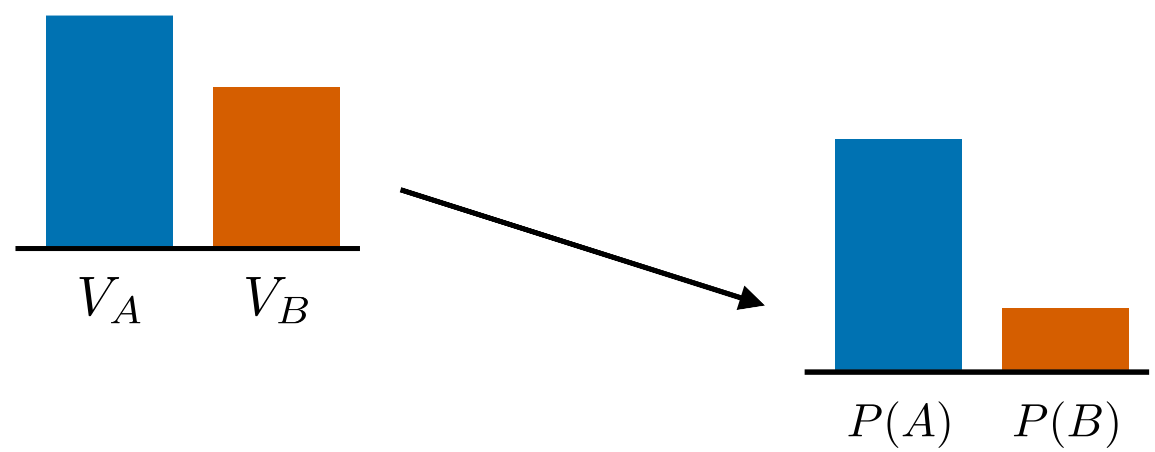
$$P(A) = \frac{exp(\beta(V_A + V_A^{soc})}{exp(\beta(V_A + V_A^{soc}) + exp(\beta(V_B - V_A^{soc})}$$



Simple models:

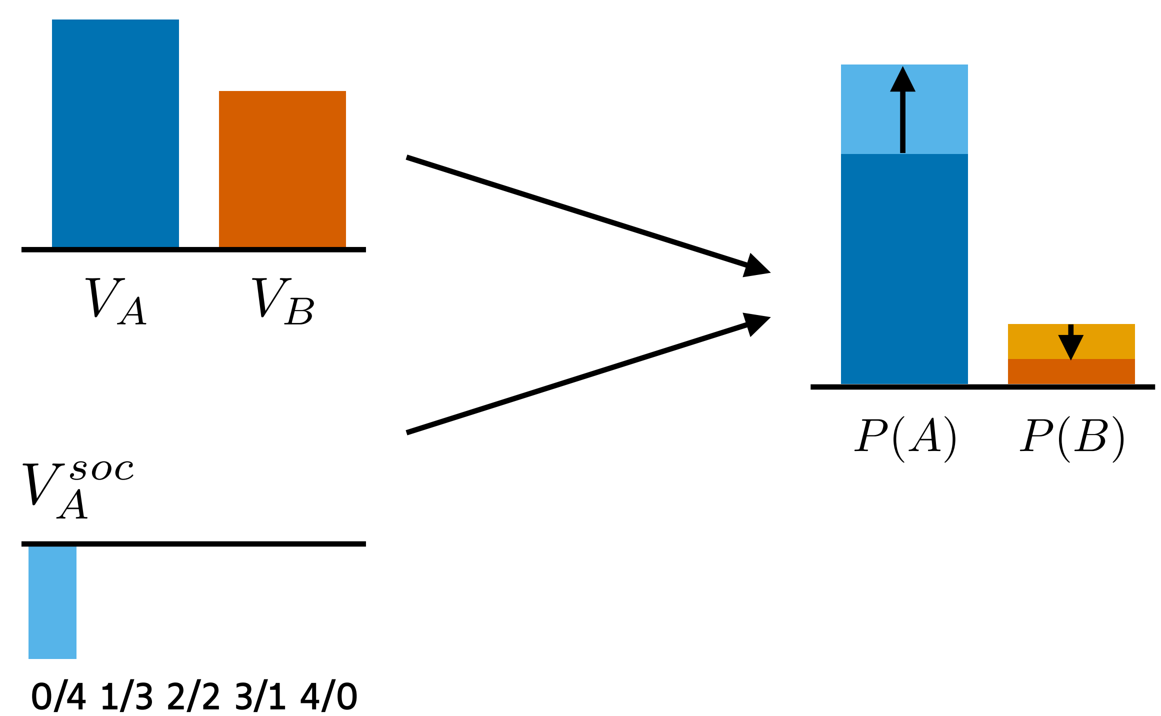
- **Value-based decision-making:**

$$P(A) = \frac{exp(\beta V_A)}{exp(\beta V_A) + exp(\beta V_B)}$$



- **Linear or threshold social modulation:**

$$P(A) = \frac{exp(\beta(V_A + V_A^{soc})}{exp(\beta(V_A + V_A^{soc}) + exp(\beta(V_B - V_A^{soc})}$$



