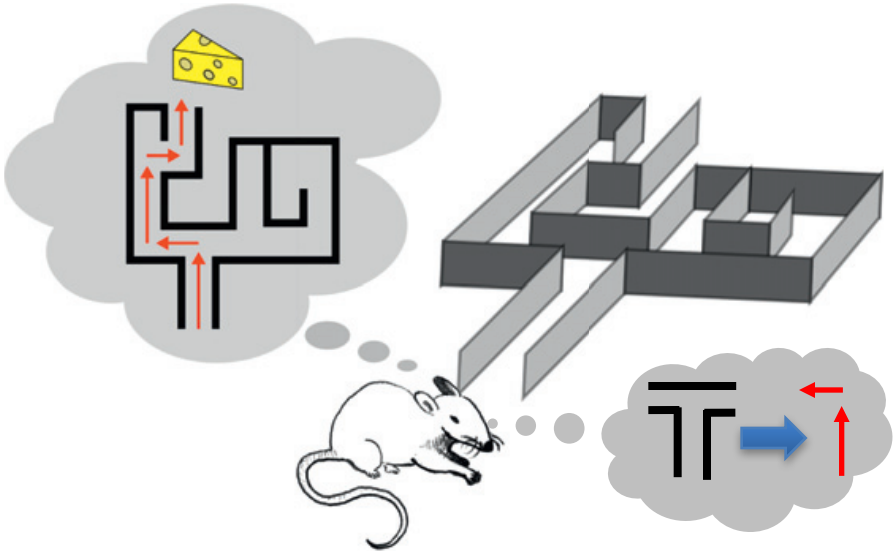
multiple decision systems and compulsion

nathaniel daw
princeton neuroscience institute &
department of psychology
princeton university

zurich, computational psychiatry

learning for decisions

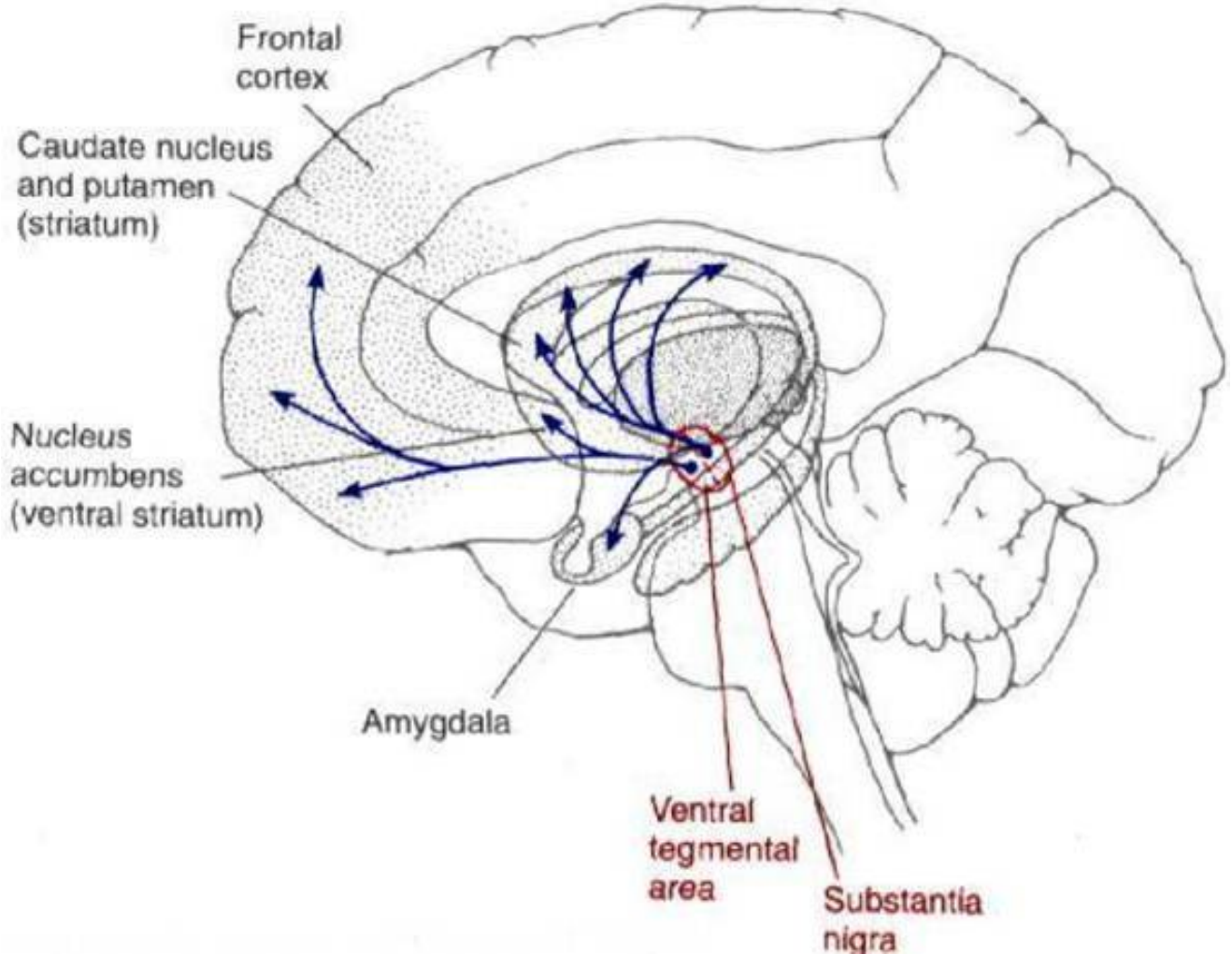


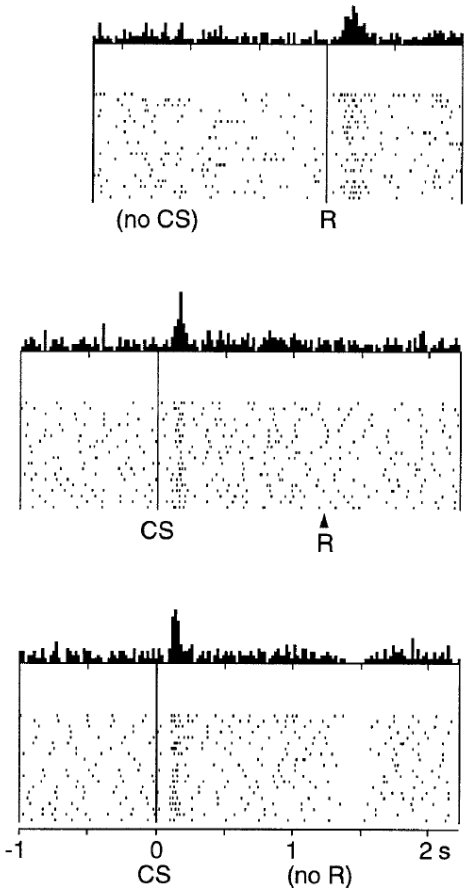
multiple systems for trial-and-error decision making: habits, slips of action, & compulsion

1. characterizing this distinction computationally via different learning strategies: model-based and model-free RL
2. are these mechanisms compromised in psychiatric disorders?

- midbrain dopamine neurons signal reward prediction errors (Schultz et al 1997)
- … driving plasticity at synapses on striatal medium spiny neurons (Reynolds & Wickens 2002)
- … influencing behavior via direct and indirect pathways (Frank et al 2004; Kravitz et al 2012)
- … implementing a temporal-difference (TD) learning rule for trial-and-error choices

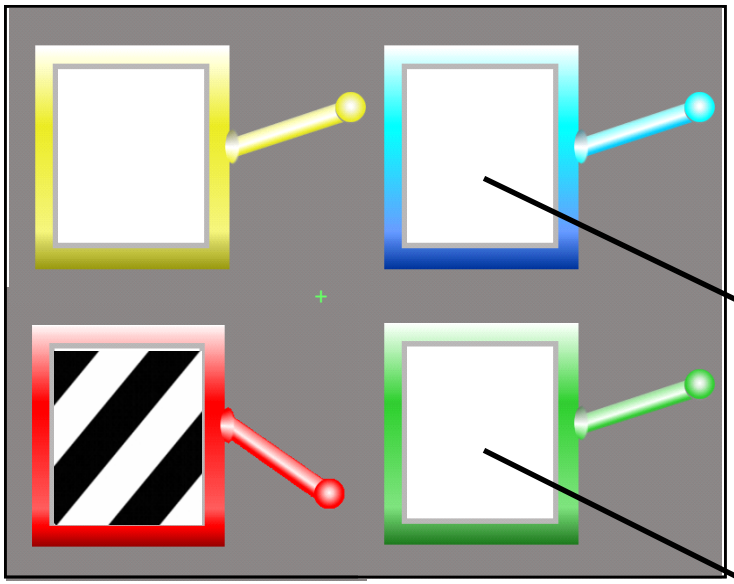
→ Learn action/reward relationships, repeat rewaded actions (Thorndike 1911)

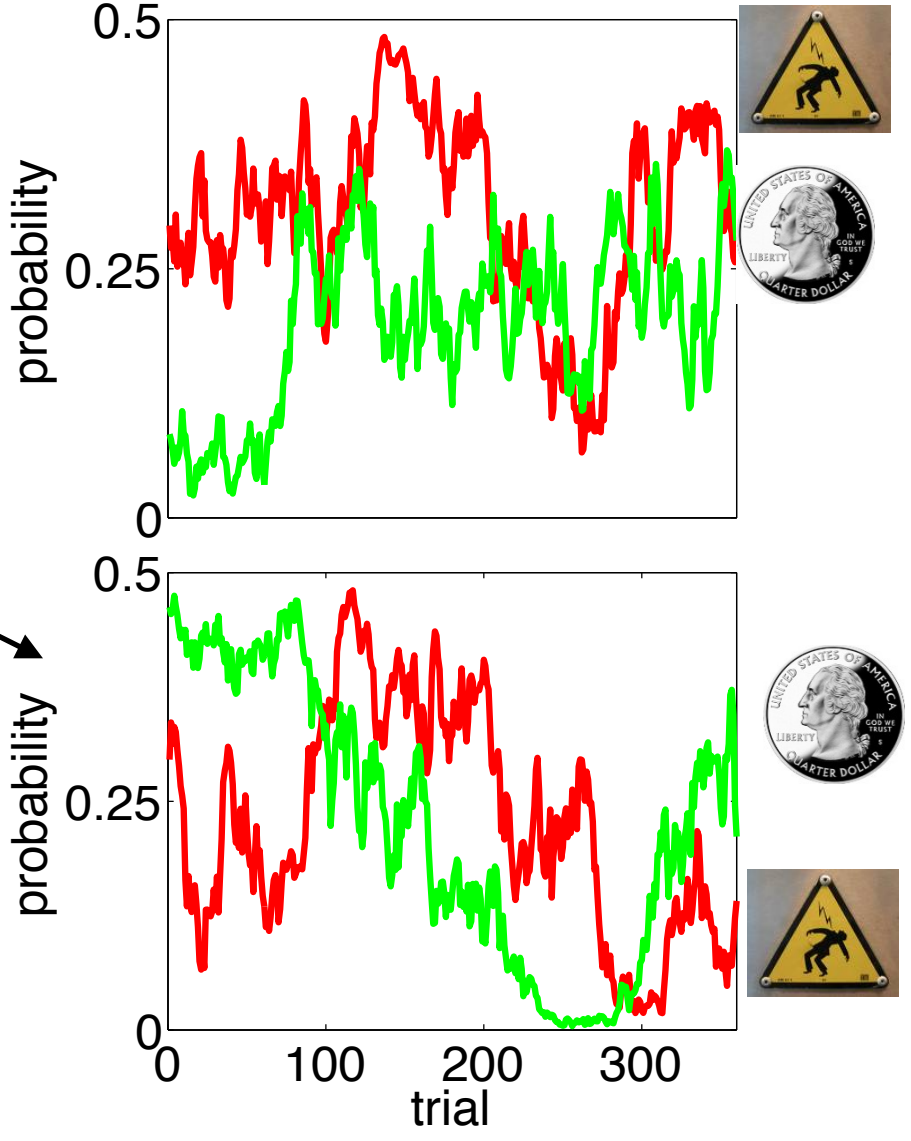




(Schultz et al 1997)

