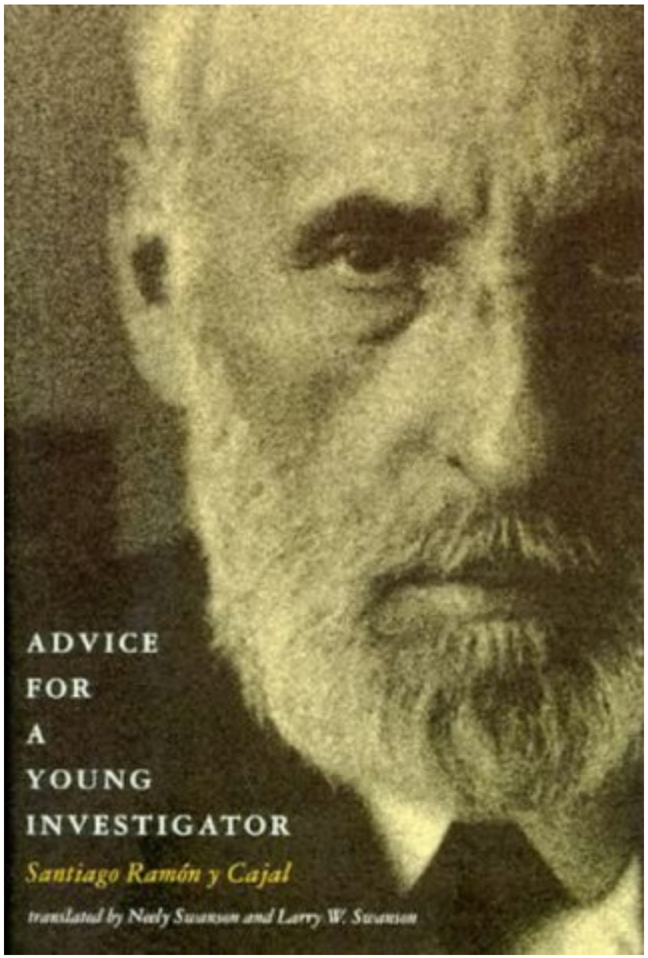
Biological Learning

Peter Dayan

Gatsby Computational Neuroscience Unit

Nathaniel Daw Sam Gershman Sham Kakade Yael Niv

5. Diseases of the Will

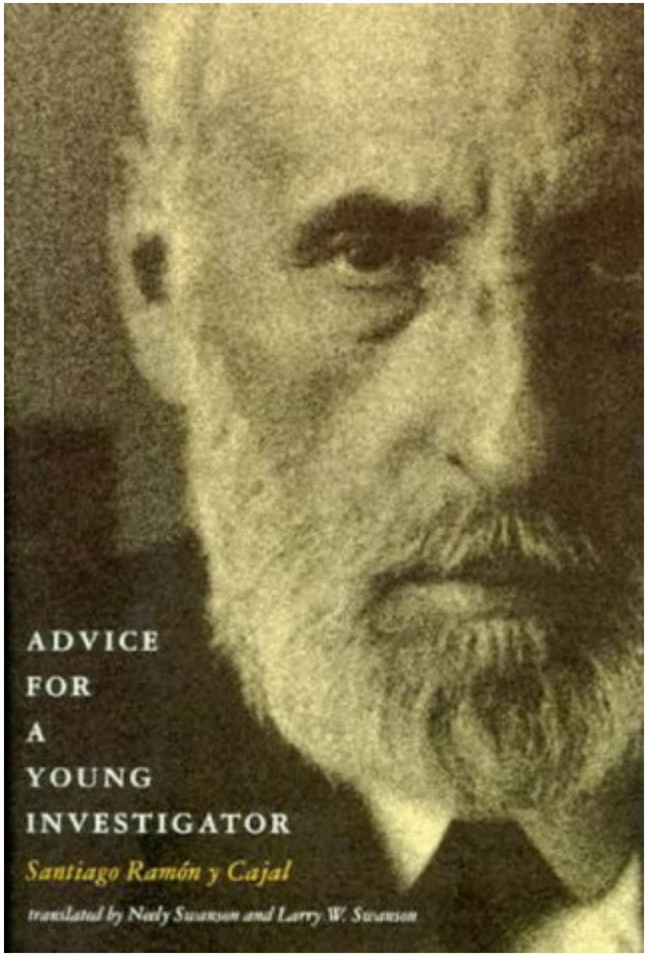


- Contemplators
- Bibliophiles and Polyglots
- Megalomaniacs
- Instrument addicts
- Misfits

Biological Learning

- error minimization/delta rule
- temporal difference learning
- Kalman filter
- Dirichlet process mixture/NPB
- Bayesian Q-learning; Bayes-adaptive MDPs
- memory-based reasoning
- particle filters for inference
- unsupervised `structural’ learning

5. Diseases of the Will



- Contemplators
- Bibliophiles and Polyglots
- Megalomaniacs
- Instrument addicts
- Misfits

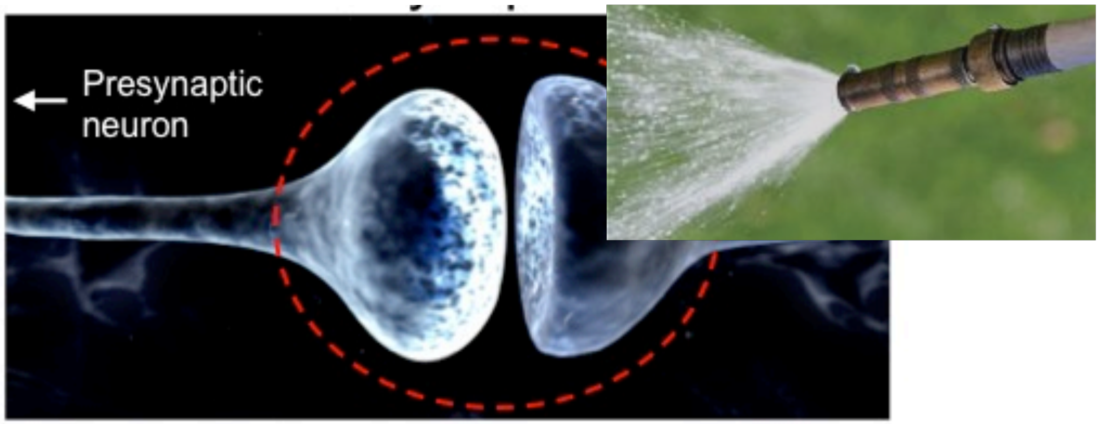
- **Theorists**

Theorists

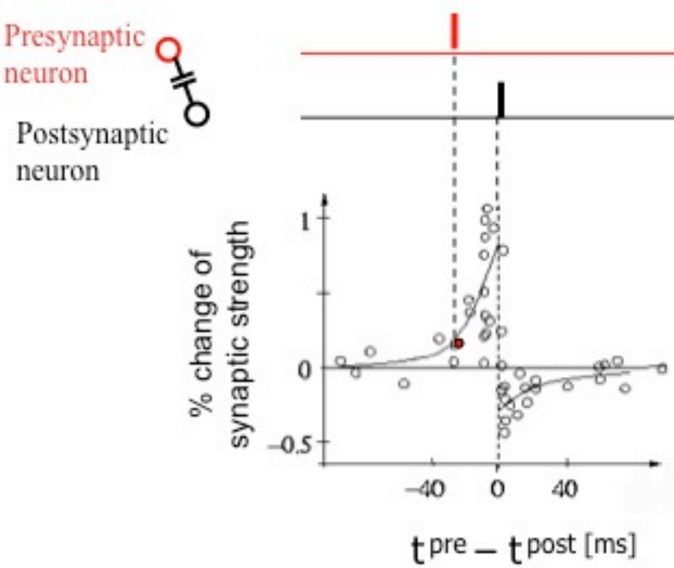
There are highly cultivated, wonderfully endowed minds whose wills suffer from a particular form of lethargy. Its undeniable symptoms include a facility for exposition, a creative and restless imagination, an aversion to the laboratory, and an indomitable dislike for concrete science and seemingly unimportant data… When faced with a difficult problem, they feel an irresistible urge to formulate a theory rather than question nature.

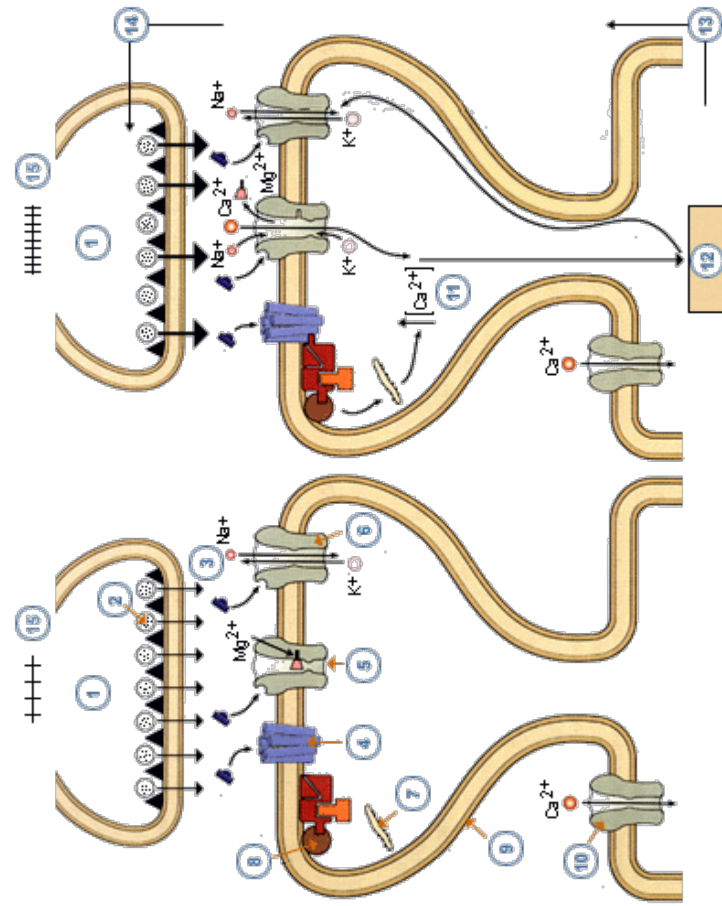
As might be expected, disappointments plague the theorist…

Neuroscience of Learning

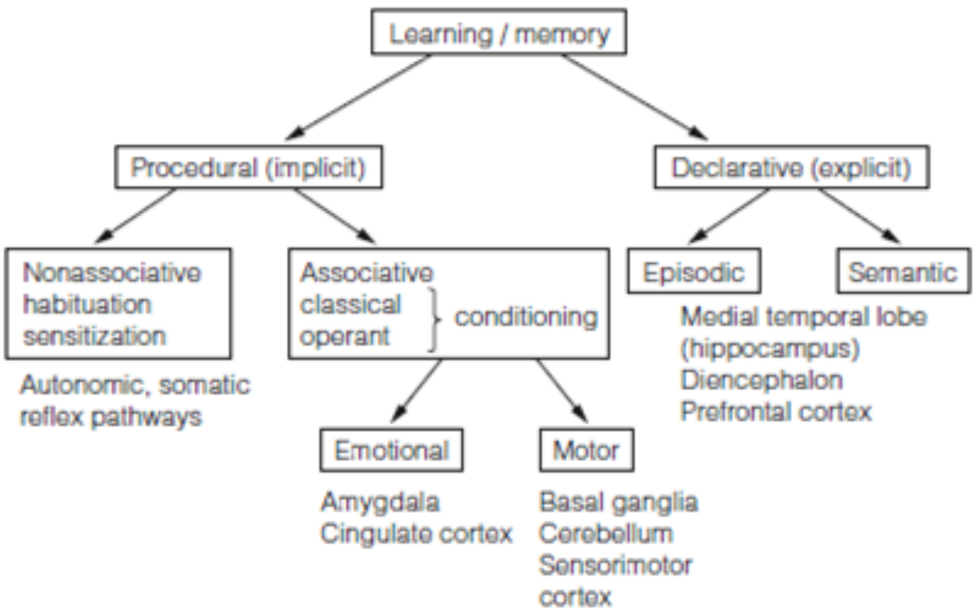


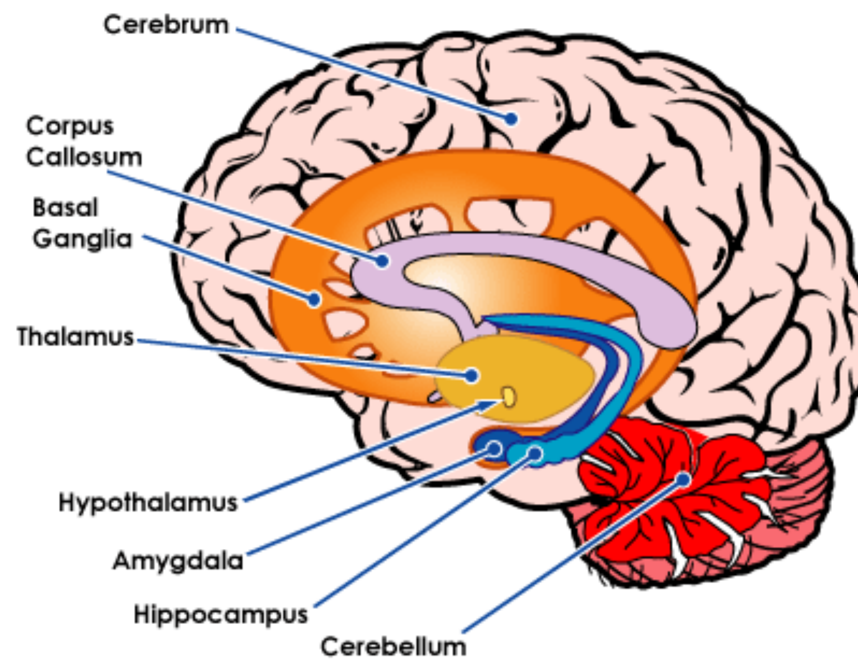
dopamine;
acetylcholine





Psychobiology of Learning



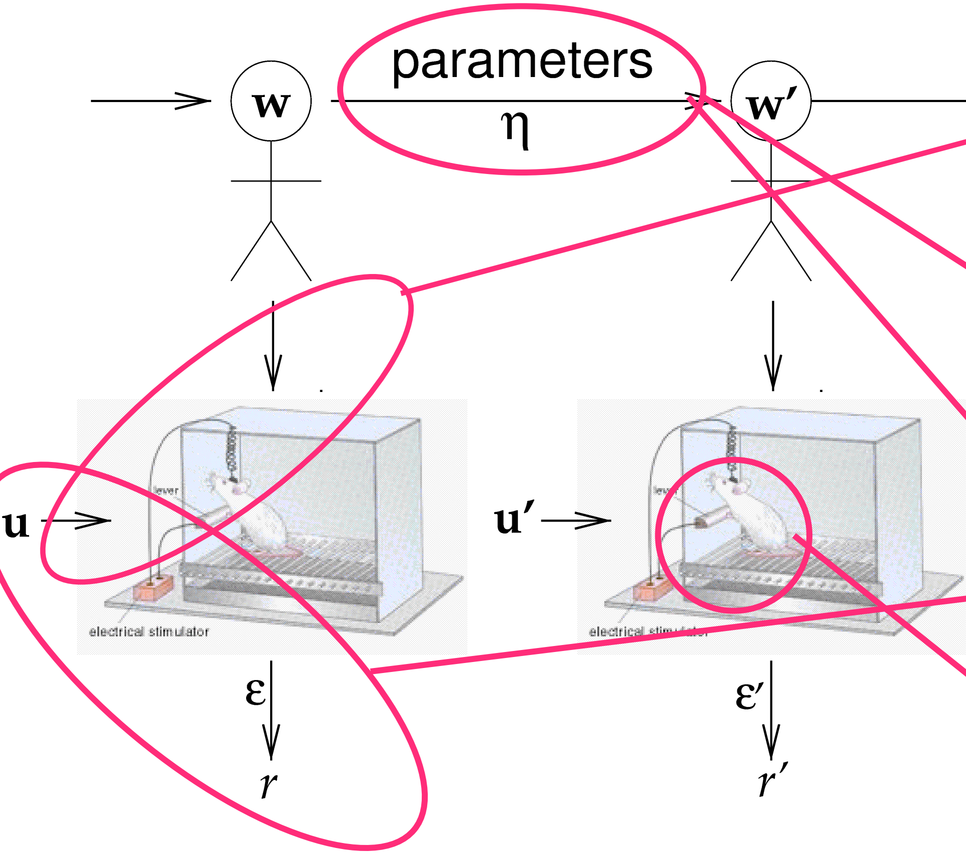


- representational learning
- ubiquitous learning of predictions
- forward/inverse models

Biological Learning

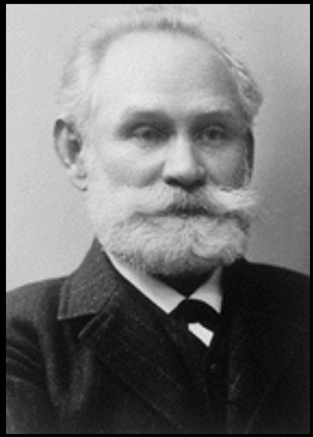
- conditioning and neural reinforcement learning
  - temporal difference learning and dopamine
  - uncertainty, acetylcholine and correlations
  - contexts and non-parametric Bayes
  - model-based, model-free and episodic RL
- representational learning
  - Hebb, PCA and infomax
  - deep learning and beyond

