Multiple decision making systems in the brain: function and dysfunction

Nathaniel Daw
New York University

MPS-UCL Symposium on Computational Psychiatry

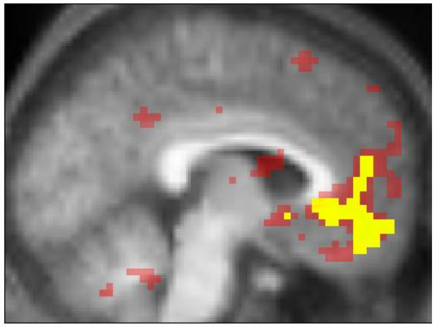
Reward and decision making

- The classic story: dopamine and the law of effect
- Why this is incomplete: multiple decision making systems, model-based and model-free
- Multiple decision systems in humans
- Implications for psychiatry

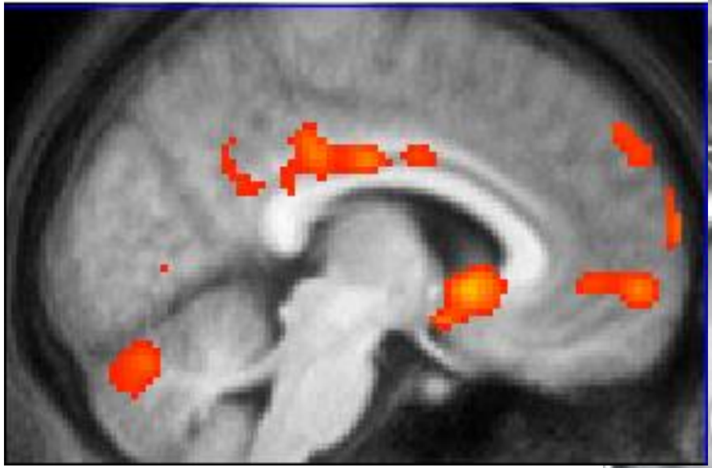
the classic story

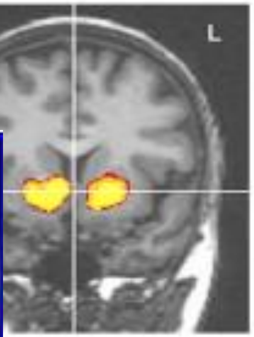
Broad findings

Reward or reward anticipation activates ventromedial prefrontal cortex & orbitofrontal cortex, striatum (sometimes midbrain)

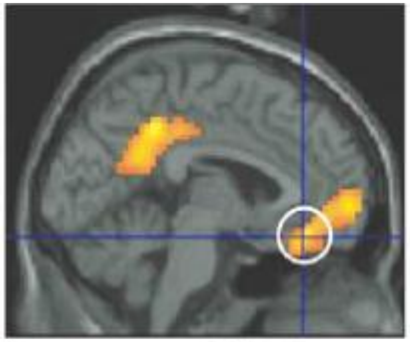


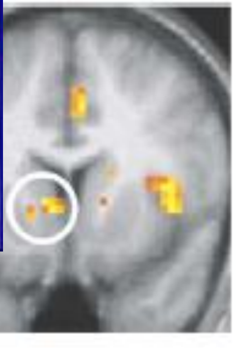
money
value predicted



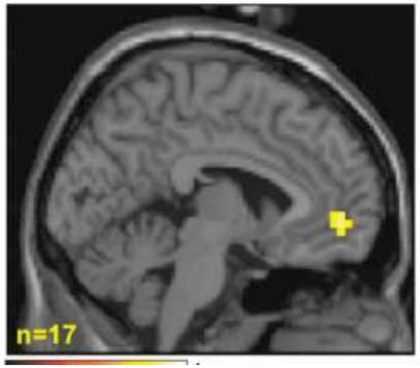


money
gain vs loss
(Kuhnen & Knutson 2005)

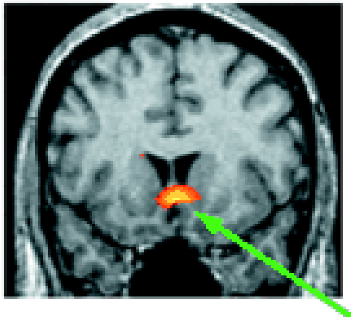




food odors
valued vs devalued
(Gottfreid et al 2003)



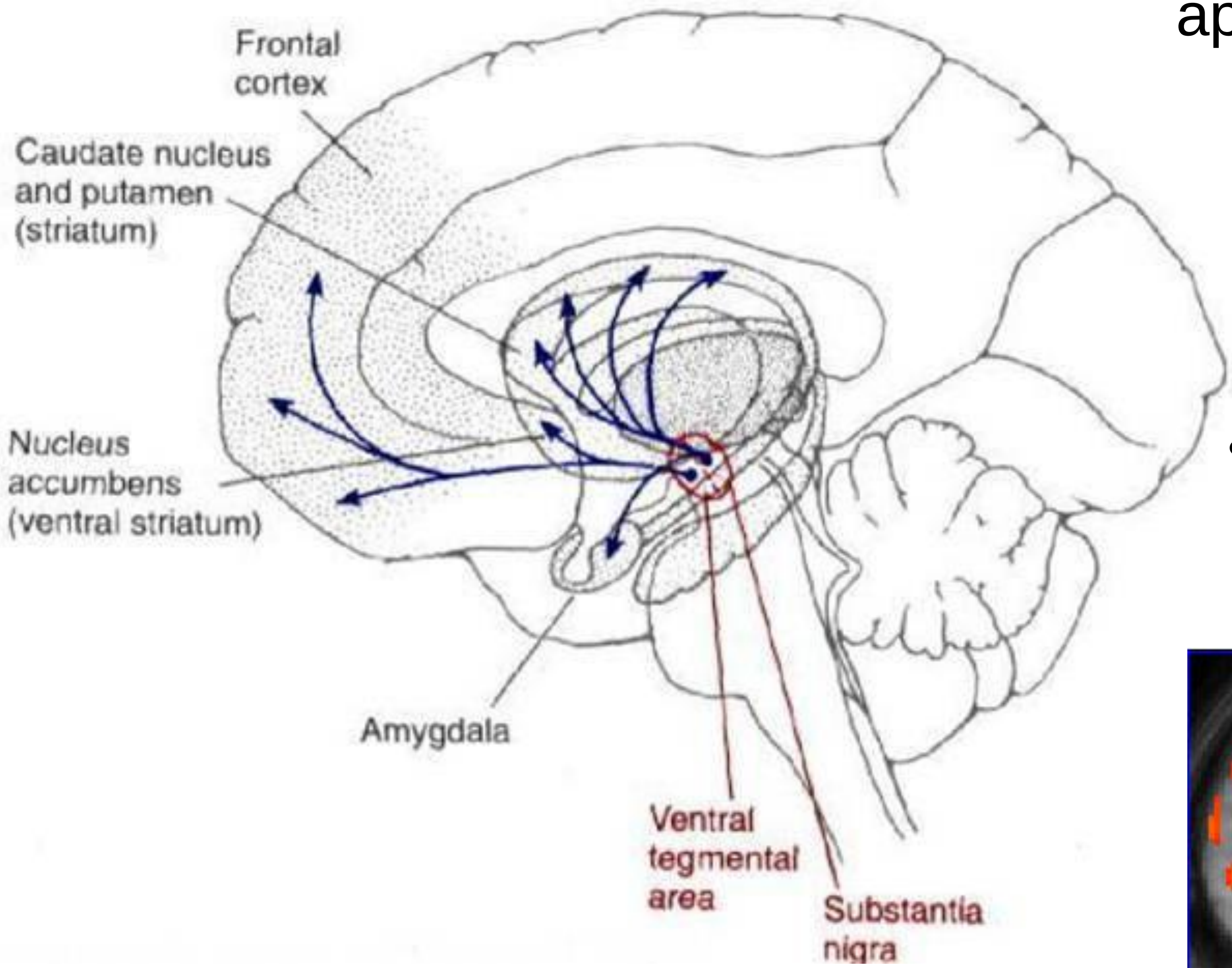
Coke or Pepsi
degree favored
(McClure et al. 2004)



juice
unpredictable vs predictable
(Berns et al 2001)

→ commonality of responding across reinforcers suggests generalized appetitive function

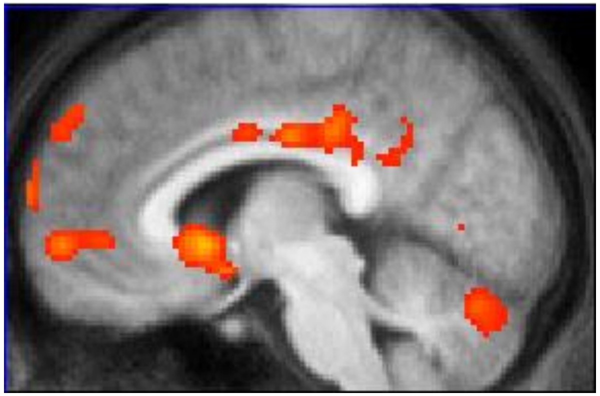
Dopamine


Frontal cortex
Caudate nucleus and putamen (striatum)
Nucleus accumbens (ventral striatum)
Amygdala
Ventral tegmental area
Substantia nigra

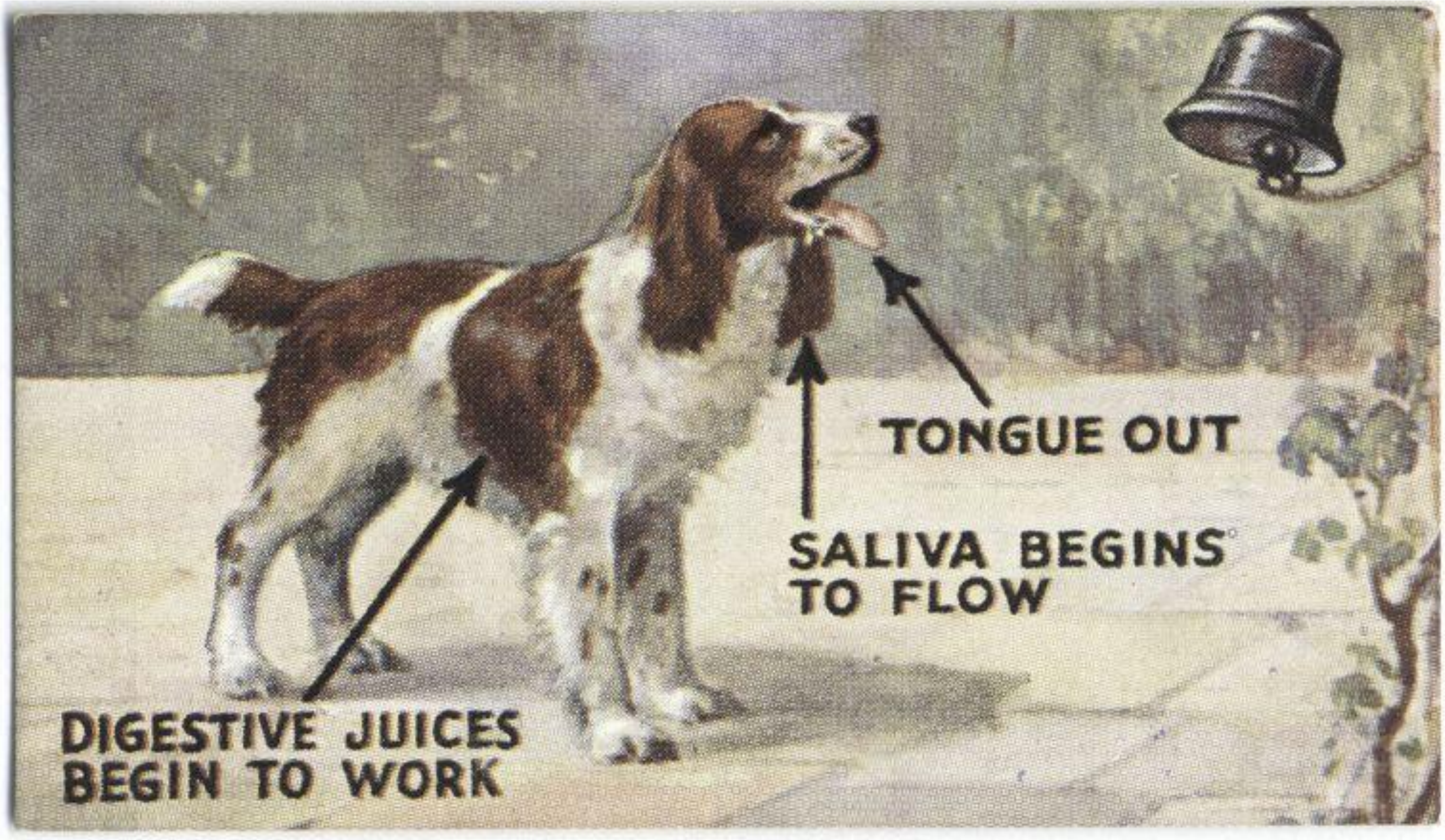
(from Kandel and Schwartz)

central tension:
appetitive vs motor

- Movement
- Reward
- Substance abuse
- Self-stimulation
- Synaptic plasticity
- Psychiatry (treatment)



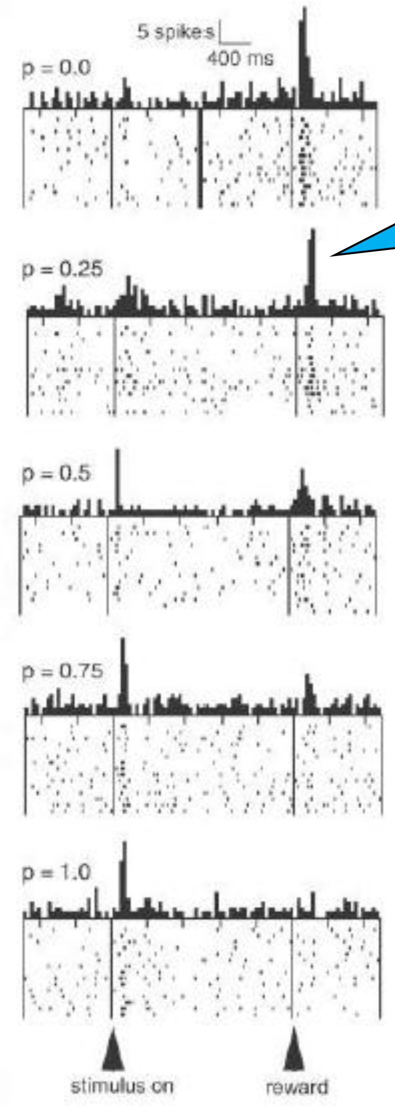
dopamine



- predictive learning is error driven

dopamine

reward following
0% predictive cue



dopamine neurons report prediction error $r_t - V_t$

reward following 50% predictive cue

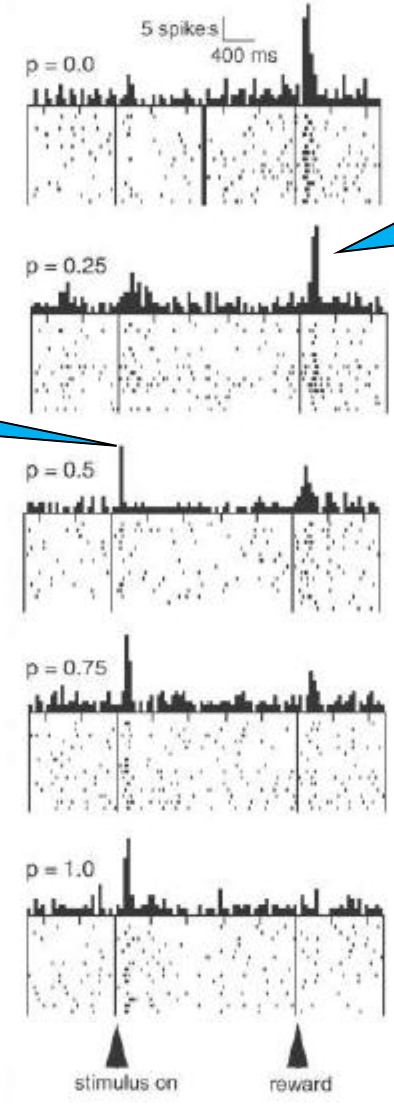
reward following 100% predictive cue

(Fiorillo et al 2003)

dopamine

cue response also a prediction error $[r_t + V_{t+1}] - V_t$

reward following 0% predictive cue



dopamine neurons report prediction error $[r_t + V_{t+1}] - V_t$

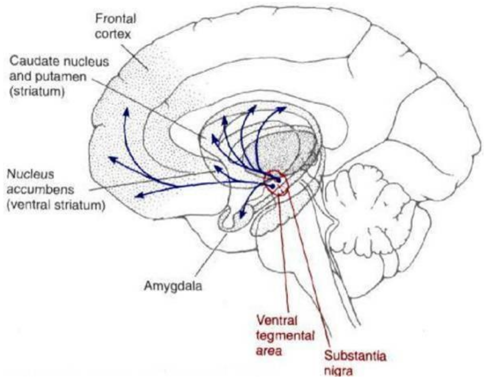
reward following 50% predictive cue

reward following 100% predictive cue

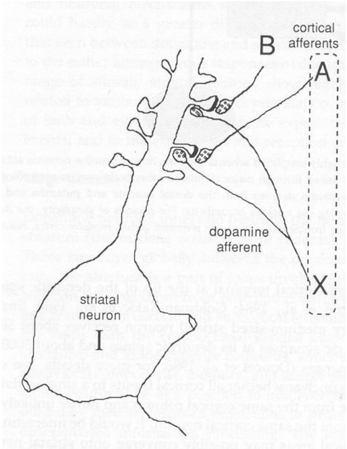
(Fiorillo et al 2003)

dopamine

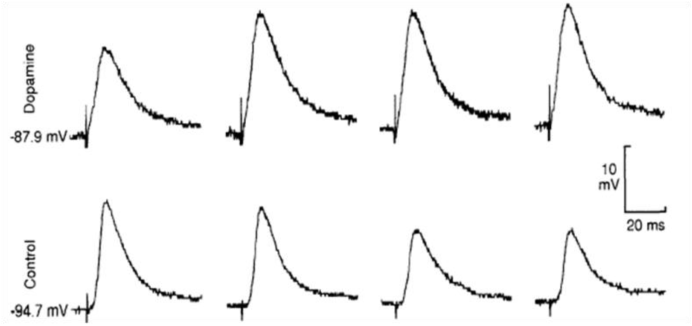
prediction errors may train predictions in striatum…


Frontal cortex
Caudate nucleus and putamen (striatum)
Nucleus accumbens (ventral striatum)
Amygdala
Ventral tegmental area
Substantia nigra