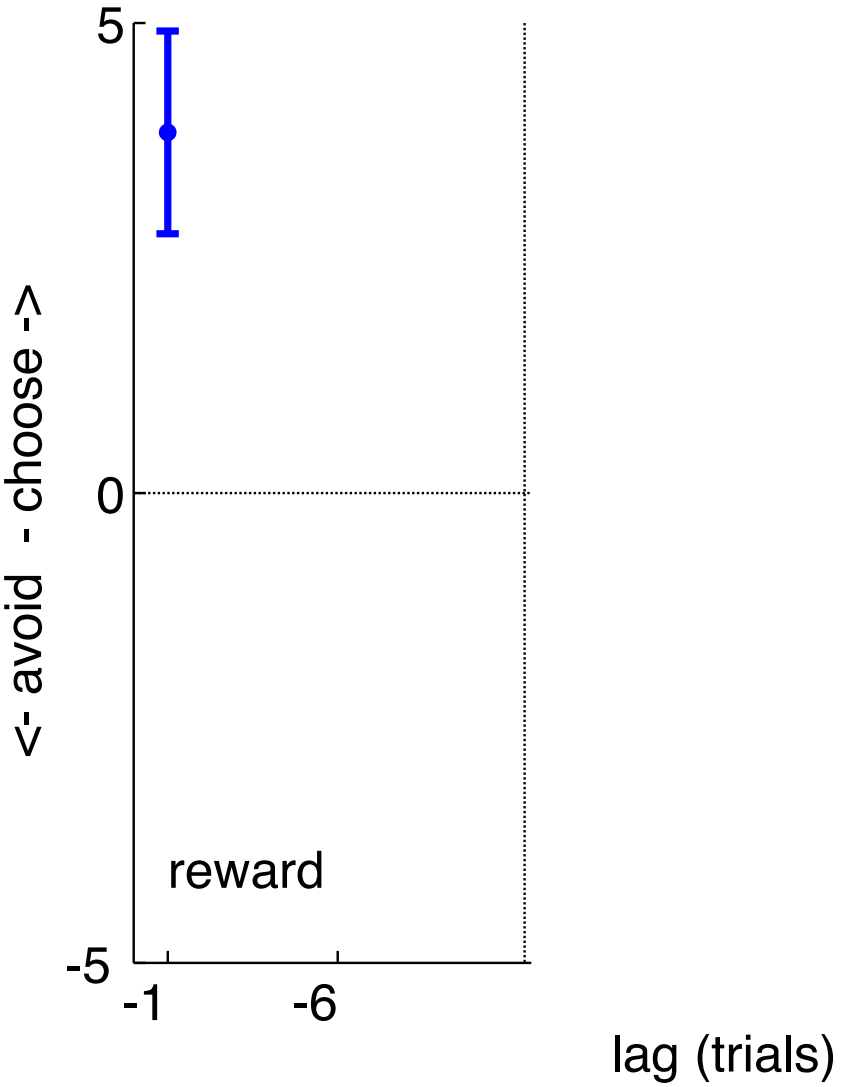
learned decision making in humans





"bandit" tasks

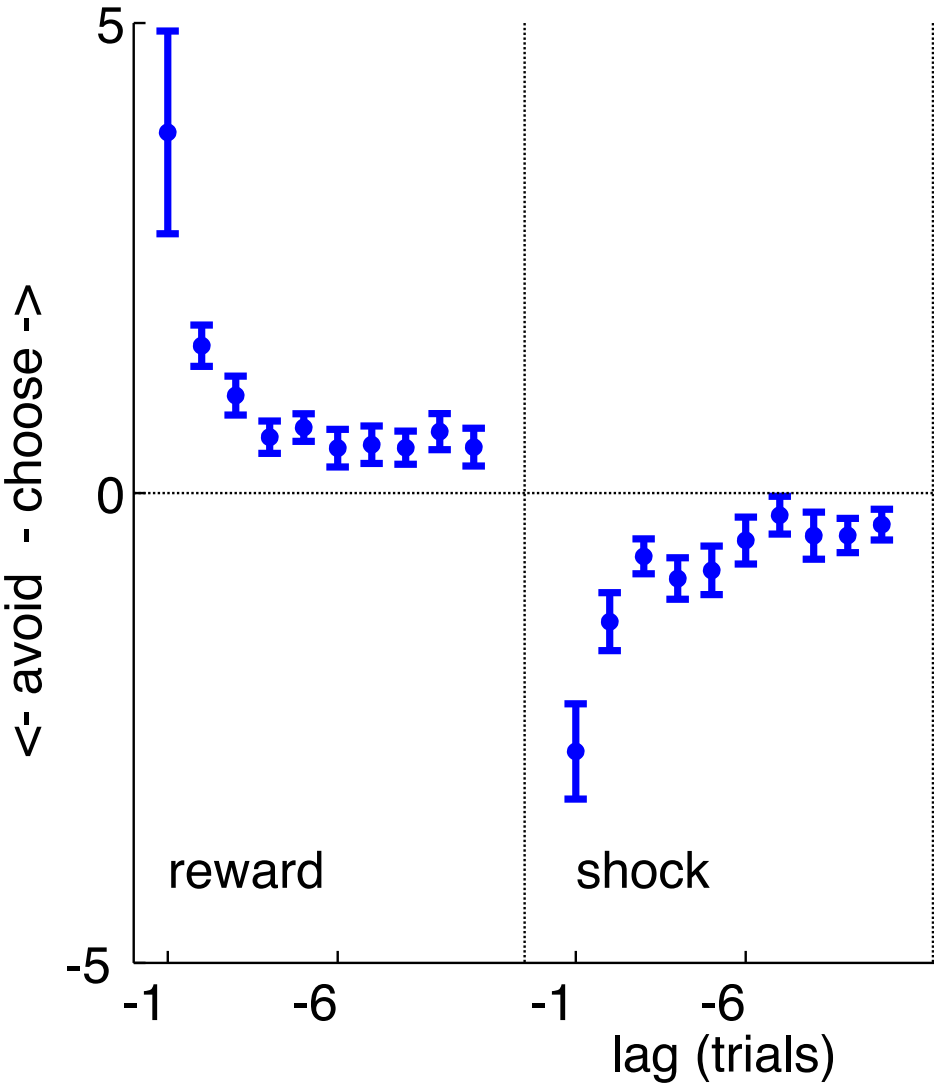
e.g. Daw et al 2006
Wittmann et al 2008
Gershman et al 2009
Schonberg et al 2010
Glascher et al 2010
Wimmer et al 2012
Seymour et al 2012
Kovach et al 2012
Madlon-Kay et al 2013



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

multinomial logistic regression: outcomes → choices

(Seymour et al. J Neuro 2012)



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

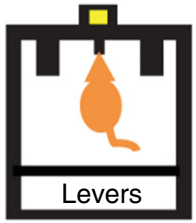
$P(choice_t = c) \propto \exp(\beta V_t(c))$

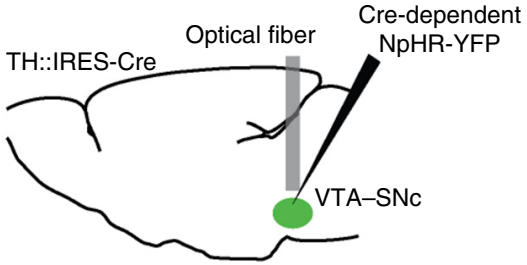
$V_{t+1}(choice_t) = V_t(choice_t) + \alpha\delta_t$

$\delta_t = reward_t - V_t(choice_t)$

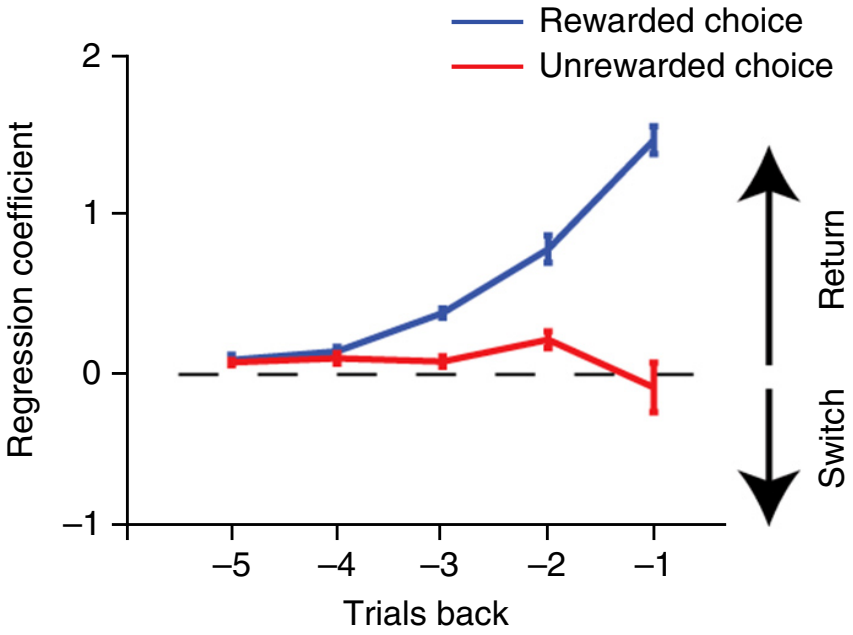
(Seymour et al. J Neuro 2012)

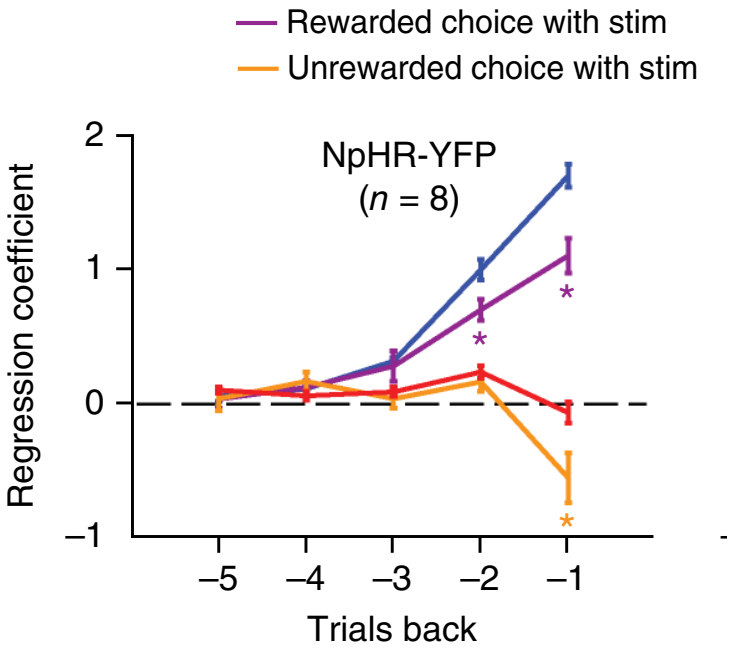
learned decision making in mice





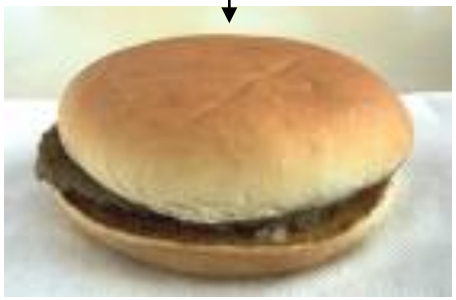
timed suppression of dopamine neurons on 10% of trials





(Parker, …, Daw & Witten, Nature Neuroscience 2016)


Nobu



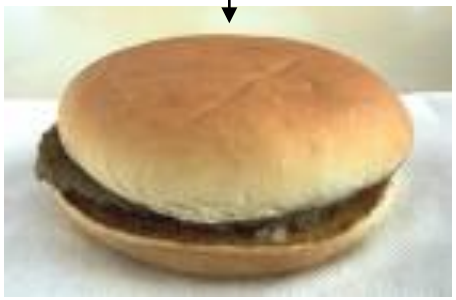
The New York Times

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 part
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.

?

<



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 part
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.

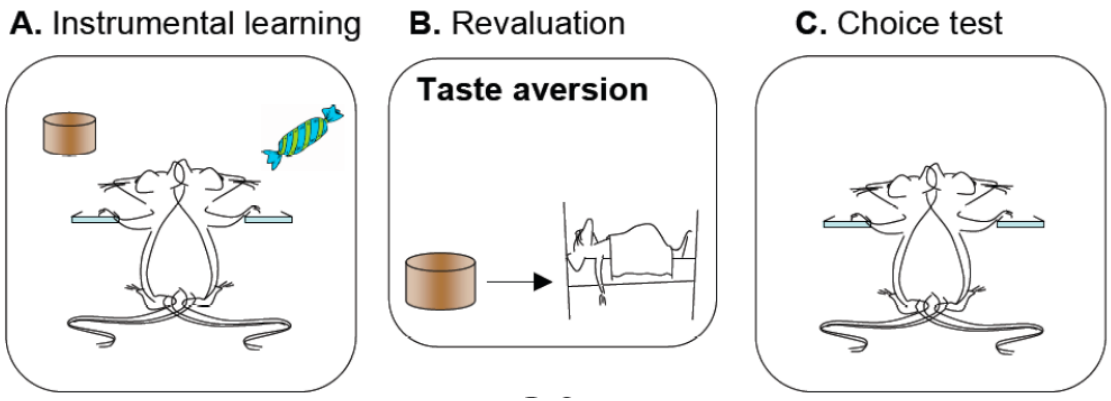
$$E[V(a)] = \Sigma_o P(o|a)\ V(o)$$

"model-free"

"model-based"

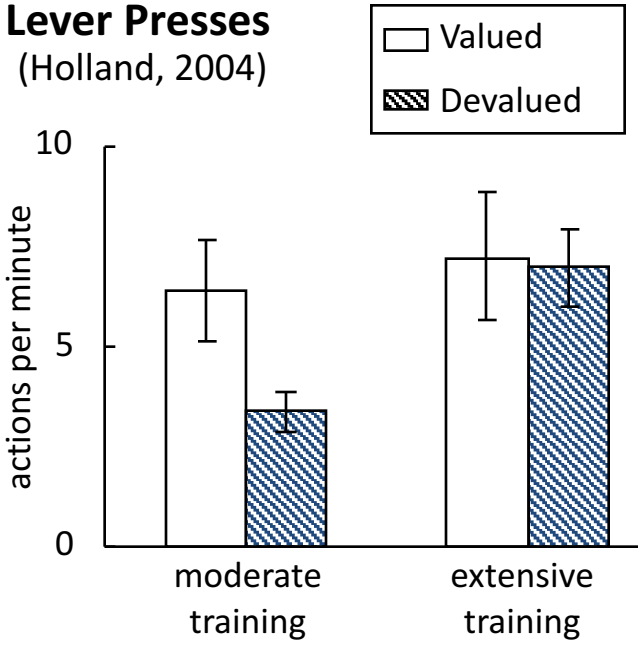
(Daw et al. Nat Neuro 2005)

rat version



(Dickinson, Balleine)

Lever Presses
(Holland, 2004)



two behavioral modes:

devaluation-insensitive ("habitual")

devaluation-sensitive ("goal directed")

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

"model-free"

"model-based"

(Daw et al. 2005)

theory

why have multiple systems?

- computational efficiency vs statistical efficiency

when to favor each?

- itself a cost-benefit tradeoff (cf Keramati et al. 2011)
- e.g. not worth deliberating when highly practiced on stable task
- new predictions, e.g. task volatility will favor deliberation
- rational cost-benefit approach to self-control, compulsion (Kurzban et al. 2012; Shenhav et al. 2013; Boureau et al. 2015)

how does the model-based system work?

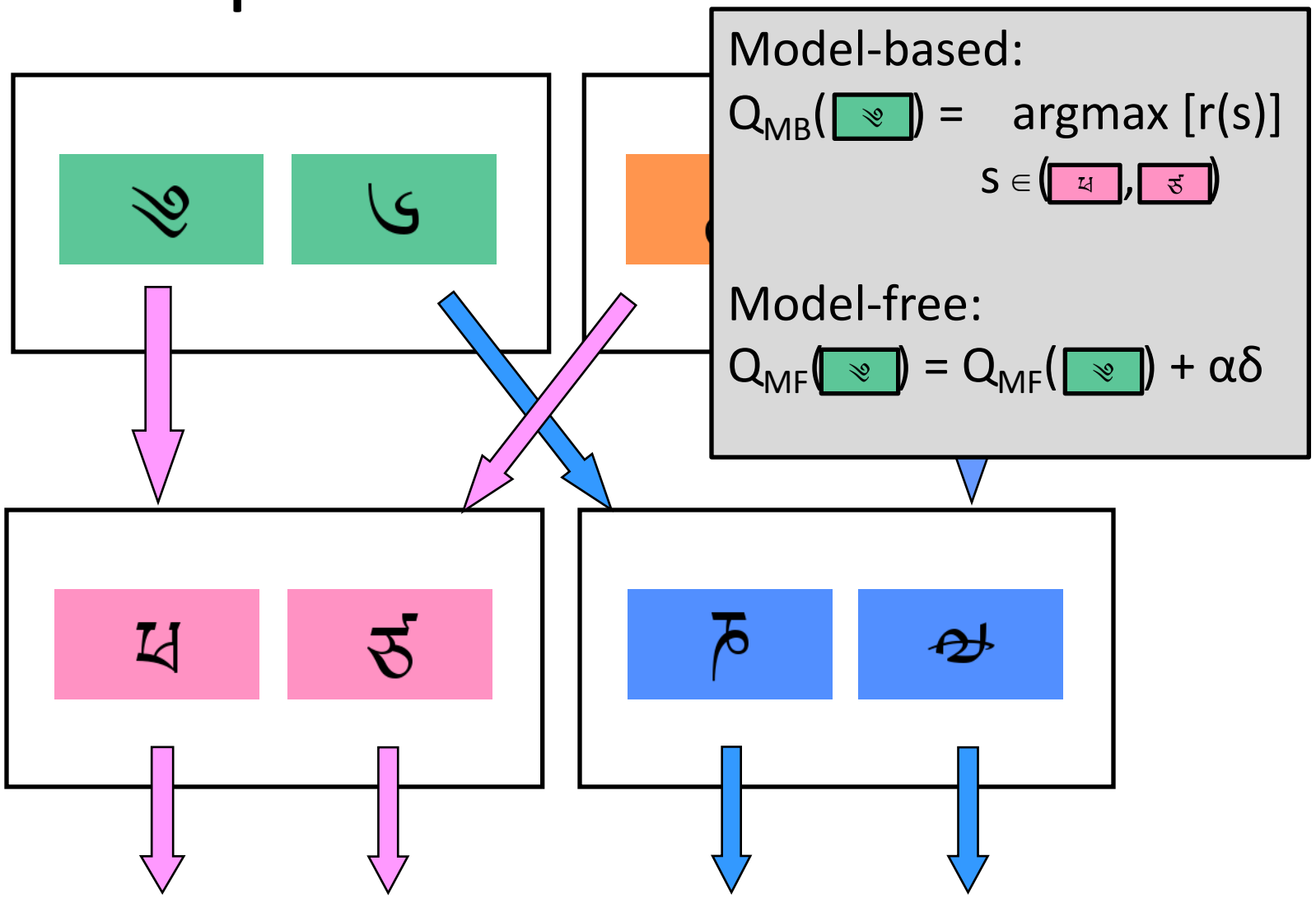
- much less known

(Daw, Niv, and Dayan, Nat Neuro 2005)

issues

1) Can we study this tradeoff with more targeted experimental designs?
2) What is the mechanism of model-based evaluation?
3) Is there a relationship between compulsion and habits?