Computational Conditioning

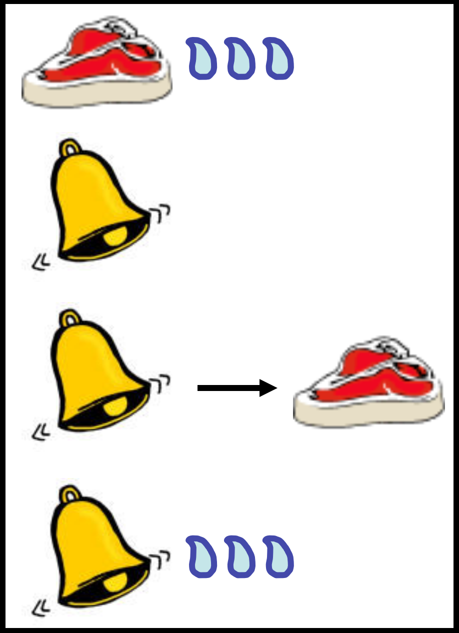


- output model
  - stimulus combination
- update model
  - expected/unexpected uncertainty
  - uncertainty-dependent learning rates
  - correlations
  - contexts
- decision model
  - action system competition
  - reinforcement learning

Layer 1: simple prediction learning



Ivan Pavlov



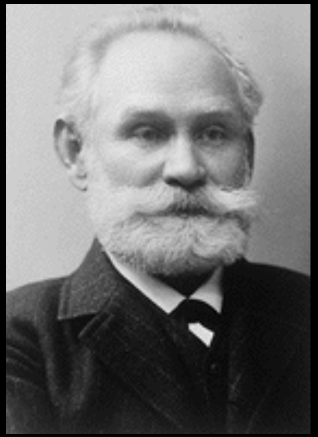
 = Unconditioned Stimulus

 = Conditioned Stimulus

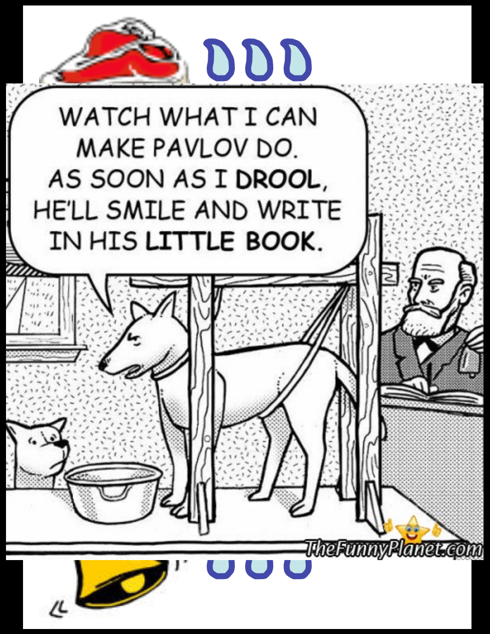
 = Unconditioned Response (reflex);
Conditioned Response (reflex)

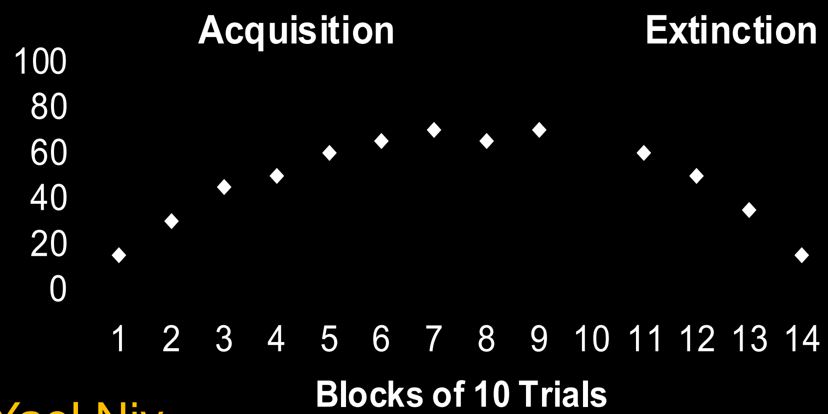
slides from Yael Niv

Animals learn predictions



Ivan Pavlov



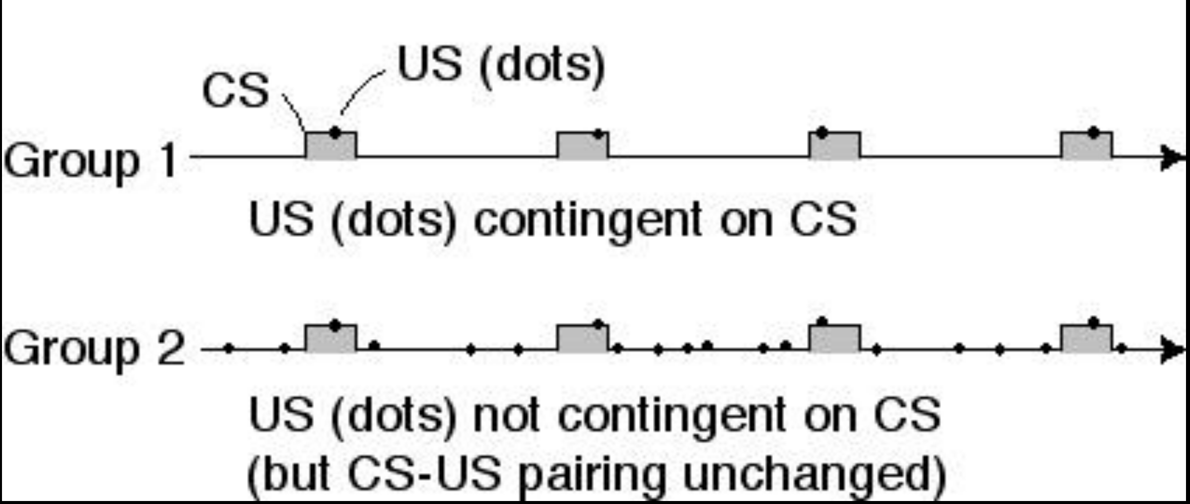


very general across species, stimuli, behaviors

slides from Yael Niv

But do they really?

1. Rescorla's control

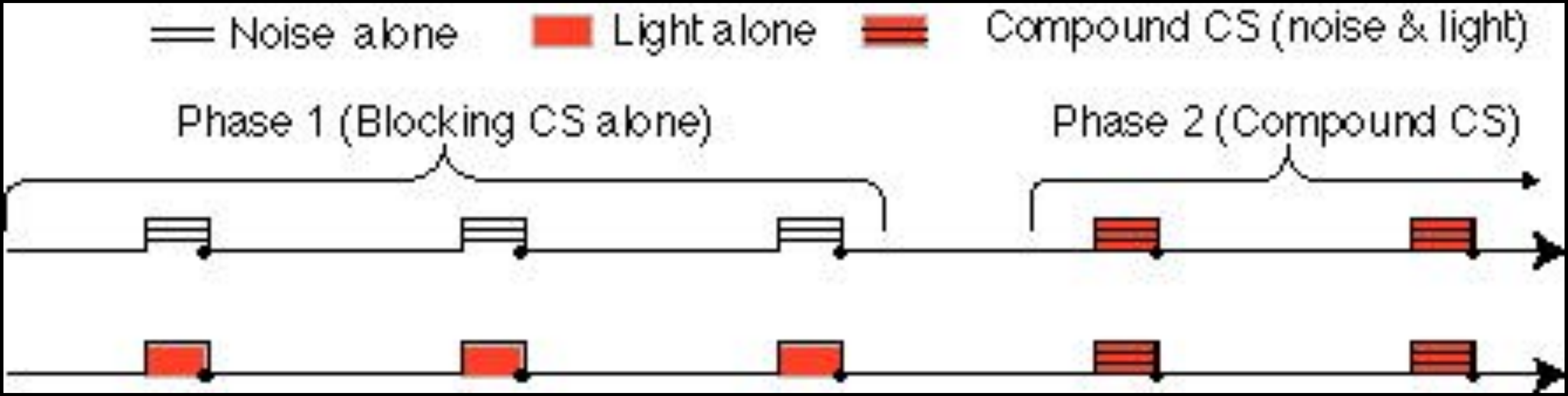


temporal contiguity is not enough - need contingency

P(food | light) > P(food | no light)

But do they really?

2. Kamin's blocking



contingency is not enough either… need surprise

slides from Yael Niv

Rescorla-Wagner

- delta rule:
  - $V(n) = \sum_i w_i u_i(n)$
  - $\delta(n) = r(n) - V(n)$
  - $\Delta w_i = \alpha_i(n)\delta(n)u_i(n)$

Assumptions:

- learning is driven by error (formalizes notion of surprise)
- summations of predictors is linear

A simple model - but very powerful!

- explains: gradual acquisition & extinction, blocking, overshadowing, conditioned inhibition, and more..
- predicted overexpectation
- associabilities

slides from Yael Niv

Rescorla-Wagner learning

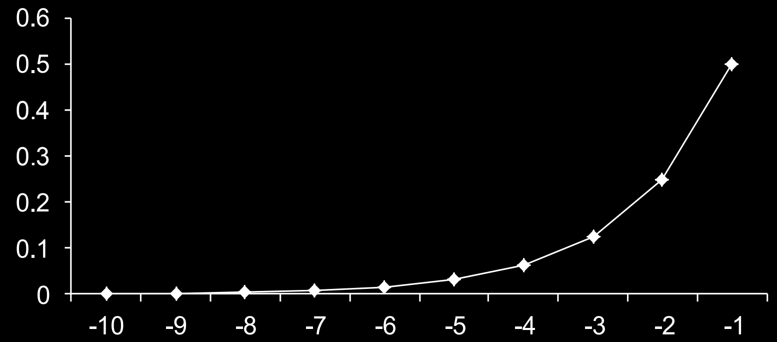
$$V_{t+1} = V_t + \eta\left(r_t - V_t\right)$$

how is the prediction on trial (t) influenced by rewards at times (t-1), (t-2), …?

$$V_{t+1} = (1-\eta)V_t + \eta r_t$$

$$V_t = \eta \sum_{i=1}^{t} (1-\eta)^{t-i} r_i$$

the R-W rule estimates expected reward using a weighted average of past rewards

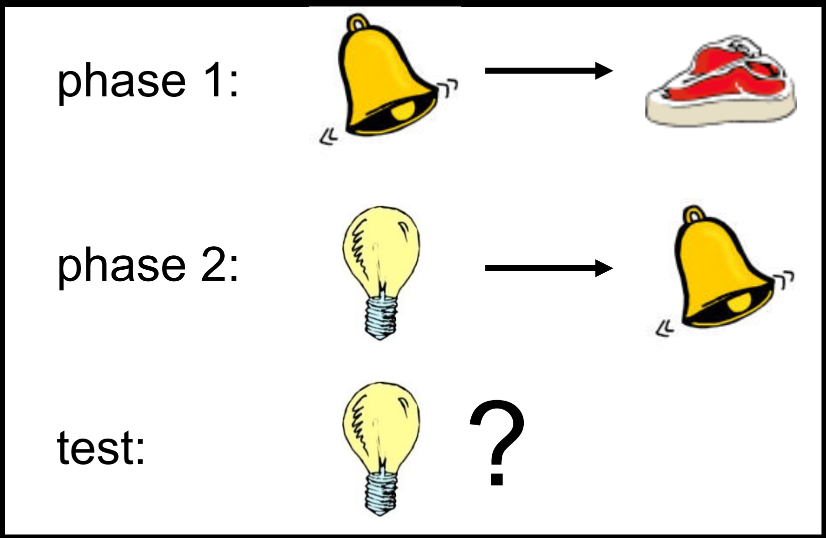


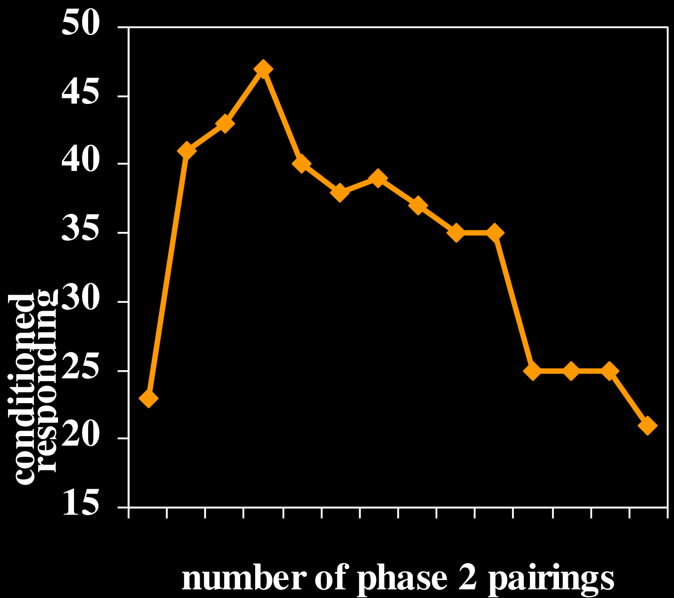
recent rewards weigh more heavily

learning rate = forgetting rate

slides from Yael Niv

But: second order conditioning





what would Rescorla-Wagner learning predict here?

animals learn that a predictor of a predictor is also a predictor of reward!
⇒ not interested solely in predicting immediate reward

slides from Yael Niv

need new formulation

Marr's 3 levels:

- The problem: optimal prediction of future reward

$$\boxed{V_t = E\left[\sum_{i=t}^{T} r_i\right]}$$

want to predict expected sum of future reward in a trial/episode

(N.B. here t indexes time within a trial)

- what's the obvious prediction error?

$$\boxed{\delta = r - V_{CS}}$$

$$\boxed{\delta_t = \sum_{i=t}^{T} r_i - V_t}$$

- what's the obvious problem with this?

slides from Yael Niv

lets start over: this time from the top

Marr's 3 levels:

- The problem: optimal prediction of future reward

$$V_t = E\left[\sum_{i=t}^{T} r_i\right]$$

want to predict expected sum of future reward in a trial/episode

$$V_t = E\left[r_t + r_{t+1} + r_{t+2} + \ldots + r_T\right]$$

Bellman eqn for policy evaluation

slides from Yael Niv

dopamine and prediction error

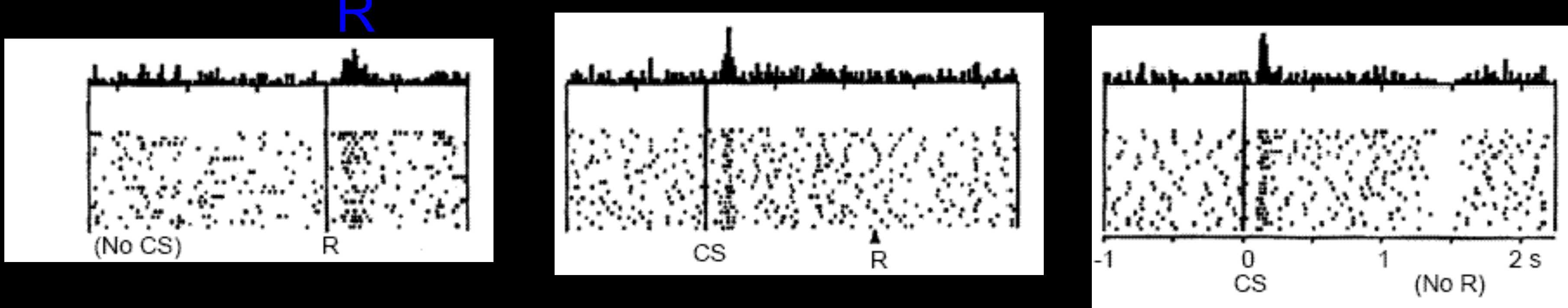
TD error $$\delta_t = r_t + V_{t+1} - V_t$$