…at the corticostriatal synapse…

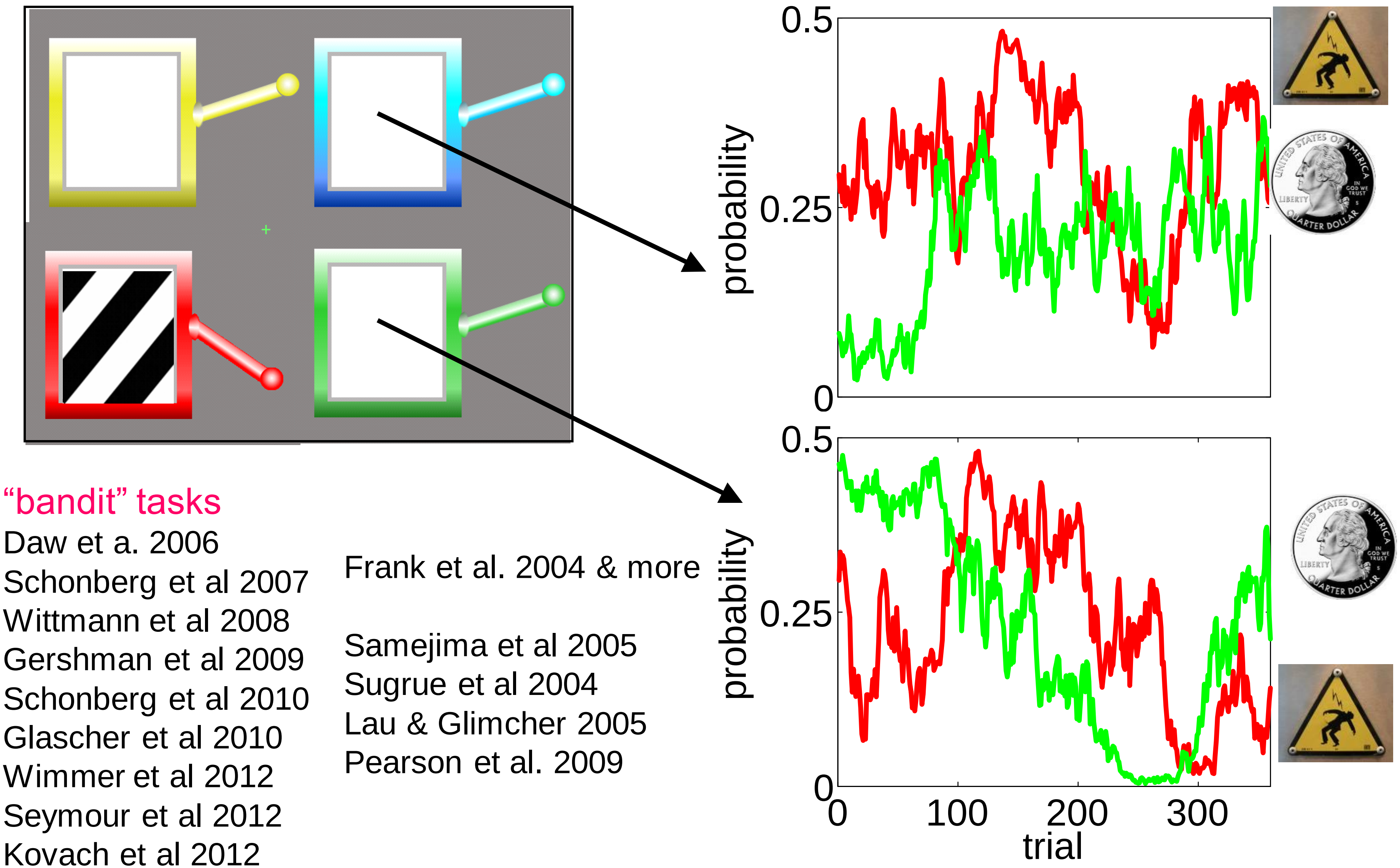


…where dopamine affects plasticity



…and neural firing promotes or opposes movement

learned decision making in humans



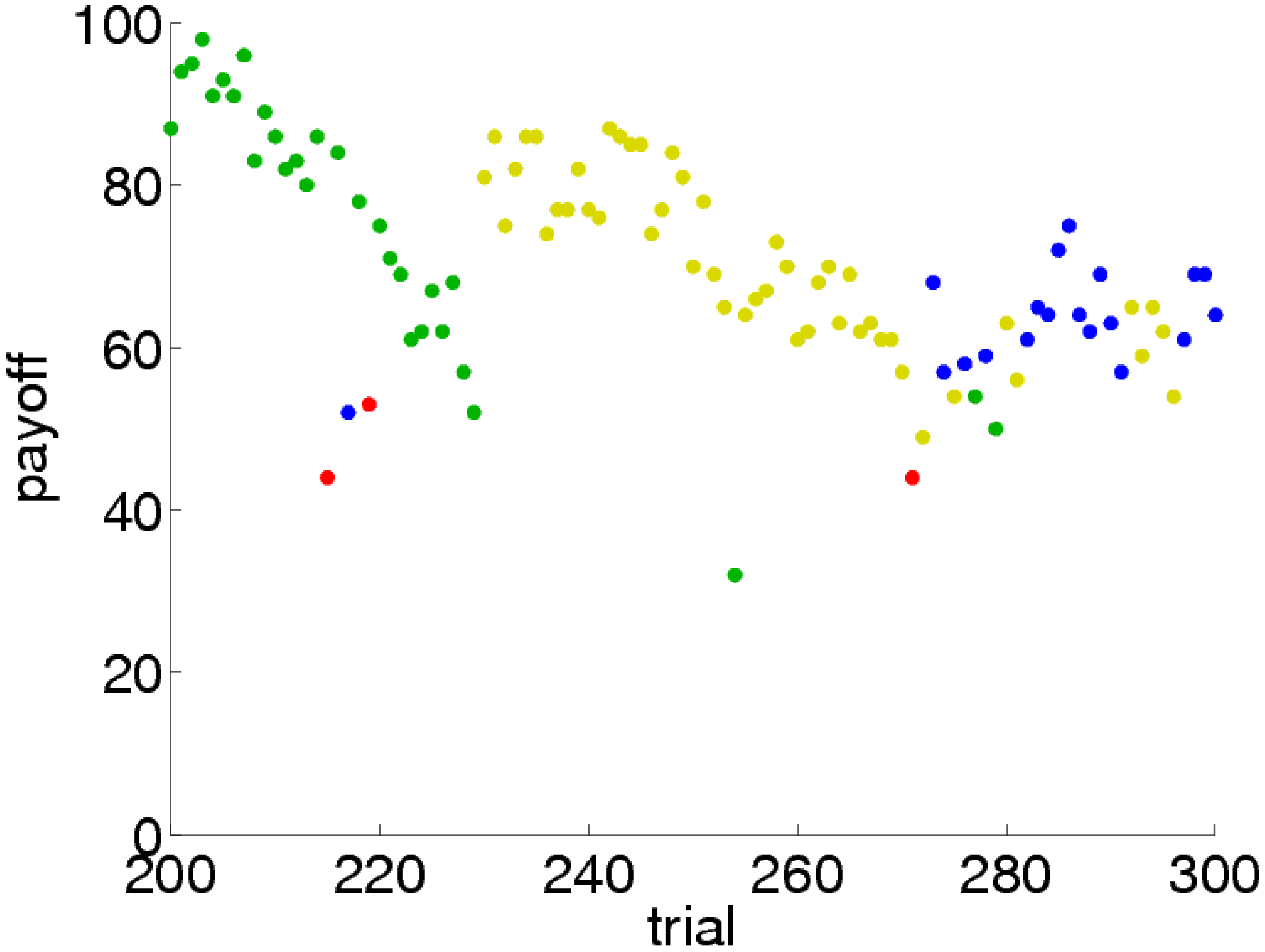
"bandit" tasks

Daw et a. 2006
Schonberg et al 2007
Wittmann et al 2008
Gershman et al 2009
Schonberg et al 2010
Glascher et al 2010
Wimmer et al 2012
Seymour et al 2012
Kovach et al 2012

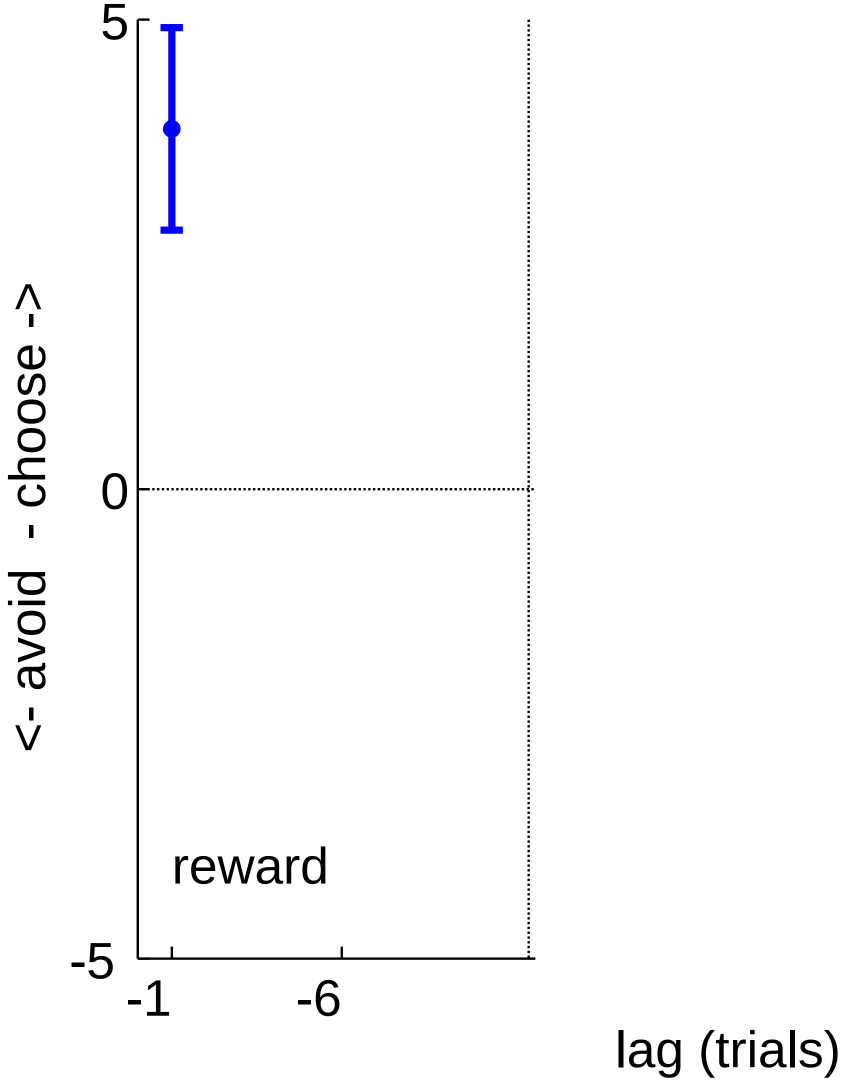
Frank et al. 2004 & more

Samejima et al 2005
Sugrue et al 2004
Lau & Glimcher 2005
Pearson et al. 2009

Behavior



(Daw et al. 2006)



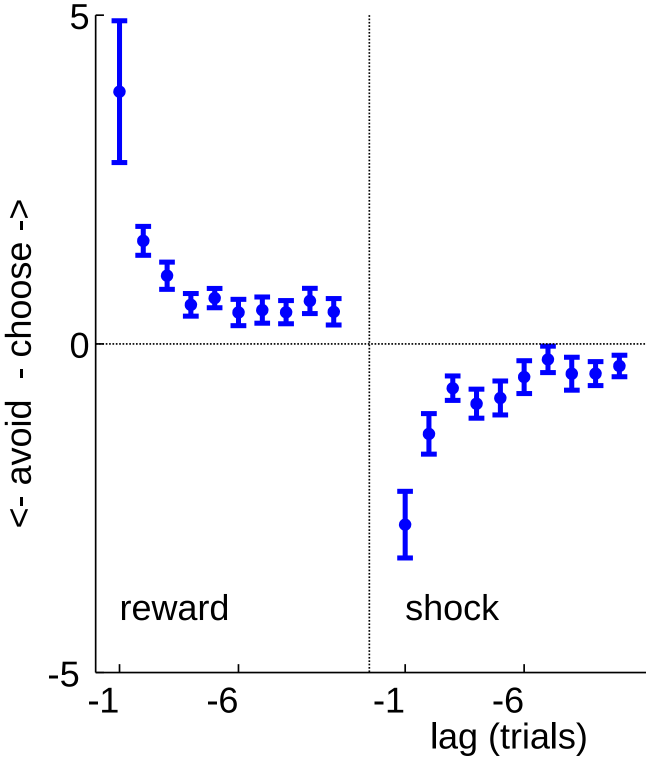
behavioral analysis: characterize the function relating outcomes to future choices (trial by trial learning model)

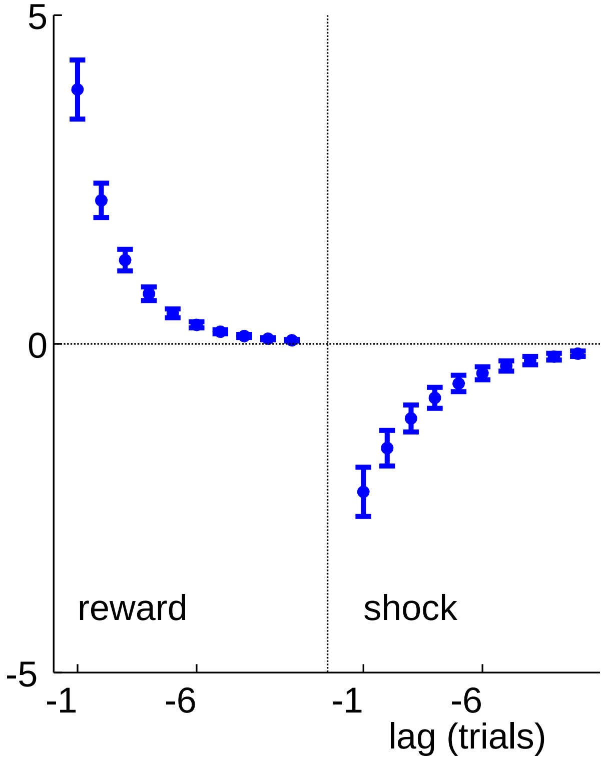
multinomial logistic regression: outcomes → choices

(Seymour et al. 2012)

Error-driven learning rules (like temporal-difference learning) predict weights should have exponential form (Lau & Glimcher 2005)

$$P(choice_t = c) \propto \exp(\beta \cdot Q_t(c))$$
$$Q_{t+1}(choice_t) = Q_t(choice_t) + \alpha \cdot \delta_t$$
$$\delta_t = reward_t - Q_t(choice_t)$$

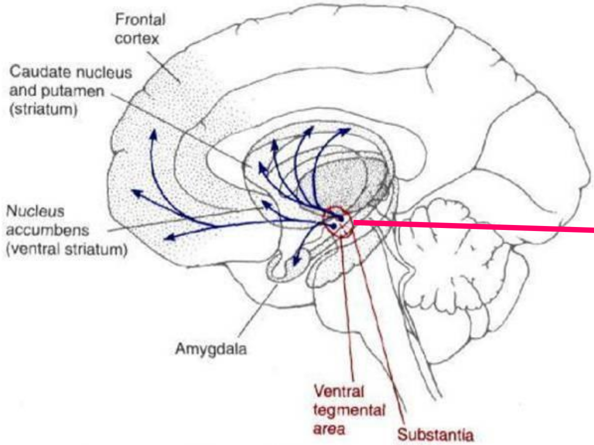


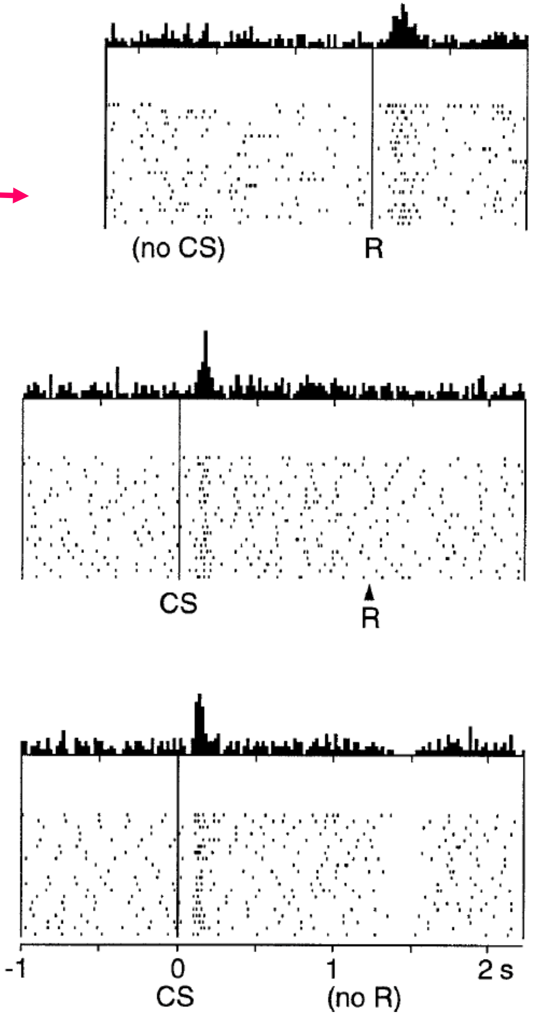


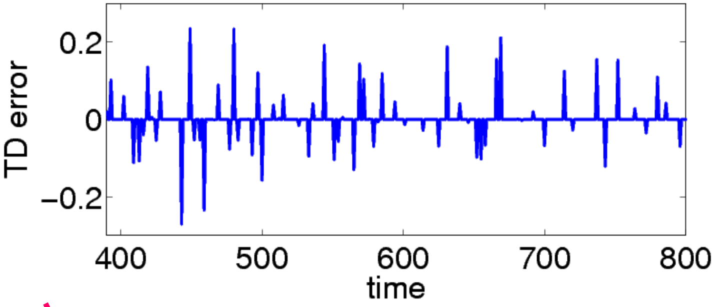
better fit (accounting for fewer free parameters)

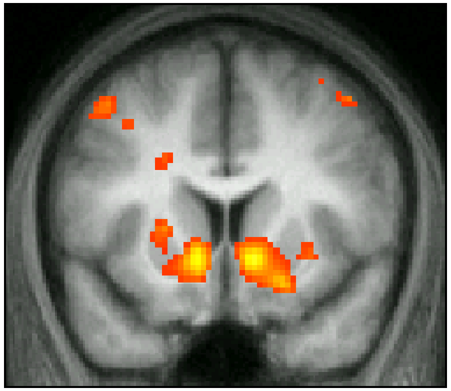
(Seymour et al. 2012)

Prediction error signals are visible at DA targets using fMRI


Frontal cortex
Caudate nucleus and putamen (striatum)
Nucleus accumbens (ventral striatum)
Amygdala
Ventral tegmental area
Substantia