sequential decision task



(*all slowly changing*)

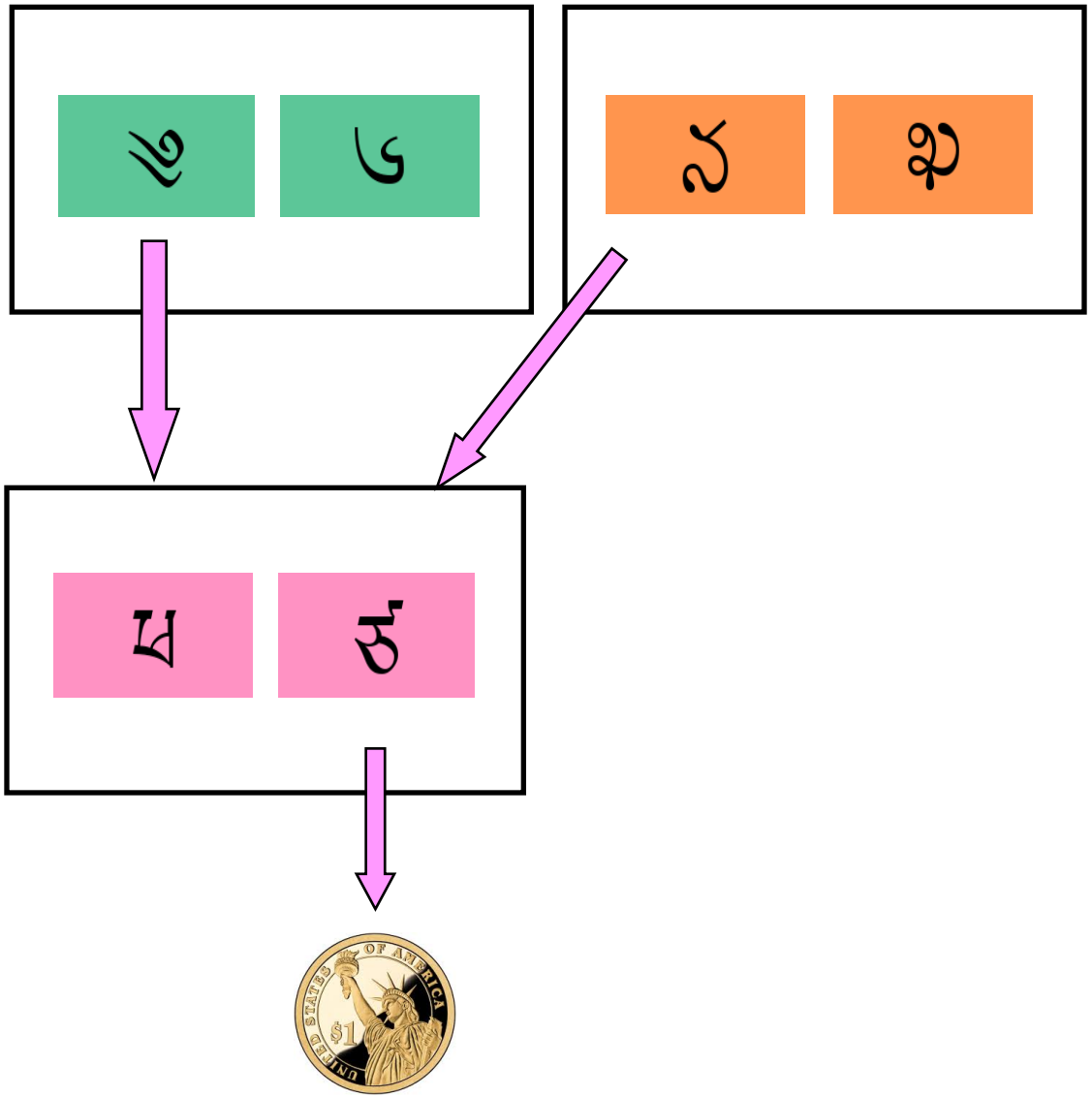
(Doll, Duncan, Simon, Shohamy & Daw *Nature Neuroscience* 2015)

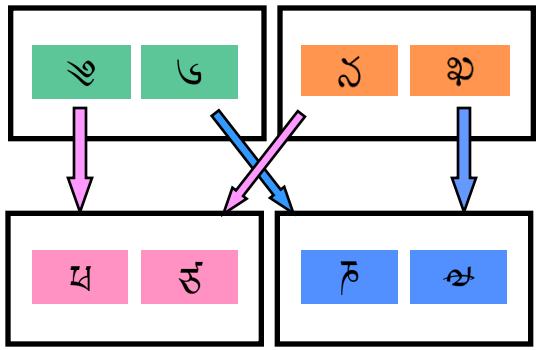
idea

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

Model-based: actions considered in terms of second-stage state
→ Feedback generalizes between equivalents

Model-free: actions reinforced by consequences
→ Feedback does not generalize

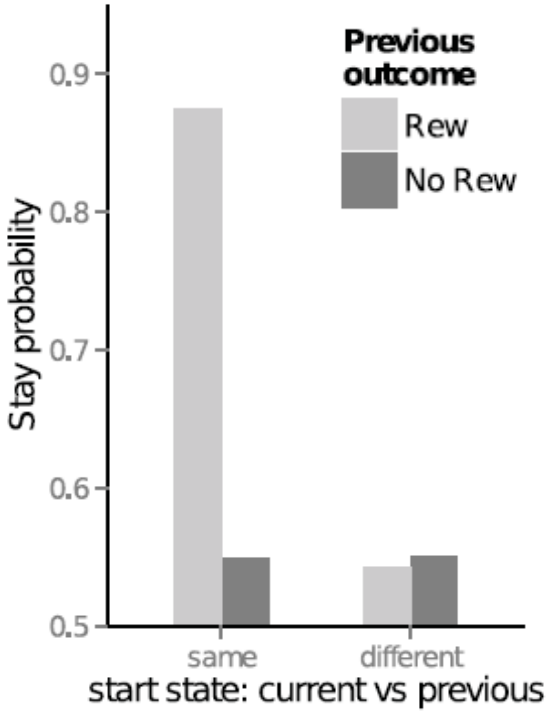




predictions

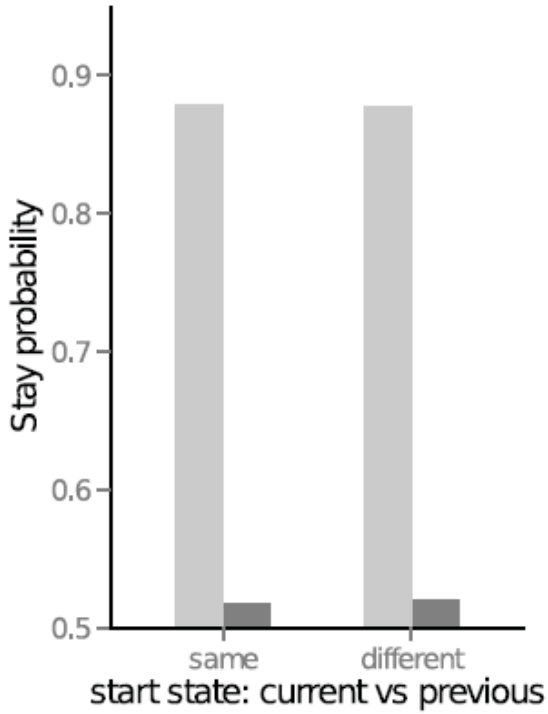
model-free

no generalization



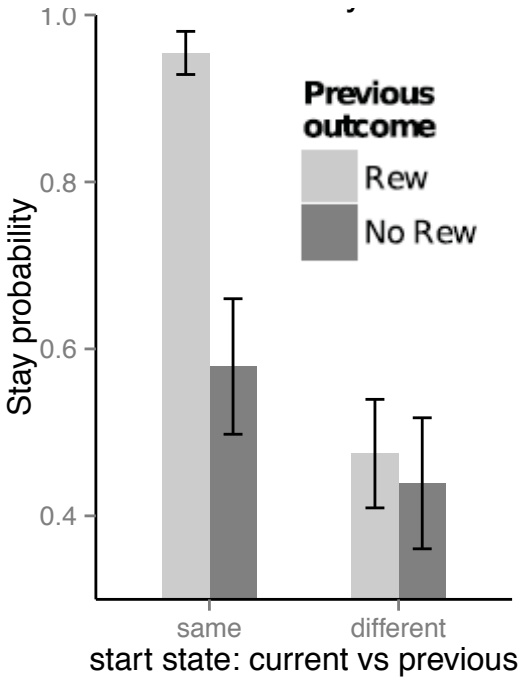
model-based

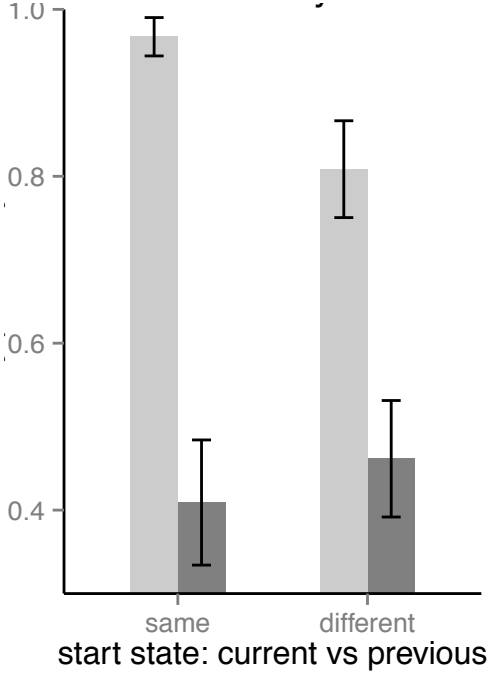
generalization

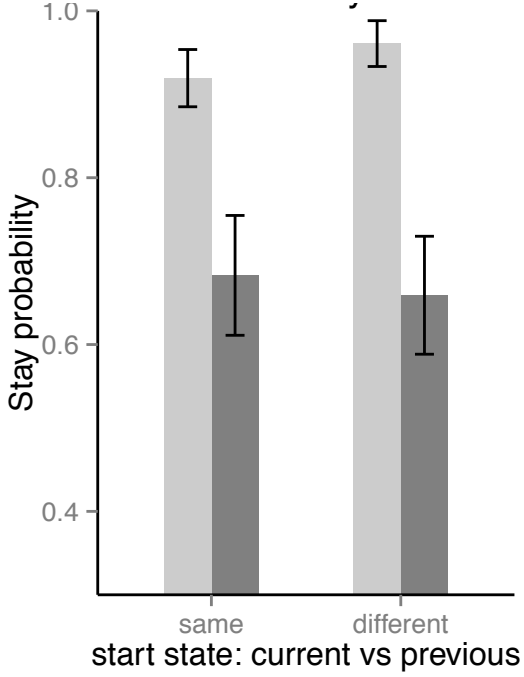


(Doll, Duncan, Simon, Shohamy & Daw *Nature Neuroscience* 2015)

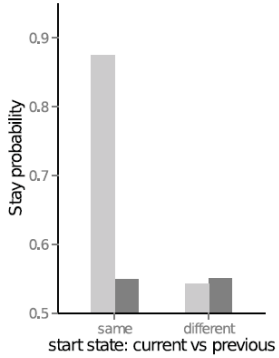
data



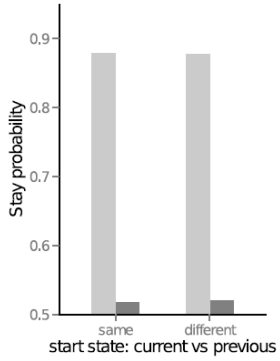




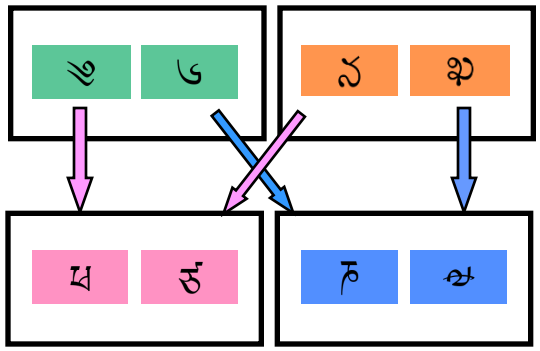
model-free



model-based

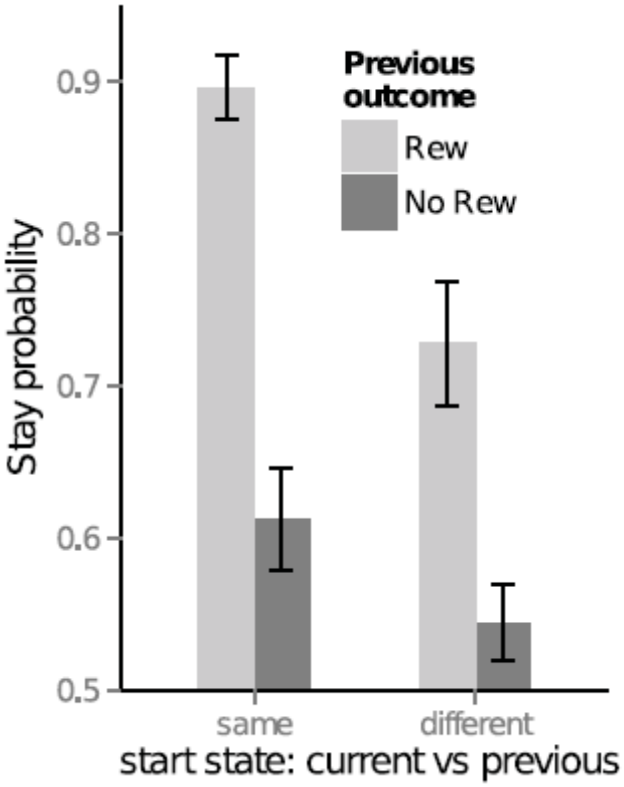


(Doll, Duncan, Simon, Shohamy & Daw *Nature Neuroscience* 2015)