L

V_t

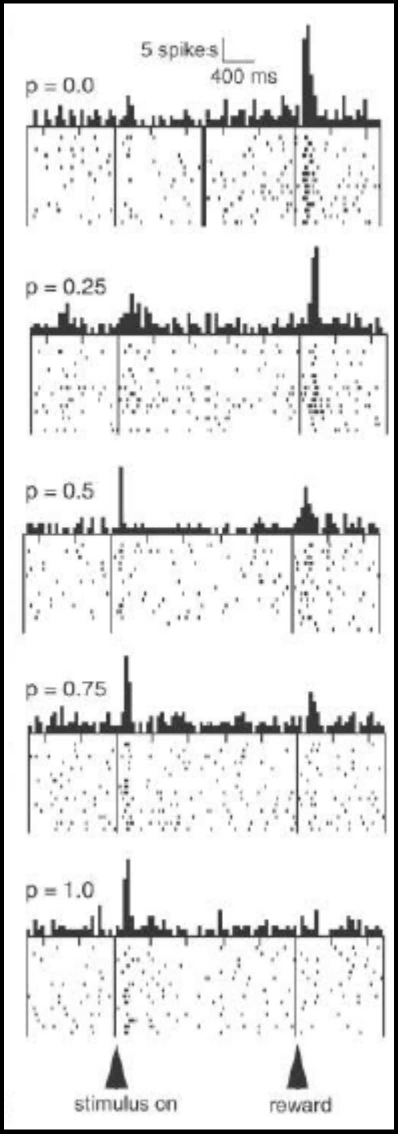
$\delta(t)$

R

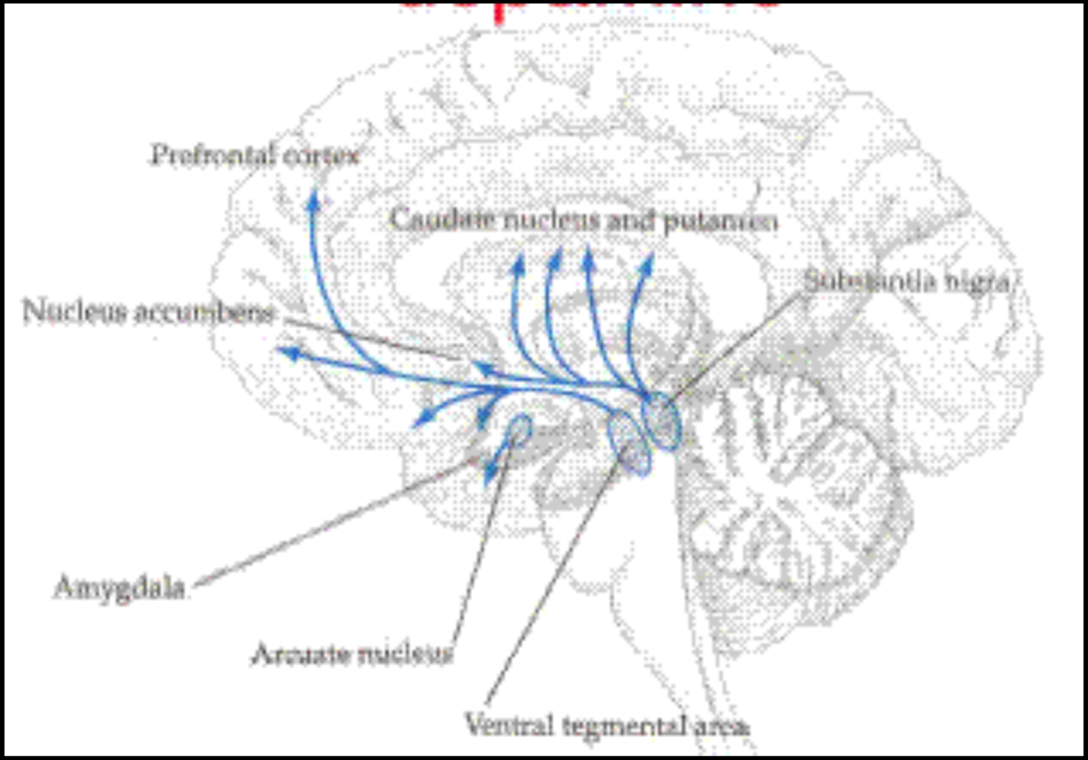


no prediction | prediction, reward | prediction, no reward

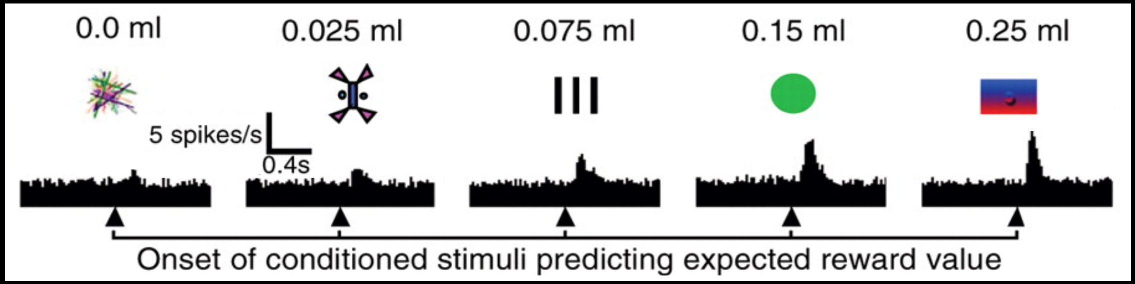
prediction error hypothesis of dopamine



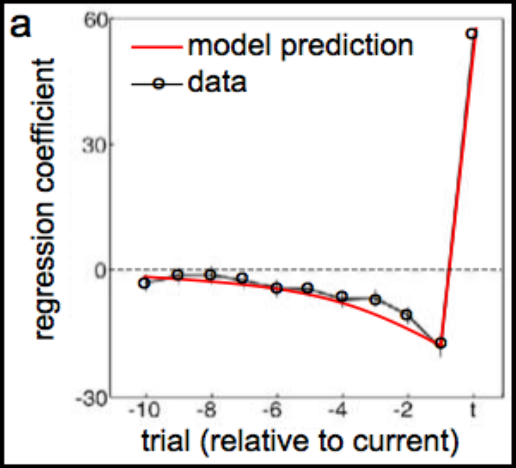
Fiorillo et al, 2003

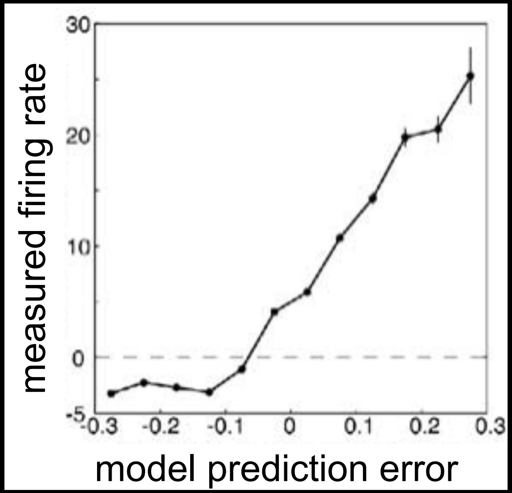


Tobler et al, 2005



prediction error hypothesis of dopamine





at end of trial: $\delta_t = r_t - V_t$ (just like R-W)

$$V_t = \eta \sum_{i=1}^{t} (1-\eta)^{t-i} r_i$$

Bayer & Glimcher (2005)

Risk Experiment

5 stimuli:

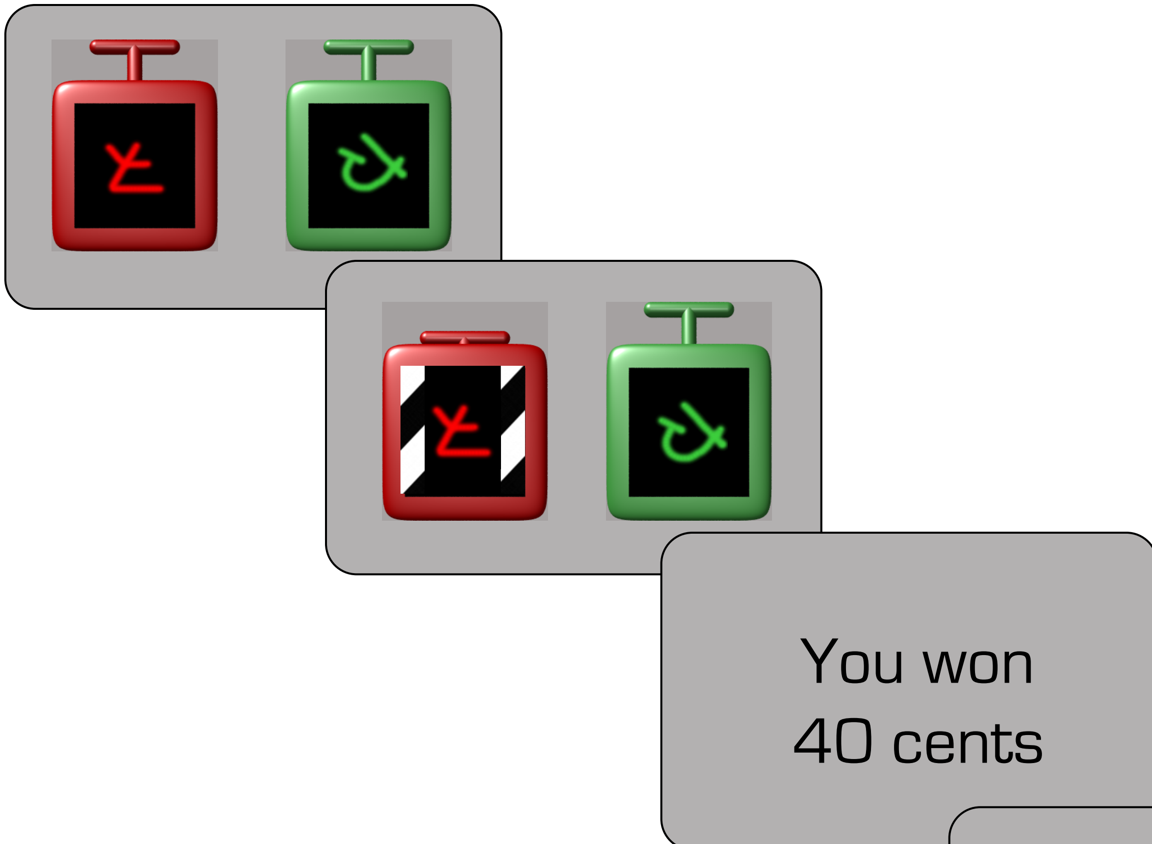
40¢

20¢

0/40¢

0¢

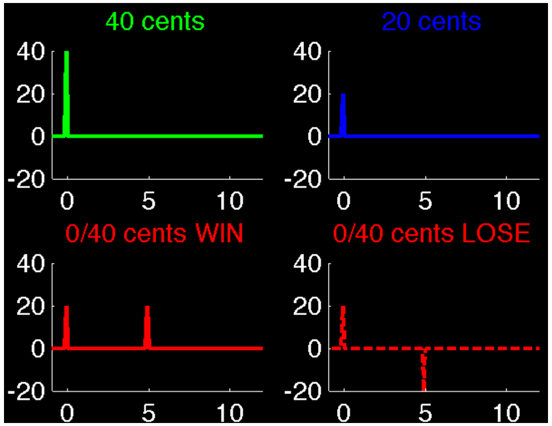
0¢

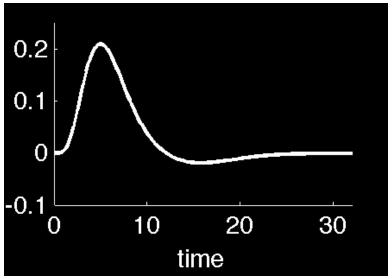


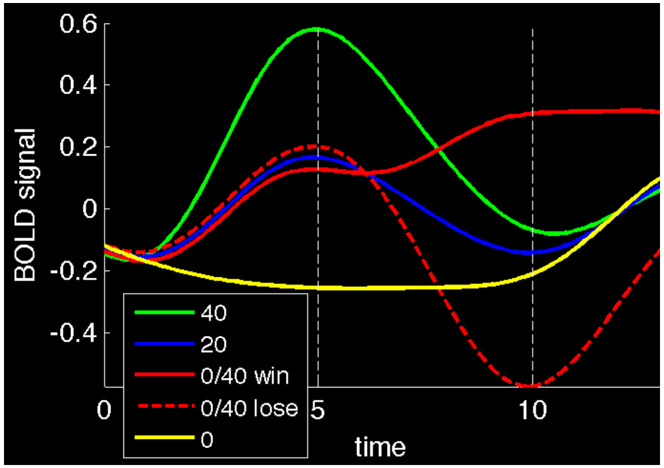
19 subjects (dropped 3 non learners, N=16)
3T scanner, TR=2sec, interleaved
234 trials: 130 choice, 104 single stimulus
randomly ordered and counterbalanced

Neural results: Prediction Errors

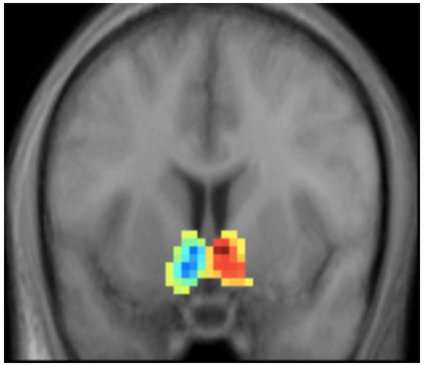
what would a prediction error look like (in BOLD)?



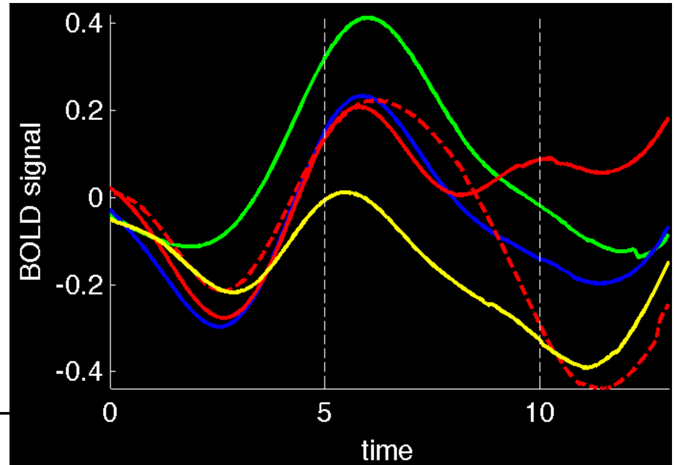


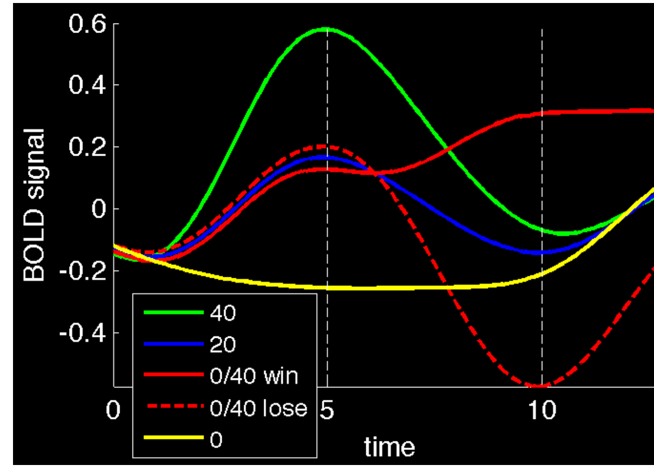


Neural results: Prediction errors in NAC



unbiased anatomical ROI
in nucleus accumbens
(marked per subject*)





raw BOLD
(avg over all subjects)

can actually decide between different neuroeconomic models of risk

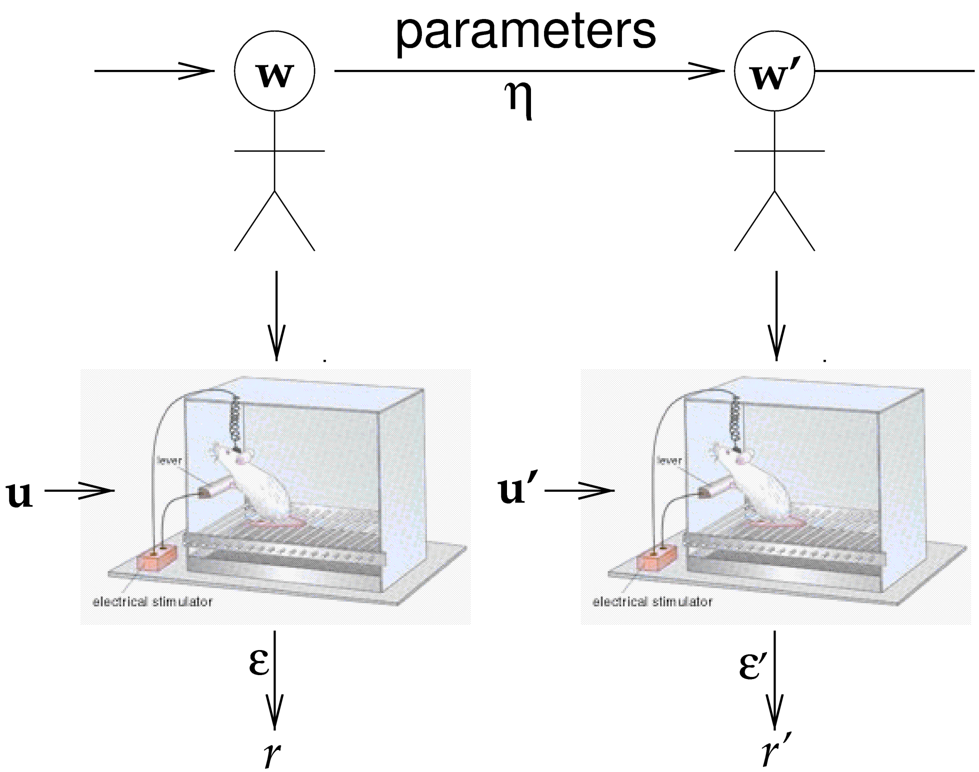
CARTOON BY MICHAEL MITTAG, WWW.COOLRISK.COM

* thanks to Laura deSouza

Biological Learning

- conditioning and neural reinforcement learning
  - temporal difference learning and dopamine
  - uncertainty, acetylcholine and correlations
  - contexts and non-parametric Bayes
  - model-based, model-free and episodic RL
- representational learning
  - Hebb, PCA and infomax
  - deep learning and beyond

Kalman Filter



expt $\mathbf{w}' = \mathbf{w} + \boldsymbol{\eta}$

reward given $r = \mathbf{w} \cdot \mathbf{u} + \epsilon$

allowable drift $\boldsymbol{\eta} \sim N[0, \sigma^2 \mathbb{I}]$

output noise $\epsilon \sim N[0, \rho^2]$

- Markov random walk (or OU process)
- no punctate changes
- additive model of combination
- forward inference

Kalman Posterior

The Kalman filter maintains uncertainty:

$$P(V) = \mathcal{N}[\hat{\mathbf{w}} \cdot \mathbf{u}, \mathbf{u} \cdot \Sigma \cdot \mathbf{u}]$$

where

ε

η

Assumed Density KF

Diagonal approx to $\Sigma = \text{diag}(\sigma_i^2)$

If $\mathbf{w} \sim \mathcal{N}\left[\hat{\mathbf{w}}, \text{diag}(\sigma_i^2)\right]$, then

$$\Delta \widehat{w}_i = \frac{\sigma_i^2}{\sum_j \sigma_j^2 + \rho^2}(r - \mathbf{u} \cdot \hat{\mathbf{w}})u_i$$

- Rescorla-Wagner error correction
- competitive allocation of learning
  - Pearce & Hall