Fitting & comparison

Confusion matrix

Best fit

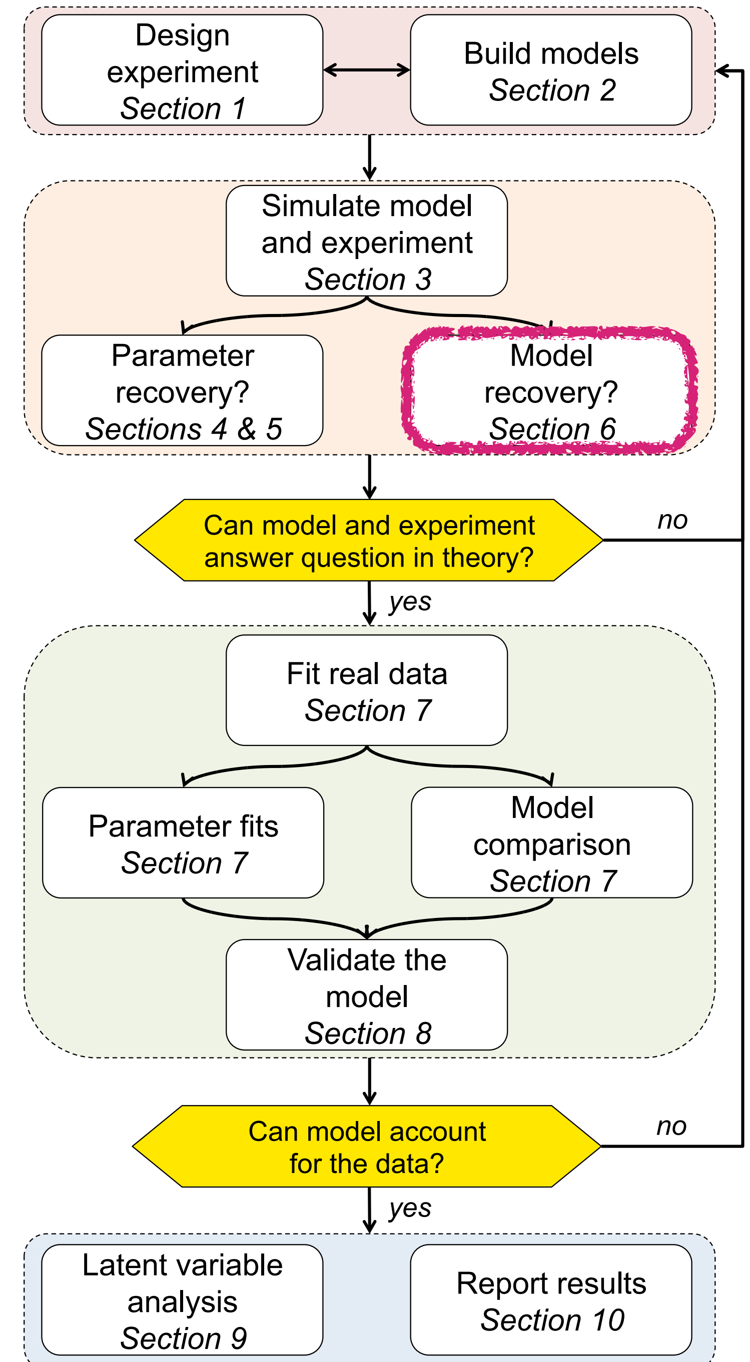
Synth. data	No	Linear	Trigger
No	0.6	0.18	0.21
Linear	0.2	0.72	0.09
Trigger	0.16	0.08	0.76