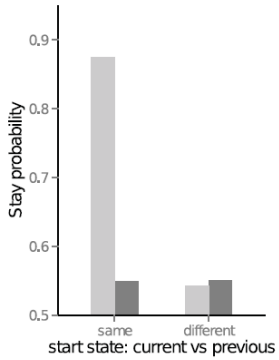
data

20 subs x 272 trials each

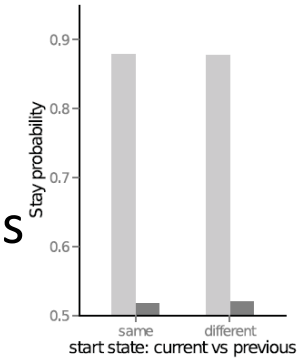


reward (MB): p<.0001
reward x same (MF) p<.005
(mixed effects logit)

model-free



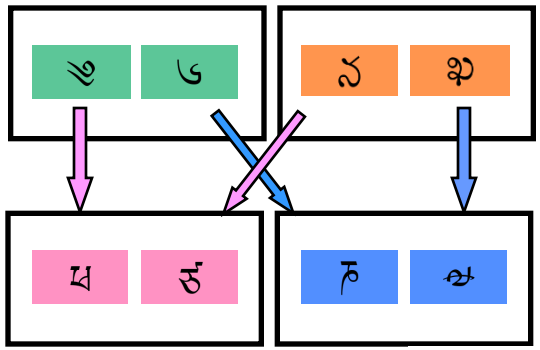
model-based



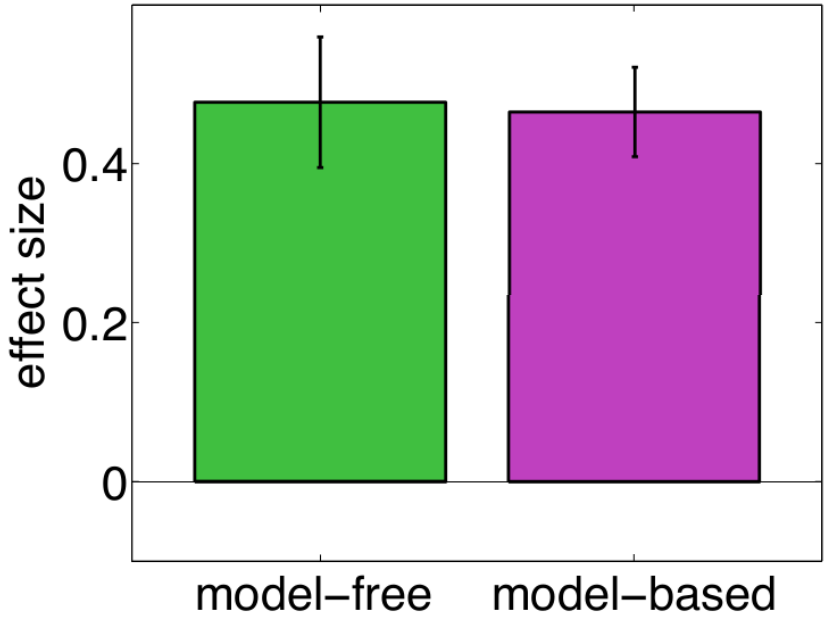
→ results reject pure reinforcement models
→ suggest mixture of planning and reinforcement processes

(Doll, Duncan, Simon, Shohamy & Daw *Nature Neuroscience* 2015)

data

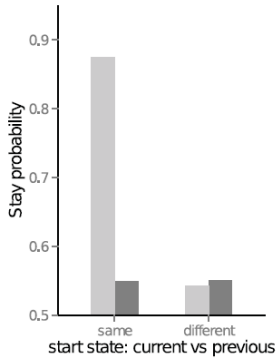


20 subs x 272 trials each

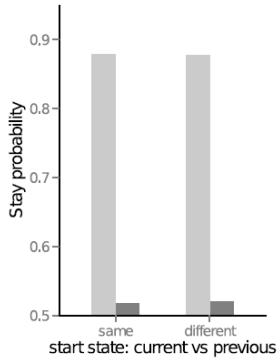


reward (MB): p<.0001
reward x same (MF) p<.005
(mixed effects logit)

model-free

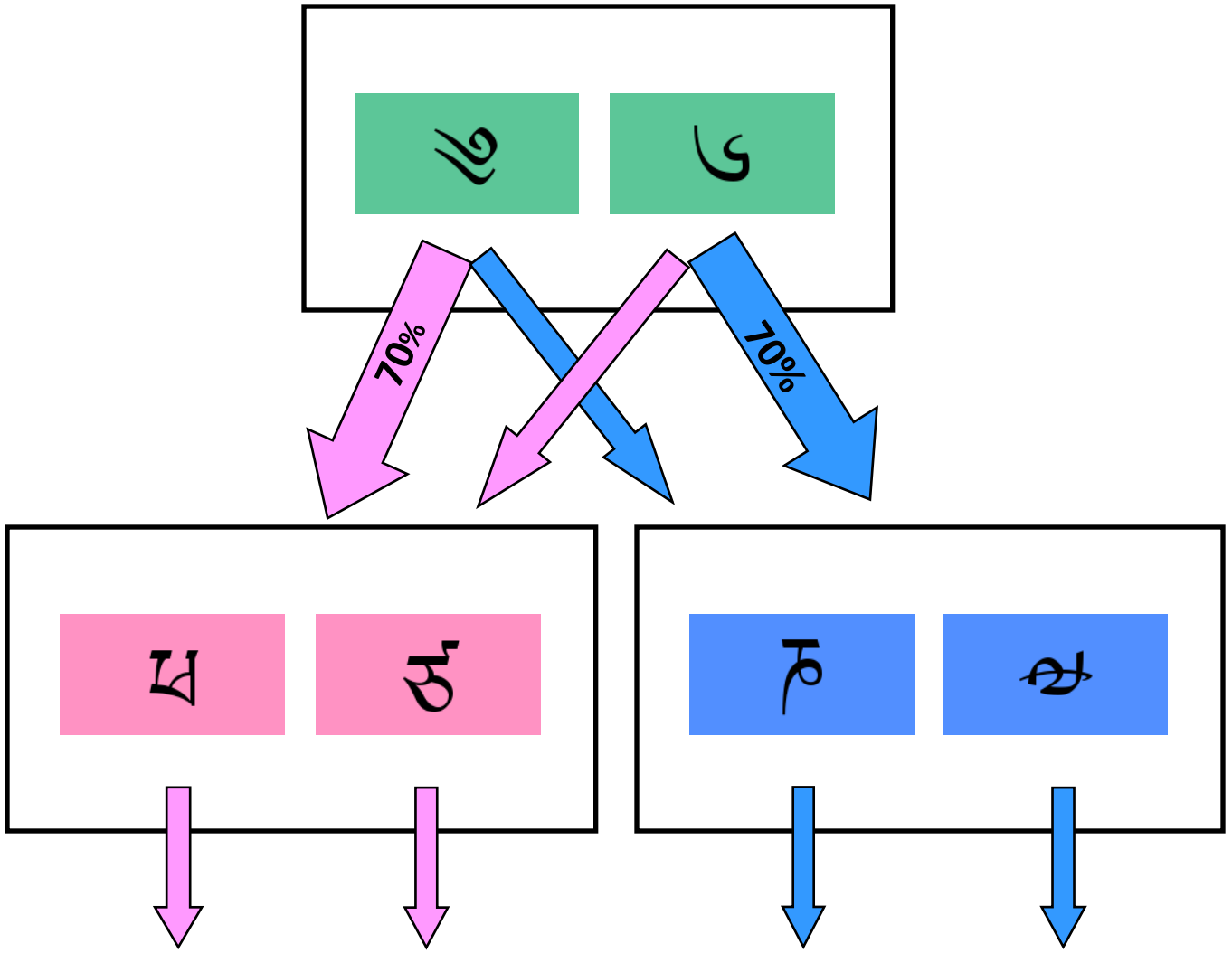


model-based



(Doll, Duncan, Simon, Shohamy & Daw *Nature Neuroscience* 2015)

variant sequential task

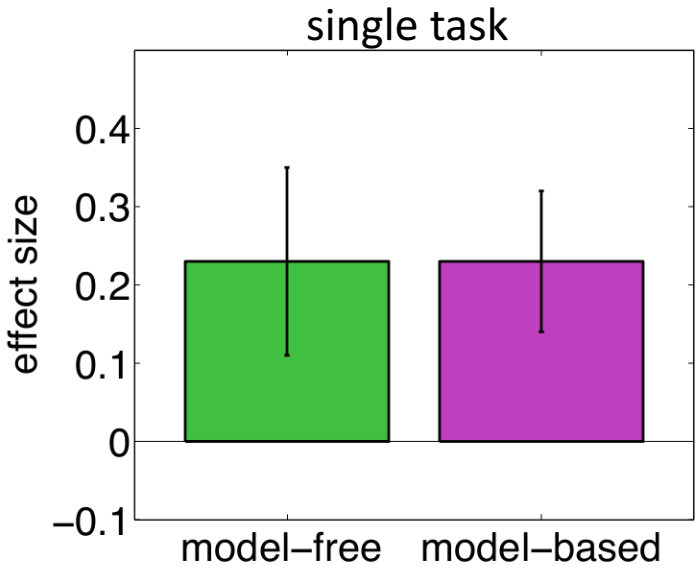


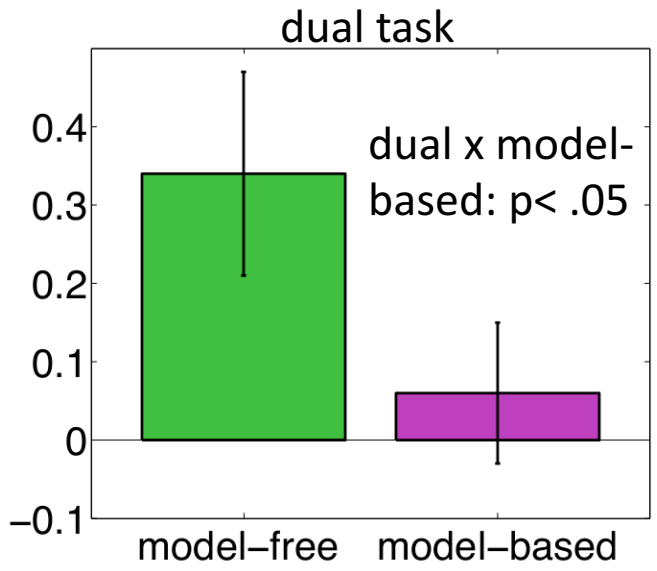
with prob: 26% 57% 41% 28%

(Daw, Gershman, Seymour, Dayan, and Dolan Neuron 2011)

What controls the tradeoff between these two sorts of learning?

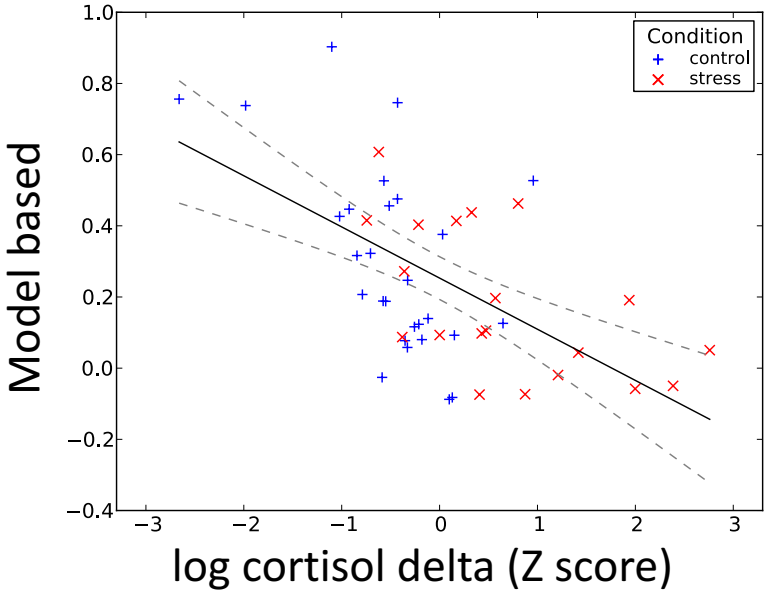
interference





(Otto et al Psych Science, 2013)

stress



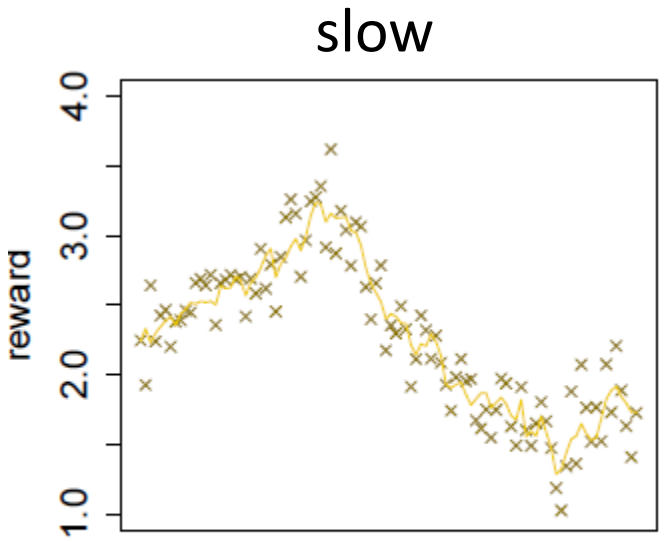
(Otto et al PNAS, 2013)

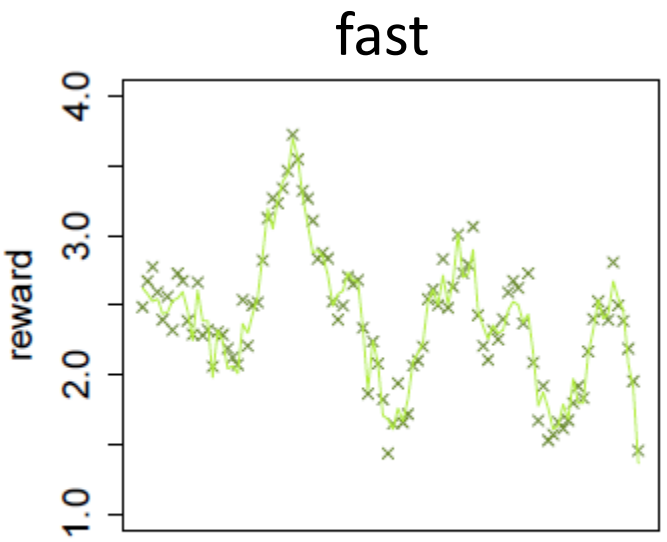
Also:

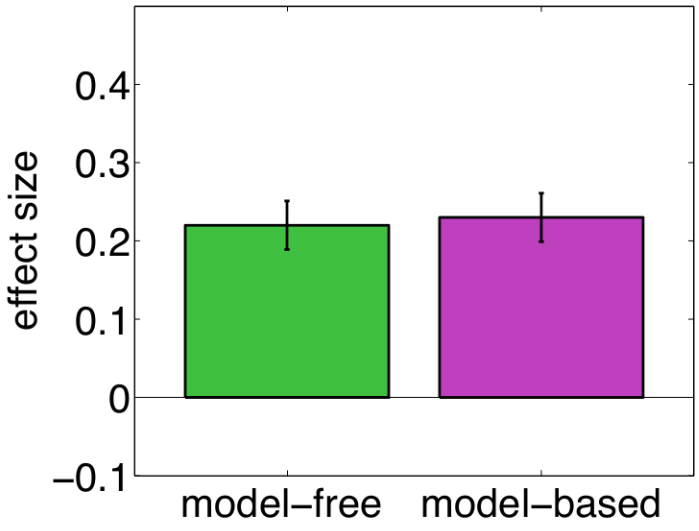
- Development & aging (Decker ea, in press; Eppinger ea 2013)
- IQ (Schad ea 2014; Gillan ea 2016)
- cognitive control (Otto ea 2015)
- PFC TMS (Smittenaar ea 2013)
- COMT (PFC DA) genotype (Doll ea 2016)
- Parkinson's disease & meds (Sharp ea 2016; Wunderlich ea 2012)
- dopamine PET (Desserno ea 2015)
- psychopathology (more later…)

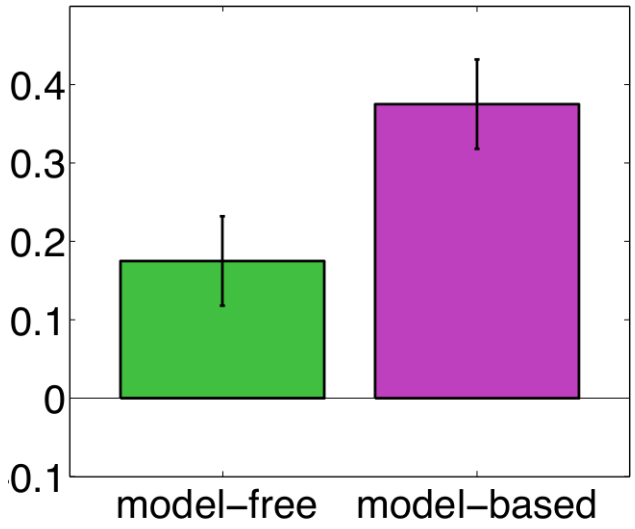
reward volatility

hypothesis (Daw et al. 2005): faster change requires more data-efficiency, promotes model-based deliberation