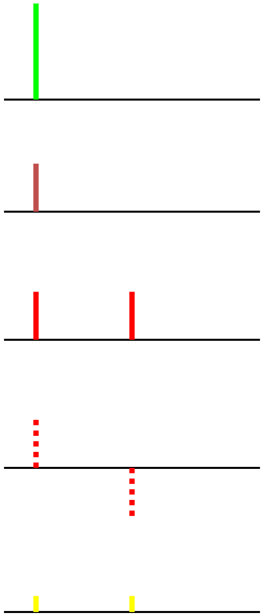




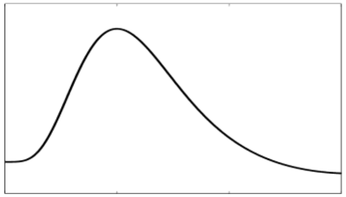


O’Doherty et al. 2004

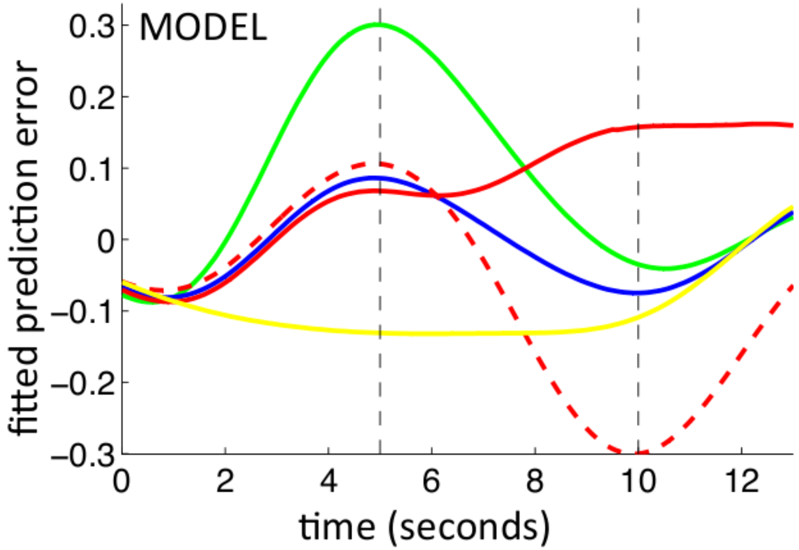
striatal BOLD and PE

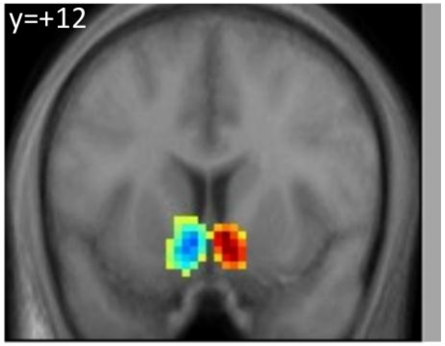


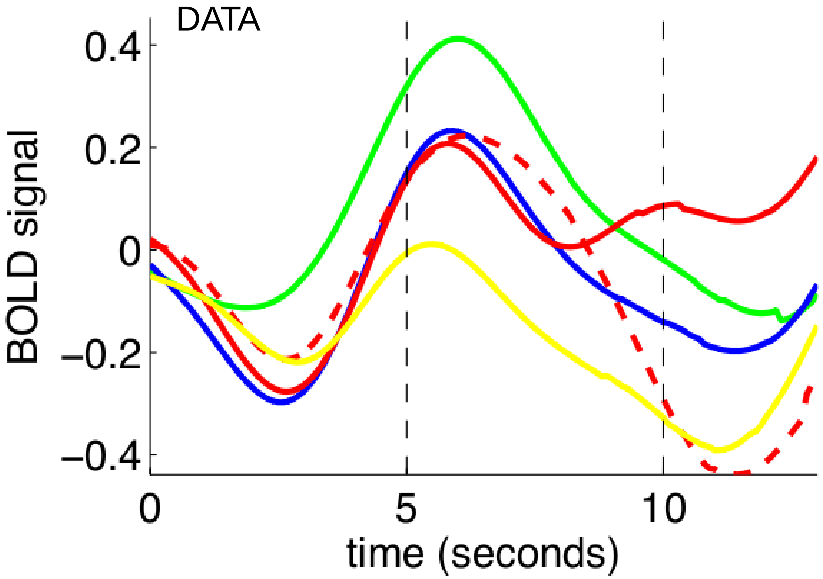
*



=



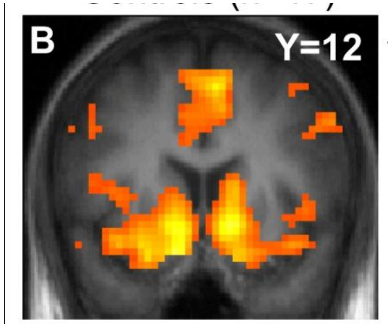




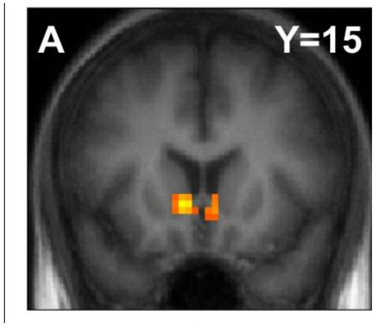
(Niv et al. 2012)

Striatal BOLD, DA, and PE

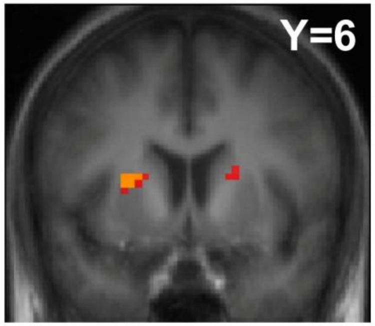
healthy control



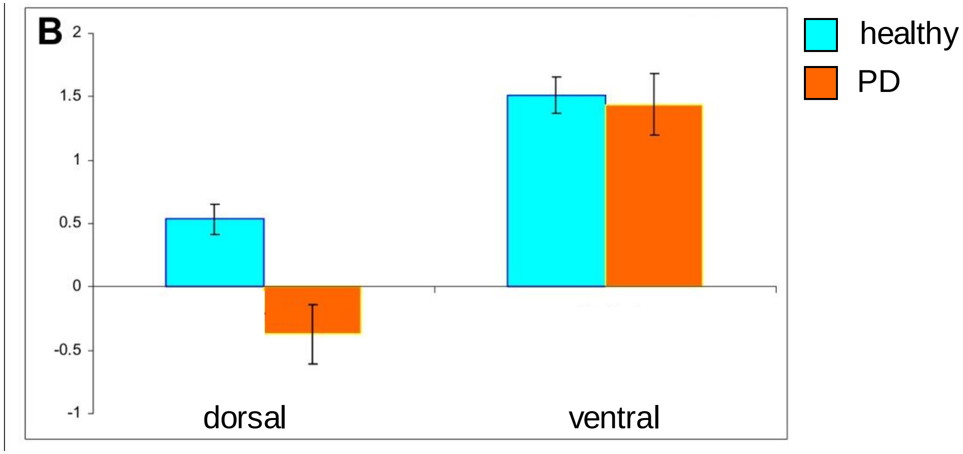
Parkinson’s disease



difference

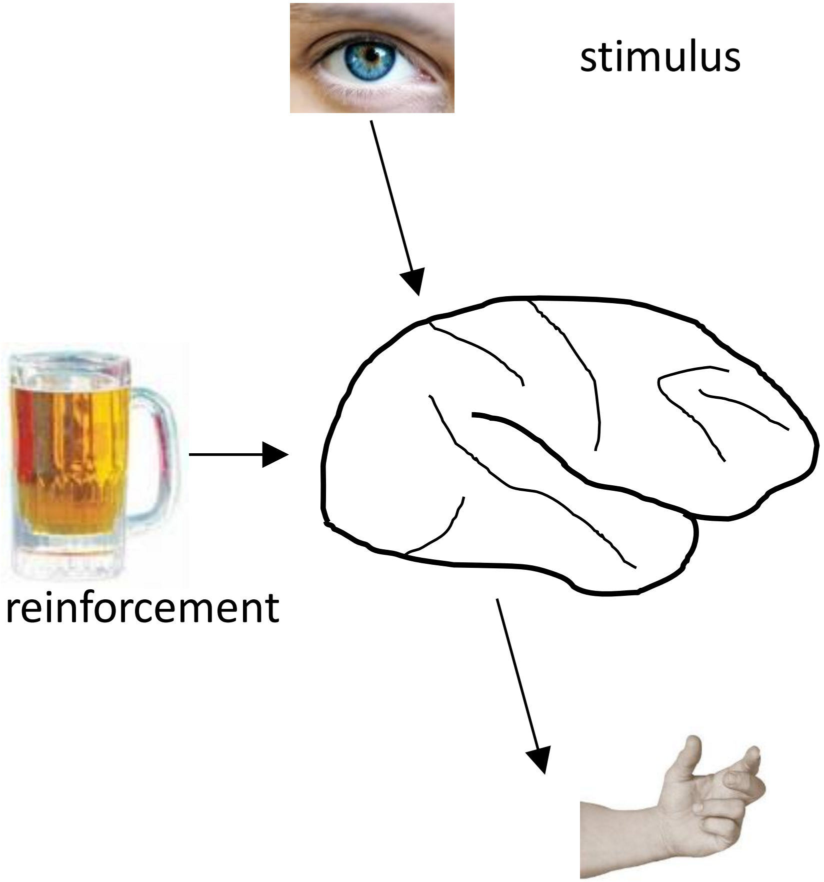


BOLD PE effect sizes



(Schonberg et al 2010; see also Pessiglione et al 2006)

the law of effect



stimulus

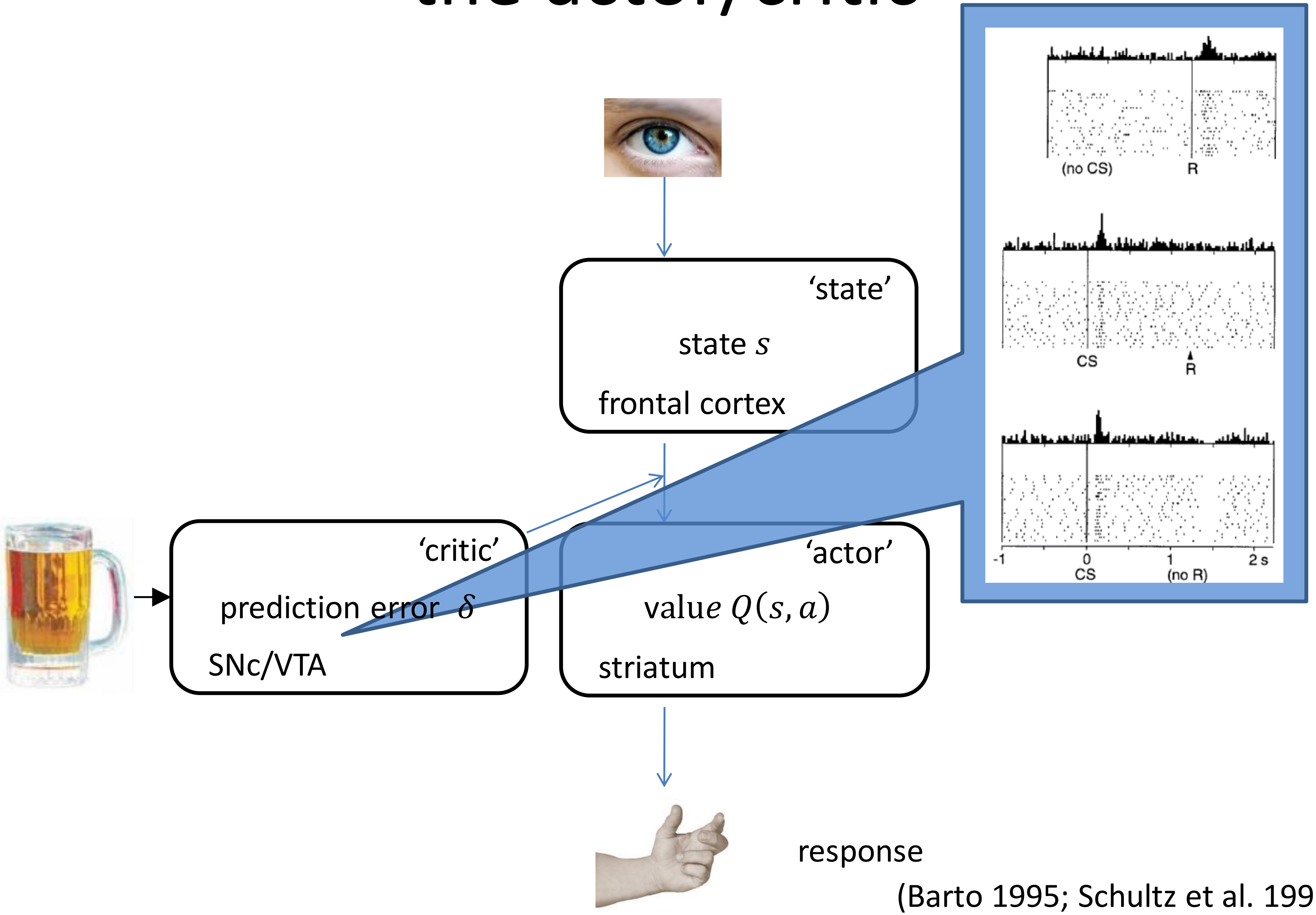
reinforcement

response

"Of several responses made to the same situation, those which are accompanied or closely followed by satisfaction to the animal will, other things being equal, be more firmly connected with the situation, so that, when it recurs, they will be more likely to recur."

Thorndike (1911)

the actor/critic



'state'

state s

frontal cortex

'critic'

prediction error δ

SNc/VTA

'actor'

value $Q(s,a)$

striatum

response

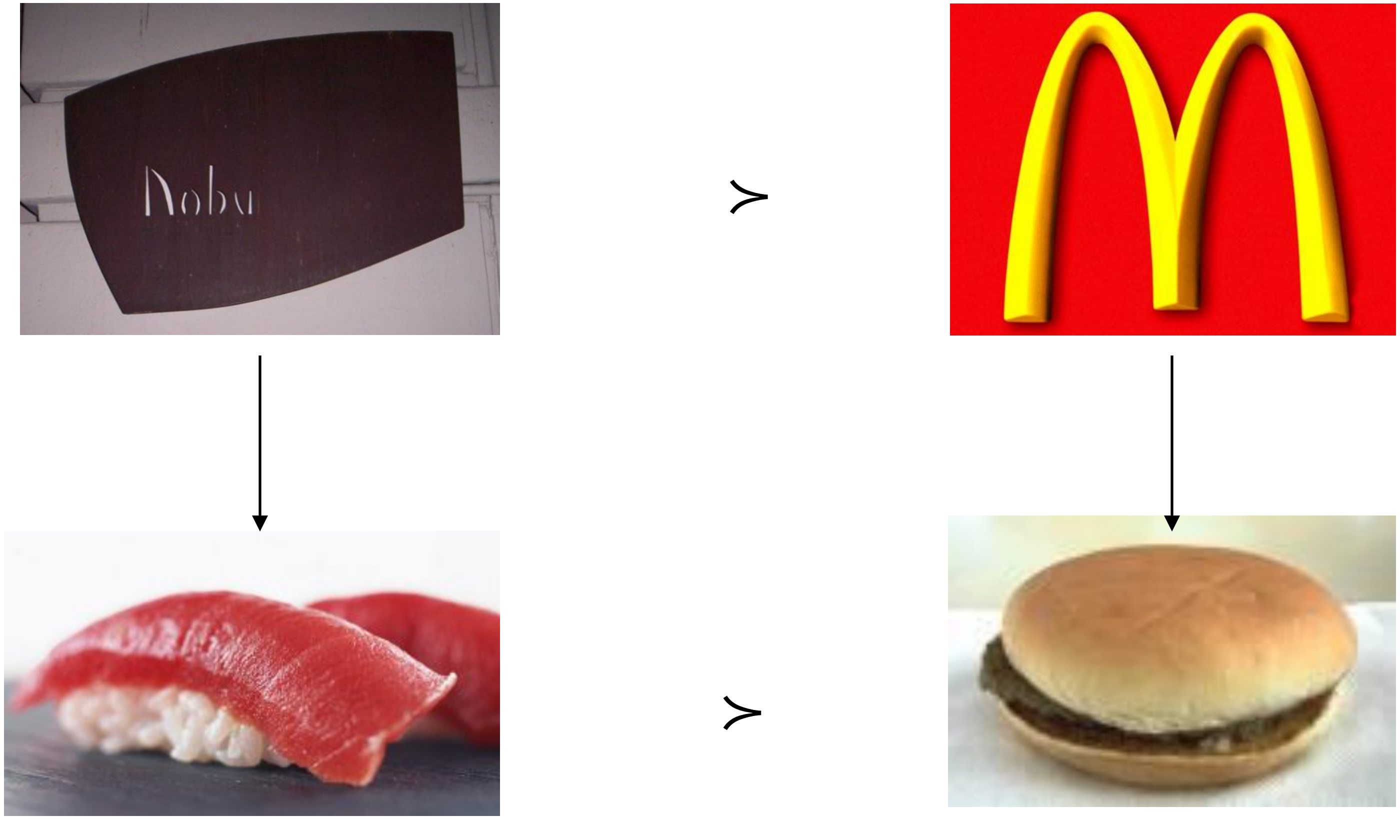
(Barto 1995; Schultz et al. 1997)

What’s wrong with all this

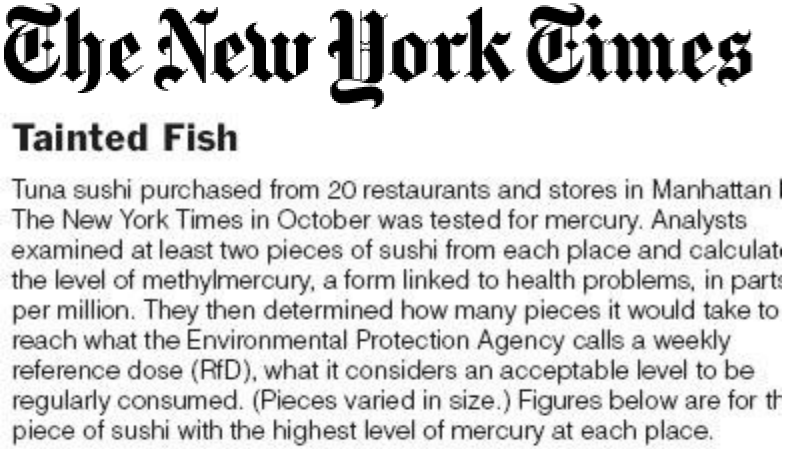
Cognitive maps

"The stimuli are not connected by just simple one-to-one switches to the outgoing responses. Rather, the incoming impulses are usually worked over and elaborated in the central control room into a tentative, cognitive-like map of the environment. And it is this tentative map, indicating routes and paths and environmental relationships, which finally determines what responses, if any, the animal will finally release."