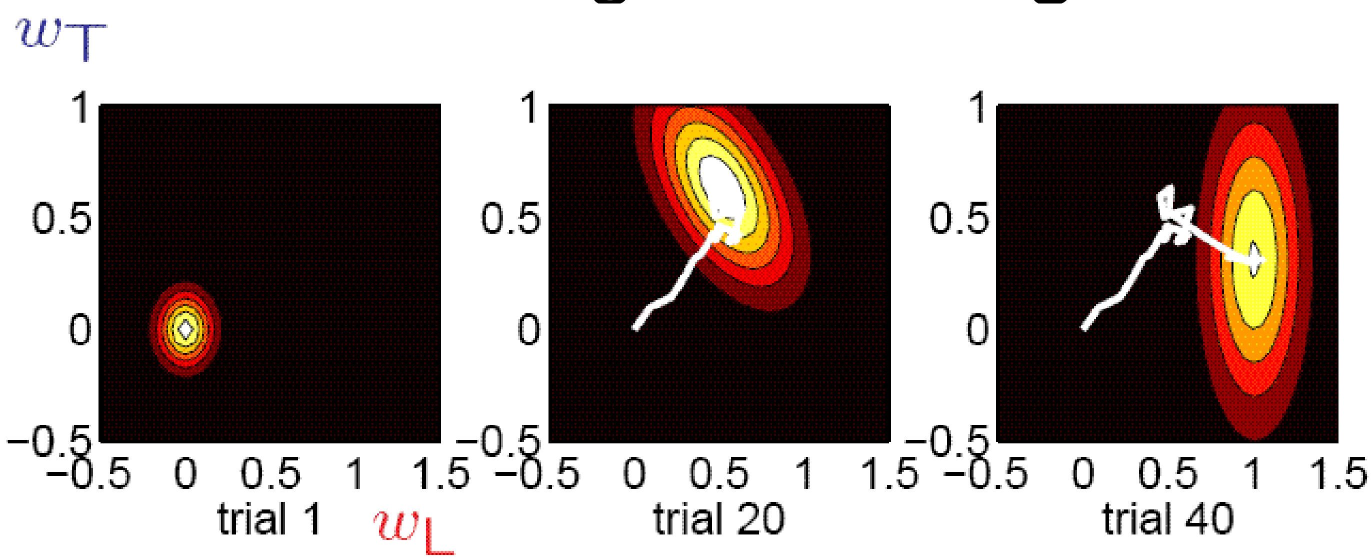
Blocking

forward	L→r	L+T→r	T→ ·
backward	L+T→r	L→r	$T \rightarrow \cdot$

- forward blocking: error correction

$$\cdot(r - \mathbf{u}\cdot\hat{\mathbf{w}})$$

- backward blocking: -ve **off-diag** $\Sigma_{LT} < 0$



Mackintosh vs P&H

- under diagonal approximation:

$$\mathrm{E}\,(r - \mathbf{u} \cdot \widehat{\mathrm{w}})^2 = \rho^2 + \sum_j \sigma_j^2 u_i^2$$

- for slow learning,

σ_j^2 changes with correlation of $(r - V)$ and u_i

  – effect like Mackintosh

Summary

- Kalman filter models many standard conditioning paradigms
- elements of RW, Mackintosh, P&H
- but:
  - downwards unblocking

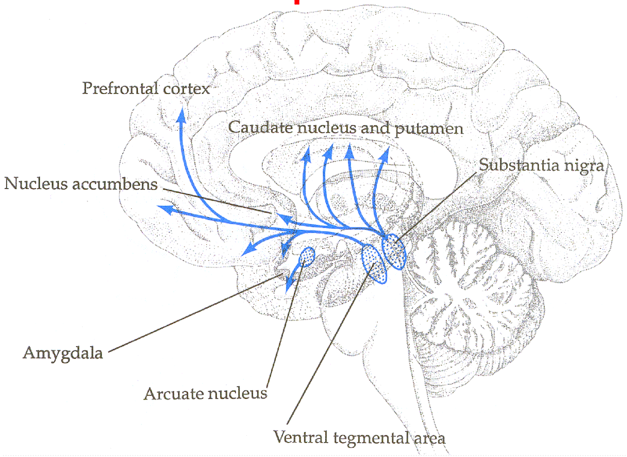
    $$L \to r\Delta r \qquad L + T \to r \qquad T \nrightarrow \pm r$$

    predictor competition

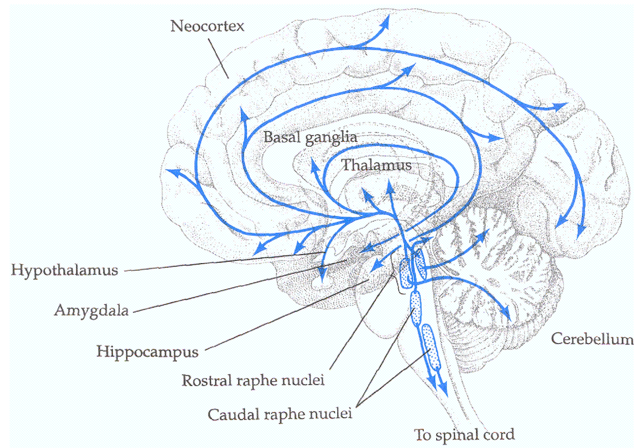
  - representational learning $L \to r$; $T \to r$; $L+T \to \cdot$
- recency vs primacy (Kruschke)

How are Learning Rates Implemented?

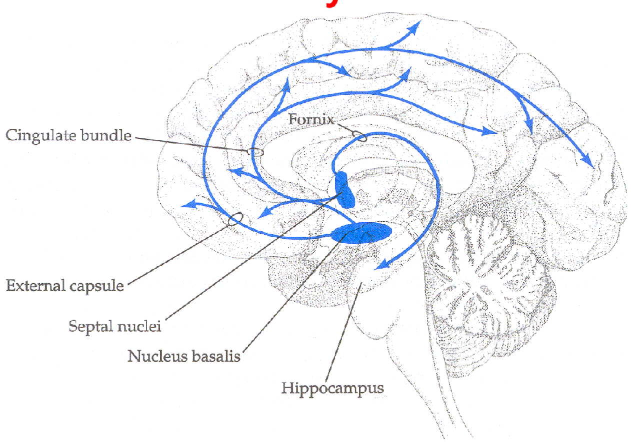
dopamine



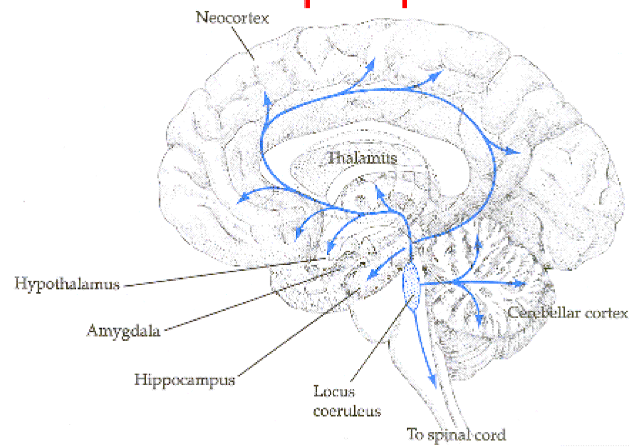
5HT



acetylcholine



norepinephrine



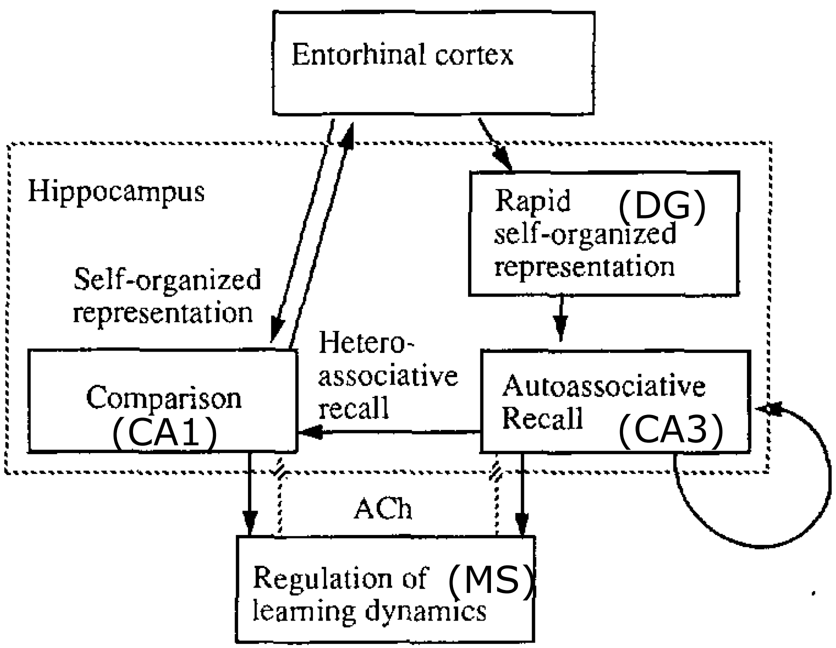
general excitability, signal/noise ratios

specific prediction errors, uncertainty signals

ACh in Hippocampus

Given ***unfamiliarity***, **ACh**:

- ***boosts*** bottom-up, ***suppresses*** recurrent processing
- ***boosts*** recurrent plasticity



(Hasselmo, 1995)

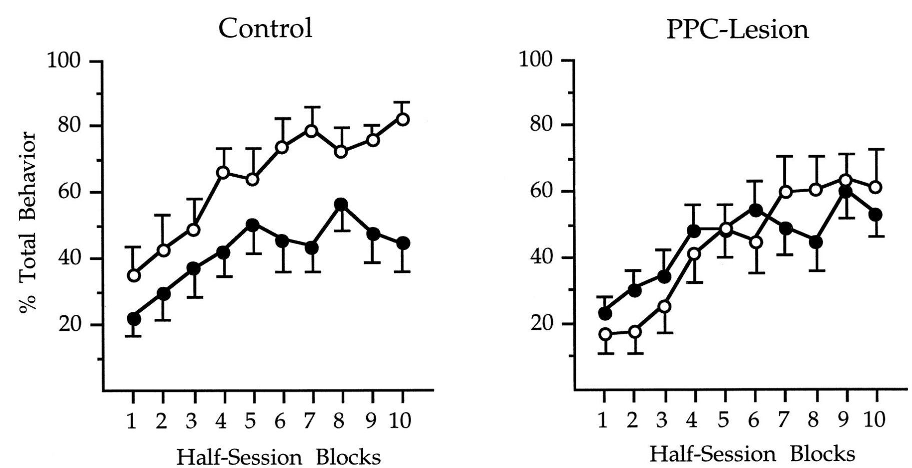
ACh in Conditioning

Given ***uncertainty***, **ACh**:

- ***boosts*** learning to stimuli of uncertain consequences

Table 1. Outline of procedures for Experiment 1

Treatment condition (groups)	Phase 1: consistent L–T relation	Phase 2: experimental change in L–T relation	Phase 3: test of conditioning to L
Consistent (CTL-C, PPC-C)	L → T → food; L → T	L → T → food; L → T	L → food
Shift (CTL-S, PPC-S)	L → T → food; L → T	L → T → food; L	L → food



(Bucci, Holland, & Galllagher, 1998)

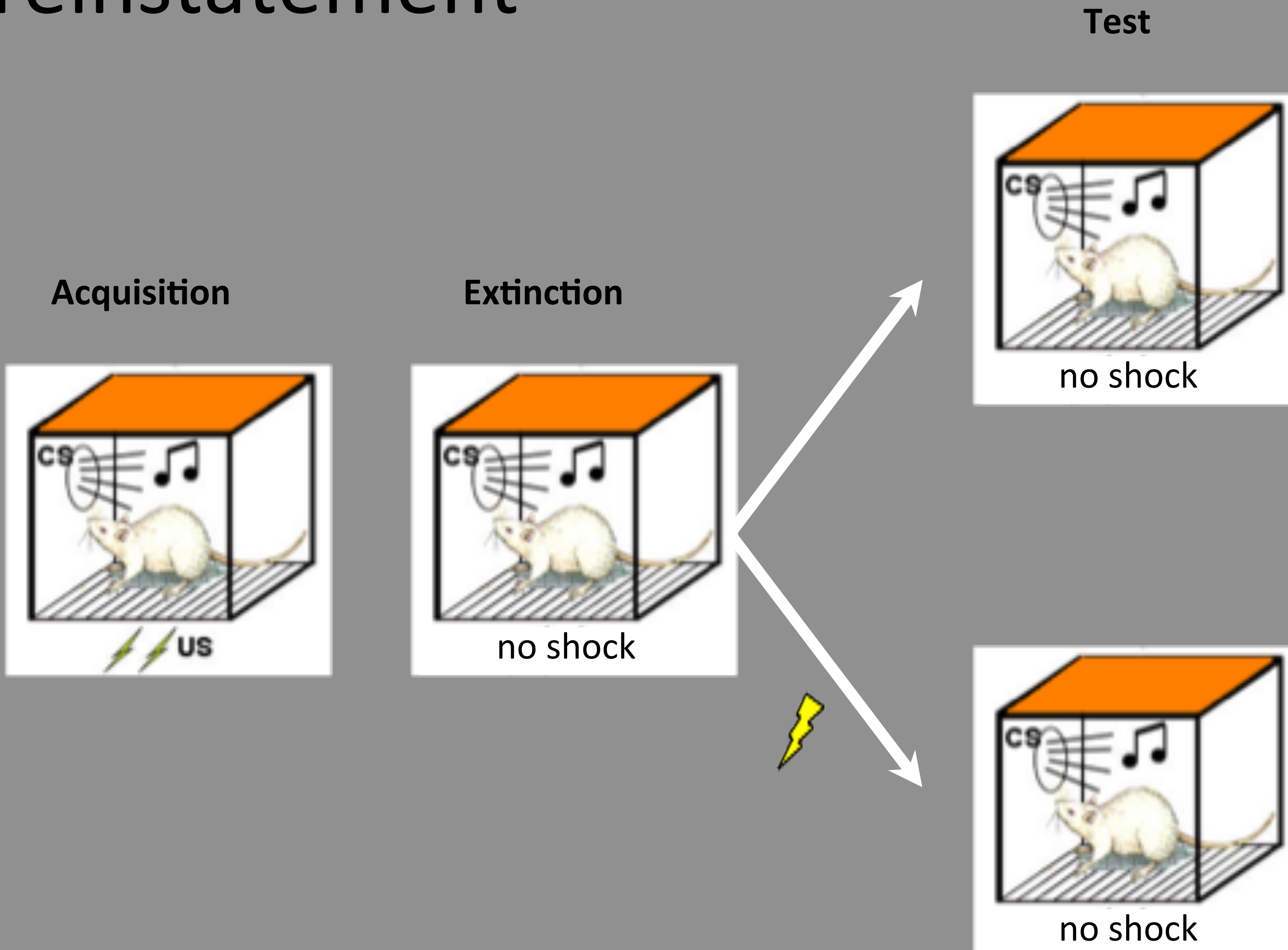
Uncertainty and Learning

- faster learning for more expected uncertainty
- cholinergic substrate – but cortical representations also
- animals seem to elide reducible and irreducible uncertainty
- what about unexpected uncertainty?

Biological Learning

- conditioning and neural reinforcement learning
  - temporal difference learning and dopamine
  - uncertainty, acetylcholine and correlations
  - contexts and non-parametric Bayes
  - model-based, model-free and episodic RL
- representational learning
  - Hebb, PCA and infomax
  - deep learning and beyond

reinstatement



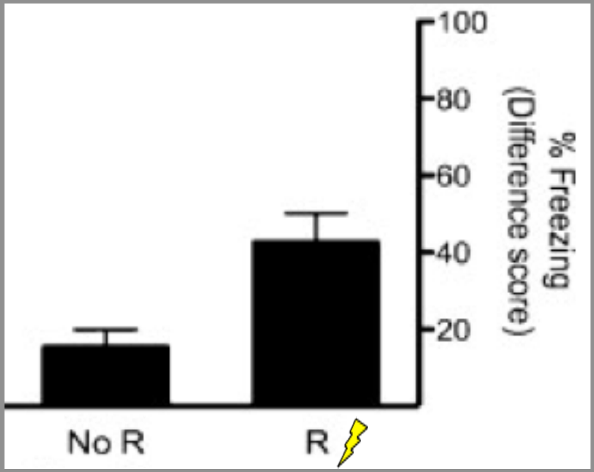
slides from Yael Niv

extinction ≠ unlearning

Acquisition

Extinction

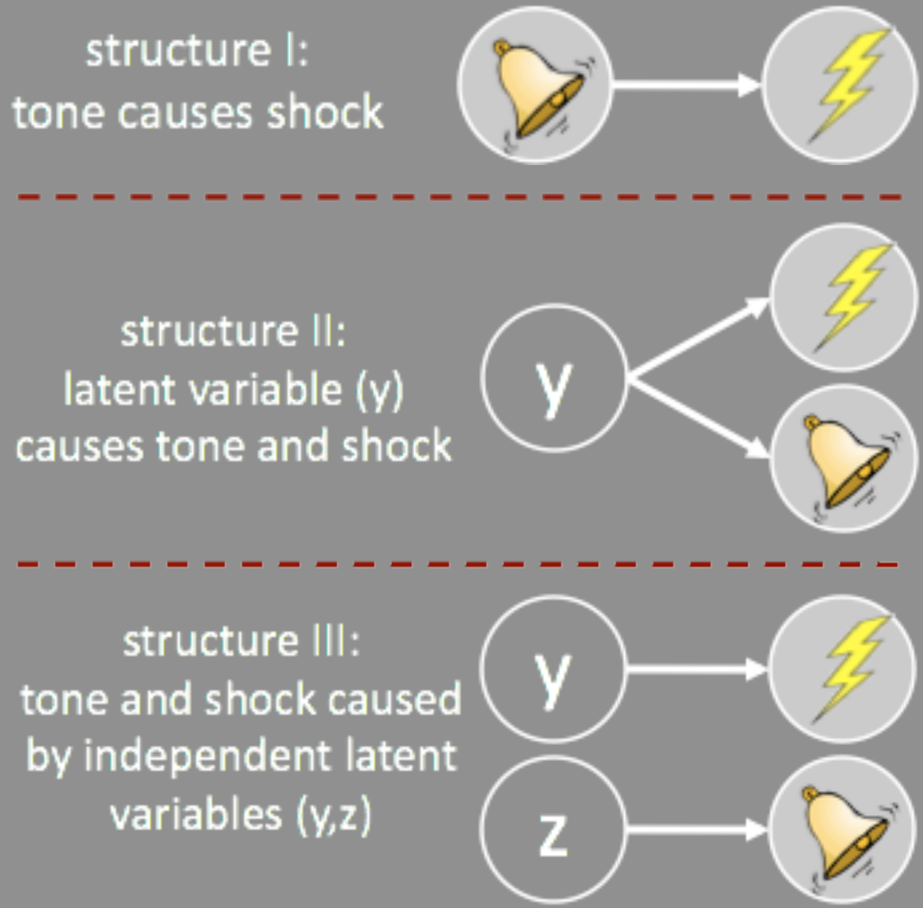
Test



Storsve, McNally & Richardson, 2012

slides from Yael Niv

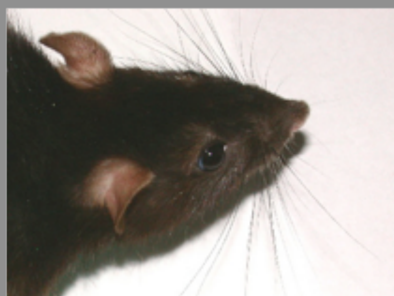
learning causal structure: Gershman & Niv