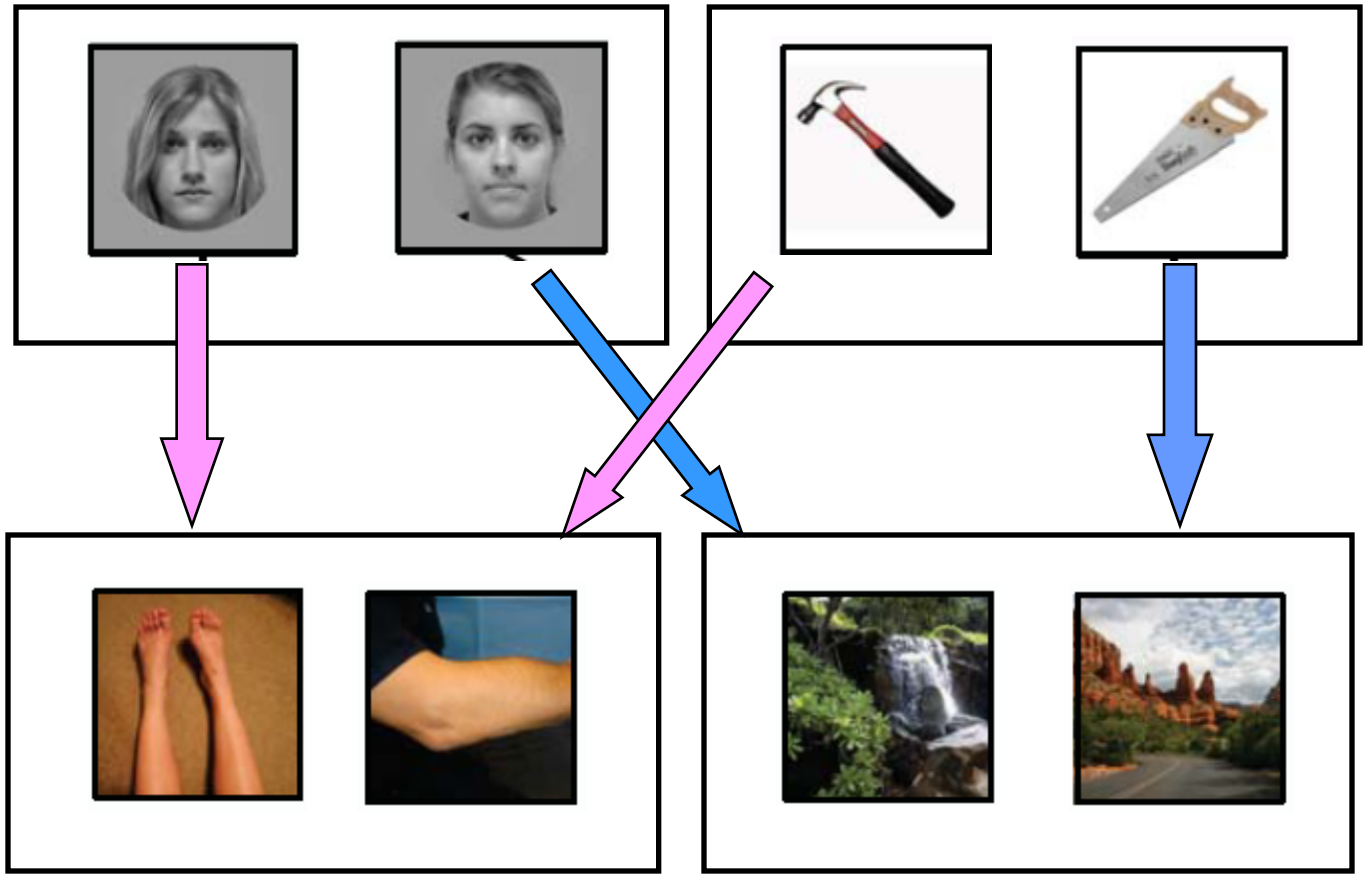
(Simon & Daw, NIPS 2011 & unpub.)

volatility x (MB – MF): $p < .05$

what are the neural mechanisms underlying this evaluation?

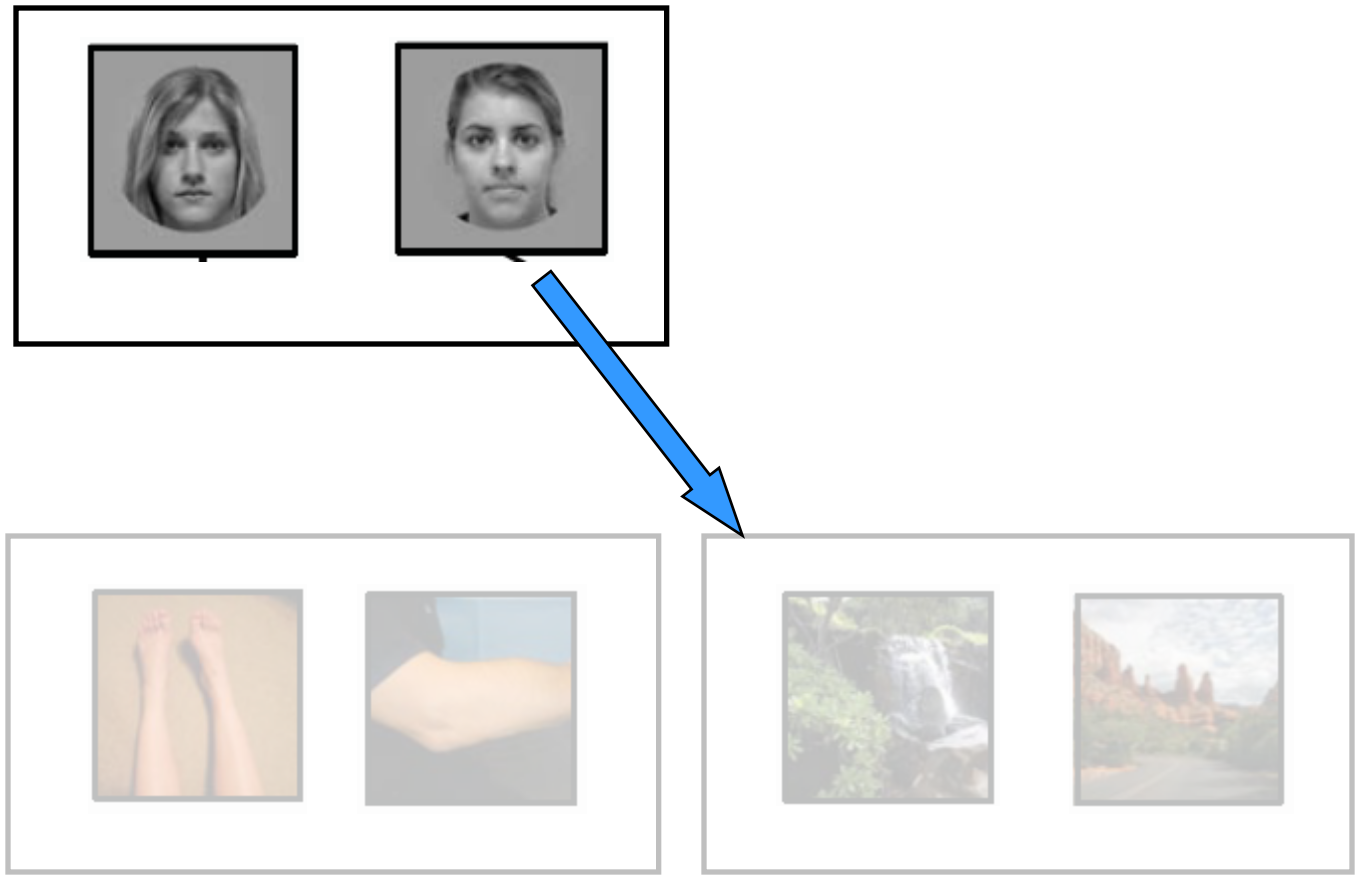
Is model-based learning really decision by simulation?

decodable stimuli



(Doll, Duncan, Simon, Shohamy & Daw *Nature Neuroscience* 2015)

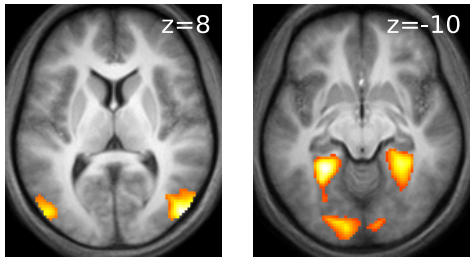
catch trials

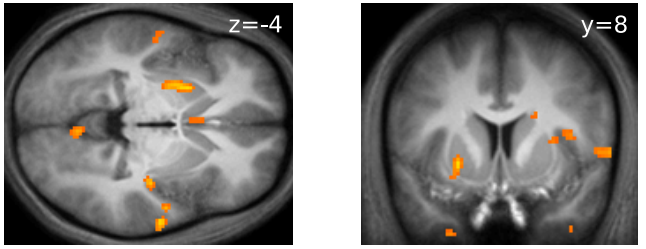


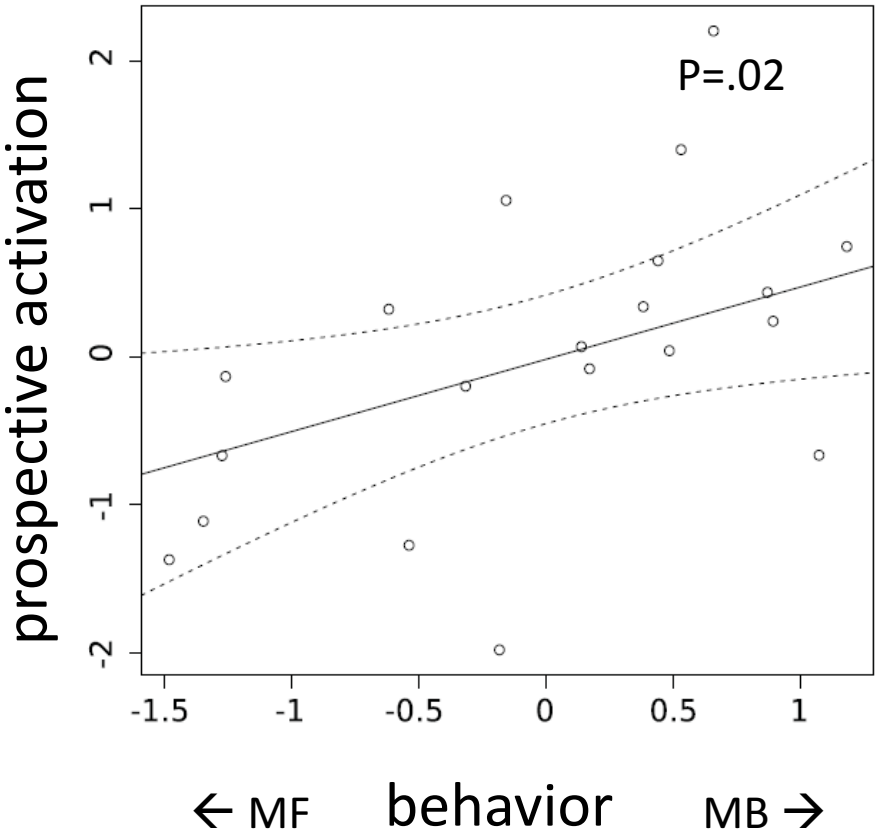
(Doll, Duncan, Simon, Shohamy & Daw *Nature Neuroscience* 2015)

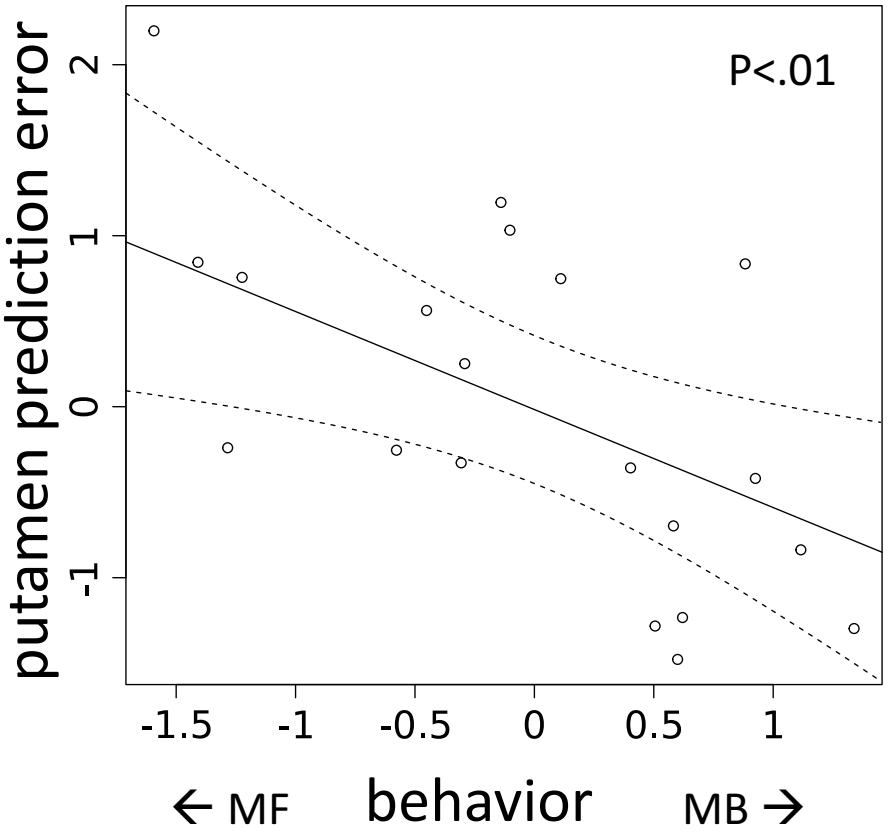
prospection (category selective ctx)

RPE (ventral putamen)









(Doll, Duncan, Simon, Shohamy & Daw *Nature Neuroscience* 2015)

Signatures of two dissociable neural evaluation mechanisms

1. forward search
2. error-driven updating

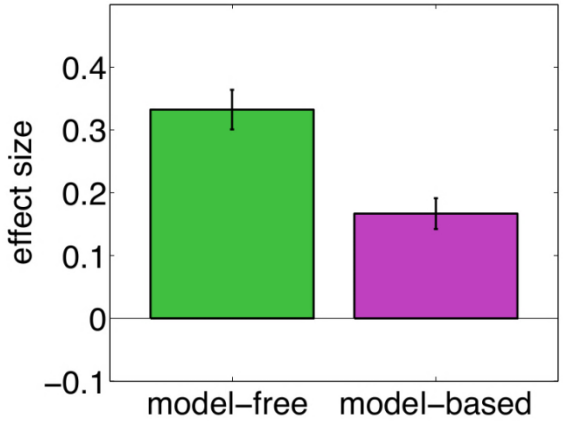
which have the expected relationships to choice behavior

→ is this really related to compulsion?

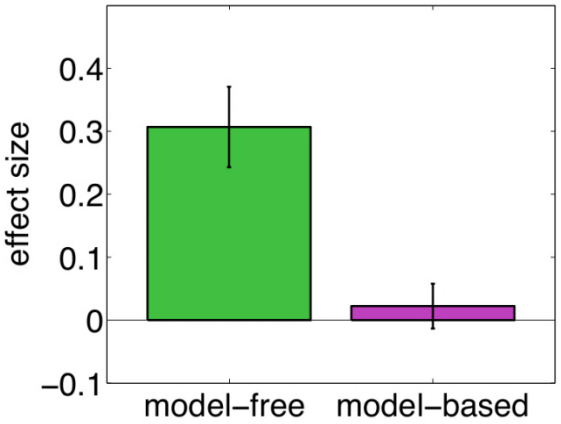
Is model-based learning related to disorders of compulsion?