Tolman (1948)


Nobu
≻
≻

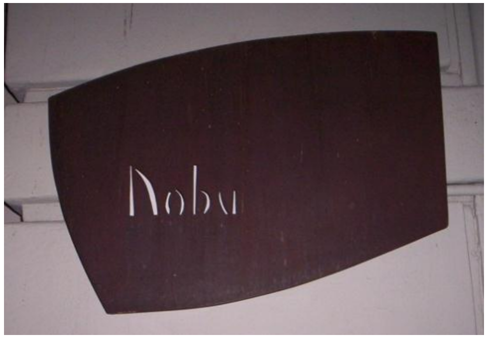
≺



The New York Times

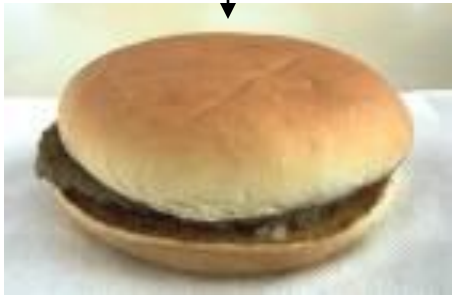
Tainted Fish

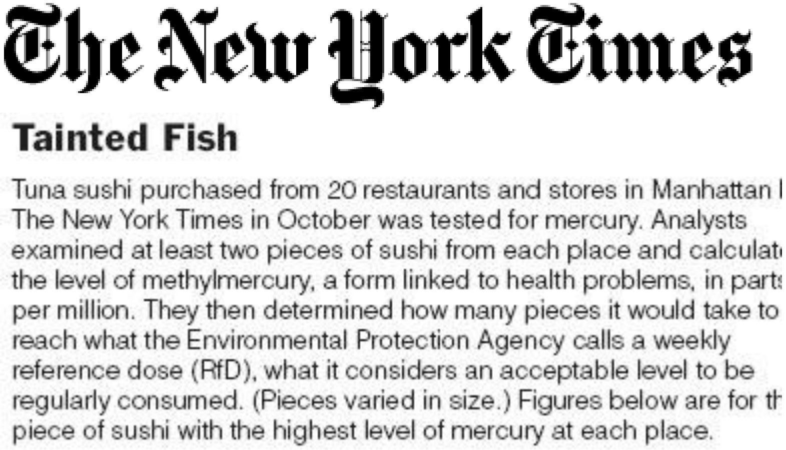
Tuna sushi purchased from 20 restaurants and stores in Manhattan
The New York Times in October was tested for mercury. Analysts
examined at least two pieces of sushi from each place and calculat
the level of methylmercury, a form linked to health problems, in parts
per million. They then determined how many pieces it would take to
reach what the Environmental Protection Agency calls a weekly
reference dose (RfD), what it considers an acceptable level to be
regularly consumed. (Pieces varied in size.) Figures below are for th
piece of sushi with the highest level of mercury at each place.



?

$\prec$





The New York Times

Tainted Fish

Tuna sushi purchased from 20 restaurants and stores in Manhattan l
The New York Times in October was tested for mercury. Analysts
examined at least two pieces of sushi from each place and calculat
the level of methylmercury, a form linked to health problems, in parts
per million. They then determined how many pieces it would take to
reach what the Environmental Protection Agency calls a weekly
reference dose (RfD), what it considers an acceptable level to be
regularly consumed. (Pieces varied in size.) Figures below are for th
piece of sushi with the highest level of mercury at each place.

$$E[U(a)] = \Sigma_o\, P(o|a)\, U(o)$$

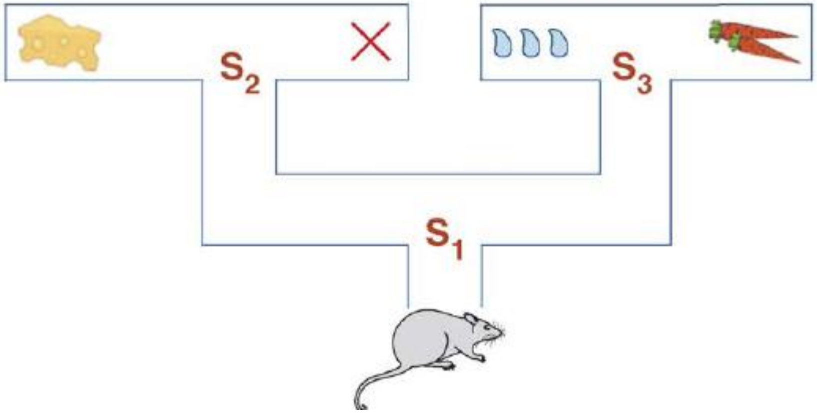
"model-free"

"model-based"

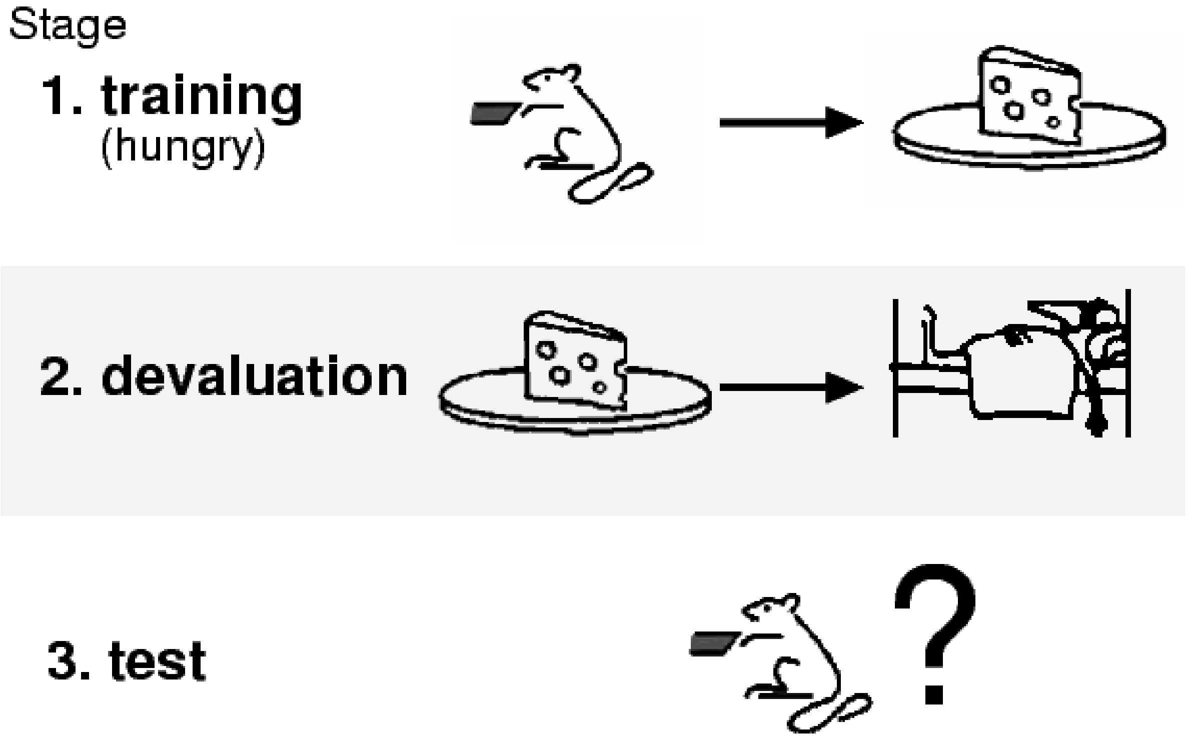
(Daw et al. 2005, Doya, 1999)

Bellman equation

$$V(s) = r(s) + \gamma \sum_{s' \in S} P(s_{t+1} = s' | s_t = s) V(s')$$



test



Stage

1. training
(hungry)

learn to leverpress for food (choose work or not)

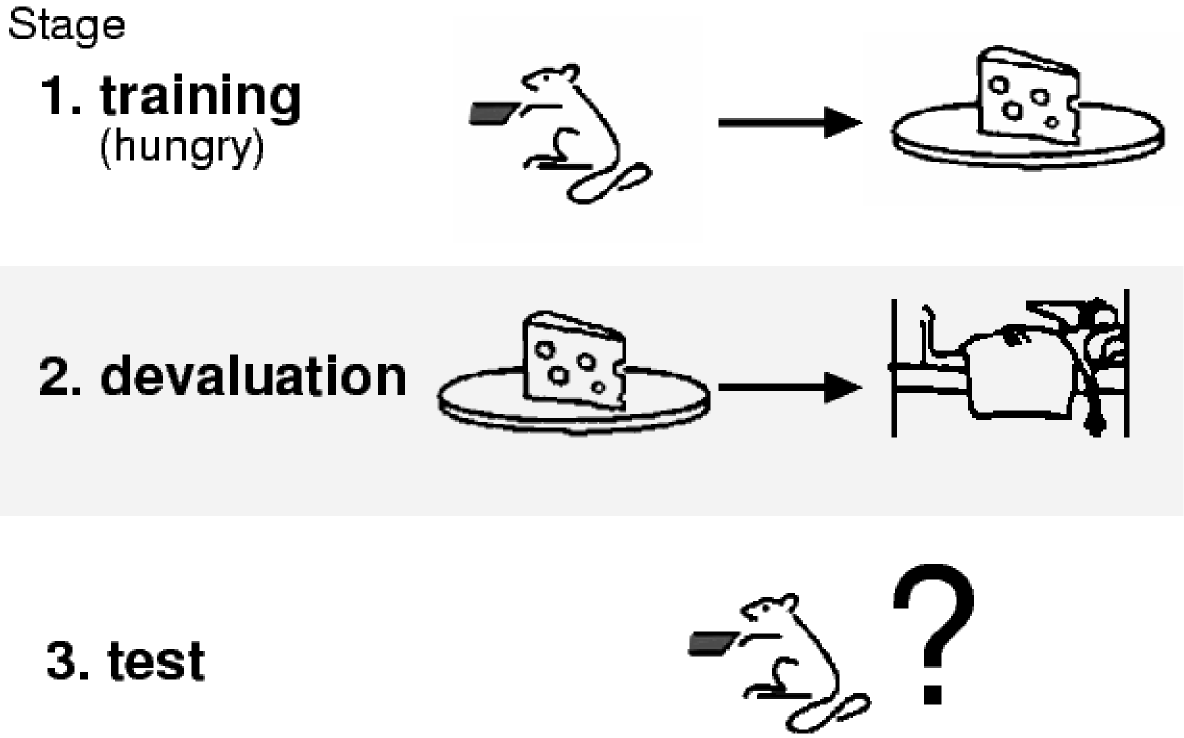
2. devaluation

pair food with illness; develop aversion (watermelon story)

3. test

will animals work for food they don’t want?

test



learn to leverpress for food (choose work or not)

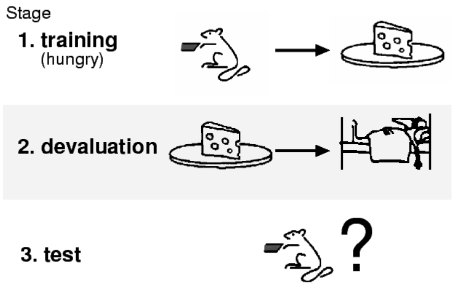
pair food with illness; develop aversion (watermelon story)

will animals work for food they don’t want?

important & confusing point:
food not delivered during test. why?

behavior compared to control group who skipped stage 2 (still want food), but also don’t get it

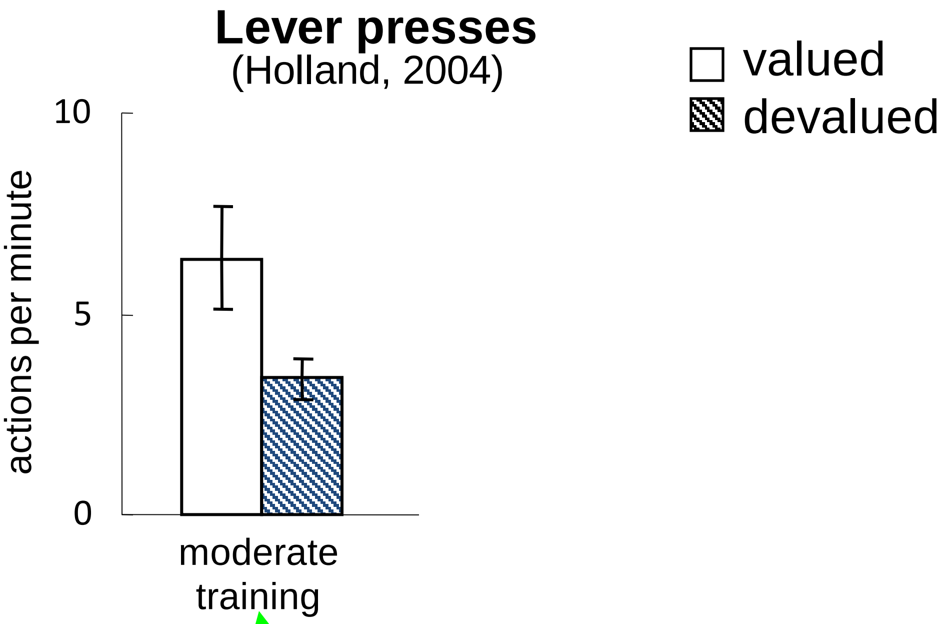
results


Stage

1. training (hungry)

2. devaluation

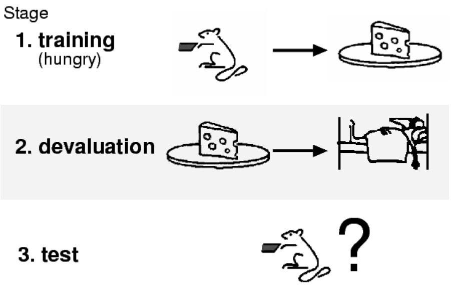
3. test

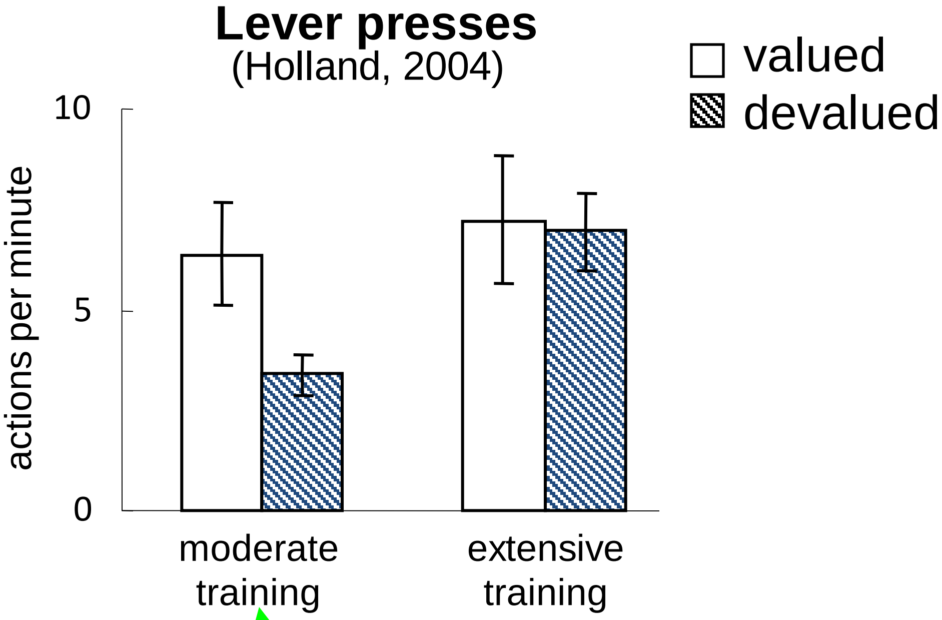

Lever presses
(Holland, 2004)

Moderate training: outcome sensitive

“goal directed”

results




Lever presses
(Holland, 2004)

Moderate training: outcome sensitive

"goal directed"

Outcome insensitive following overtraining

"habitual" like TD

Animals will work for food they don't want, sometimes

→ familiar counterpart: actions become automatic with repetition

Lesions

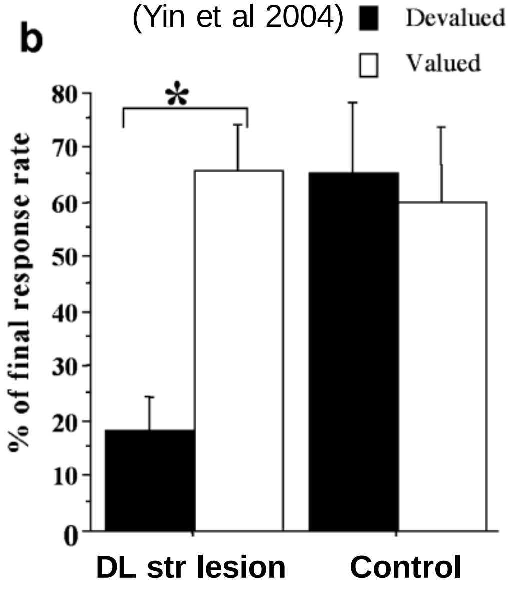
- With lesion of dorsolateral striatum (also its DA input) rats acquire normally but never form habits: perpetually devaluation sensitive

- Prefrontal areas, also dorsomedial striatum produce opposite pattern: even undertrained rats are habitual (devaluation insensitive)

→ Behavior arises from dissociable neural systems

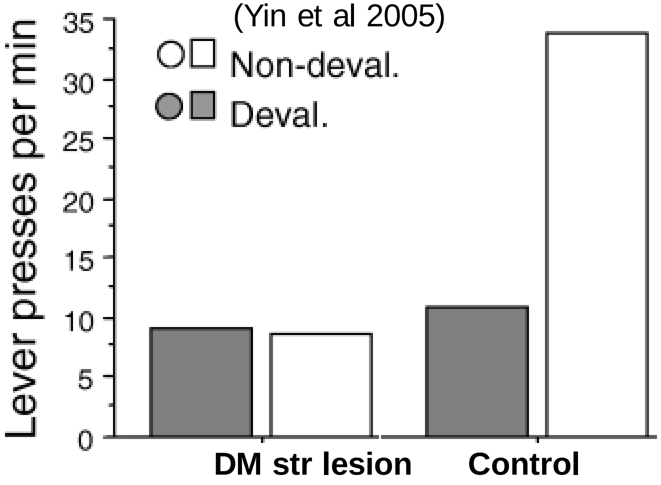
Overtrained

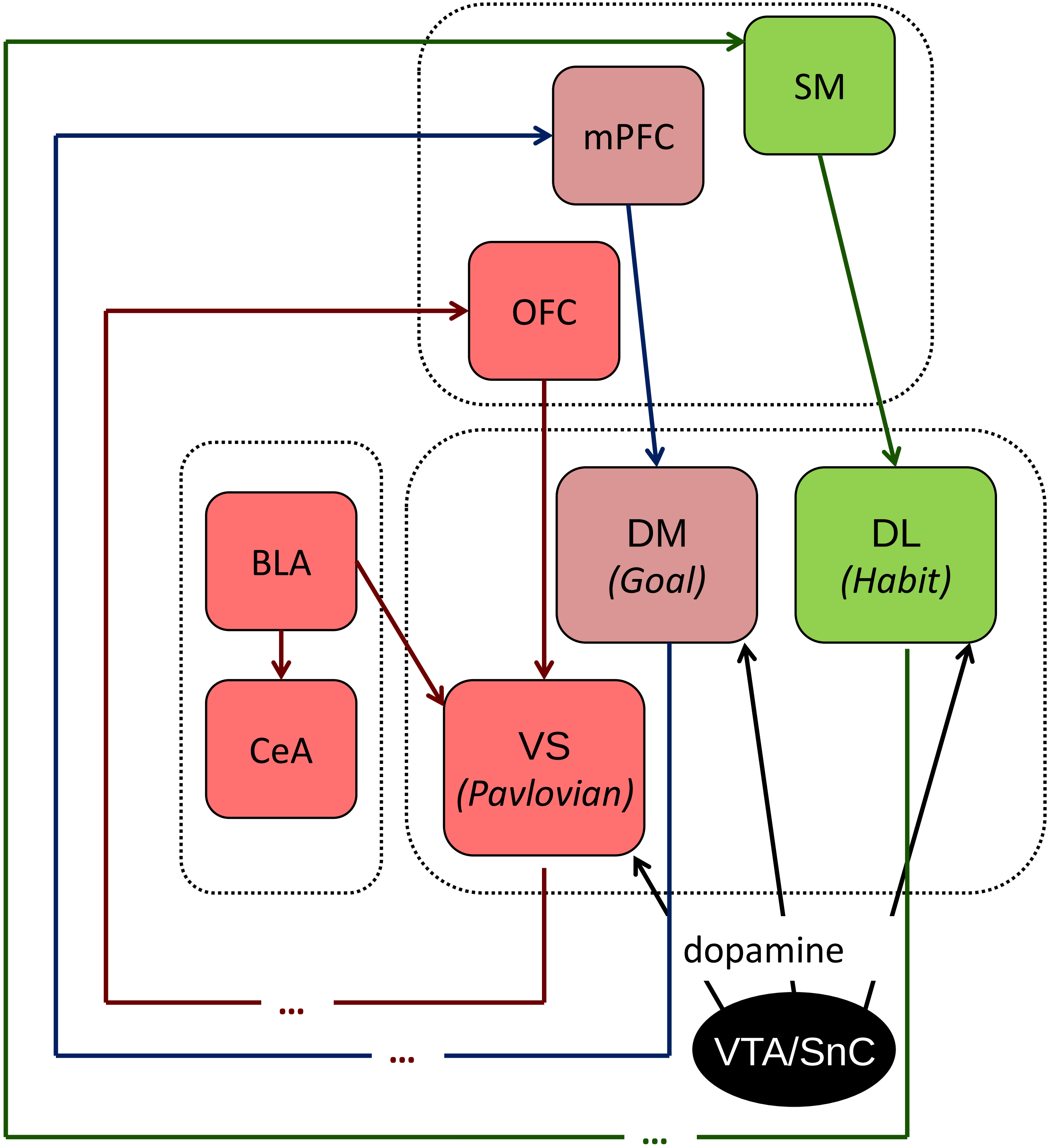
(Yin et al 2004)