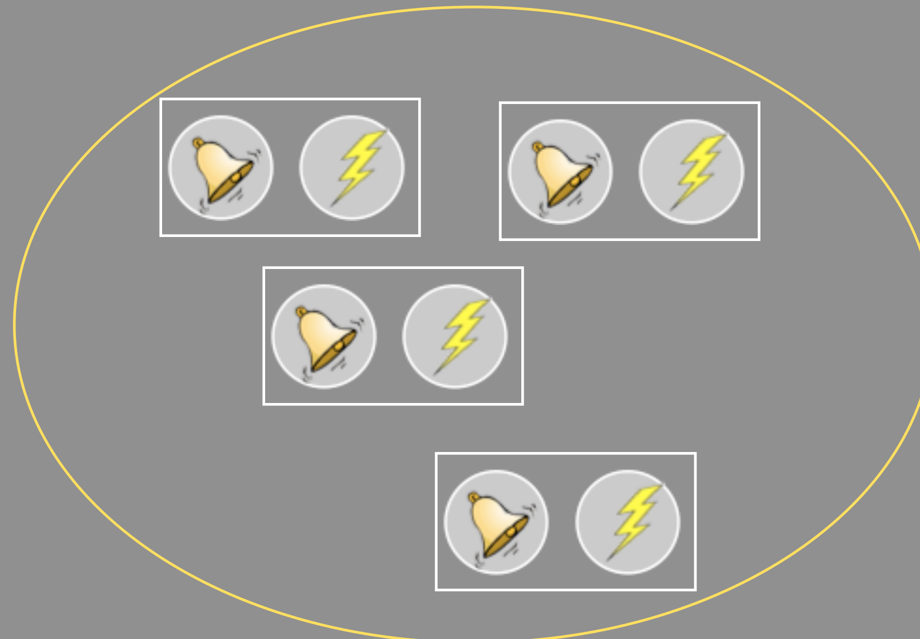


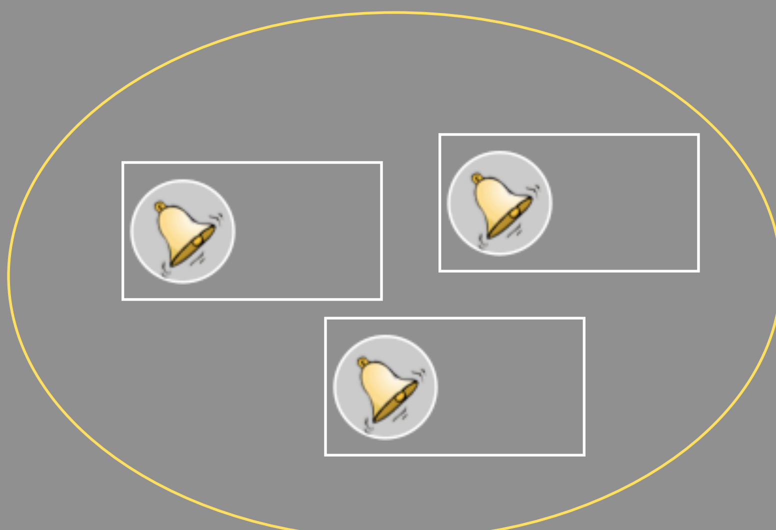
Sam Gershman

conditioning as clustering: DPM

Gershman & Niv;
Daw & Courville; Redish



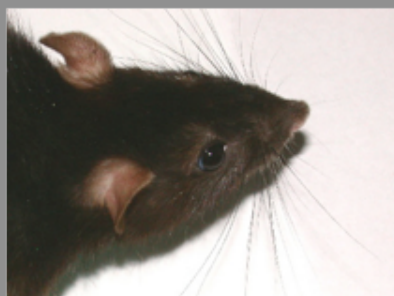


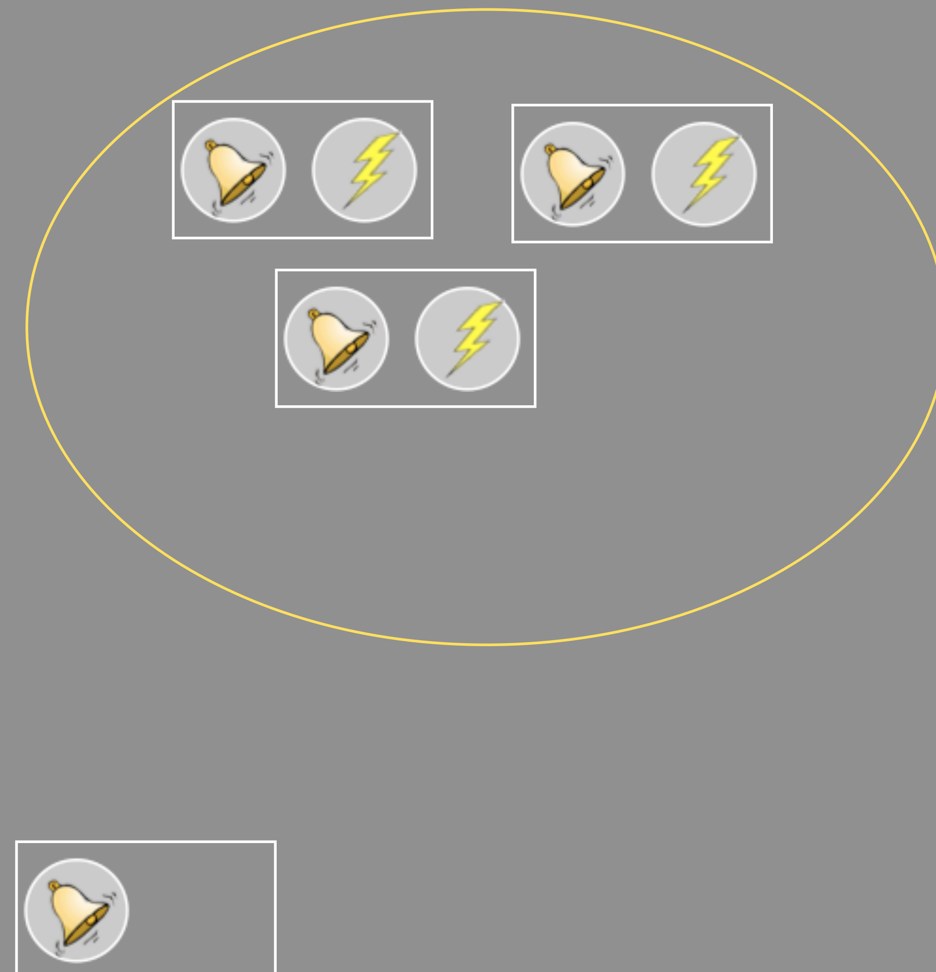


Within each cluster:
“learning as usual”
(Rescorla-Wagner, RL etc.)

associative learning versus state learning

Gershman & Niv





structural learning

(create new state)

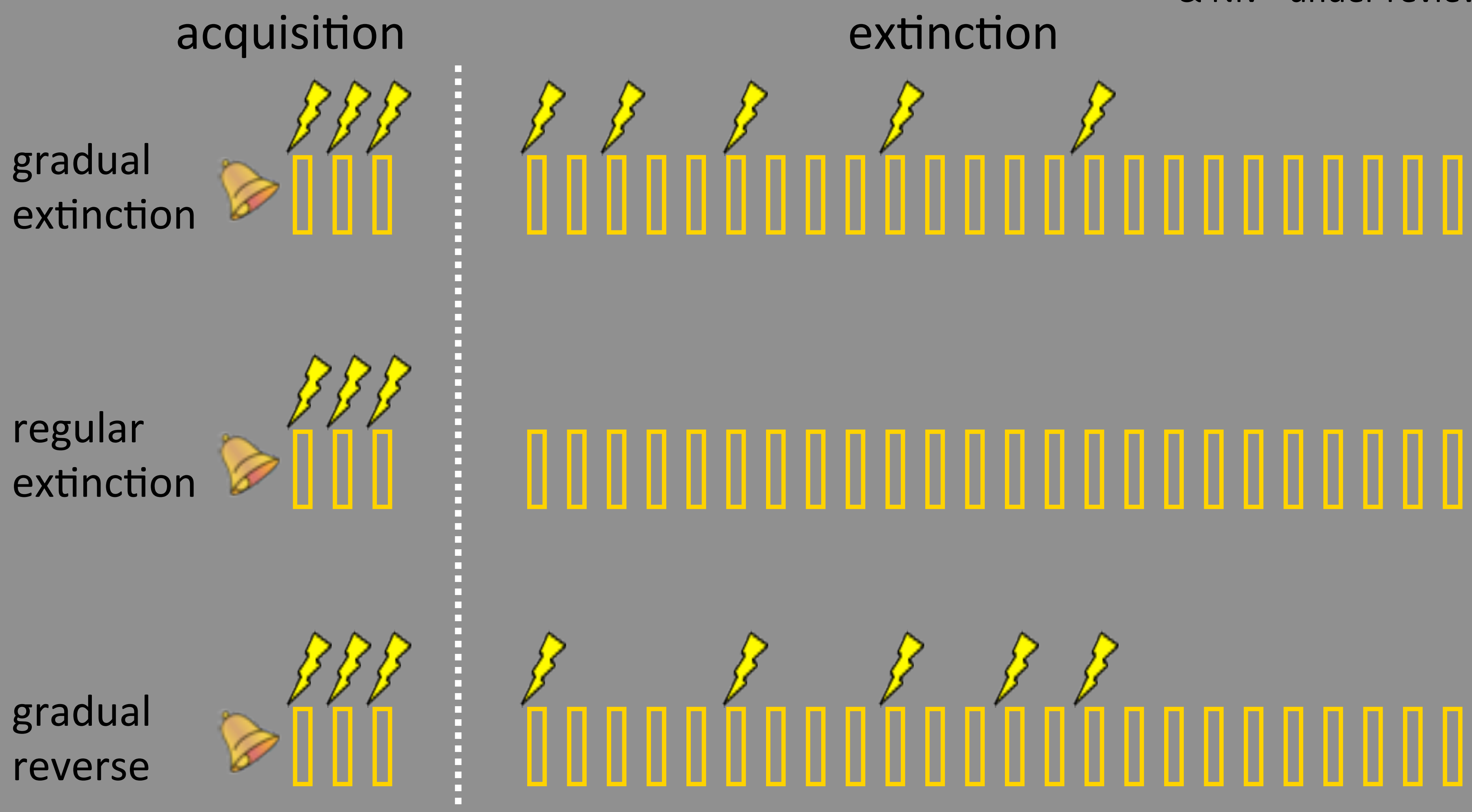
how to erase a fear memory

hypothesis: prediction errors (dissimilar data) lead to new states

what if we make extinction a bit more similar to acquisition?

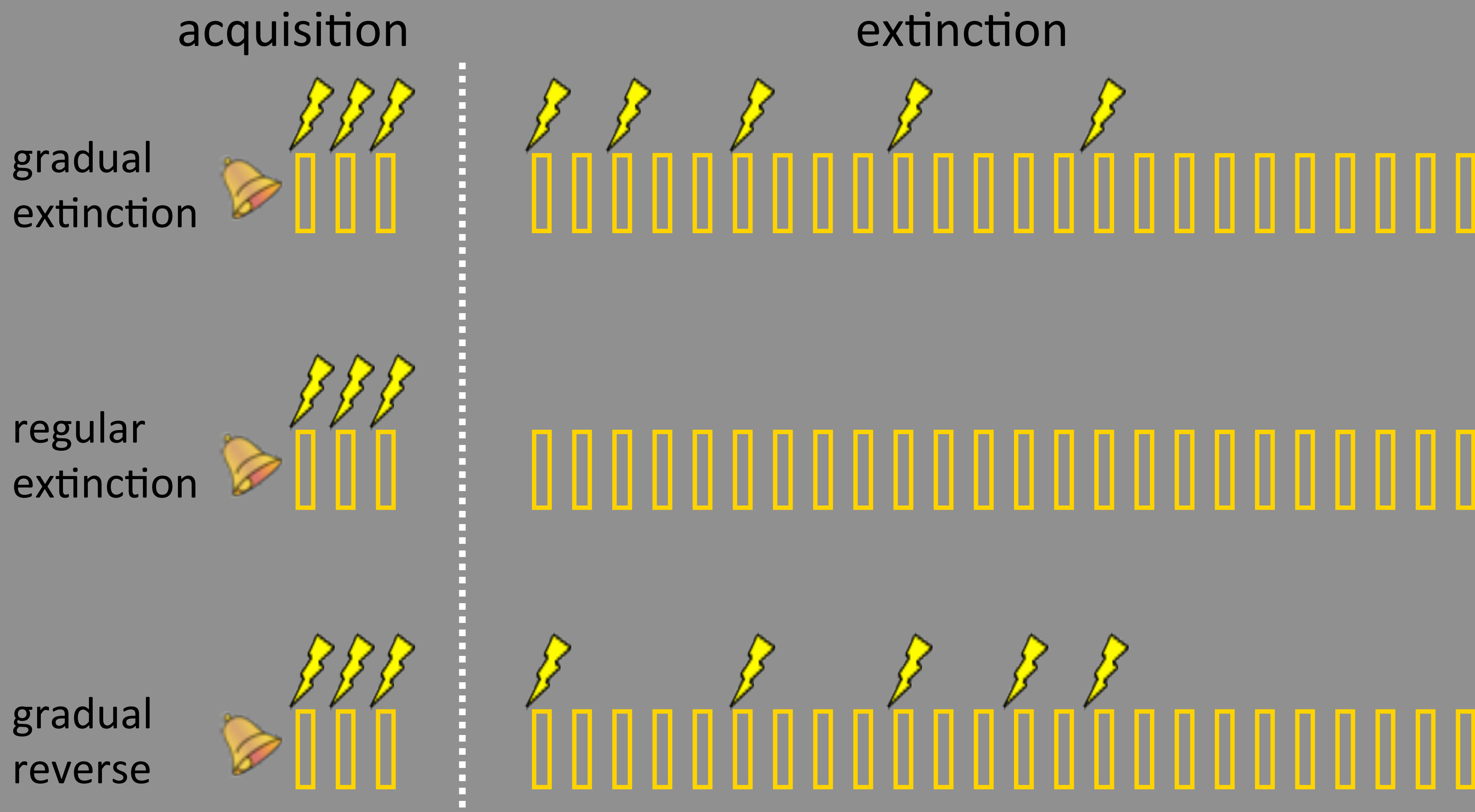
gradual extinction

Gershman, Jones, Norman, Monfils & Niv - under review



gradual extinction

Gershman, Jones, Norman, Monfils & Niv - under review

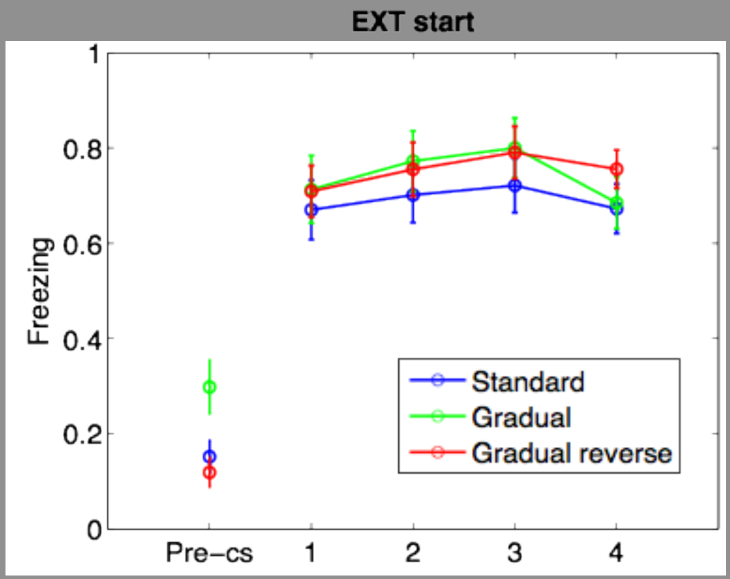


test one day (reinstatement) or 30 days later (spontaneous recovery)