Claire Gillan

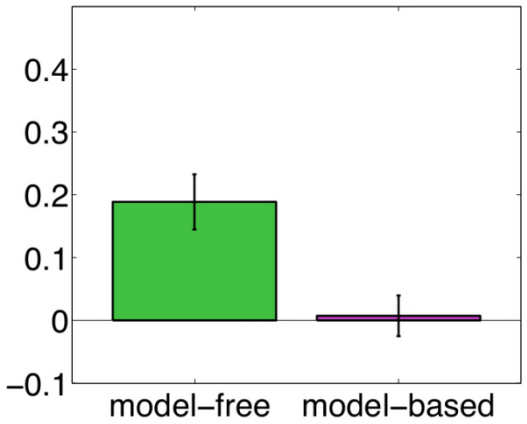
Healthy volunteers, n=106



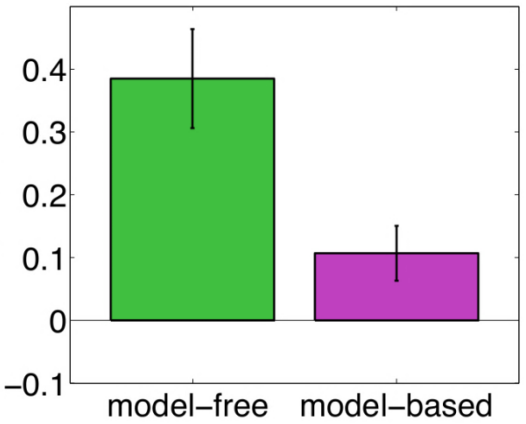
Binge eating disorder, n=30



Stimulant abusers, n=36



OCD, n=35



Methamphetamine/cocaine
Abstinent at least 1 wk

(Voon et al., Biological Psychiatry, 2014)

however...

OPEN ACCESS Freely available online

PLOS ONE

Impairments in Goal-Directed Actions Predict Treatment Response to Cognitive-Behavioral Therapy in Social Anxiety Disorder

Gail A. Alvares, Bernard W. Balleine, Adam J. Guastella*

Brain & Mind Research Institute, The University of Sydney, Sydney, New South Wales, Australia

Biological Psychiatry

Archival Report

Corticostriatal Control of Goal-Directed Action Is Impaired in Schizophrenia

Richard W. Morris, Stephanie Quail, Kristi R. Griffiths, Melissa J. Green, and Bernard W. Balleine

the crisis in psychiatry

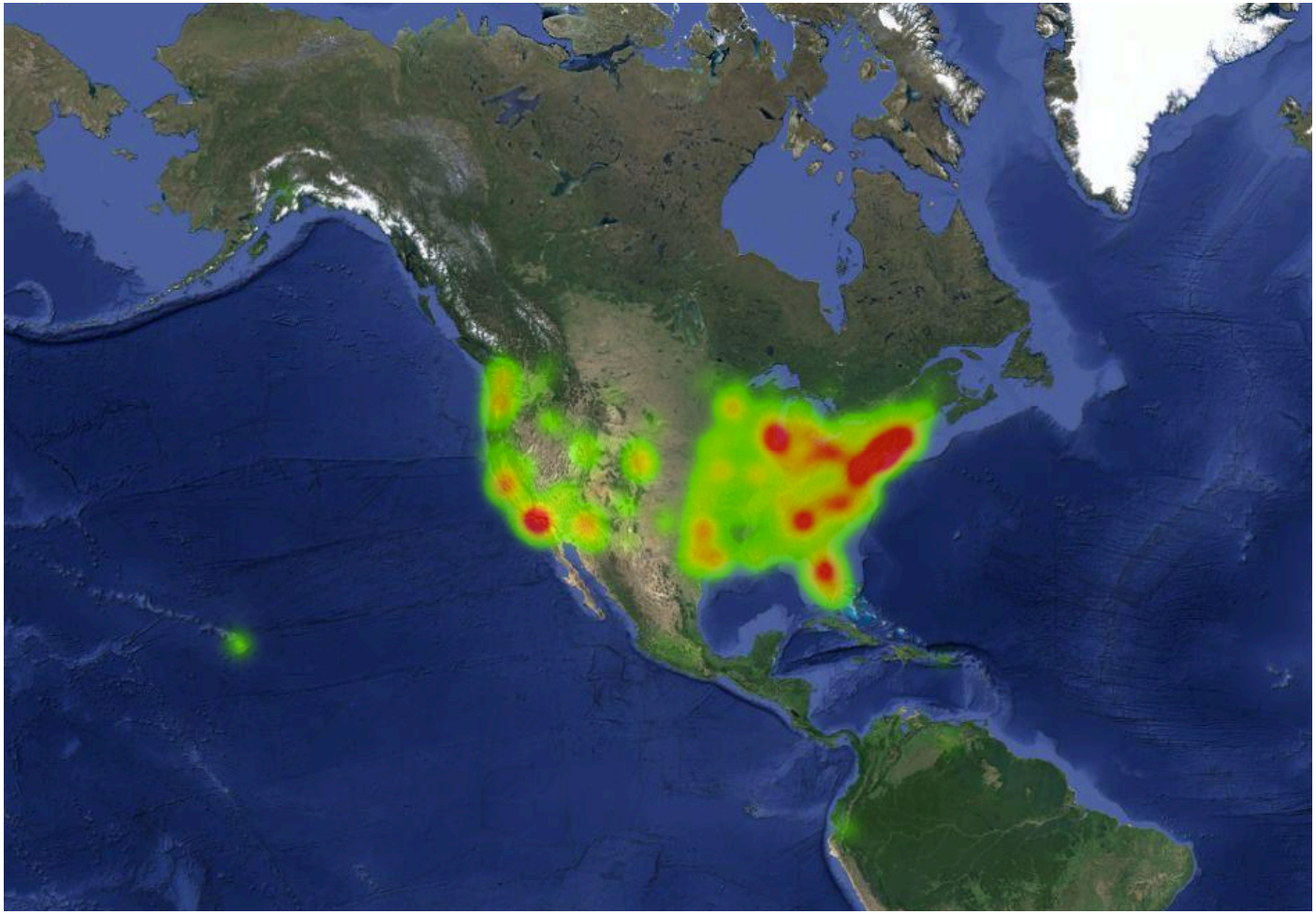
this may reflect a more general problem with psychiatric research – and psychiatric diagnoses

- co-morbidity, heterogeneity
- push toward dimensional, symptom-based view
- hope this will clarify etiology, neural basis

in a general population sample, look for evidence that this relationship is:

- graded/dimensional
- generalizes across diagnoses (“transdiagnostic”)
- yet is also specific to compulsive aspect

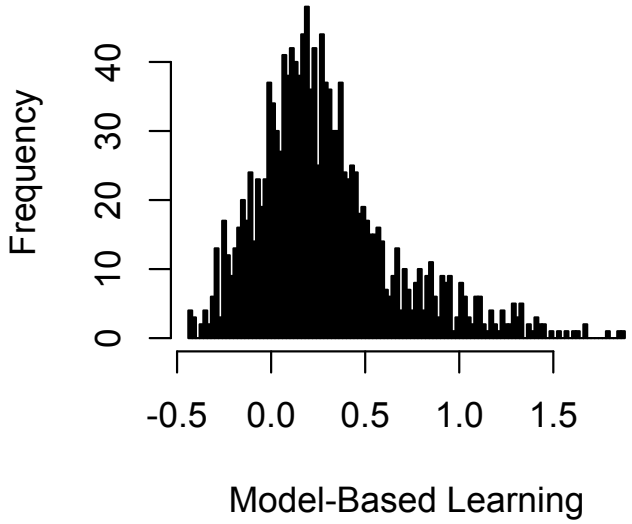
Large-scale online testing



Amazon Mechanical Turk

Experiment

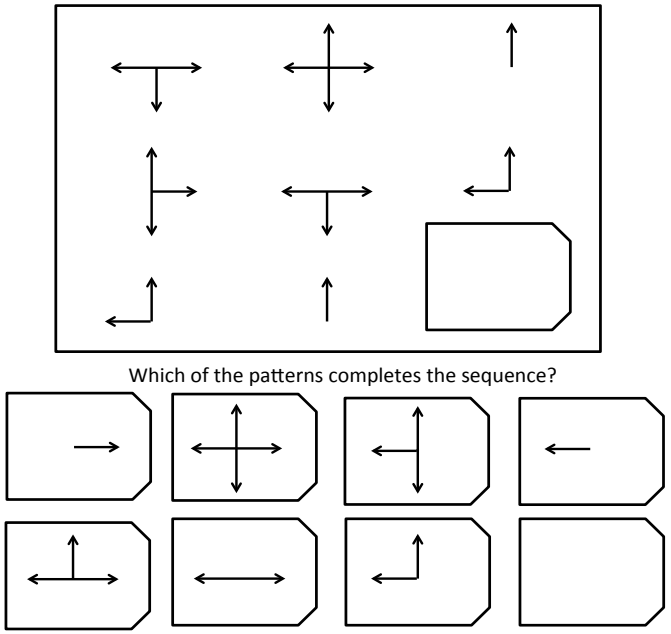
1. Model-based learning task



2. Self-Report Clinical Scales

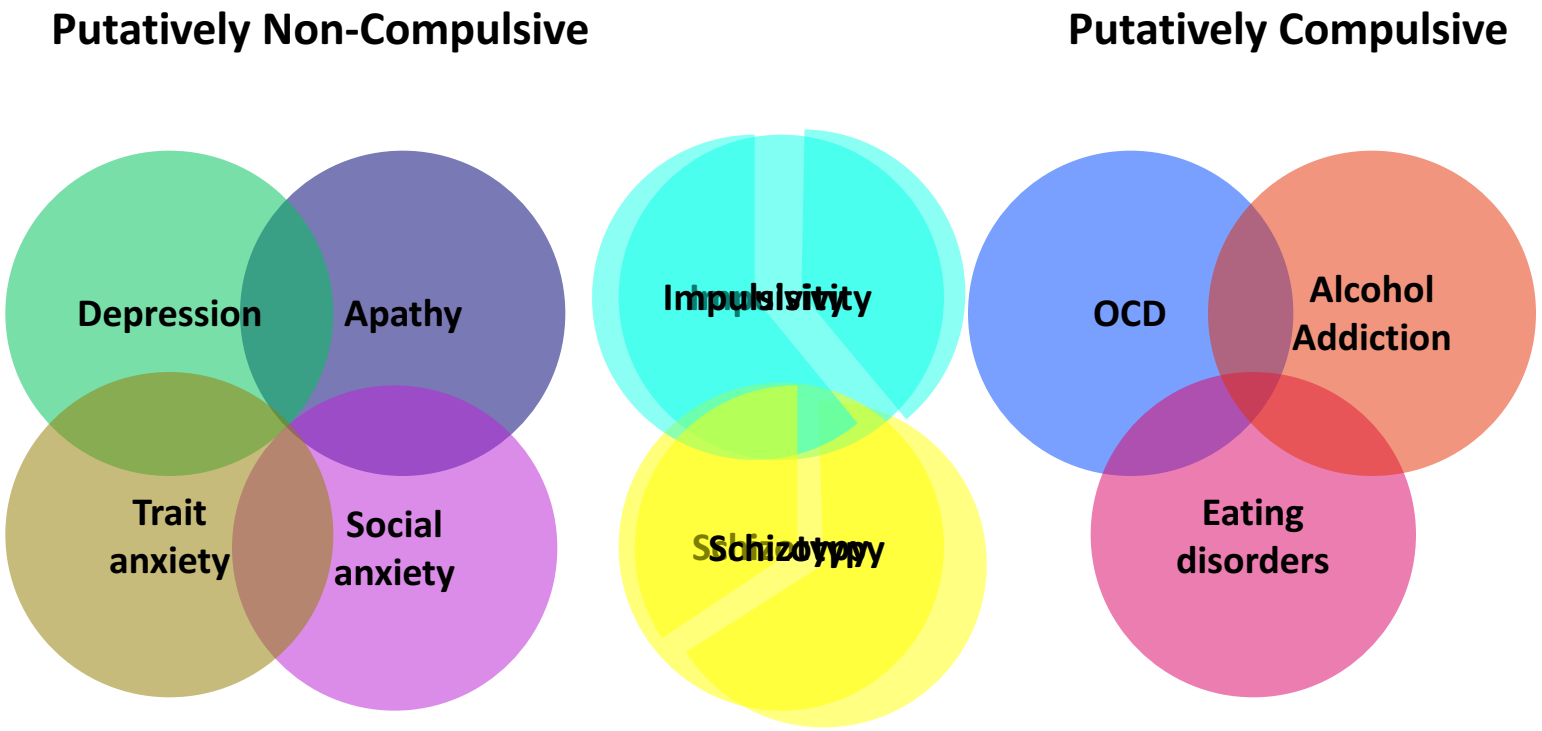
OCD: OCI-r (foa et al, 2002)
Depression: SDS (Zung, 1965)
Anxiety: STAI-trait (Spielberger, 1983)
…