Moderate training

(Yin et al 2005)

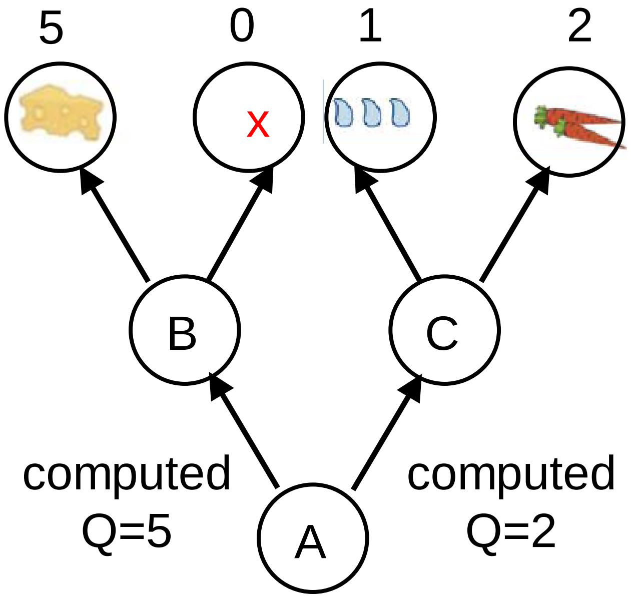


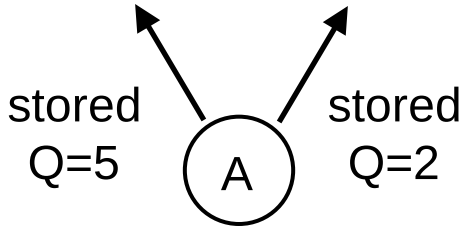

SM
mPFC
OFC
BLA
CeA
DM
(Goal)
DL
(Habit)
VS
(Pavlovian)
dopamine
VTA/SnC
...
...
...

outcome sensitivity

model-based:
can immediately adapt to value shifts
like goal-directed

model-free:
cannot immediately adapt
like habits



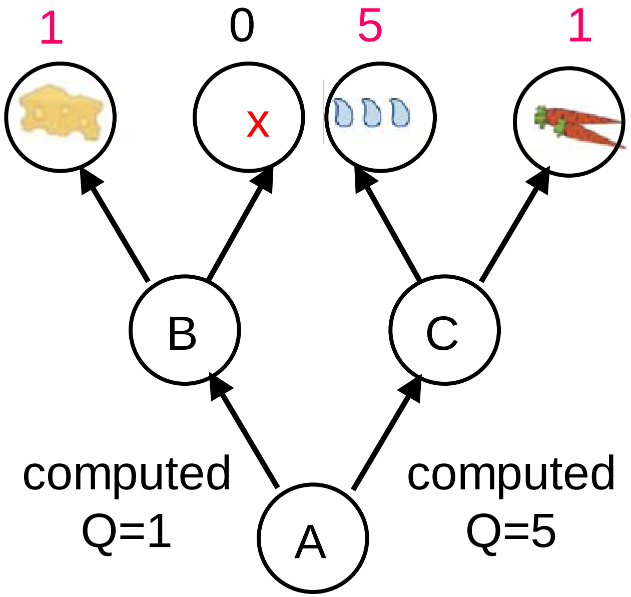


(Daw et al 2005)

outcome sensitivity

model-based:
can immediately adapt to value shifts
like goal-directed

model-free:
cannot immediately adapt
like habits



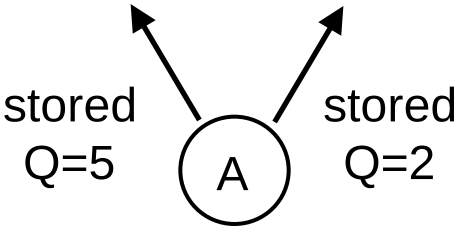
1 0 5 1

B C

computed
Q=1

A

computed
Q=5



stored
Q=5

A

stored
Q=2

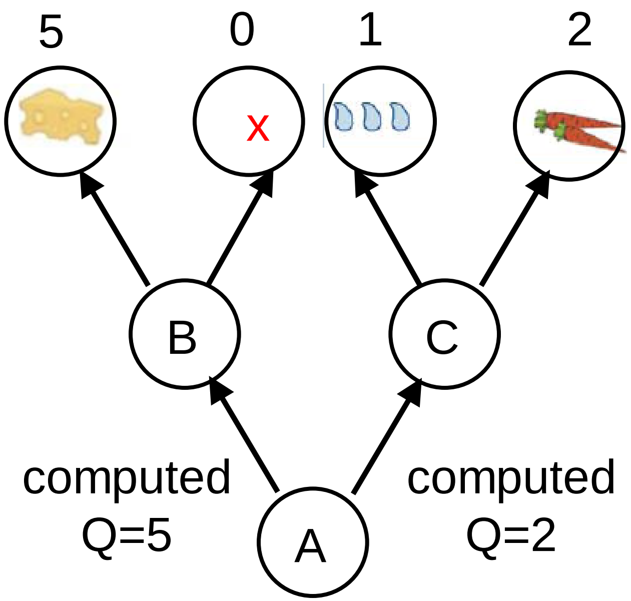
(Daw et al 2005)

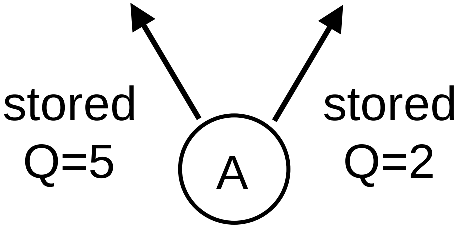
Why multiple systems

outcome sensitivity

model-based:
can immediately adapt to value shifts
like goal-directed

model-free:
cannot immediately adapt
like habits



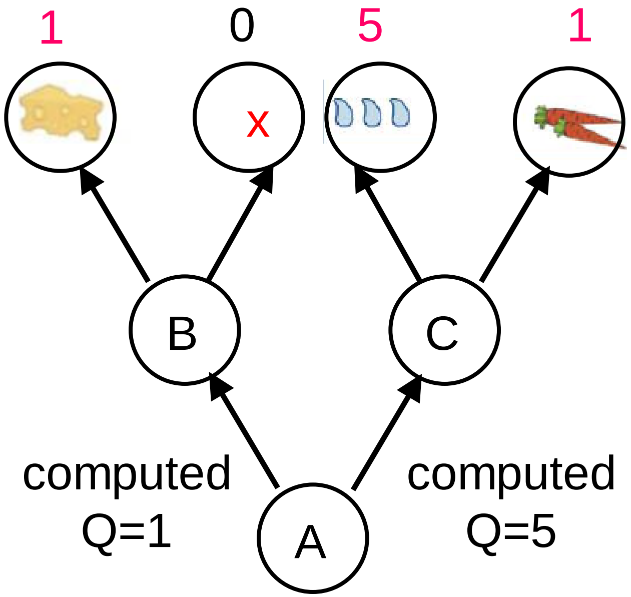


(Daw et al 2005)

outcome sensitivity

model-based:
can immediately adapt to value shifts
like goal-directed

model-free:
cannot immediately adapt
like habits



1 0 5 1

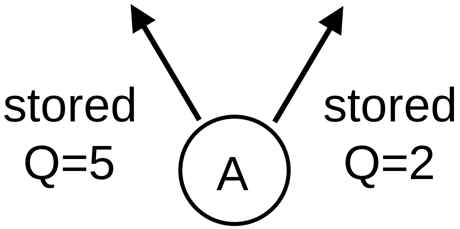
x

B C

computed
Q=1

A

computed
Q=5



stored
Q=5

A

stored
Q=2

(Daw et al 2005)

theory

why have multiple systems?

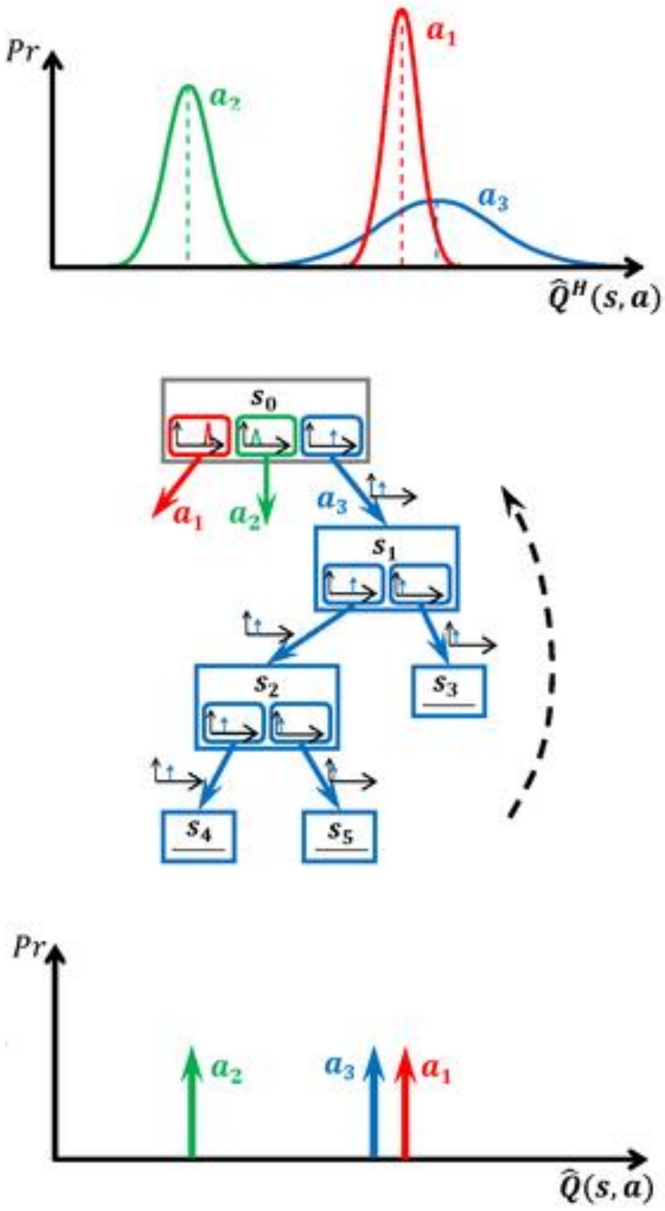
- computational efficiency vs statistical efficiency

when to favor each?

- itself a decision-theoretic tradeoff (cf Keramati et al. 2011)
- e.g. little value to deliberating when highly practiced on a stable task
- this model explains lots of data on what circumstances favor each system

how does the model-based system work?

(Daw et al. 2005)



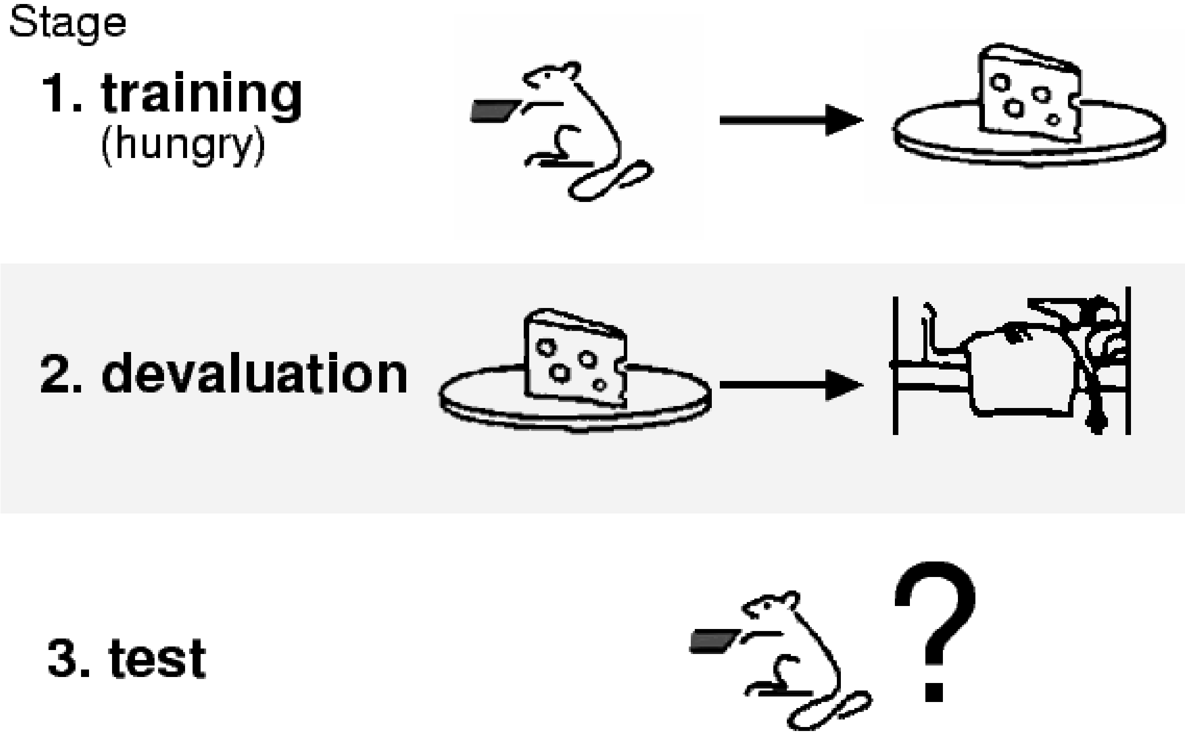
$$Gain_{s,a}(Q^*(s,a)) = \begin{cases} \hat{Q}^H(s,a_2) - Q^*(s,a) \\ \quad \text{if } a = a_1 \text{ and } Q^*(s,a) < \hat{Q}^H(s,a_2) \\ Q^*(s,a) - \hat{Q}^H(s,a_1) \\ \quad \text{if } a \neq a_1 \text{ and } Q^*(s,a) > \hat{Q}^H(s,a_1) \\ 0 \quad \text{otherwise} \end{cases}$$

$$\begin{aligned} VPI(s,a) &= E[Gain_{s,a}(Q^*(s,a))] \\ &= \int_{-\infty}^{\infty} Gain_{s,a}(x) Pr[Q^H(s,a) = x]\, dx \end{aligned}$$

(Keramati et al. 2011)

Human analogues

Unappealing approach



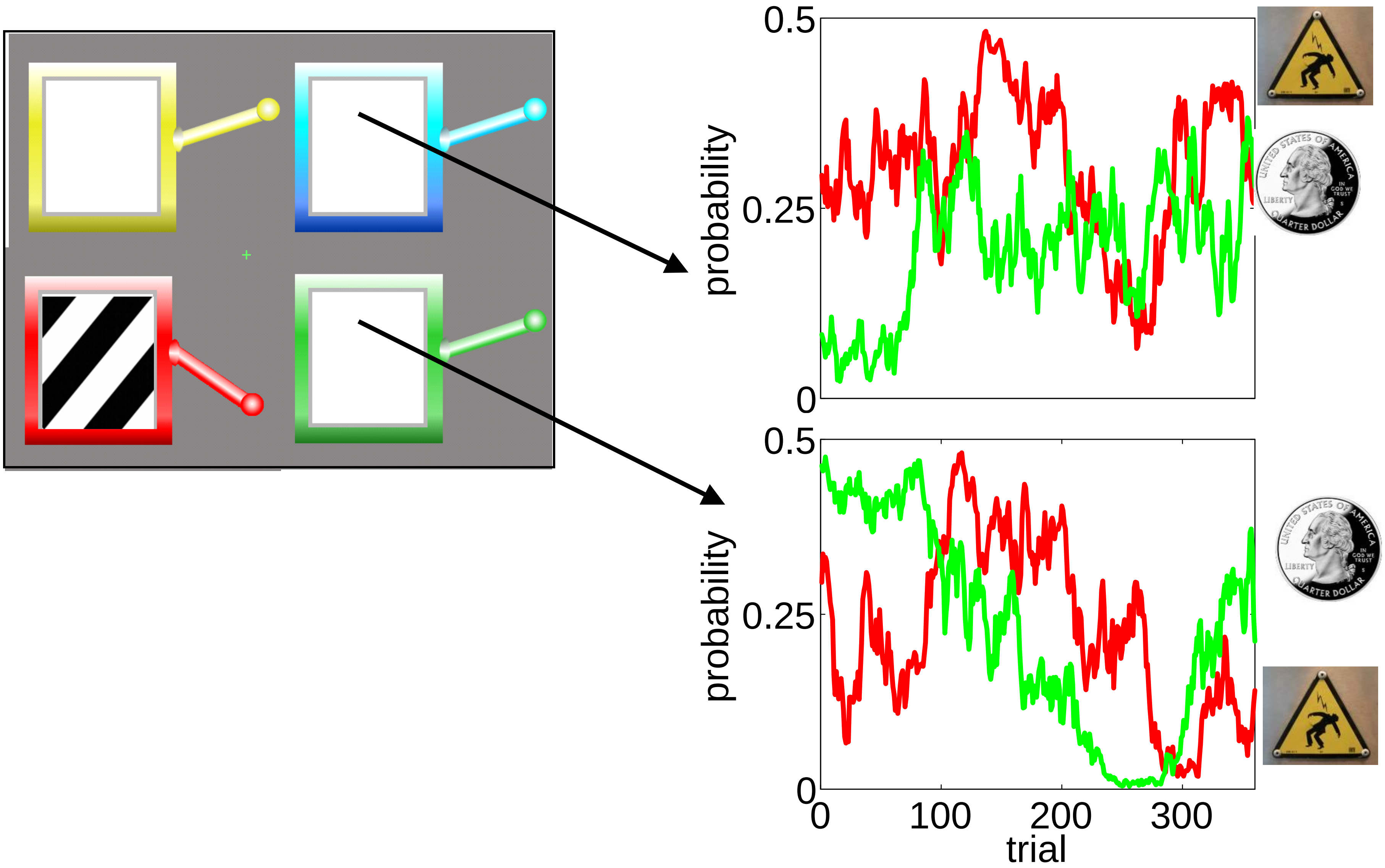
Stage

1. **training**
   (hungry)