Inversion matrix

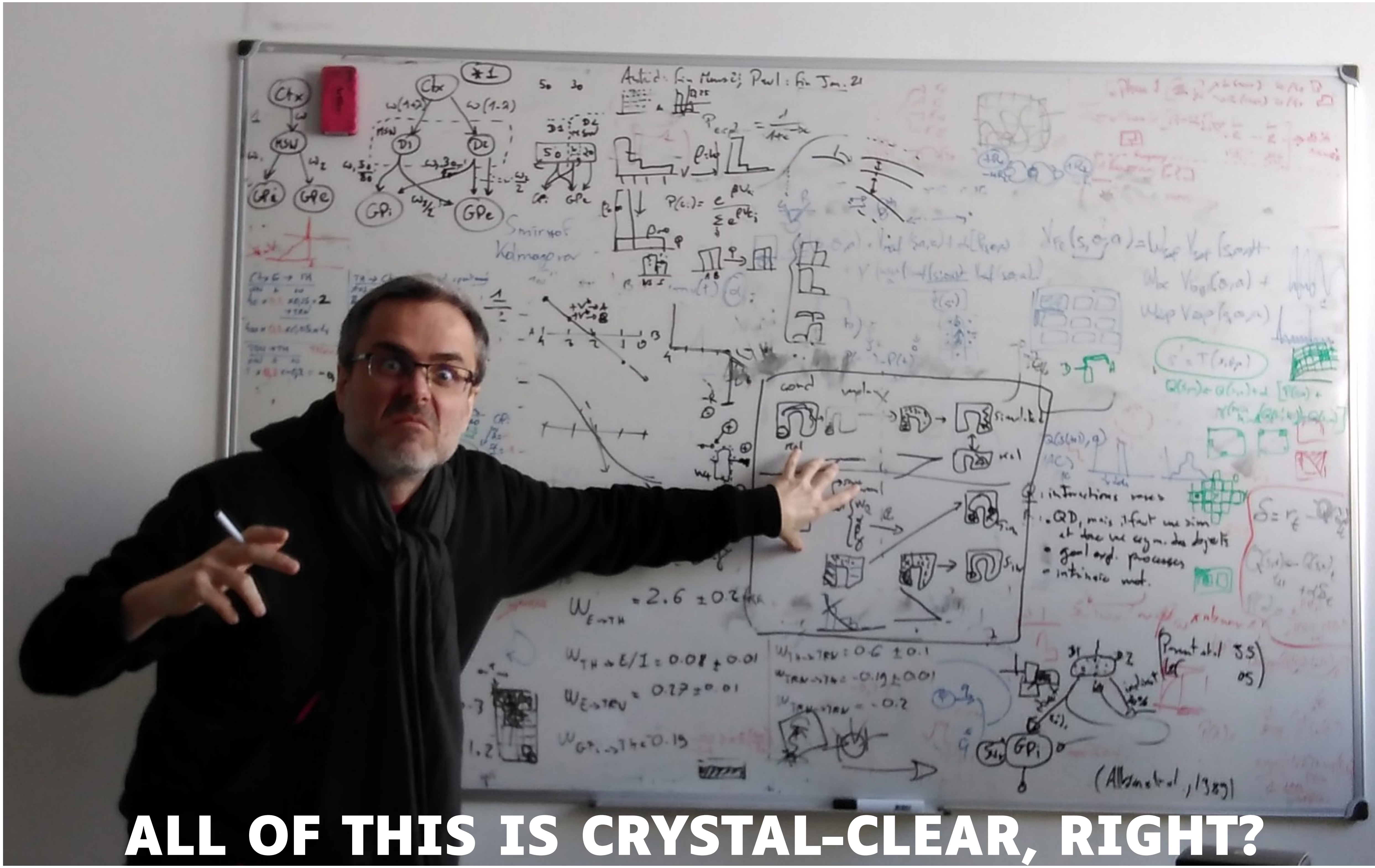
Best fit

Synth. data	No	Linear	Trigger
No	0.63	0.19	0.2
Linear	0.21	0.73	0.08
Trigger	0.16	0.08	0.72

- Problems with parameter and model recovery: lack of points!



~~Adapted~~ from (Wilson & Collins, 2019)



ALL OF THIS IS CRYSTAL-CLEAR, RIGHT?

Conclusion

- Compute Log $\mathcal{L}(\theta|x)$!
- Grab as many trials as you can!
- Pre θ optimization:
  Generate synthetic data.
  Check that parameter recovery is OK.
  Check that you models can be discriminated!
- Post θ optimization and model selection:
  Free wheel simulations.
  Check that your less crappy model is not too crappy...

References

- Anderson, D., & Burnham, K. (2002). Model selection and multi-model inference. Second edition. *NY: Springer-Verlag*.
- Deb, K. (2001). Multi-Objective Optimization Using Evolutionary Algorithms. *John Wiley & Sons, Inc.*
- Palminteri, S., Wyart, V., & Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in cognitive sciences*, 21(6), 425-433.
- Wilson, R. C., & Collins, A. G. (2019). Ten simple rules for the computational modeling of behavioral data. *Elife*, 8, e49547.