3. IQ, age and gender



N=1413

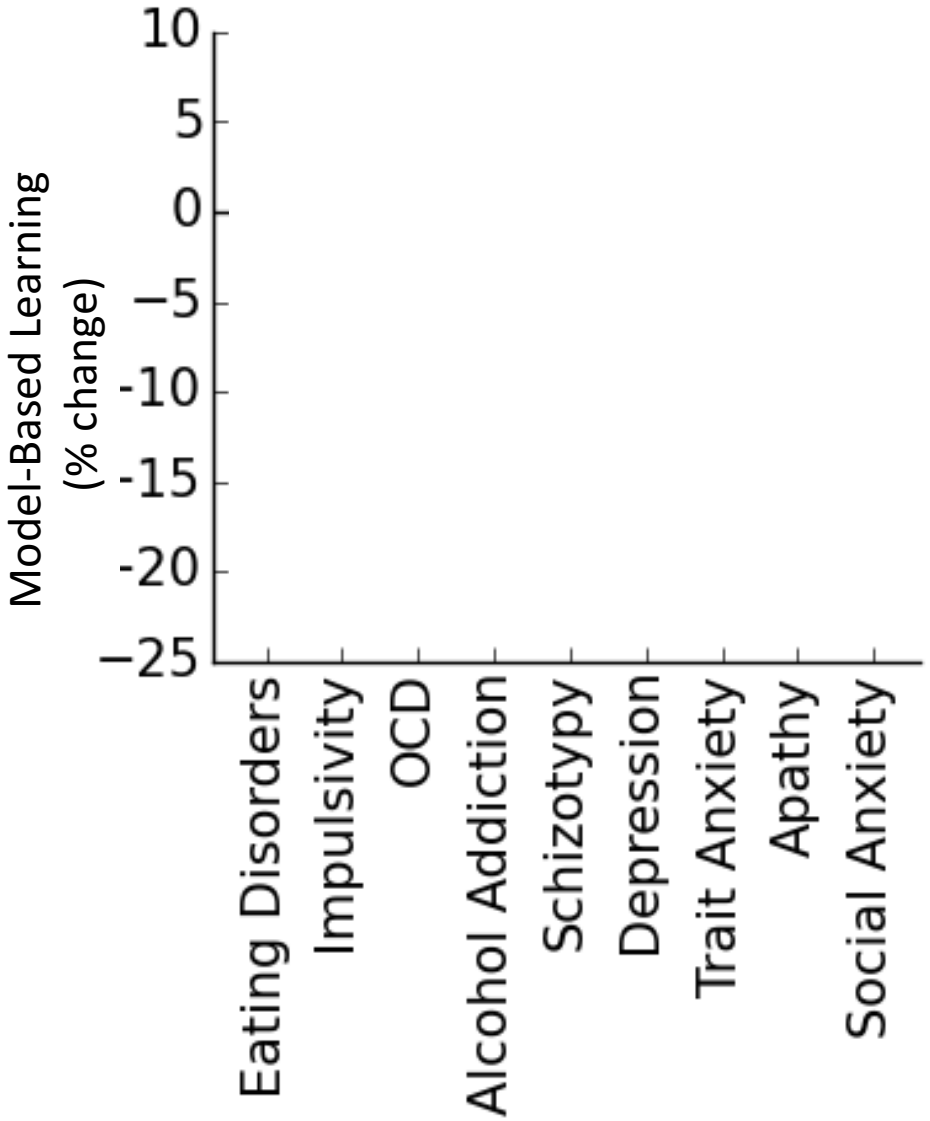
Measures



Gillan, Kosinski, Whelan, Phelps & Daw, under review

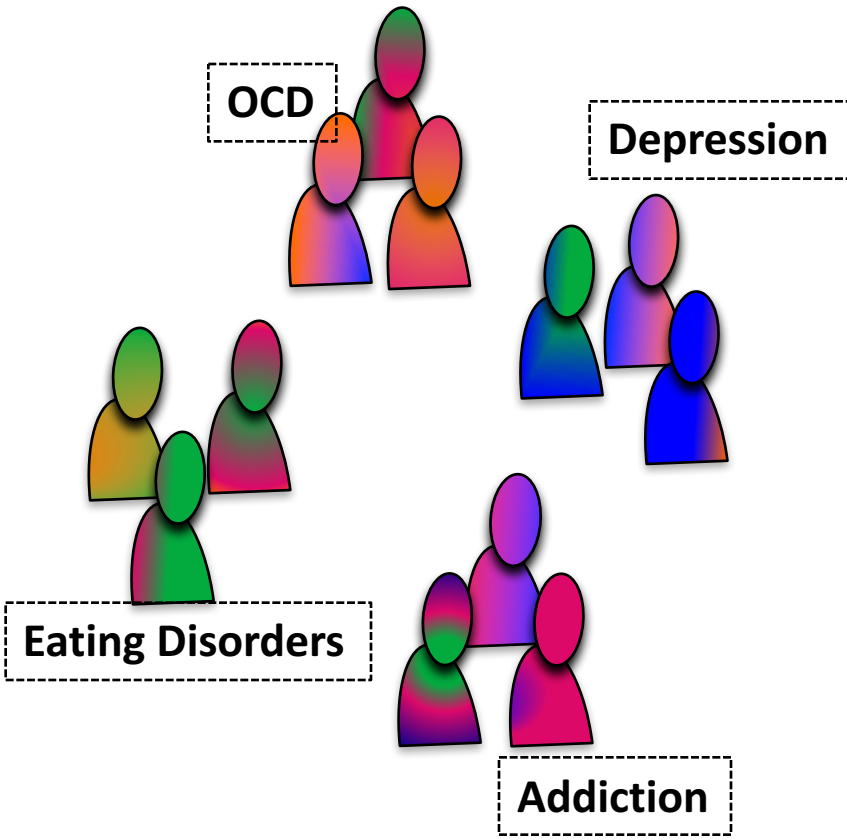
N=1413

Experiment 2



*p<.05 ** p<.01 ***p<.001

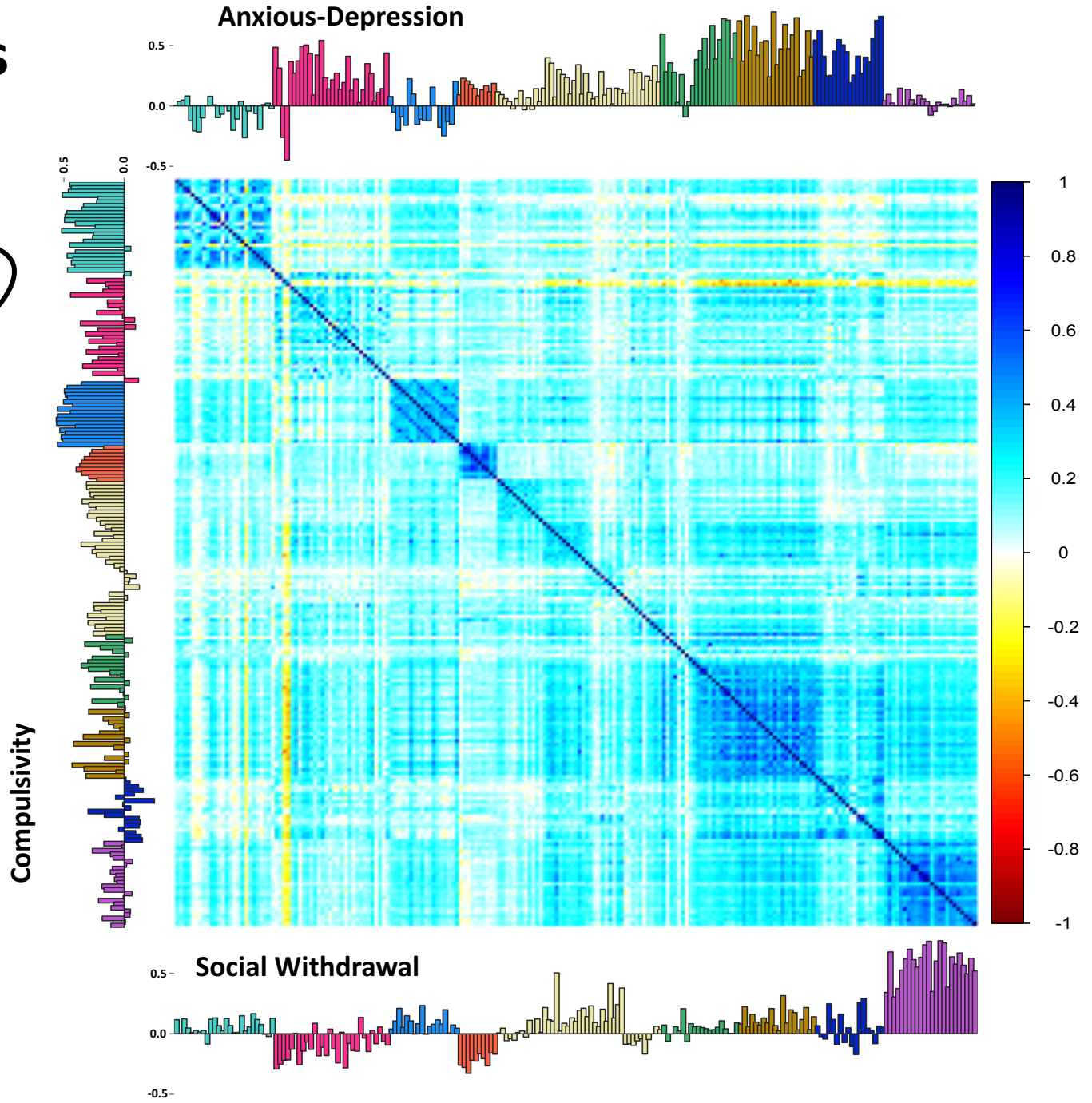
But… this just illustrates the categorization problem



Gillan, Kosinski, Whelan, Phelps & Daw, under review

Factor Analysis

Inter-correlation of 209 individual self-report items



Questionnaire

- Eating Disorders
- Impulsivity
- OCD
- Alcohol Misuse
- Schizotypy
- Depression
- Trait Anxiety
- Apathy
- Social Anxiety

Factor 2: Compulsivity

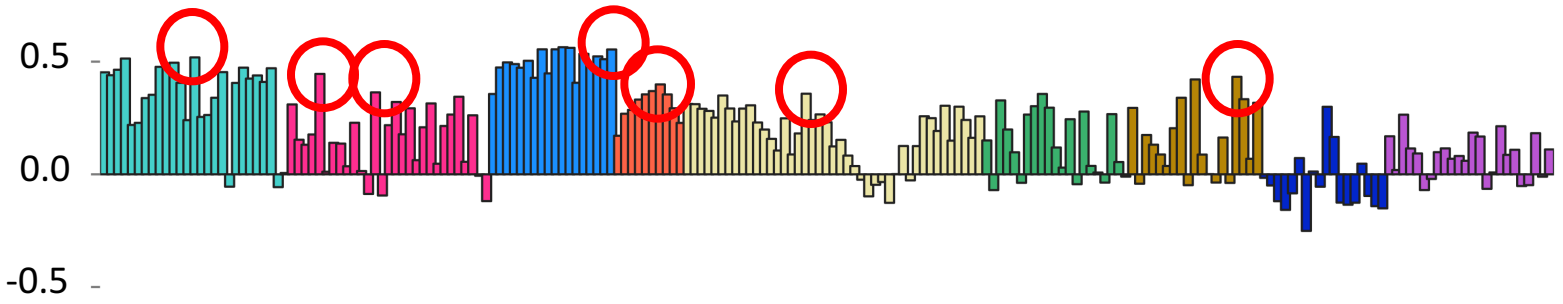
Questionnaire
- Eating Disorders
- Impulsivity
- OCD
- Alcohol Misuse
- Schizotypy
- Depression
- Trait Anxiety
- Apathy
- Social Anxiety

"I am preoccupied with the thought of having fat on my body"

"I have racing thoughts"

"I have disturbing thoughts"

Do you often have difficulty in controlling your thoughts?



"I repeatedly check doors, windows, drawers, etc."

"I have gone on eating binges where I feel that I may not be able to stop"

"I buy things on impulse"

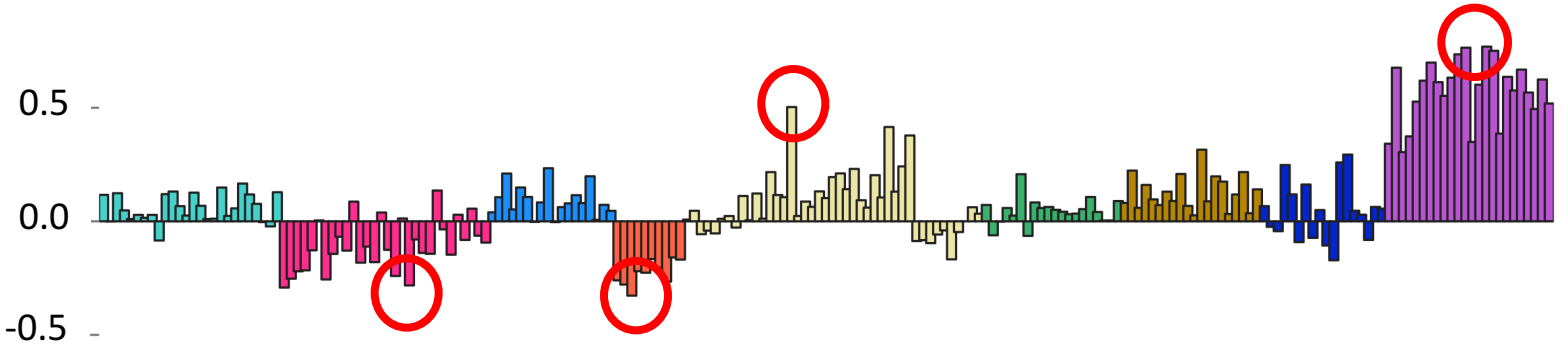
"How often … have you needed a first drink in the morning to get yourself going …?"

Factor 3: Social Withdrawal

Questionnaire
- Eating Disorders
- Impulsivity
- OCD
- Alcohol Misuse
- Schizotypy
- Depression
- Trait Anxiety
- Apathy
- Social Anxiety

"How often do you have 6 or more drinks on one occasion?"

"I do not plan tasks carefully"

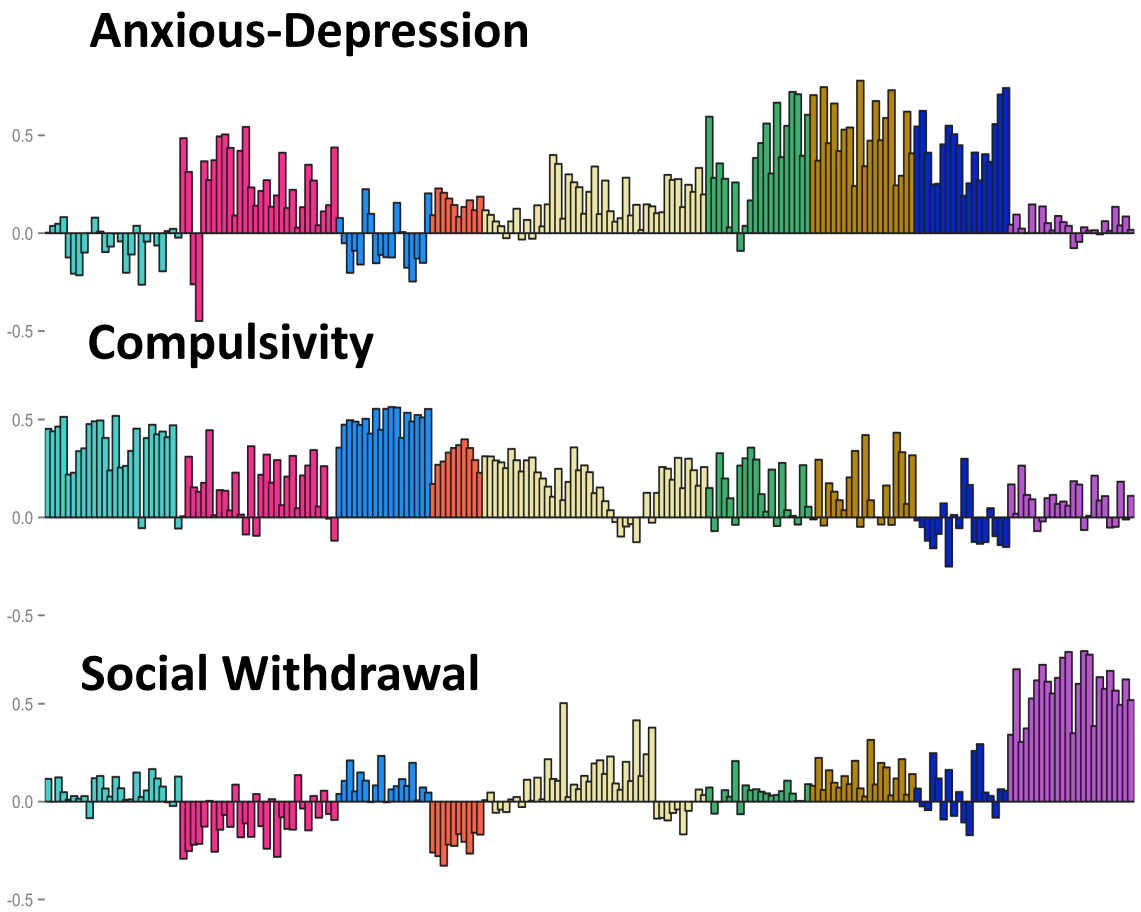


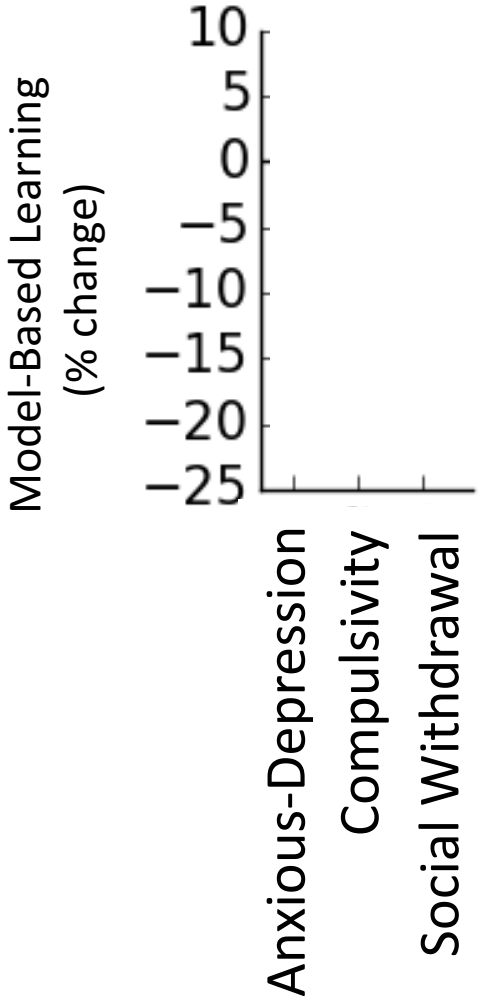
"Meeting strangers"

"Being the center of attention"

Do you dread going into a room when other people have gathered and are talking?