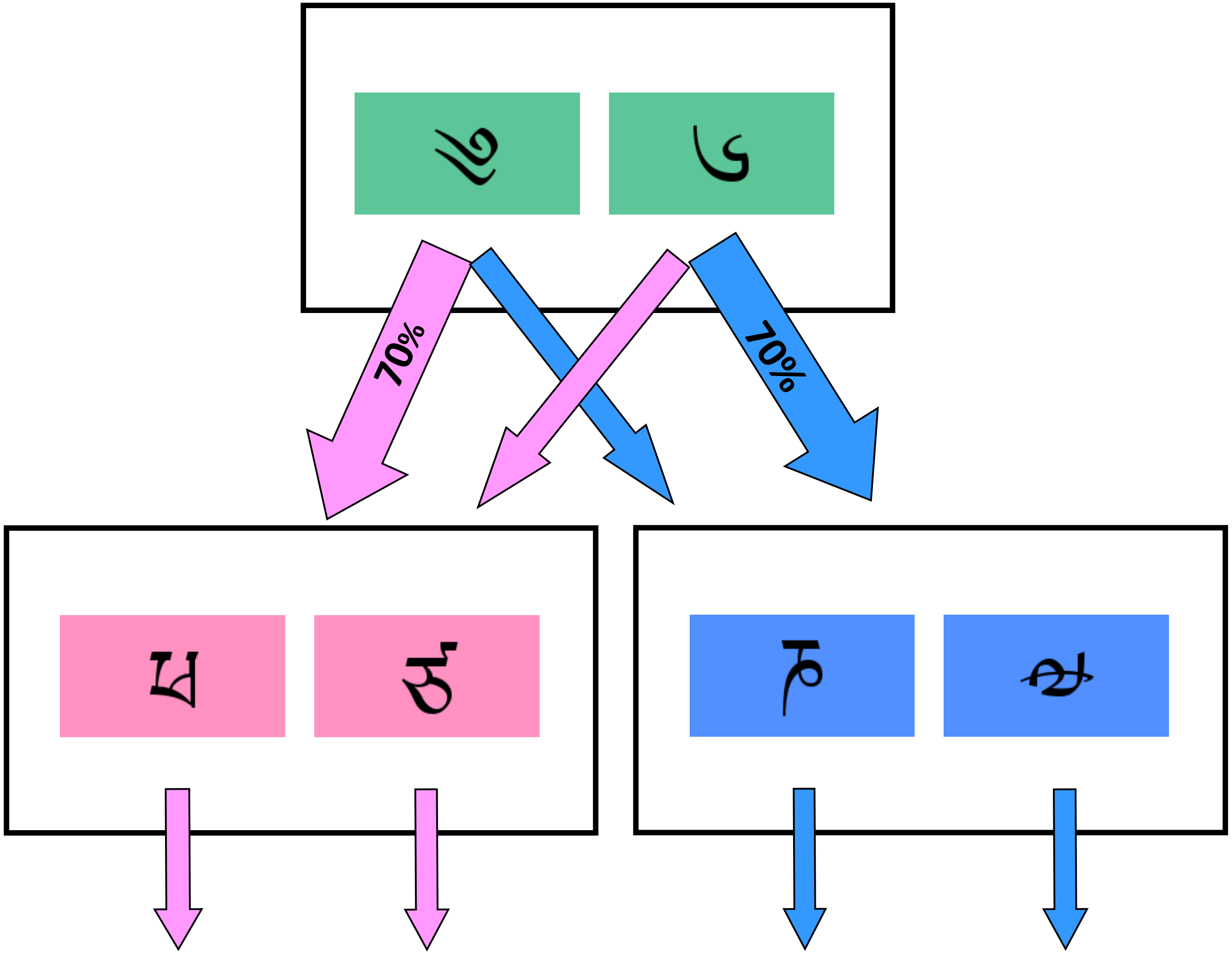
2. **devaluation**

3. **test**

learned decision making in humans



sequential decision task



with prob: 26% 57% 41% 28%

(*all slowly changing*)

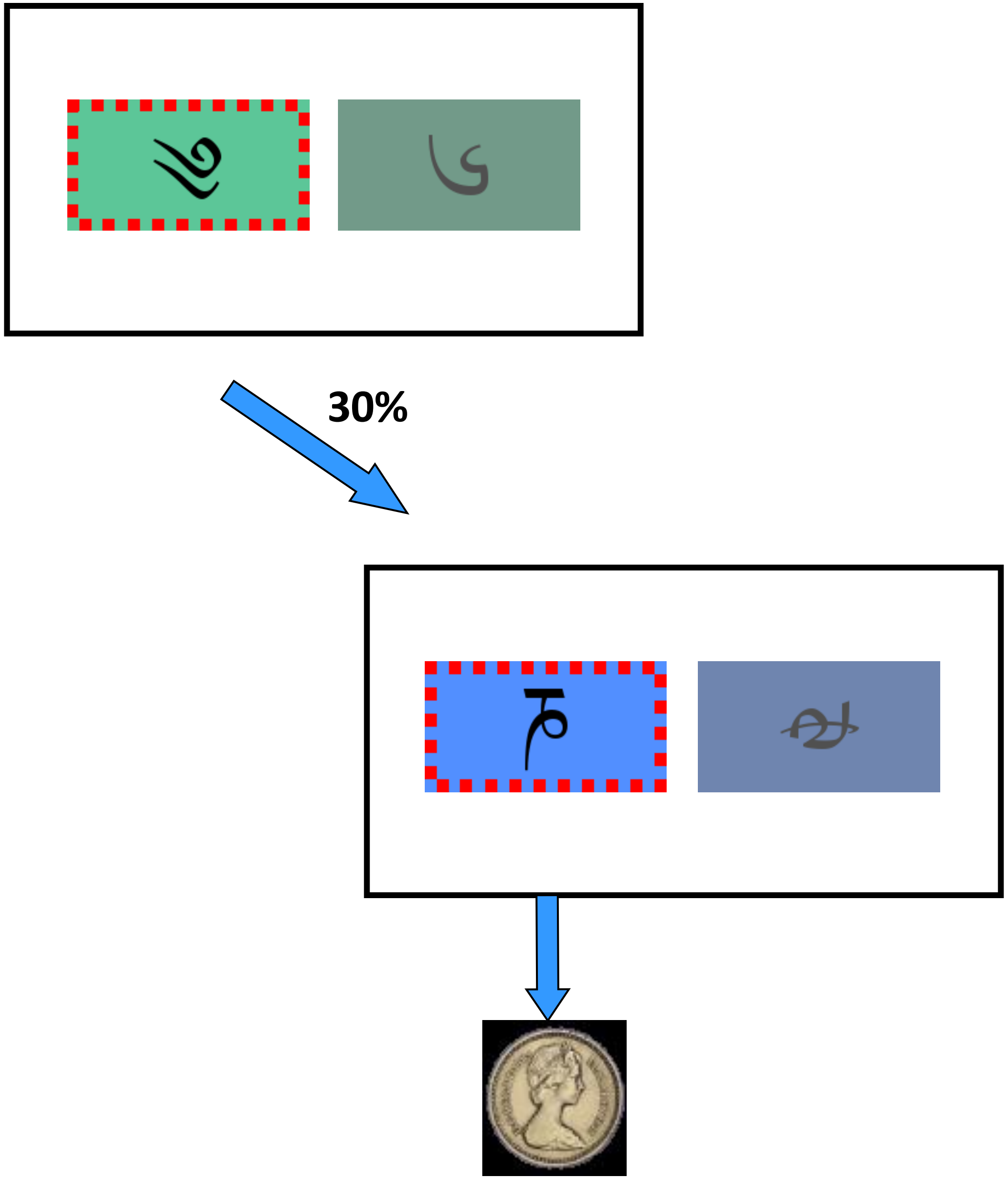
(Daw et al Neuron 2011)

idea

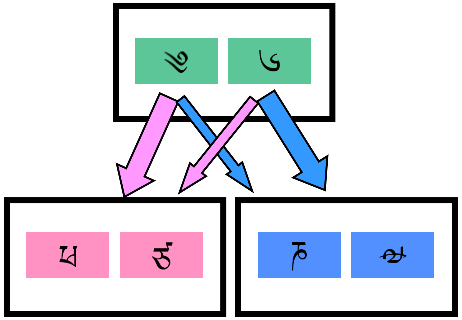
How does bottom-stage feedback affect top-stage choices?

Example: rare transition at top level, followed by win

- Which top-stage action is now favored?

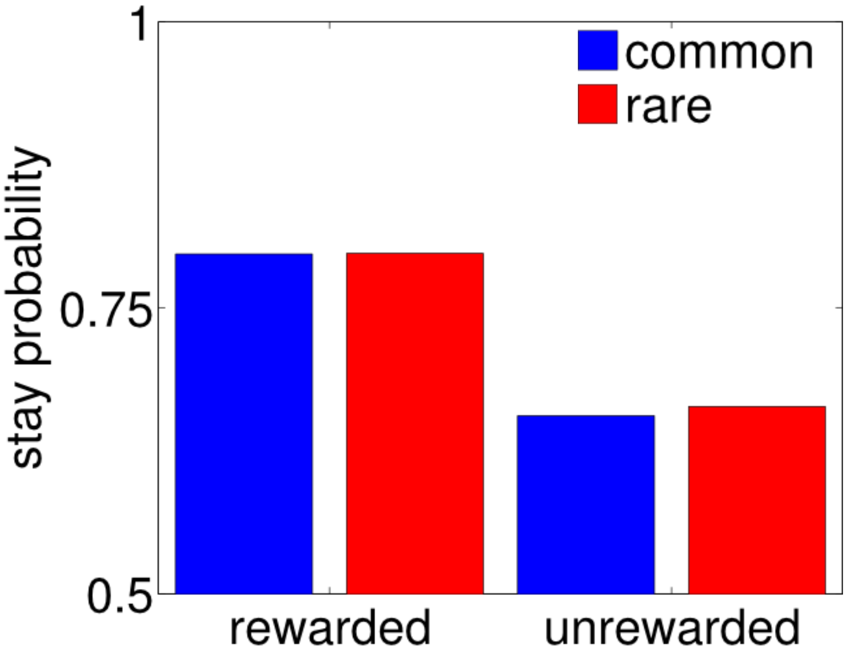


predictions



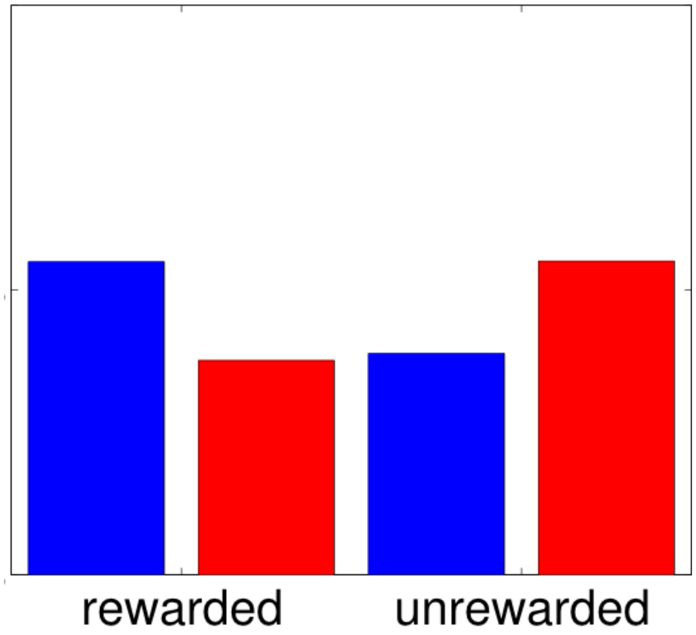
direct reinforcement

ignores transition structure

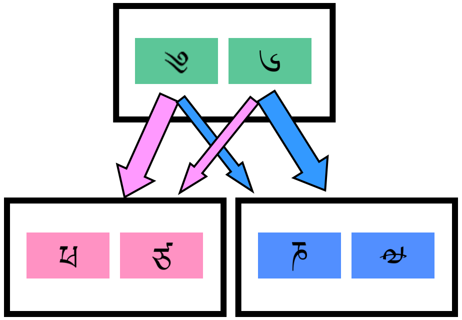


model-based planning

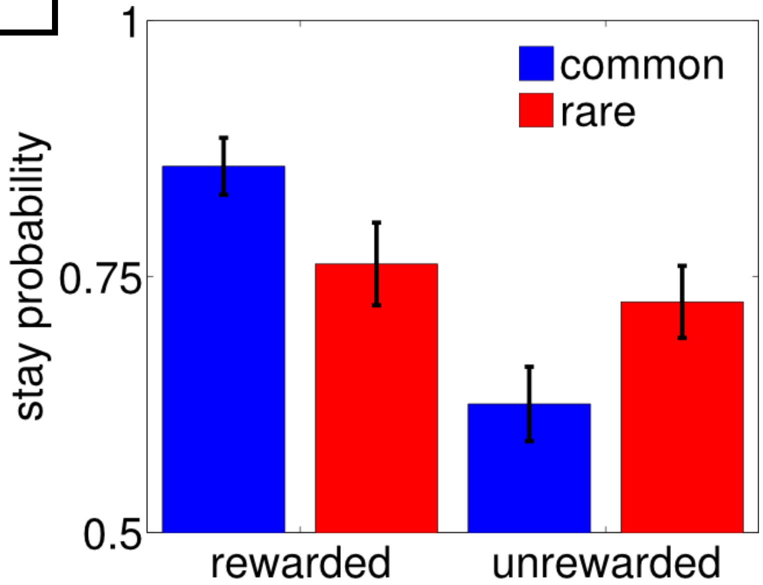
respects transition structure



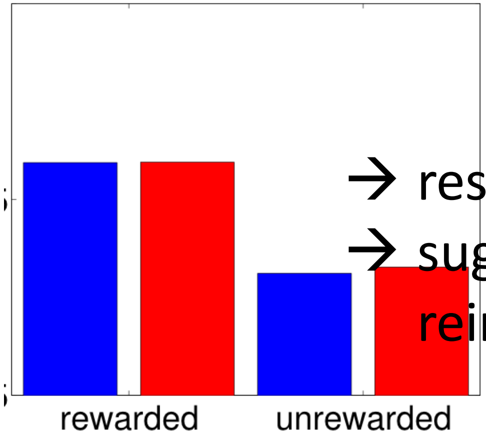
data



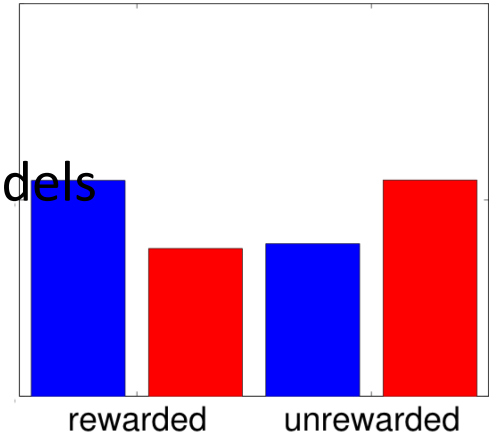
17 subs x 201 trials each



reward: p<1e-8
reward x rare: p<5e-5
(mixed effects logit)



reinforcement



planning

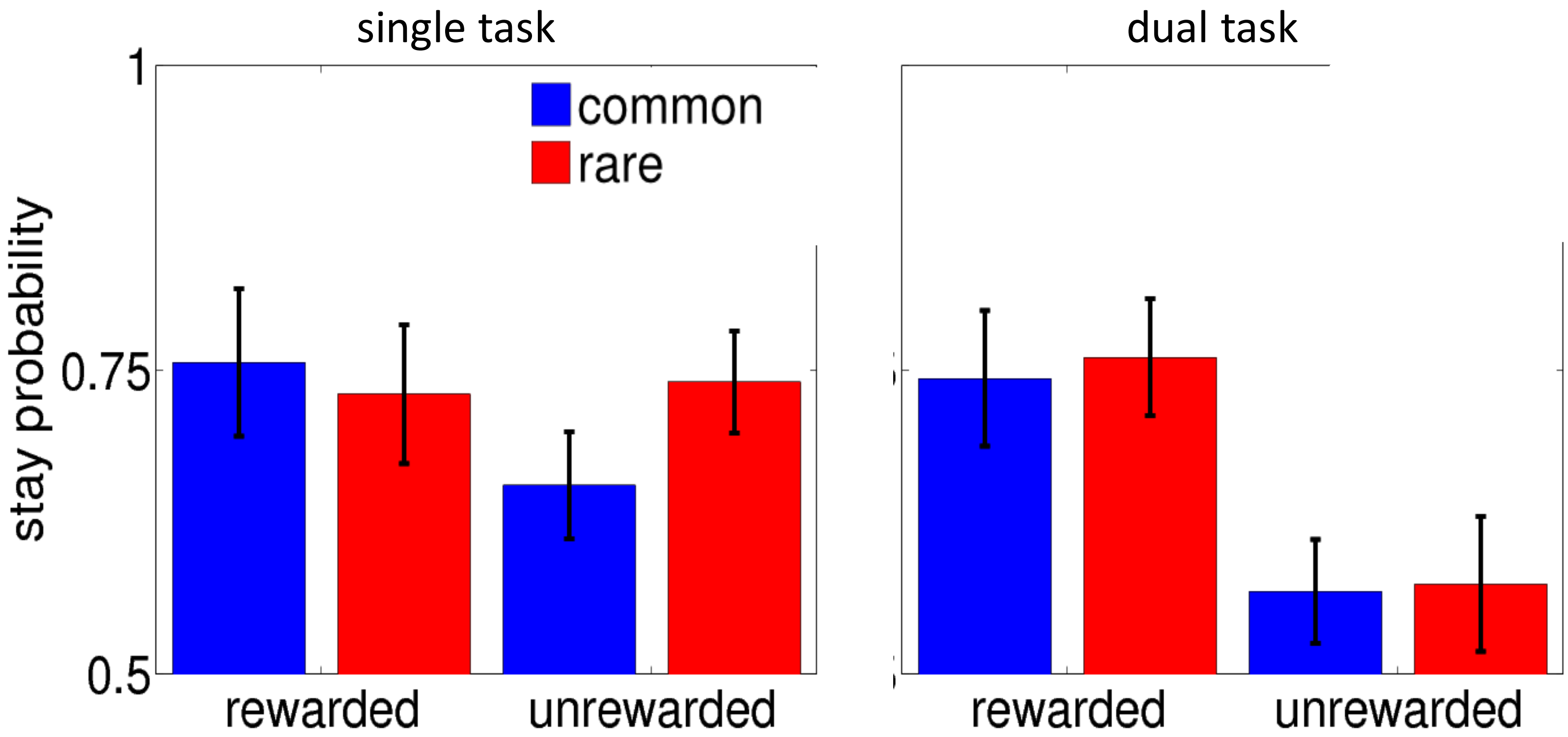
→ results reject pure reinforcement models

→ suggest mixture of planning and reinforcement processes

(Daw et al Neuron 2011)

Does this distinction track traditional measures of automaticity?