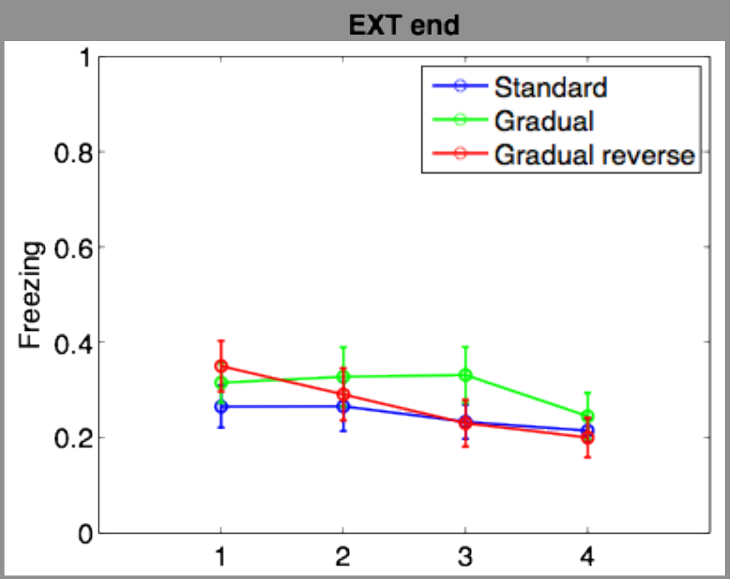
gradual extinction

Gershman, Jones, Norman, Monfils & Niv - under review

first trials of extinction

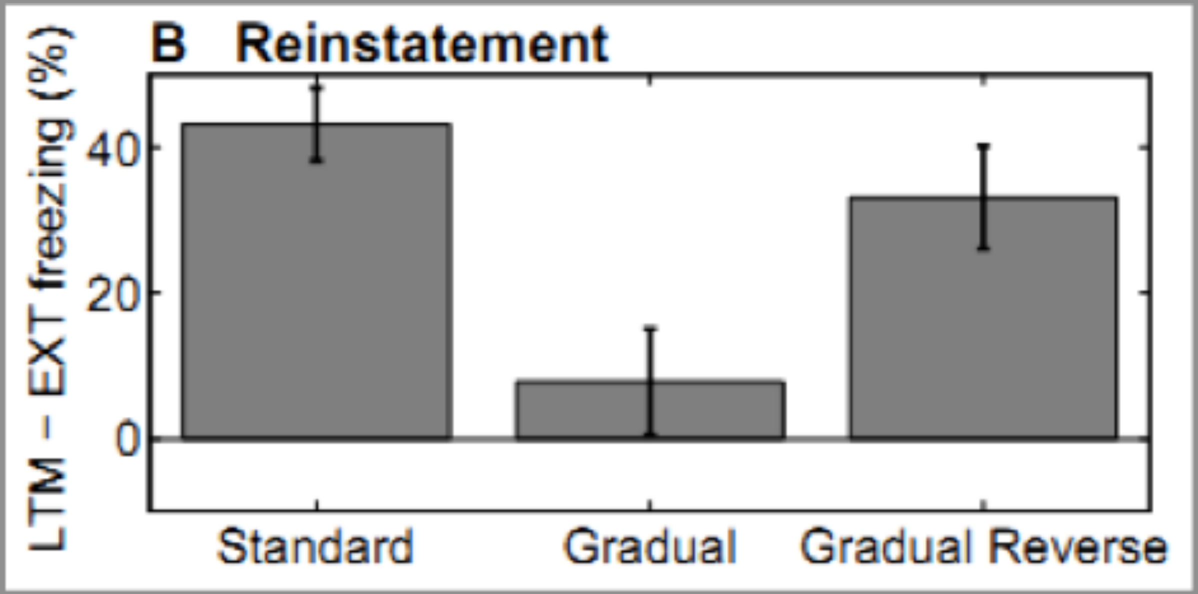


last trials of extinction



gradual extinction

Gershman, Jones, Norman, Monfils
& Niv - under review



only gradual extinction group shows no reinstatement

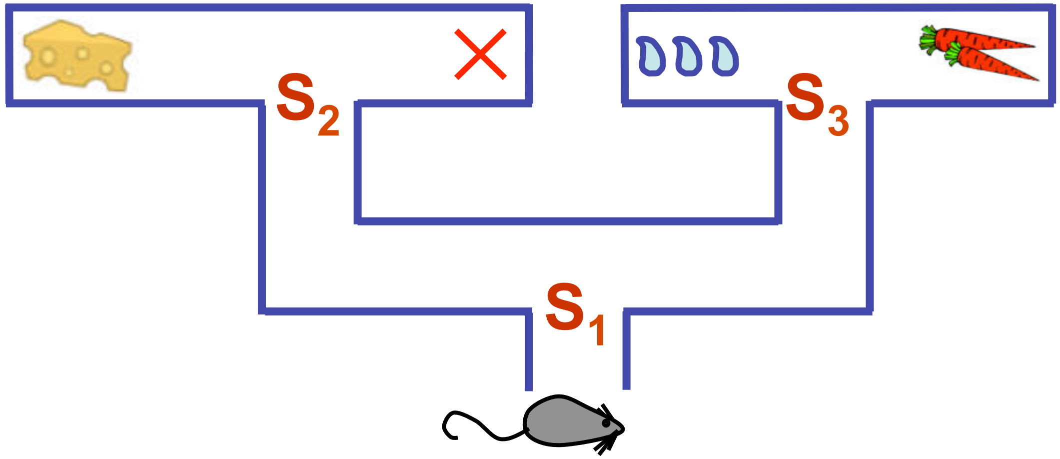
unexpected uncertainty

- stability-plasticity dilemma (Grossberg)
  - solved by clustering
- realization:
  - norepinephrine – neural interrupt
  - orbitofrontal cortex
- NPB – sensitive to prior for novel context
- explains surprising effects in extinction, reconsolidiation, etc

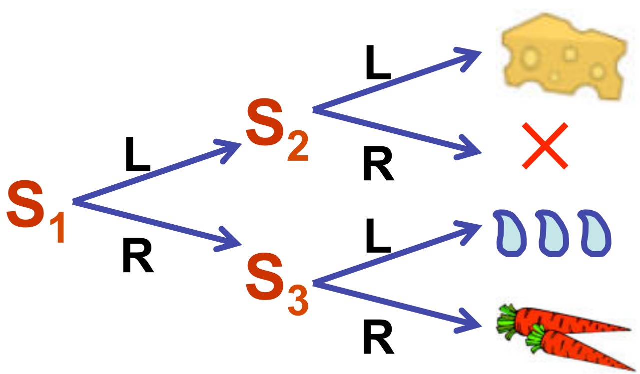
Biological Learning

- conditioning and neural reinforcement learning
  - temporal difference learning and dopamine
  - uncertainty, acetylcholine and correlations
  - contexts and non-parametric Bayes
  - model-based, model-free and episodic RL
- representational learning
  - Hebb, PCA and infomax
  - deep learning and beyond

Reinforcement Learning



forward model (goal directed)



Hunger	Thirst
= 4	= 2
= 0	= 0
= 2	= 4
= 3	= 1

caching (habitual)

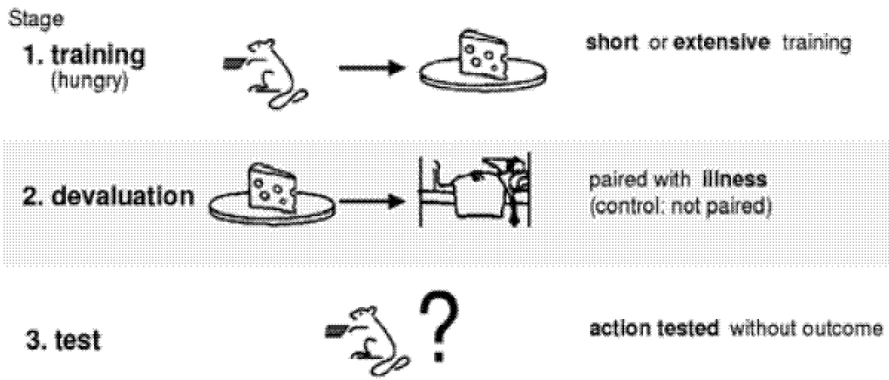
(NB: trained hungry)

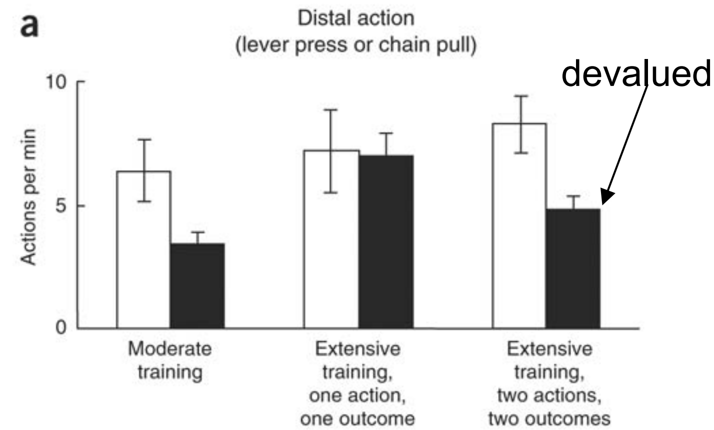
- $H;S_1,L \rightarrow 4$
- $H;S_1,R \rightarrow 3$
- $H;S_2,L \rightarrow 4$
- $H;S_2,R \rightarrow 0$
- $H;S_3,L \rightarrow 2$
- $H;S_3,R \rightarrow 3$

acquire with simple learning rules

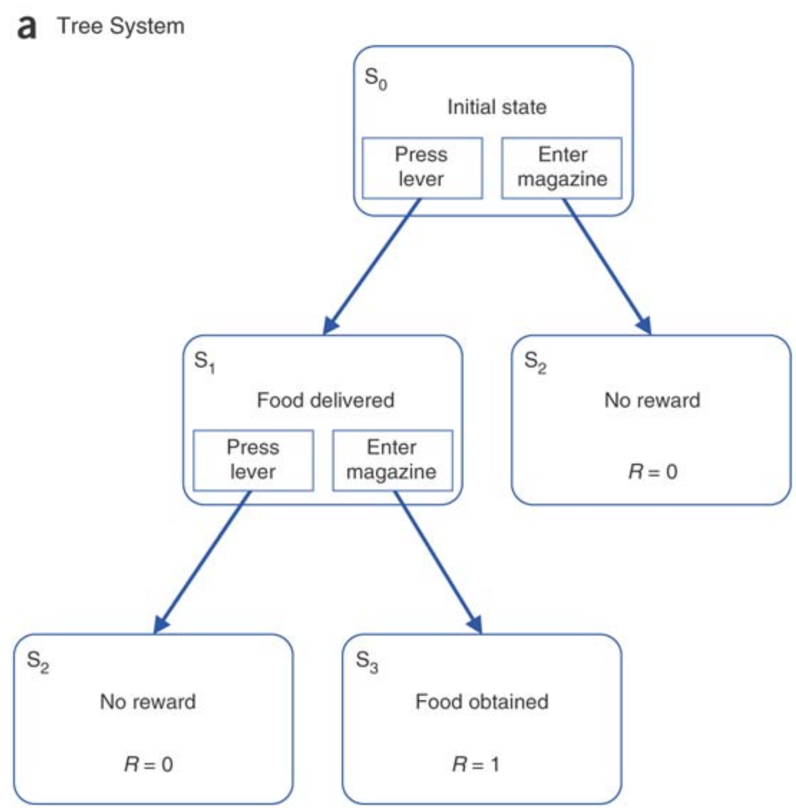
acquire recursively

Two Systems:

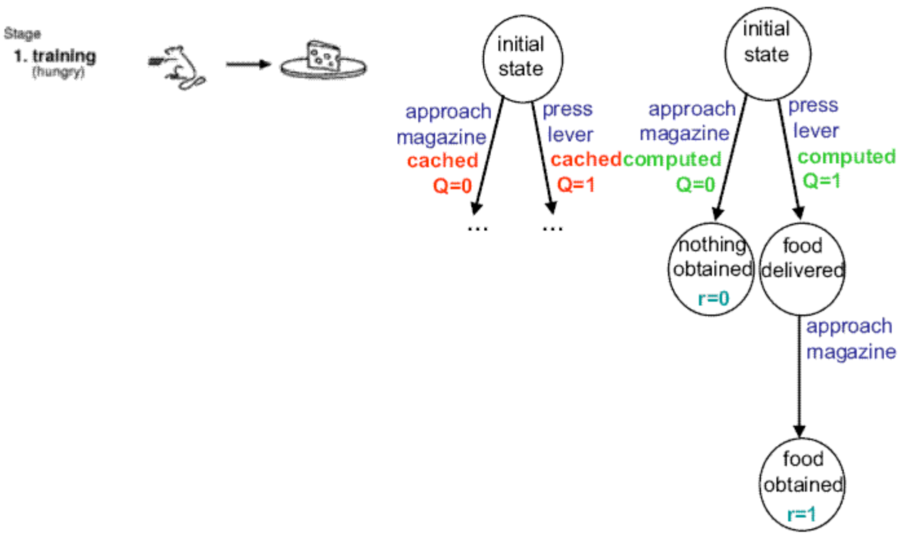

Stage
1. training
(hungry)
short or extensive training
2. devaluation
paired with illness
(control: not paired)
3. test
?
action tested without outcome

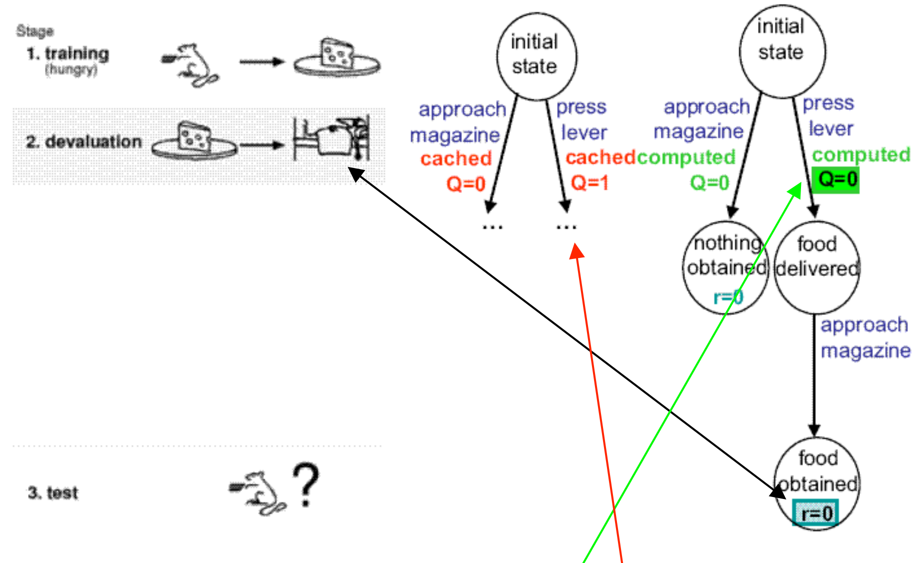

a
Distal action
(lever press or chain pull)
devalued
Actions per min
10
5
0
Moderate training
Extensive training, one action, one outcome
Extensive training, two actions, two outcomes


b Cache system
S0
Initial state
Press lever Q = 1
Enter magazine Q = 0
S1
Food delivered
Press lever Q = 0
Enter magazine Q = 1
S2
No reward
Q = 0
S3
Food obtained
Q = 1


a Tree System
S0
Initial state
Press lever
Enter magazine
S1
Food delivered
Press lever
Enter magazine
S2
No reward
R = 0
S2
No reward
R = 0
S3
Food obtained
R = 1

Behavioural Effects





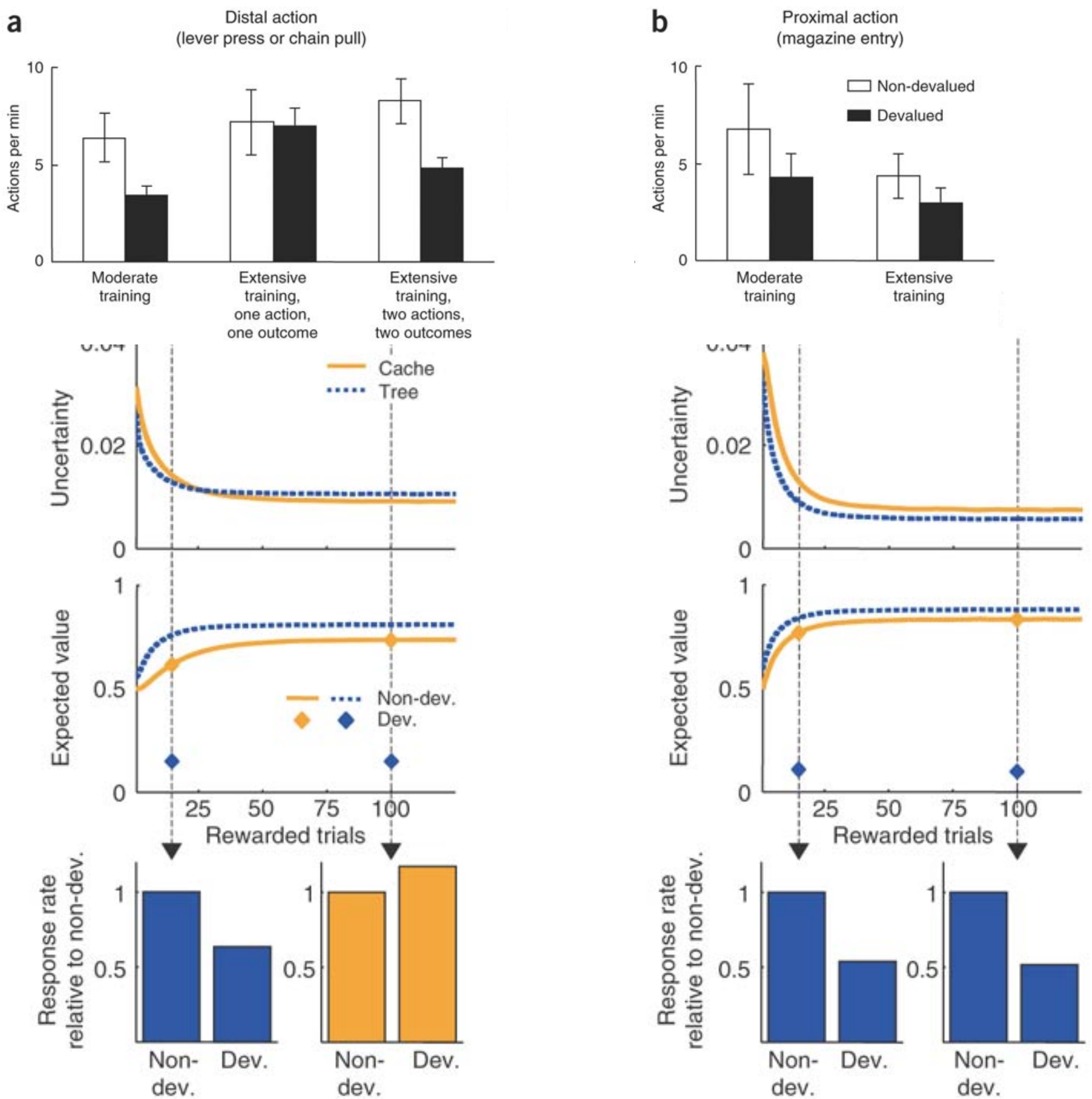
- Actions based on model will decline
- Actions based on model-free will persist