*p<.05 ** p<.01 ***p<.001





Gillan, Kosinski, Whelan, Phelps & Daw, under review

MB learning is selectively linked to compulsion, across diagnoses

- similar results from fully supervised, item-level analysis

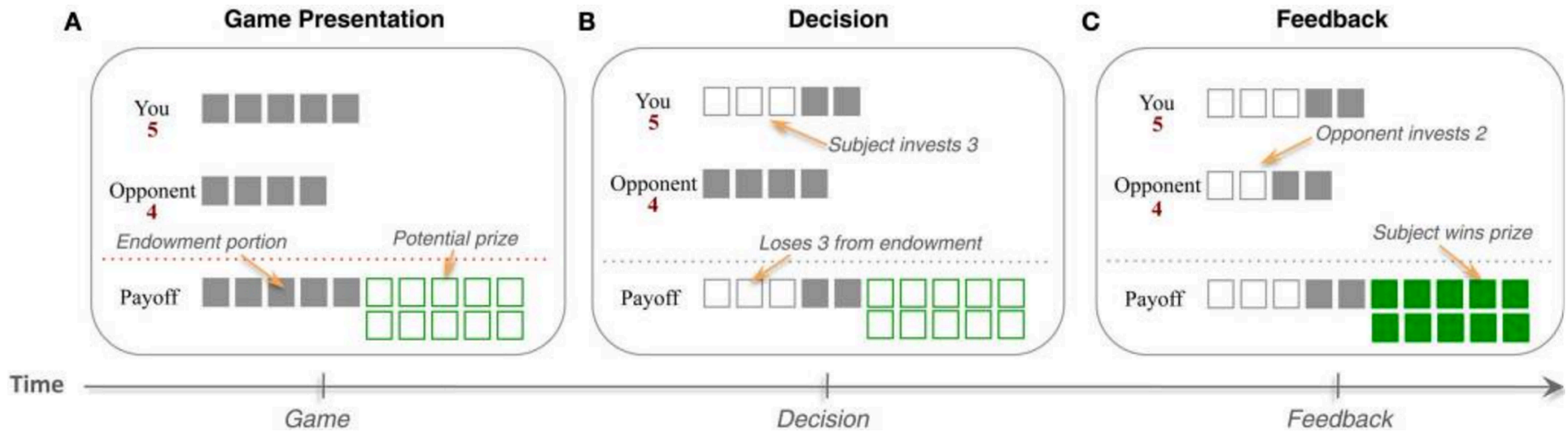
Compulsive thoughts and behaviors cluster in factor analysis

- relevant to obsessions vs. compulsions?

Of course these are just some symptom scales, and just one behavioral task

- progressively refine both sides
- promise of large-scale online testing more broadly

patent race game

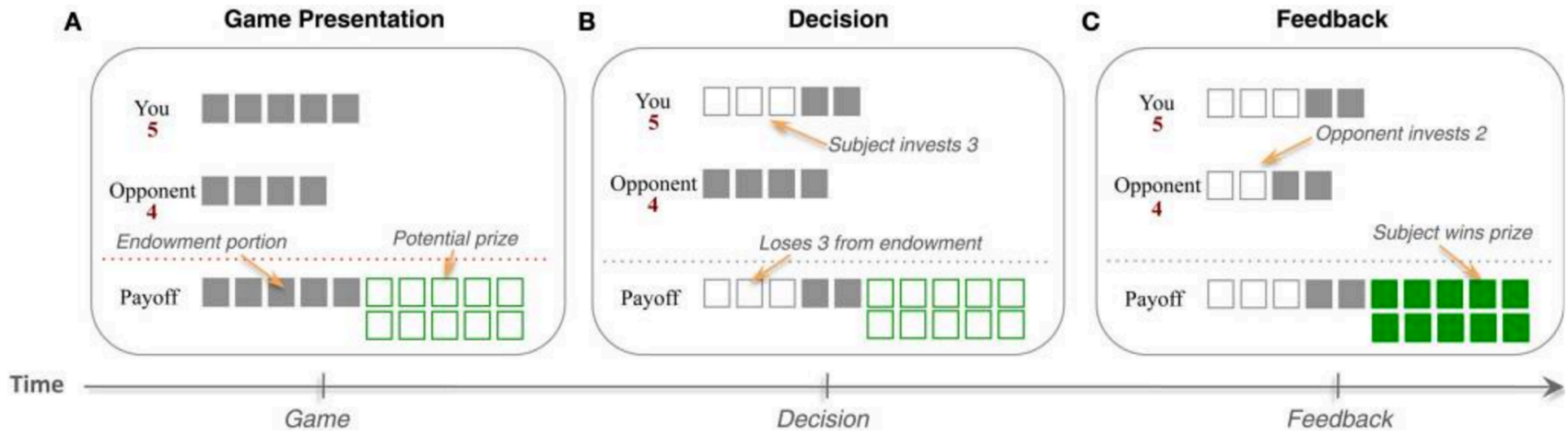


invest a portion of endowment, win prize if you invest more than opponent

- repeated play (80 trials) against replayed investments from previous subjects
- mixed strategy equilibrium
- learning (e.g. about opponents' move distribution, or which moves work)

Elana Meer; Lindsay Hunter; Ming Hsu

patent race game



theory, EWA (Camerer & Ho, 1999) nests:

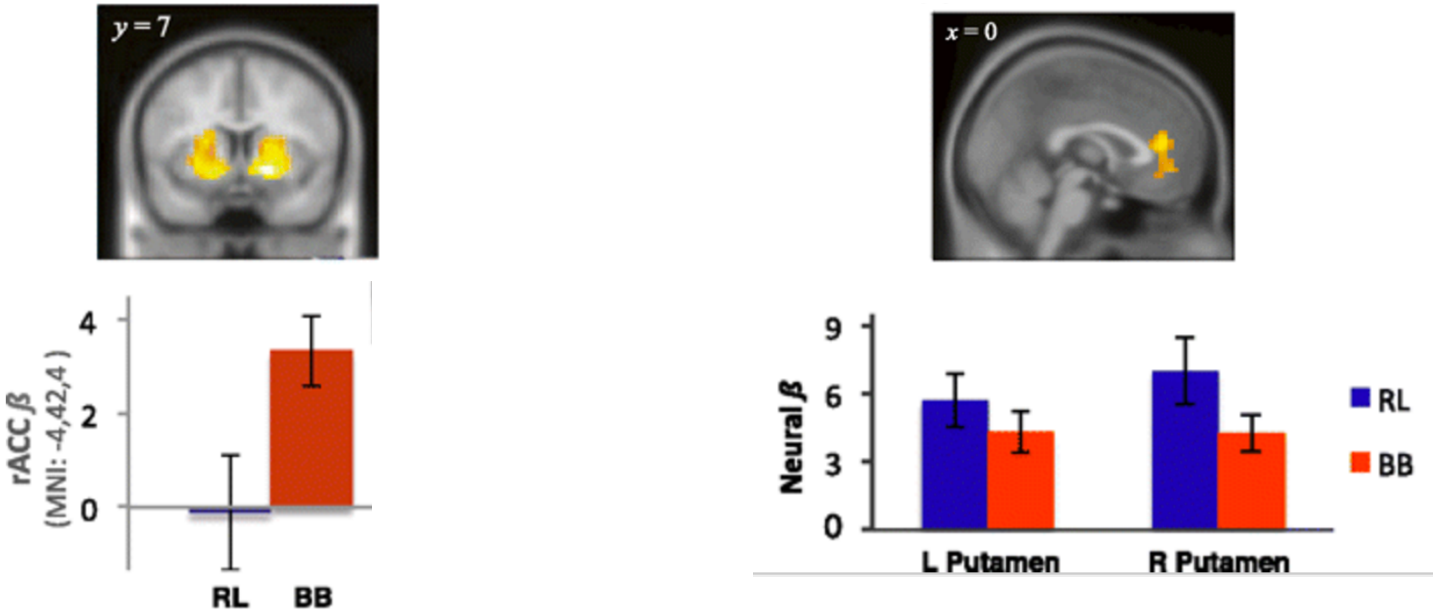
1. (model-free) "reinforcement learning", about reward received (or not) after actions
2. (model-based) "belief learning" about opponents' likely strategies, (& best-respond)

in this setting, (2) is algebraically equivalent to counterfactual learning about foregone rewards, governed by free parameter δ:

$$Q_{t+1}(c_t) = \phi \cdot Q_t(c_t) + r(c_t) \qquad \textit{for chosen action}$$
$$Q_{t+1}(u_t) = \phi \cdot Q_t(u_t) + \delta \cdot r(u_t) \qquad \textit{for unchosen actions}$$

Elana Meer; Lindsay Hunter; Ming Hsu

patent race game



(Zhu et al., PNAS 2012)

In a series of papers using EWA and games like this, Ming Hsu & colleagues (2012, 2014, 2015) have shown evidence for a similar two-system story as with MDPs

- fMRI dissociation between reward and belief learning (striatum, PFC)
- individual differences (striatal vs PFC dopamine genes, aging)

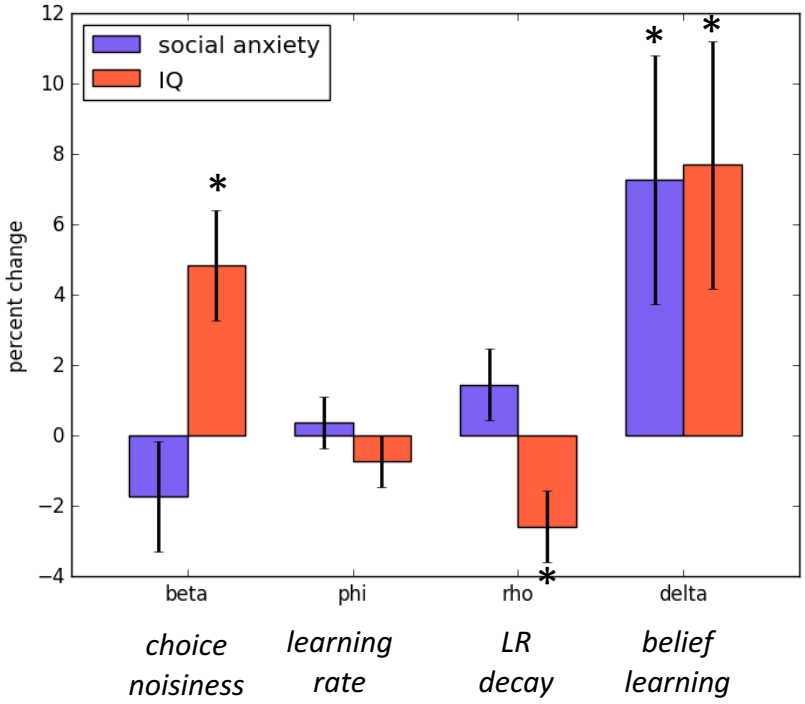
Due to social framing, this seemed like a strong candidate to follow up on social anxiety effects on model-based learning

preliminary results

N=366, Turk sample

- social anxiety, IQ (ravens matrices), 80 trials of patent race
- fit EWA model

parameter δ (rel. strength of MB) increasing in anxiety (p<.05)



Elana Meer; Lindsay Hunter; Ming Hsu

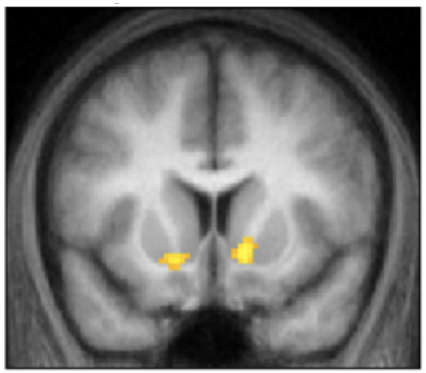
Question: how to account for the goal-directed nature of compulsion?

MB/MF interactions

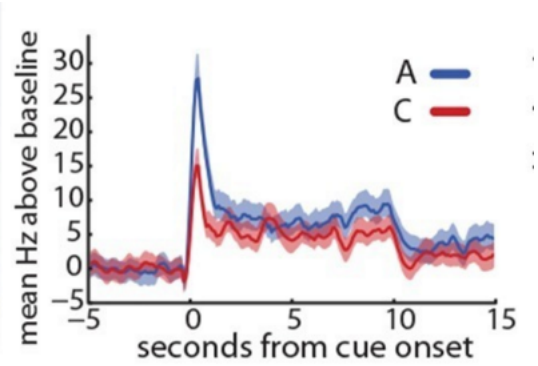
- Dyna & replay (Gershman et al. 2014)
- Pruning/truncation (Simon & Daw 2011)
- Successor representation (Daw & Dayan 2015)
- MF goal selection (Cushman and Morris, in press)
- MB as reoriented toward object of compulsion, rather than generally deficient (Voon et al. 2015)

interactions

MB valuations → Dopamine (& PEs)

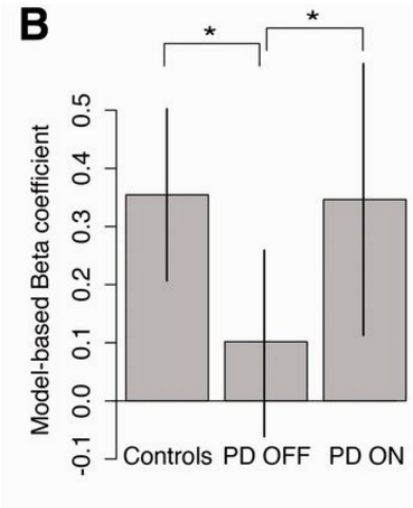


(Daw, Gershman et al, 2011)



(Sadacca et al., 2016)

Dopamine → MB valuation



- Parkinson's disease & meds (Sharp et al., 2016)
- COMT genotype (Doll et al., 2016)
- PET (Deserno et al., 2015)
- L-Dopa (Wunderlich et al., 2012)

conclusions

1. distinguish two reinforcement learning computations in the human brain
   - linked with two distinct neural mechanisms
   - forward search vs error-driven updating
   - fills in detail behind important dual-system models

2. model-based learning is linked to compulsion (& tentatively, social anxiety)
   - generalizes across disorders but is specific to a subset
   - broad usefulness of large scale online testing in psychiatry

3. many future questions
   - can we understand neural mechanism for model-based computation in finer detail? (animals!)
   - how does interaction work? (important e.g. for drugs)
   - does this give us a handle on other dual-system phenomena and frameworks, e.g. self-control, time discounting?

Lab:
Bradley Doll
Claire Gillan
Ross Otto
Elana Meer
Lindsay Hunter
Evan Russek
Oliver Vikbladh

Collaborators:
Liz Phelps
Daphna Shohamy
Sam Gershman
Ming Hsu
Peter Dayan
Valerie Voon
Dylan Simon

Funding:
NIMH
NIDA
NINDS
McDonnell Foundation
Templeton Foundation
Google DeepMind