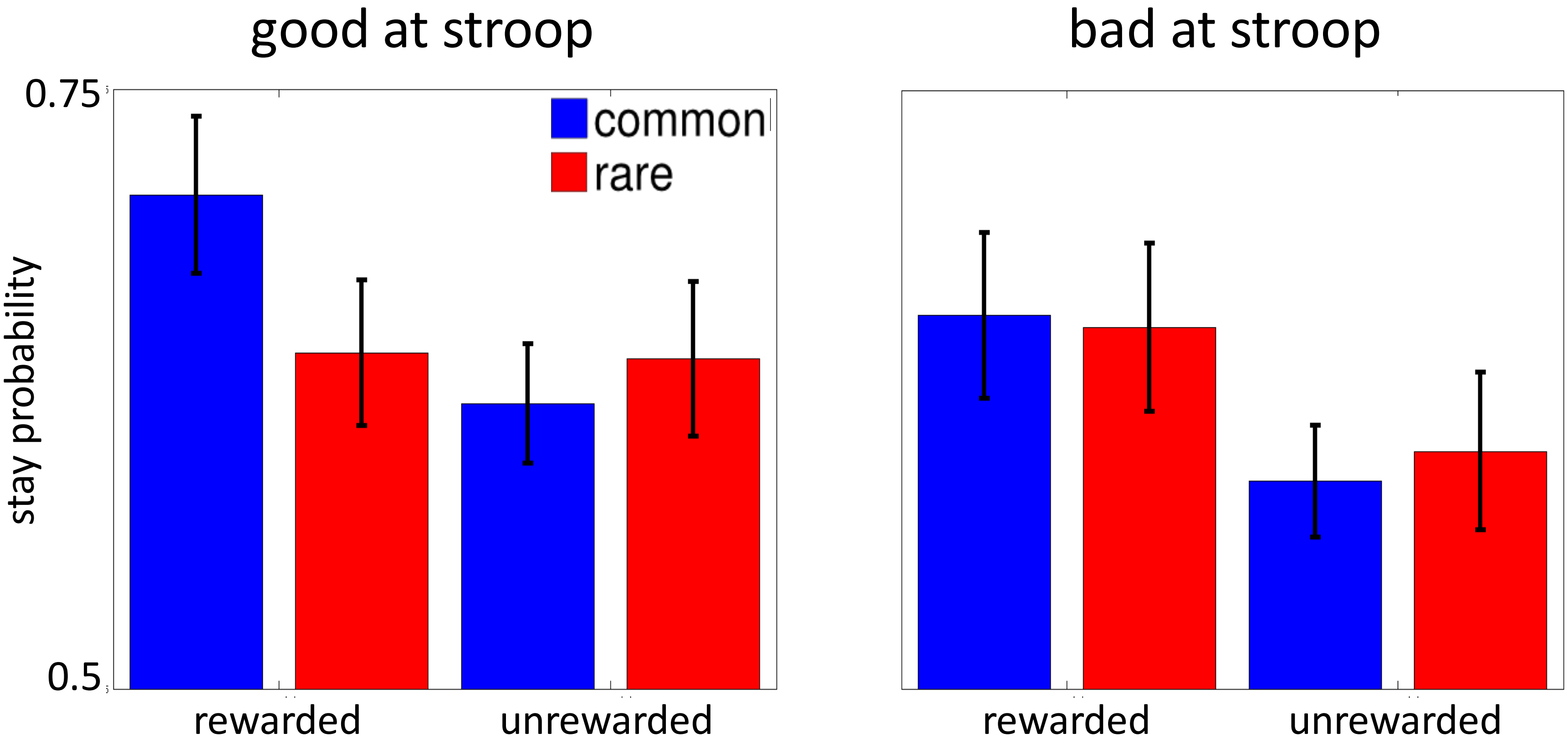
dual task



dual x reward: p < 5e-7
dual x reward x rare: p< .05

(Otto et al. in press)

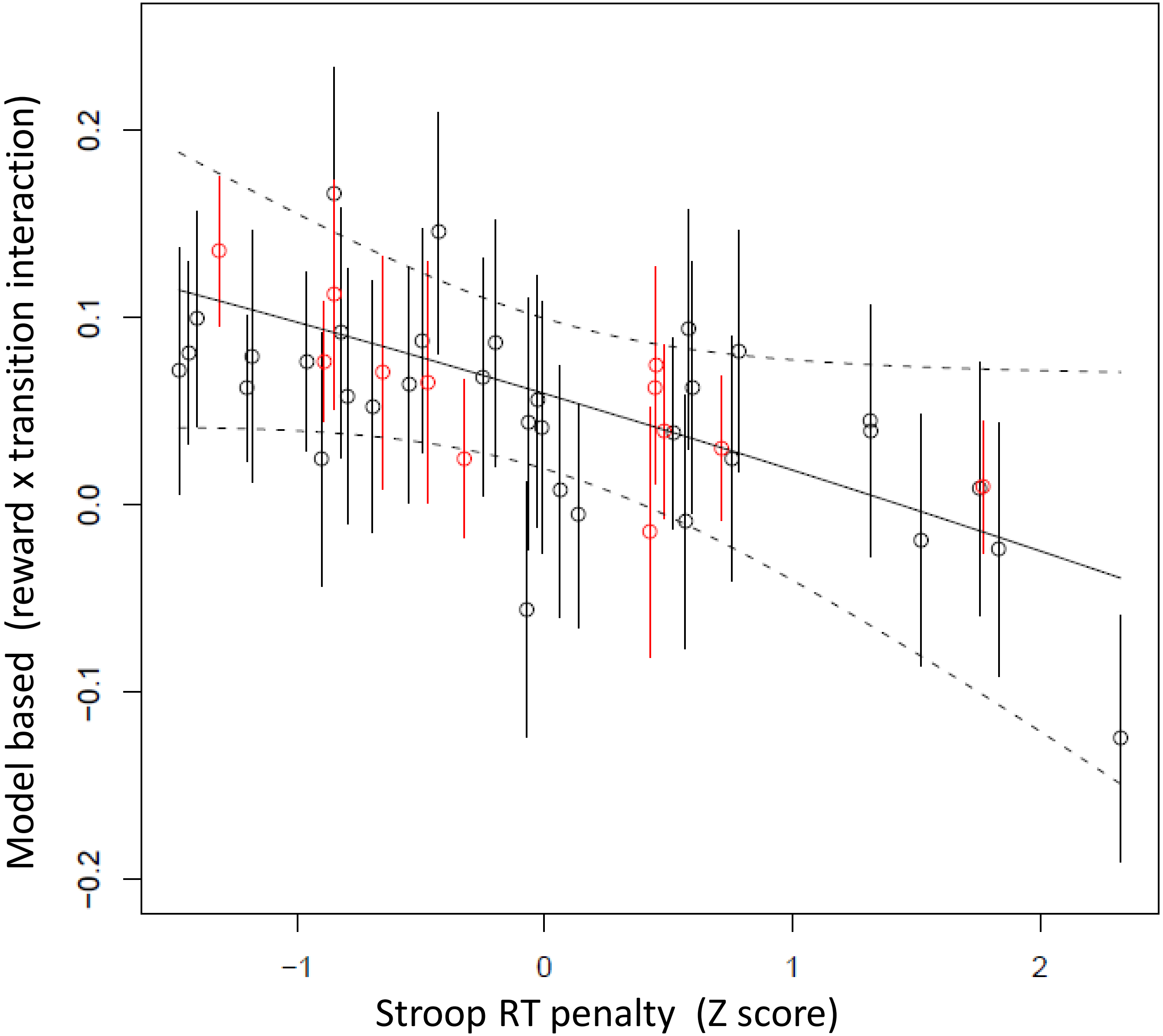
RED



good at stroop

bad at stroop

(Skatova et al in prep)



Degree of model-based learning increases with good cognitive control (P<.05)
→ suggests mechanism for arbitration

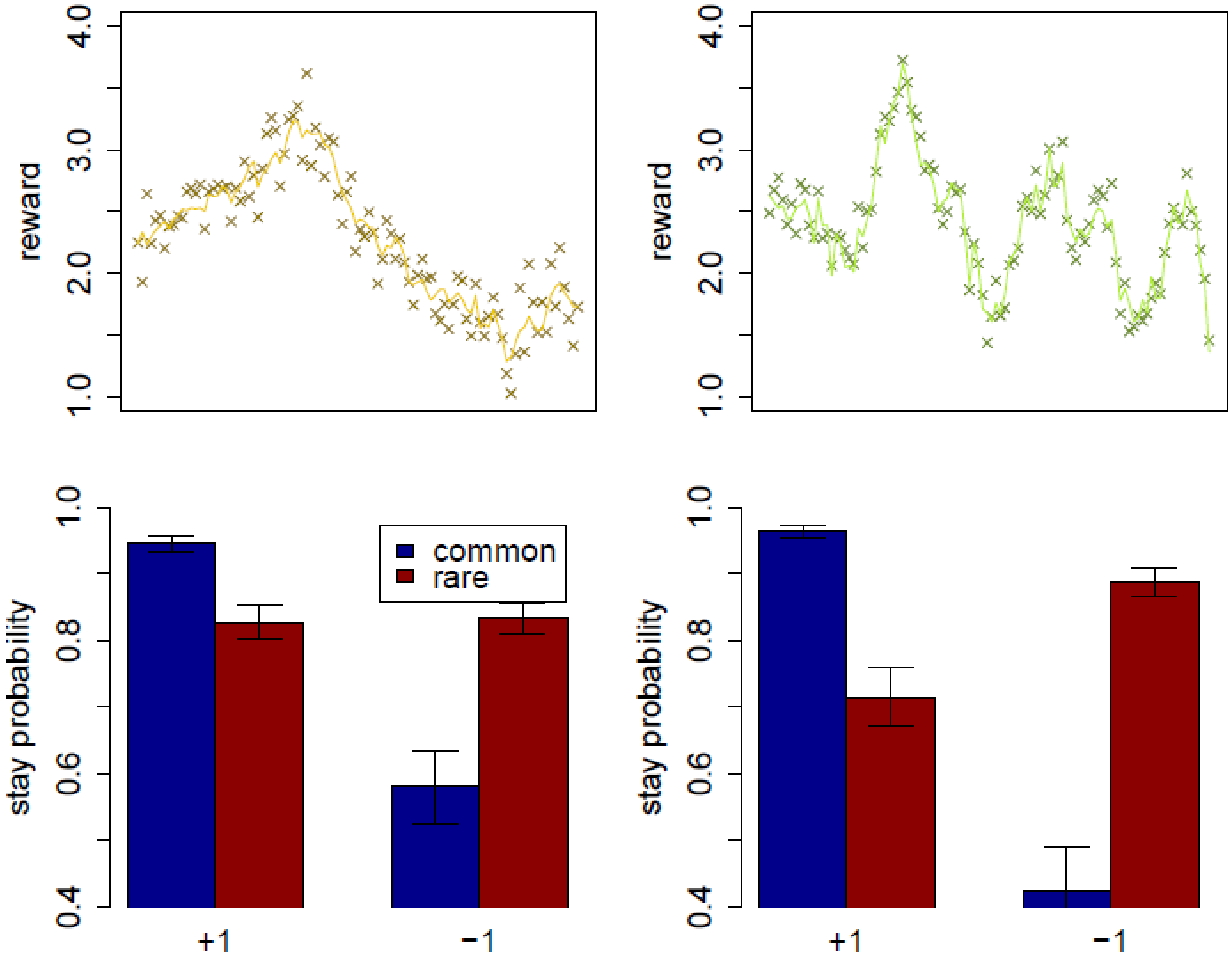
(Skatova et al in prep)

Can we modulate the tradeoff between these two sorts of learning?

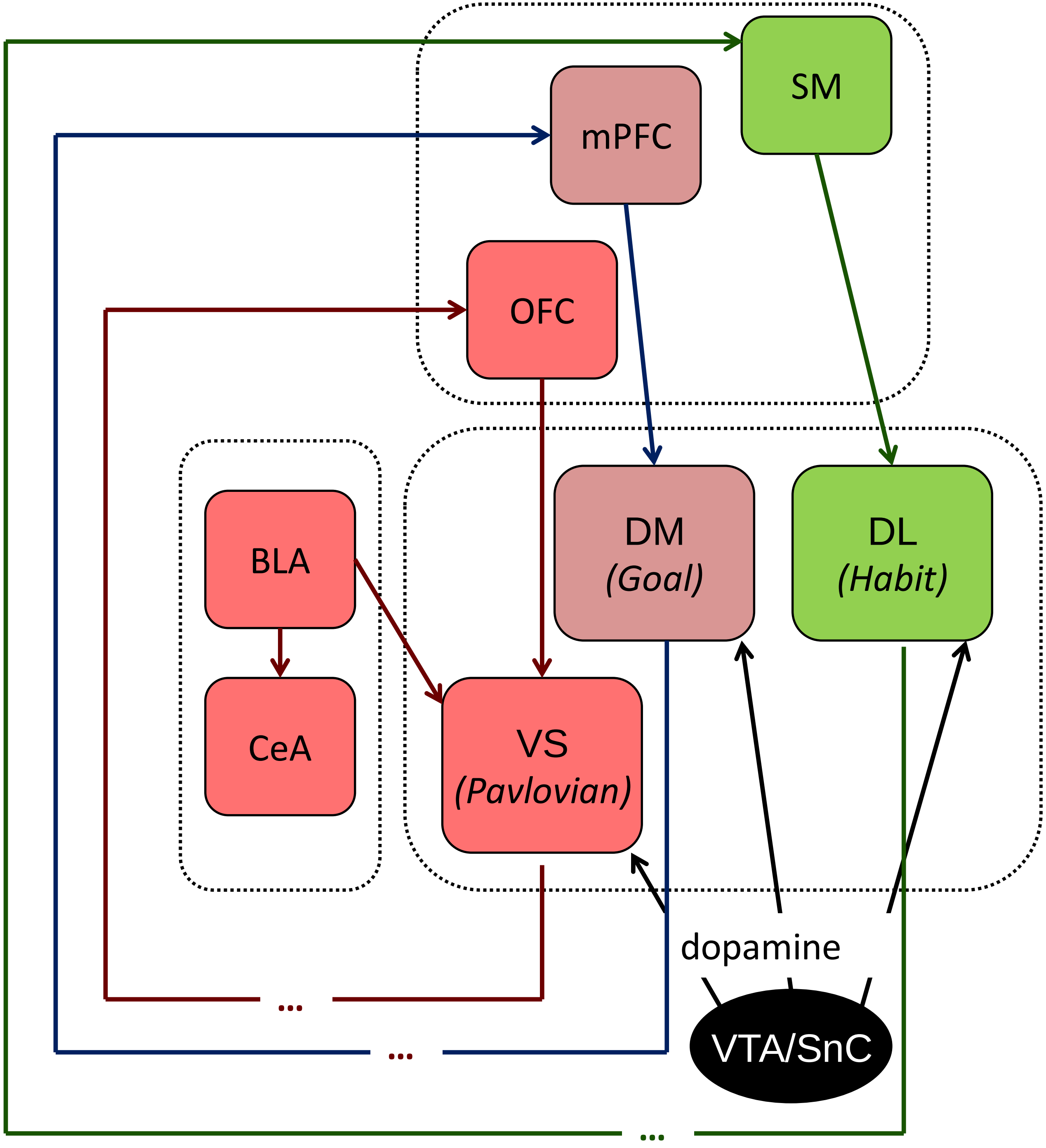
reward volatility

Idea (Daw et al. 2005): tradeoff between statistical efficiency (model based) and computational simplicity (model free)

→ hypothesis: faster change requires more data-efficiency, promotes model-based

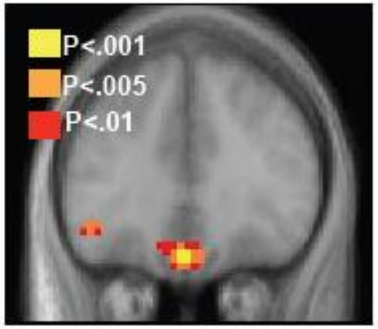


(Simon & Daw NIPS 2011 & in prep)


SM
mPFC
OFC
BLA
CeA
VS
(Pavlovian)
DM
(Goal)
DL
(Habit)
dopamine
VTA/SnC
...
...
...

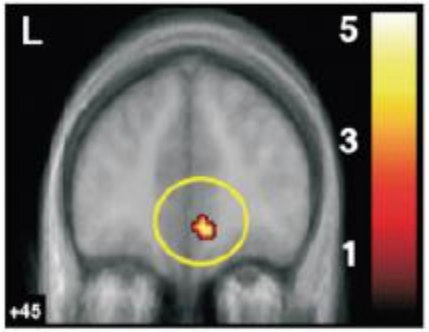
model-based regions in humans

devaluation

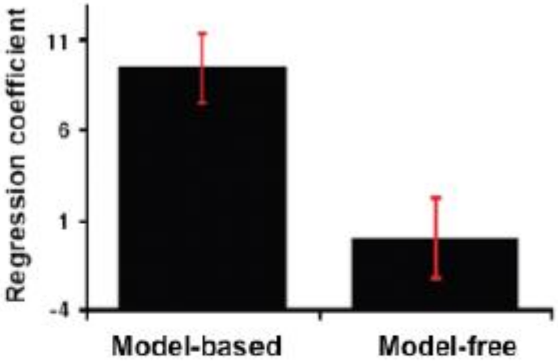


Valentin et al 2007

serial reversal

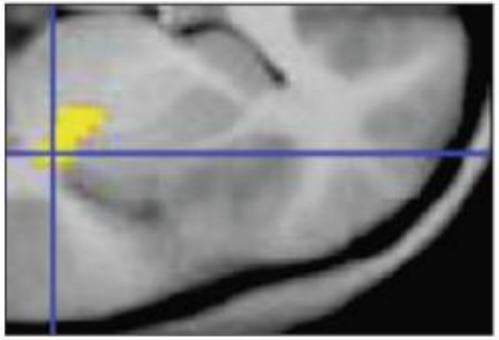


Hampton et al. 2006



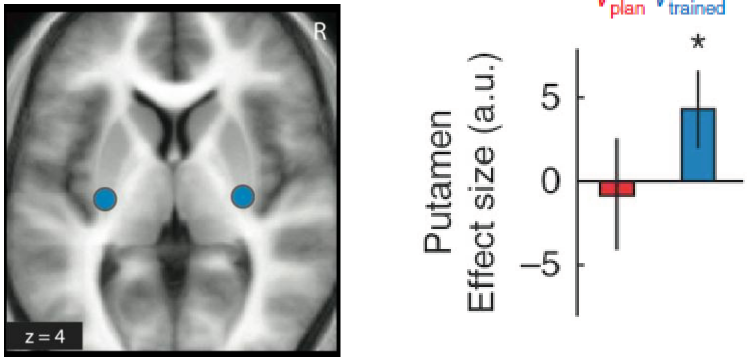
overtraining regions in humans (model free?)

devaluation



Tricomi et al. 2009

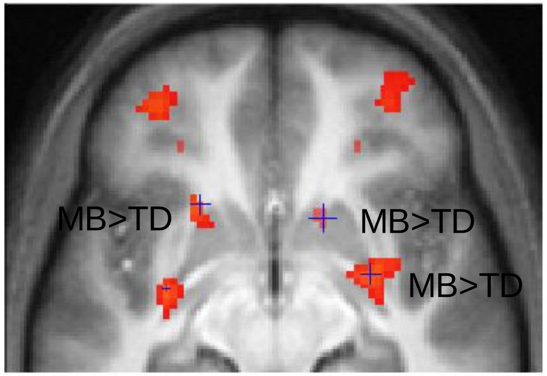
sequential RL



Wunderlich et al. 2012

maze navigation

But:



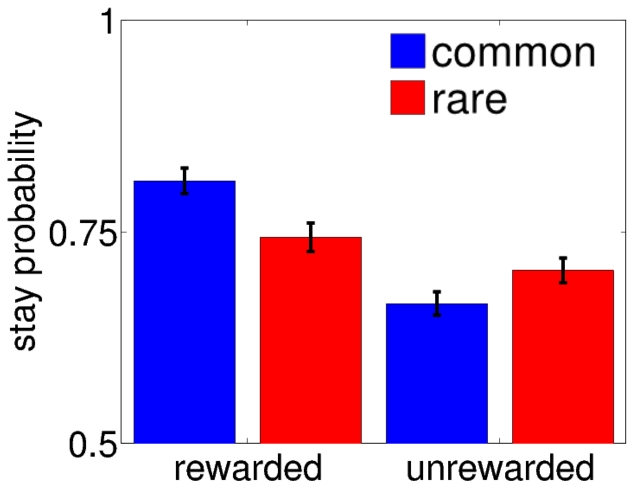
Simon & Daw 2011

Psychiatric implications

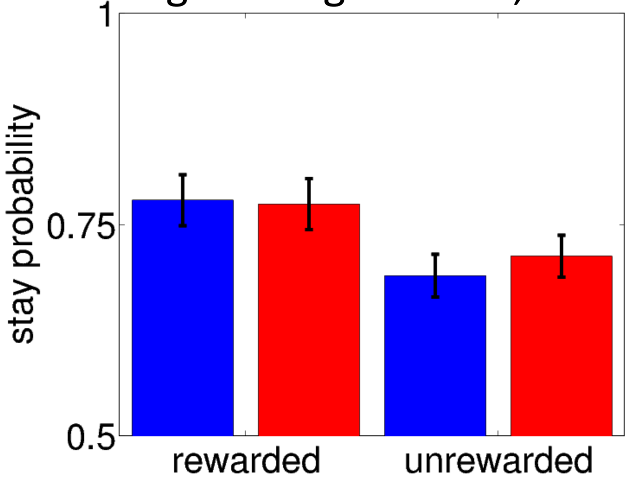
Psychiatric implications

1. Compulsion: widely assumed that model free system is automatic, and may underlie compulsion as in drug abuse, dieting etc.