Learning

- uncertainty-sensitive learning for both systems:
  - model-based:
    - data efficient
    - computationally ruinous
  - model-free:
    - data inefficient
    - computationally trivial
  - uncertainty-sensitive control migrates from actions to habits

One Outcome

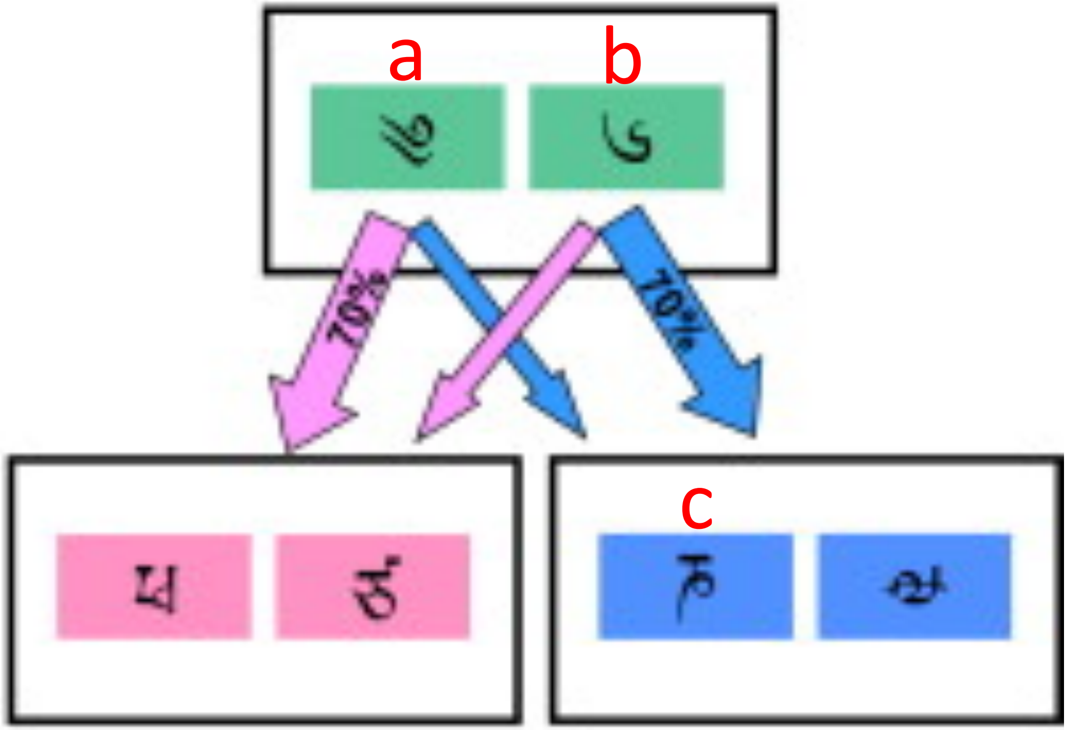
uncertainty-sensitive learning



shallow tree implies goal-directed control wins

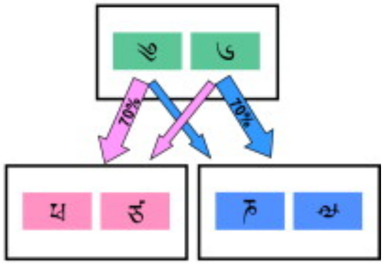
Daw, Niv, Dayan

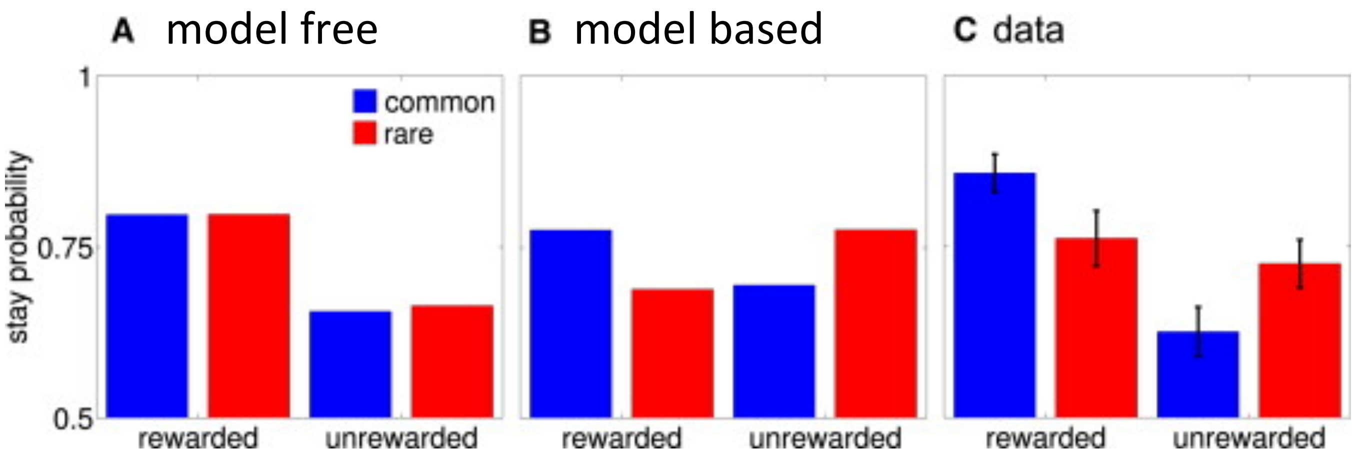
Human Canary...



- if a → c and c → £££ , then do more of a or b?
  - MB: b
  - MF: a (or even no effect)

Behaviour



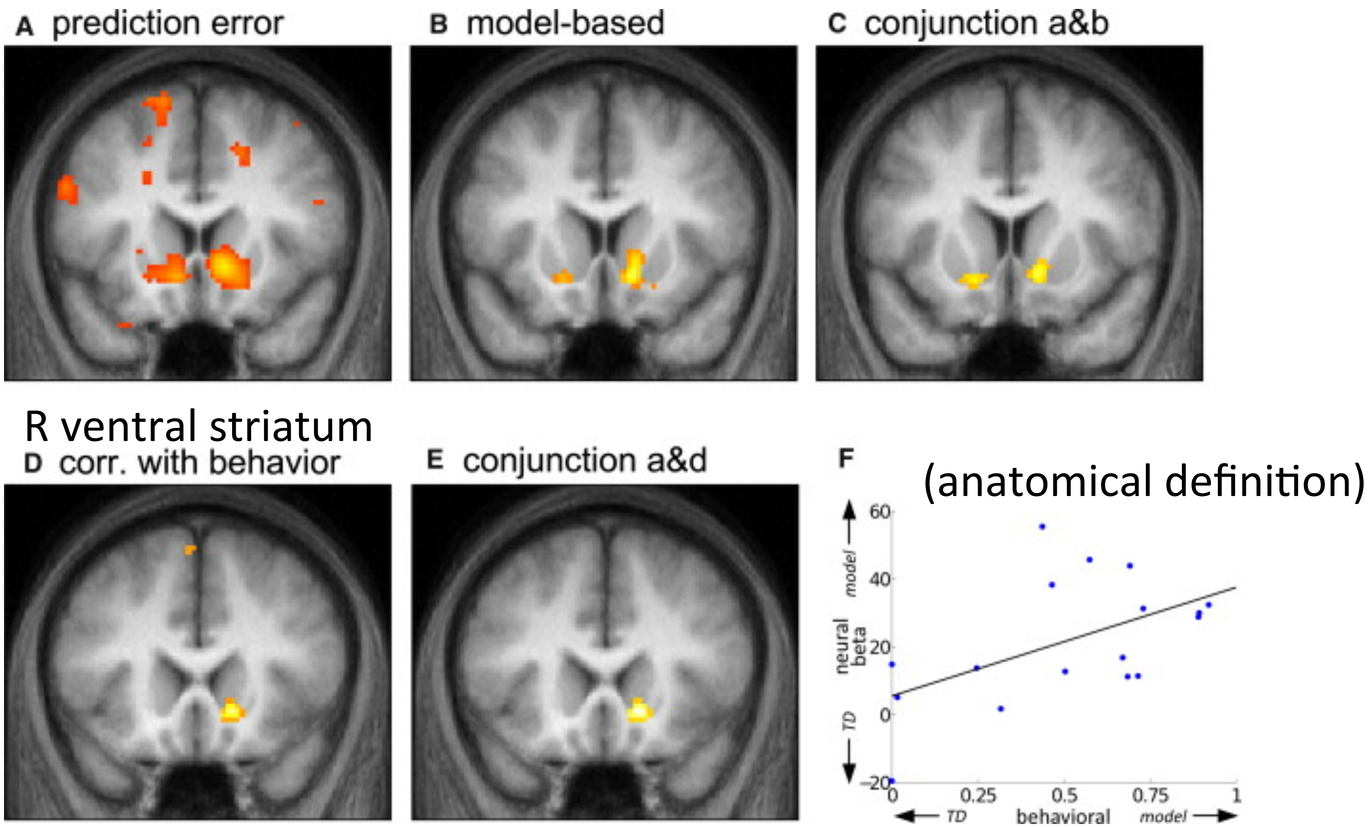


- assume a mix

$$Q_{tot}(x,a) = (1-\beta)Q_{MF}(x,a) + \beta Q_{MB}(x,a)$$

- expect that β will vary by subject (but be fixed)

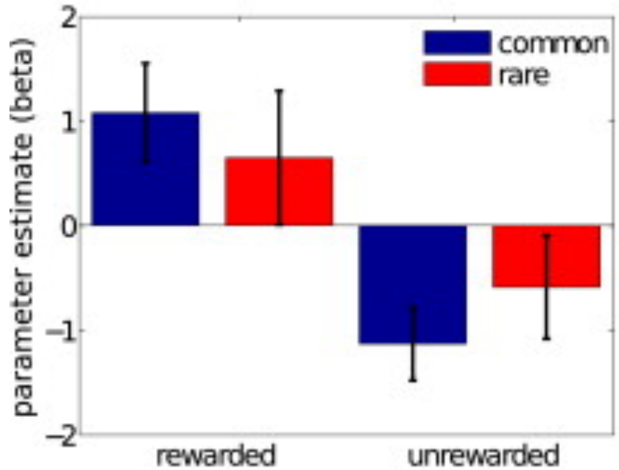
Neural Prediction Errors (1→2)

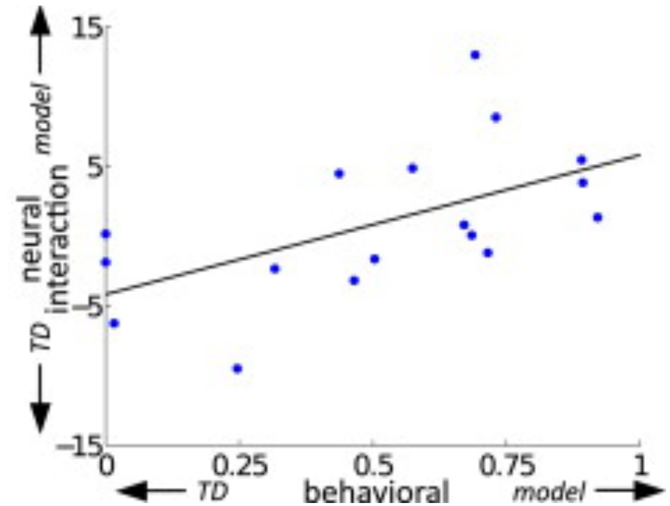


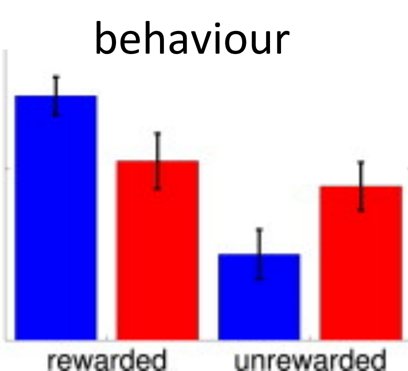
- note that MB RL does not use this prediction error – training signal?

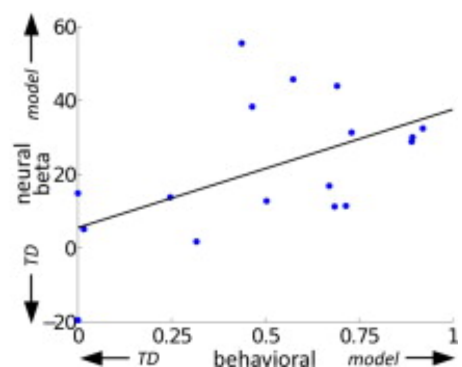
Neural Prediction Errors (1)

- right nucleus accumbens









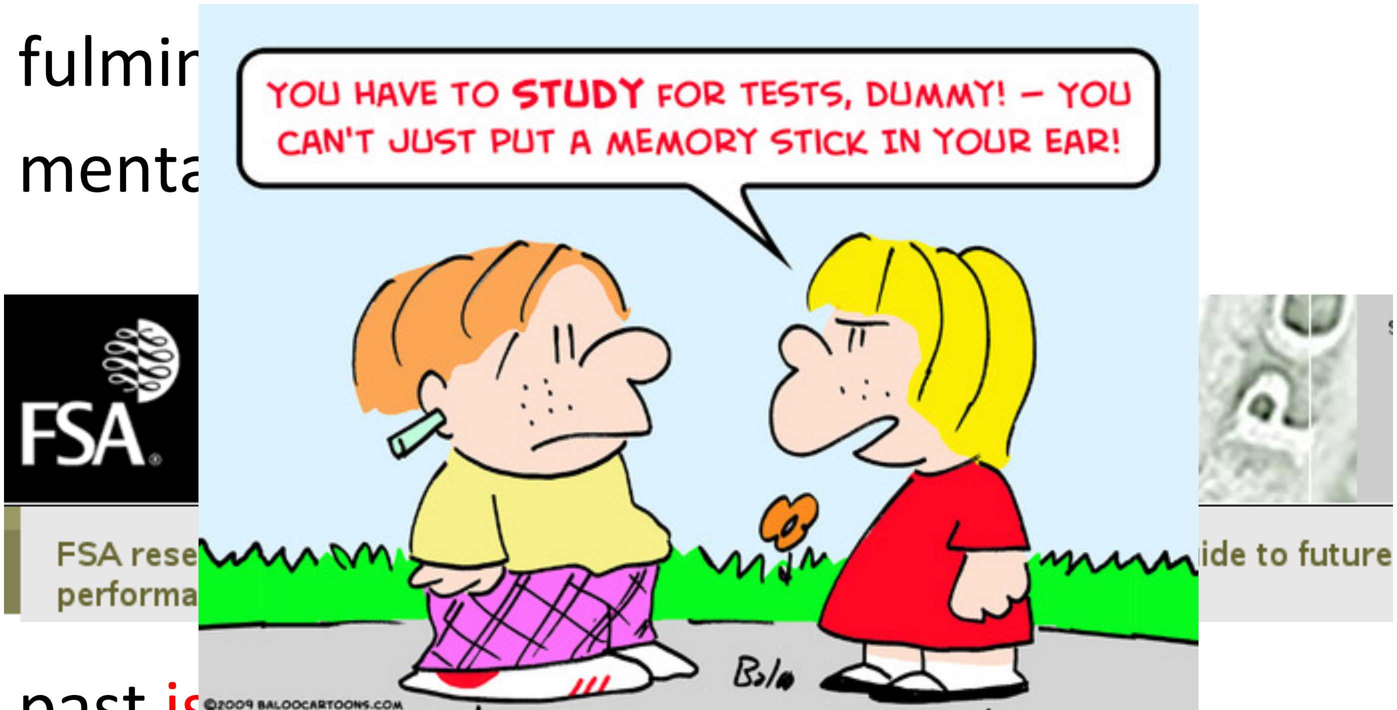
1-2, not 1

Model-based and Model-free

- categories justified by statistical/ computational costs
- separate neural substrates
- but:
  - more integrated than we thought
  - process account for MB (DYNA-2)?
  - related to many other dichotomies
  - MB priors?