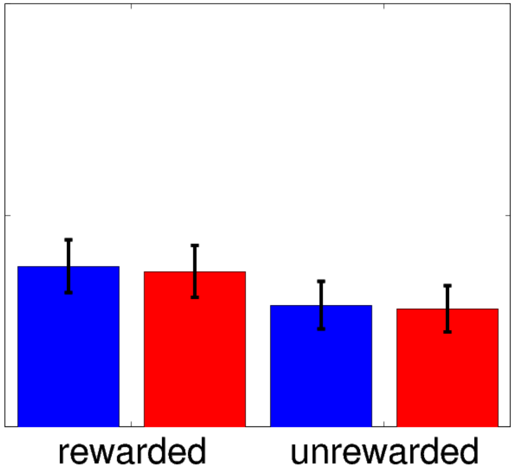
Healthy volunteers, n=106



Binge eating disorder, n=30

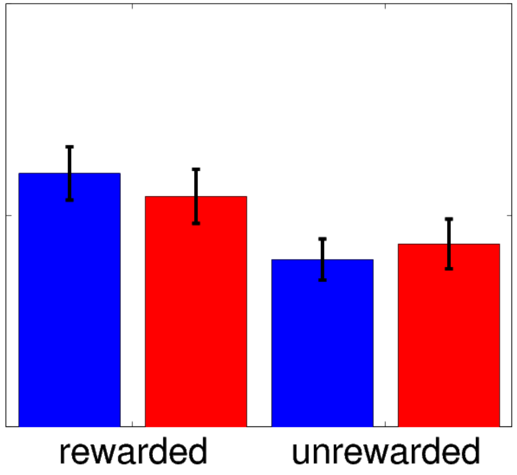


Stimulant abusers, n=36



Methamphetamine/cocaine
Abstinent at least 1 wk

OCD, n=35



Valerie Voon et al., under review

Psychiatric implications

1. Compulsion: widely assumed that model free system is automatic, and may underlie compulsion as in drug abuse, dieting etc.
2. Theory of mind: In multiplayer interactions, model-based RL amounts to learning a model of the opponents' beliefs. This may have relevance to autism etc.

p-beauty contest

- Write down your initials and an integer between 0 and 100, inclusive
- we will average all entries. The contestant who picks closest to 2/3 of the average wins the prize (a drink)
- Prize split in case of tie

- what did you choose?

- why?

- what do you think your colleagues chose?

Why is this called a p-beauty contest?

- Keynes (1936):

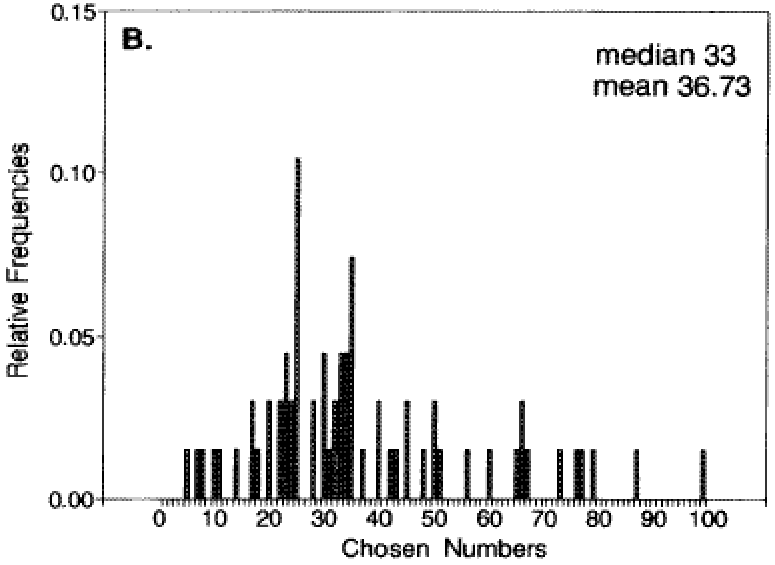
> It is not a case of choosing those [faces] which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth and higher degrees.

- Economists are fond of old quotes.

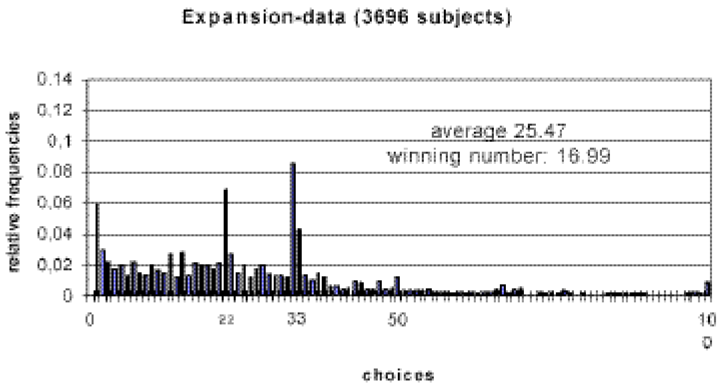
Beverly Christensen
Which will You elect Miss Rheingold 1957?
Pick the girl who'll win a contract worth $50,000! Vote at any Rheingold store or tavern!
Rheingold
EXTRA DRY
Margie McNally
Diane Baker

Results

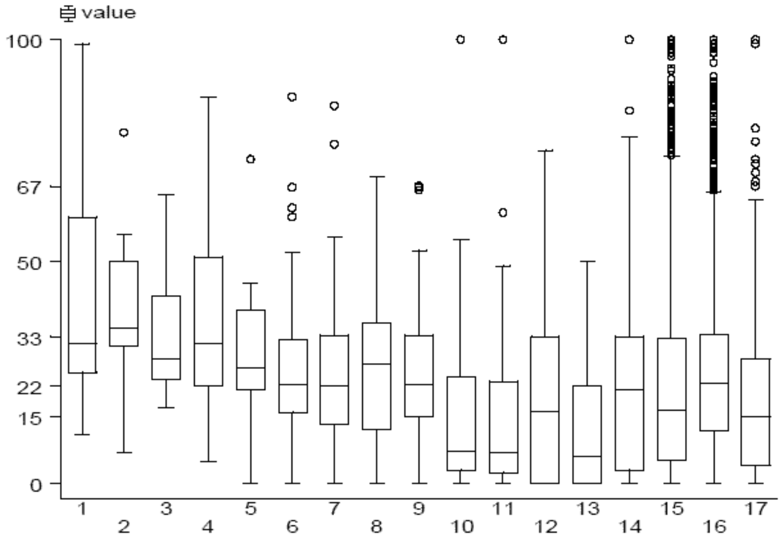
- Mean around 25-40; win around 16-27
- Suggests 0-3 rounds of iterated reasoning



German undergrads - Nagel 1995



Spanish newspaper - Nagel et al. 1999



Many studies - Nagel et al. 1999

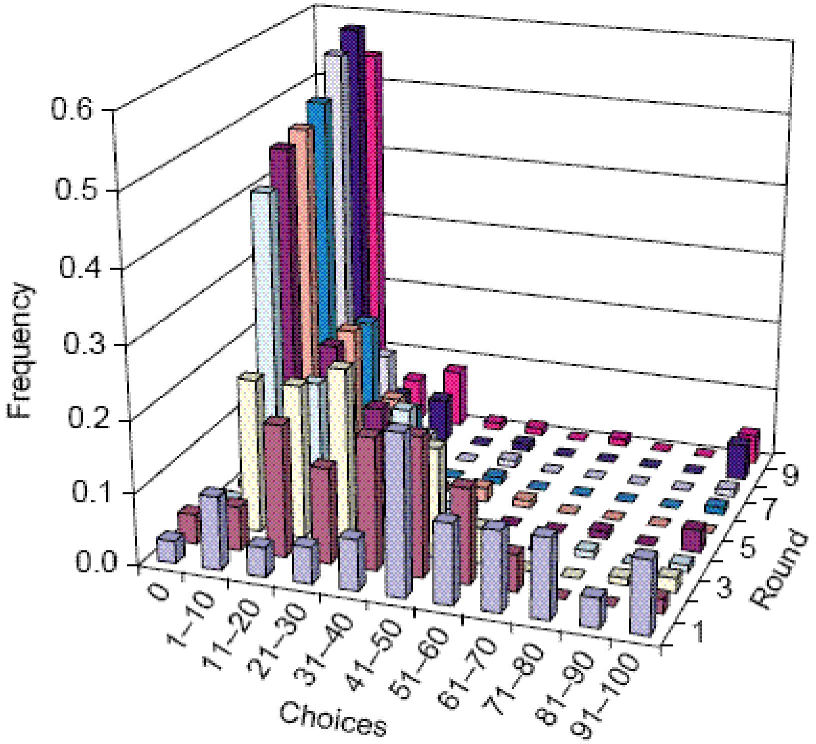
learning in p-beauty contest

- how does learning look with repeated play in p-beauty contest?
- do subjects approach equilibrium?
- how does this learning relate to the mechanisms and principles we talked about yesterday?

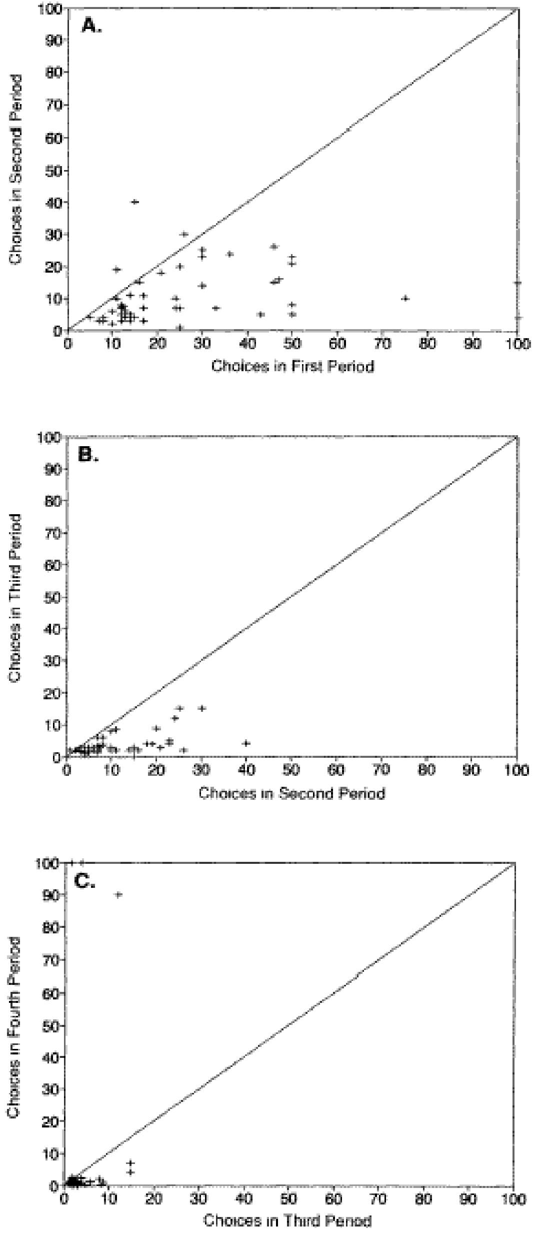
Which will You elect Miss Rheingold 1957?
Pick the girl who'll win a contract worth $50,000! Vote at any Rheingold store or tavern!
Rheingold
EXTRA DRY

equilibration

- fast approach to equilibrium with repeated play
  - 0 a bad guess initially but a good guess pretty soon



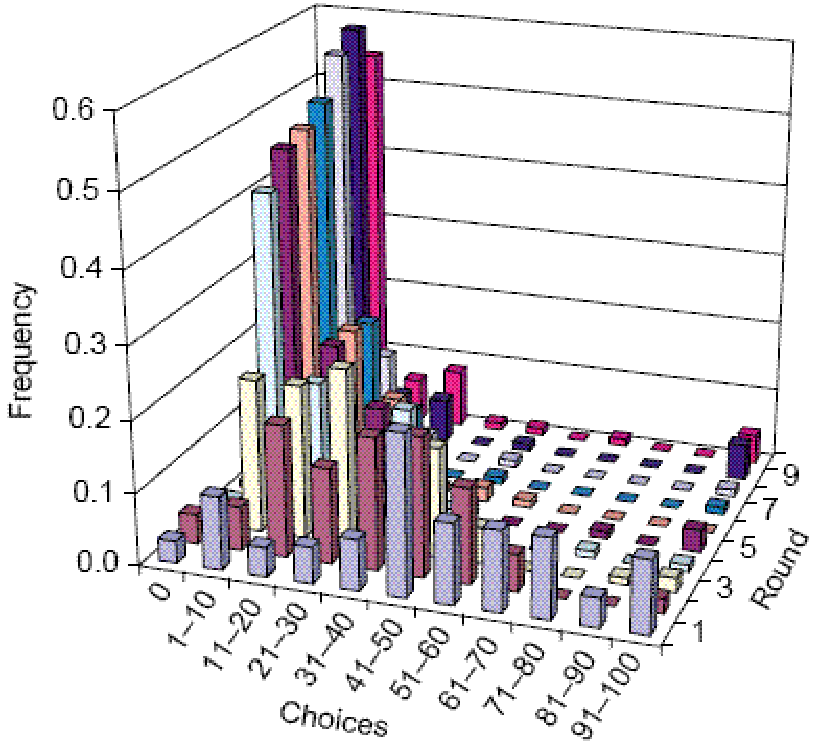
Singaporean undergrads – Ho et al. 1998



German undergrads – Nagel 1995

equilibration

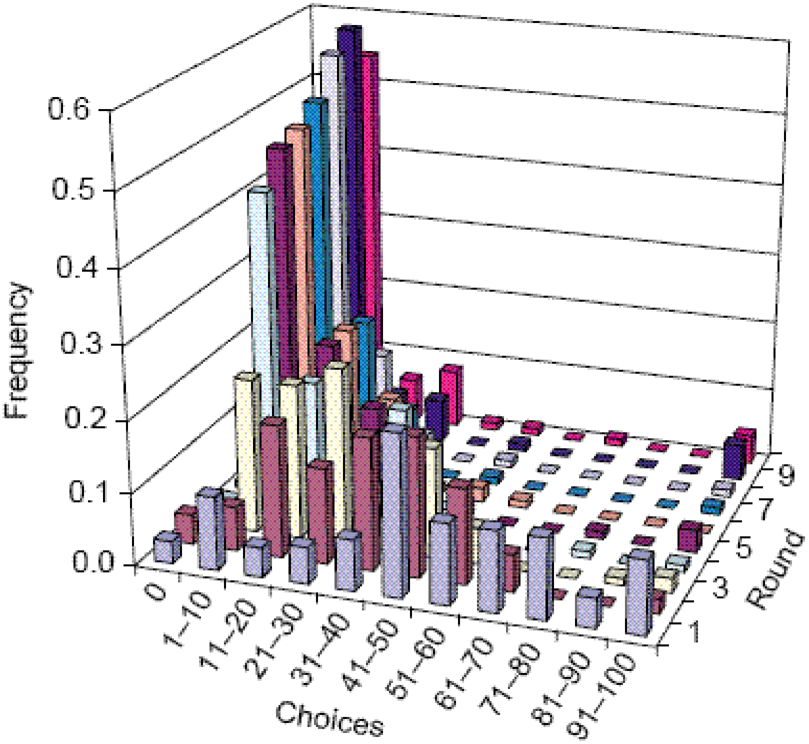
- what does law of effect (simple TD, etc) predict about p-BC learning?
- what's the problem here?



Singaporean undergrads – Ho et al. 1998

cognitive maps

- what is the counterpart of a cognitive map in this sort of task?



- EWA theory (Camerer & Ho) treats learning in games as weighted sum of model-based (belief learning, iterative reasoning) and model-free
- Different games (& different individuals) produce different levels of model-basedness

Psychiatric implications

1. Compulsion: it is widely assumed that model free system is automatic, and may underlie compulsion as in drug abuse, dieting etc.
2. Theory of mind: In multiplayer interactions, model-based RL amounts to learning a model of the opponents' beliefs. This may have relevance to autism etc.
3. Reward processing & motivation: while many have noted that, e.g. schizophrenia, involves impaired associative learning and reward processing, it is not known which sort

Open questions

- Are the systems really separate or interacting? How to understand this computationally?
- Are there more than two systems (e.g. a separate episodic or spatial controller)
- Why do people use more or less belief learning in different games?
- How do these ideas map onto other dual-process models throughout psychology and neuroscience

NYU:

Sam Gershman (*now Princeton*)
Ross Otto
Dylan Simon
Seth Madlon-Kay
Aaron Bornstein
Sara Constantino
Nick Gustafson
Y-Lan Boureau
Daniel Campbell-Meiklejohn
Brad Doll
Steve Fleming
Jian Li
Hanneke den Ouden
Mattia Rogatti

Elsewhere:

Yael Niv
Ben Seymour
Peter Dayan
Ray Dolan
Anya Skatova
Valerie Voon

Funding:

NIMH
NIDA
NINDS
NARSAD
HFSP
McDonnell Foundation
McKnight Endowment