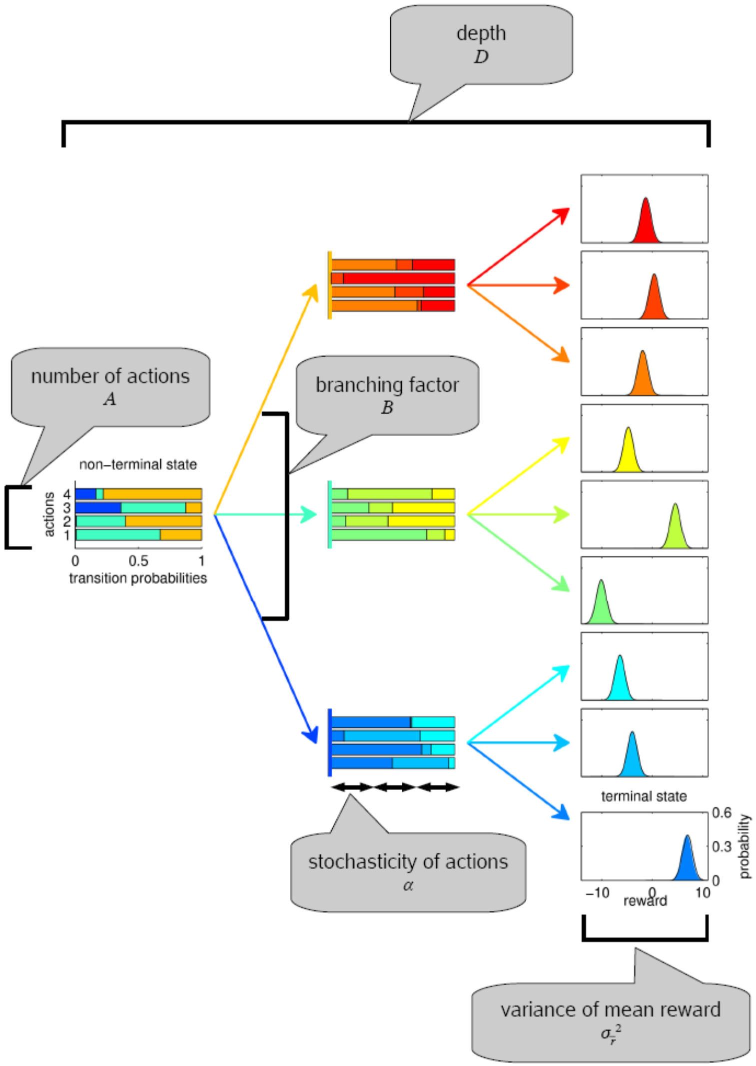
Why have Episodic memory?

- fulmir
- menta



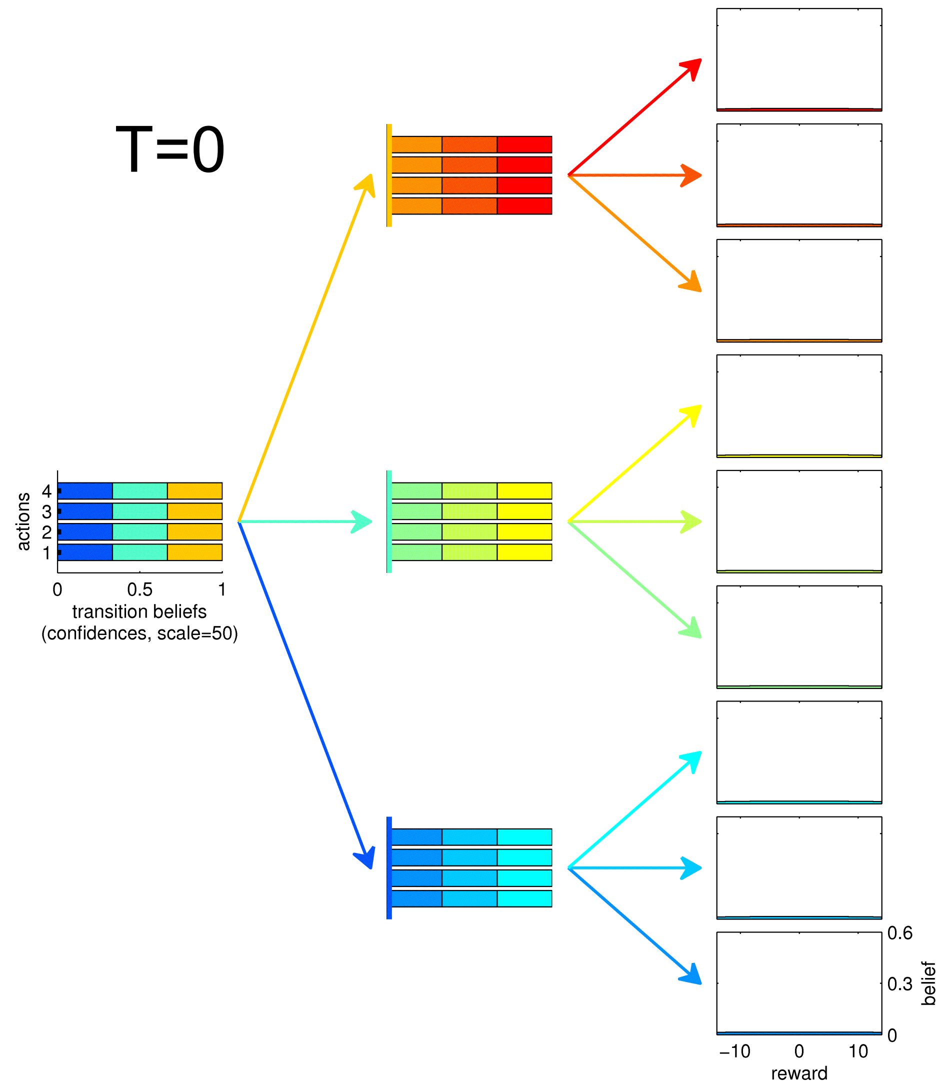
- past is a guide to the future
  - why single events and not statistics?
  - role of hippocampus in control?

The Third Way

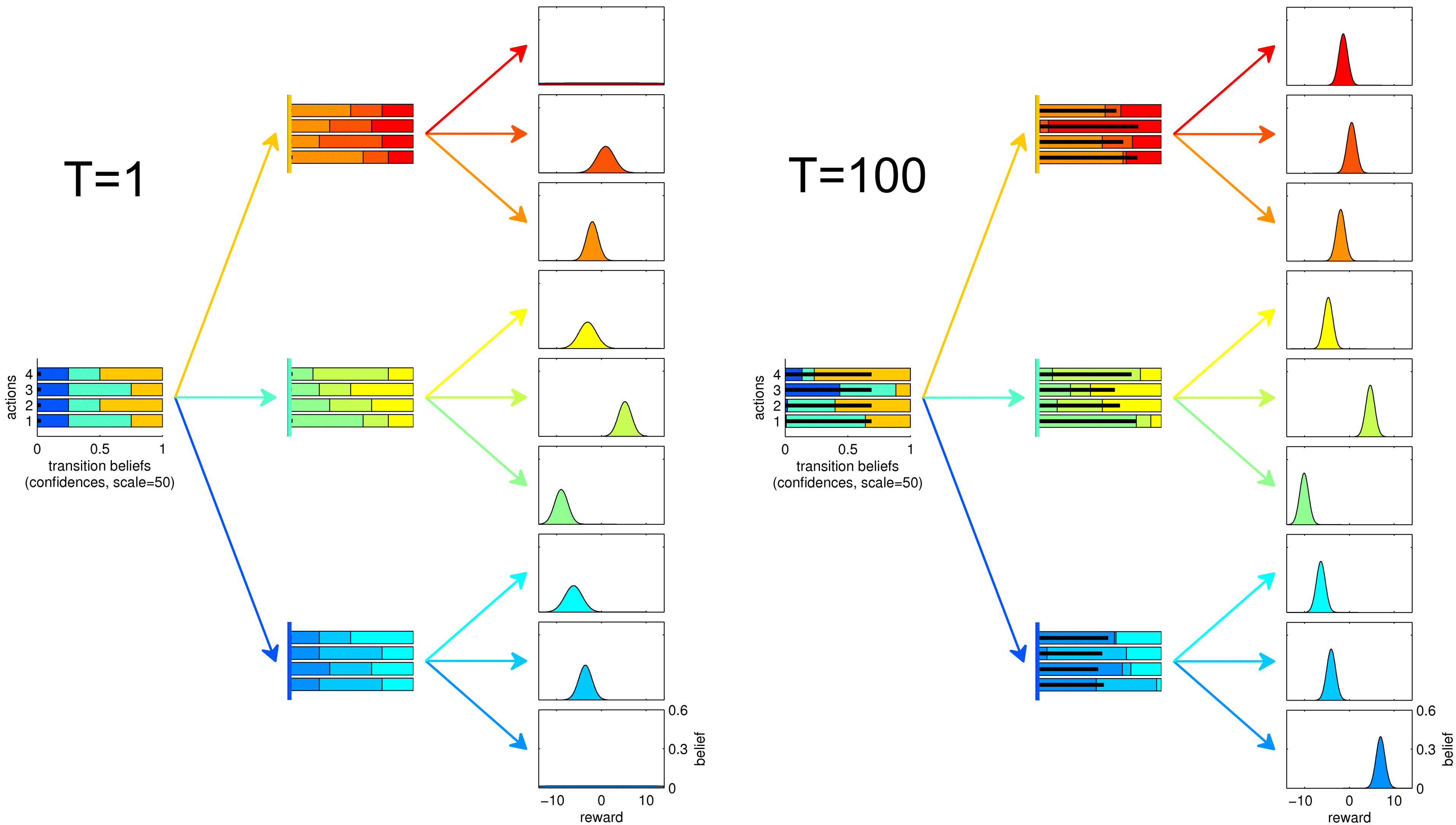


- simple domain
- **model-based** control:
  - build a tree
  - evaluate states
  - count cost of uncertainty
- **episodic** control:
  - store conjunction of states, actions, rewards
  - if reward > expectation, store all actions in the whole episode (Düzel)
  - choose rewarded action; else random

Semantic Controller



Semantic Controller

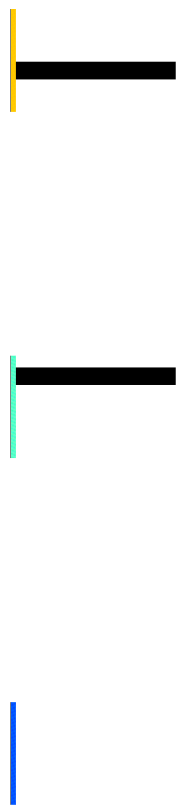


Episodic Controller

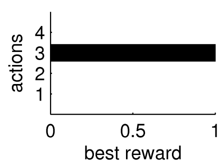
T=0

best
reward

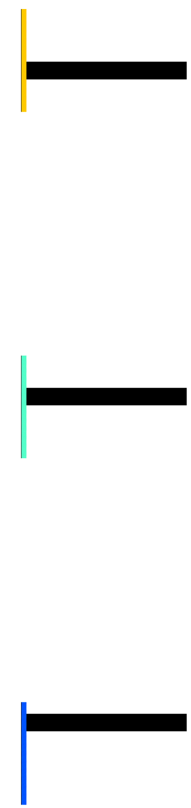
Episodic Controller



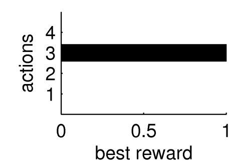
T=1



best
reward

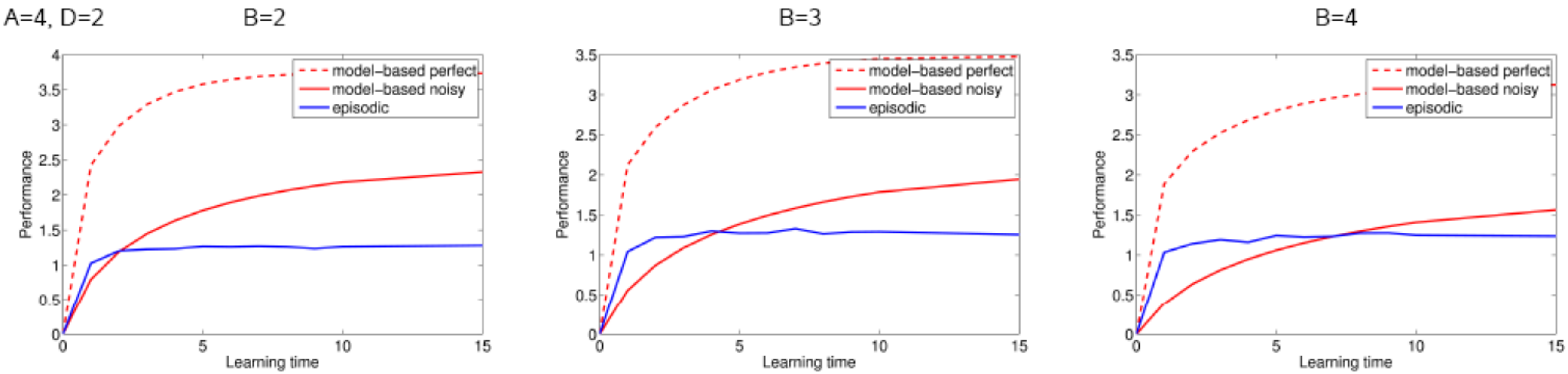


T=100



best
reward

Performance



- episodic advantage for early trials
- lasts longer for more complex environments
- can’t compute statistics/semantic information

Neural Reinforcement Learning

- error minimization/delta rule
- temporal difference learning
- Kalman filter
- Chinese restaurant process/NPB
- Bayesian Q-learning; Bayes-adaptive MDPs
- memory-based RL
- mixture models for attention
- particle filter for inference
- unsupervised learning random effects models for individual differences

Other Issues

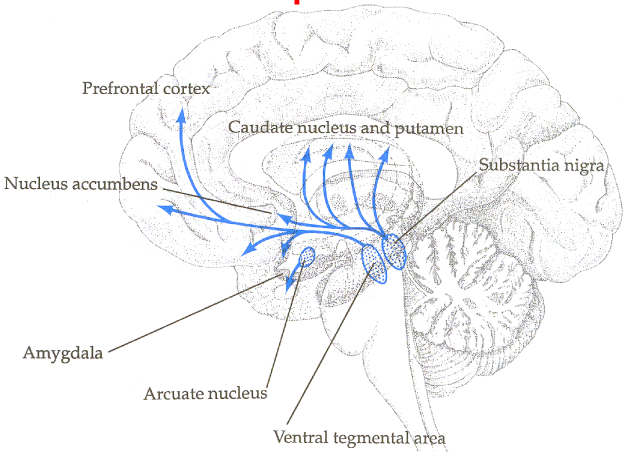
- active learning
  - exploration/exploitation
- priors over decision problems
  - controllability
  - hierarchy
- learning about others: game theory
- representational learning

Biological Learning

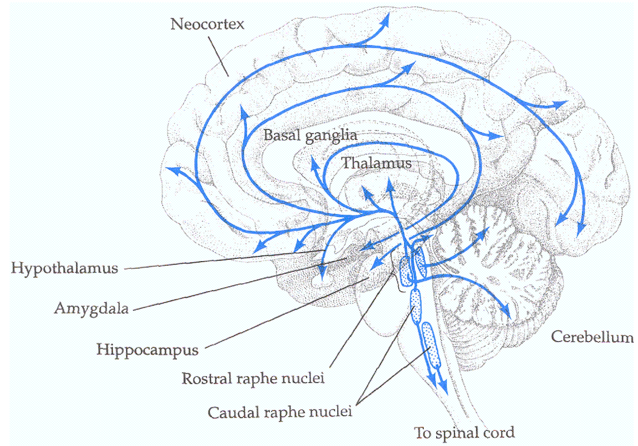
- error minimization/delta rule
- temporal difference learning
- Kalman filter
- Dirichlet process mixture/NPB
- Bayesian Q-learning; Bayes-adaptive MDPs
- memory-based reasoning
- particle filters for inference
- unsupervised `structural' learning

Computational Neuromodulation

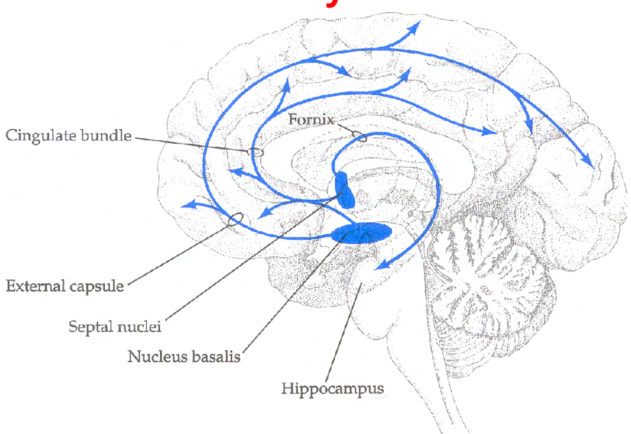
dopamine



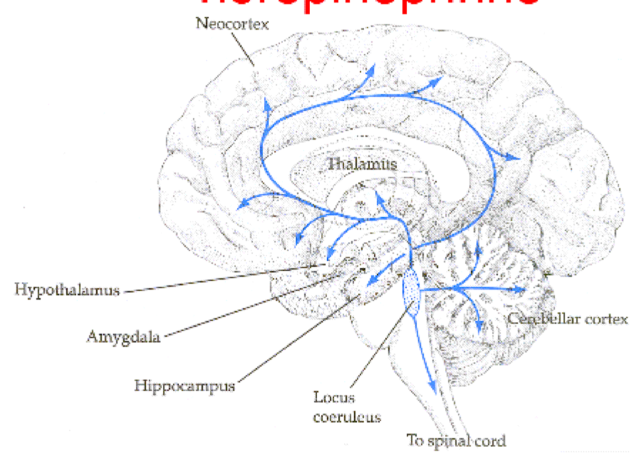
5HT



acetylcholine



norepinephrine



general: excitability, signal/noise ratios

specific: prediction errors, uncertainty signals