Examples of Reinforcement Learning Models applied to Computational Psychiatry:

Models of Substance Addiction

Peggy Seriès, IML
Informatics, University of Edinburgh, UK

pseries@inf.ed.ac.uk

CCN Lecture 10

A New Model for Mental Illness

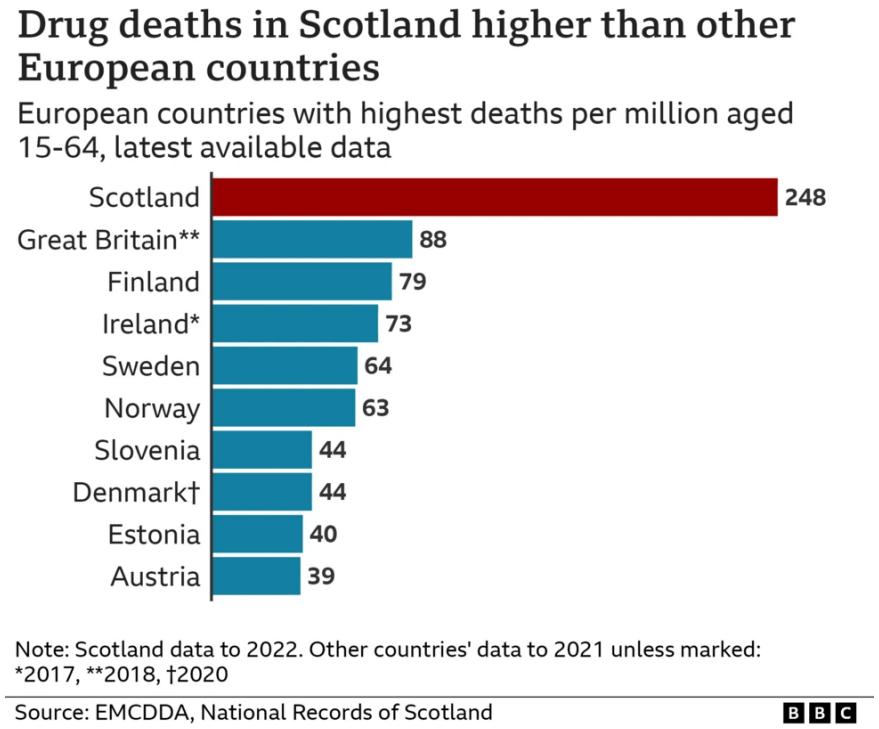
Mental illness is the result of an impairment in prediction, due to having a distorted internal model of the world, possibly due to an impairment in learning.

Applications of RL models to Computational Psychiatry

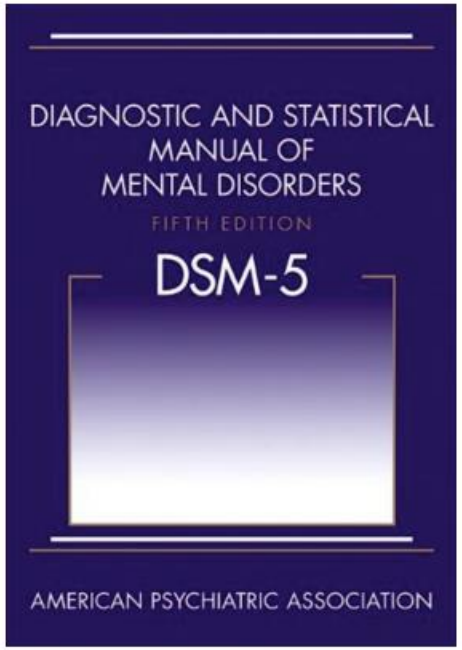
- RL models have been used to model almost all psychiatric disorders.
- **idea**: disorder can be understood as impairment in learning/decision-making.

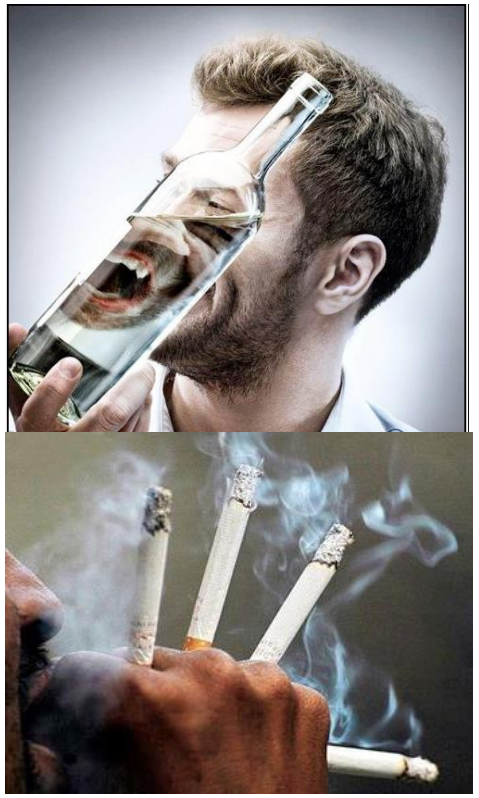
- In the following 2 lectures, two examples:
- Substance Addiction
- Anxiety and Depression

- Nearly **23 million Americans—almost one in 10**—are addicted to alcohol or other drugs.
- More than two-thirds of people with addiction abuse alcohol.
- The top three drugs causing addiction are marijuana, opioid (narcotic) pain relievers, and cocaine.

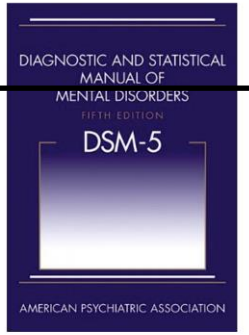


Substance Addiction: Diagnosis





Substance Addiction: Diagnosis



Person must demonstrate two of the following criteria within a 12-month period:

- regularly consuming **larger amounts of a substance than intended** or for a longer amount of time than planned
- often attempting to or expressing a **wish to moderate** the intake of a substance without reducing consumption
- spending long periods trying to get hold of a substance, use it, or recover from use
- **craving** the substance, or expressing a strong desire to use it
- **failing to fulfil** professional, educational, and family obligations
- regularly using a substance in spite of any social, emotional, or personal issues it may be causing or making worse
- giving up pastimes, passions, or social activities as a result of substance use
- consuming the substance in places or situations that could cause physical injury
- continuing to consume a substance despite being aware of any **physical or psychological harm** it is likely to have caused
- **increased tolerance**, meaning that a person must consume more of the substance to achieve intoxication
- **withdrawal symptoms**, or a physical response to not consuming the substance that is different for varying substances but might include sweating, shaking and nausea

> 2= mild; > 4=moderate; > 6=severe

Addiction

Controlled Drug Use

↓ ?

Loss of Behavioural Control

Systems involved: the Reward System

- Mesolimbic Dopaminergic system – drug consumption characterized by increase of dopamine release

- DA system: originates in the ventral tegmental area (VTA) of the midbrain, and projects to the nucleus accumbens (NA - ventral striatum).

- The amygdala (A), hippocampus (HC) and medial prefrontal cortex (PFC) send excitatory projections to NA. Drug seeking behaviour induced by Glutamatergic projections from the prefrontal cortex to the NA.

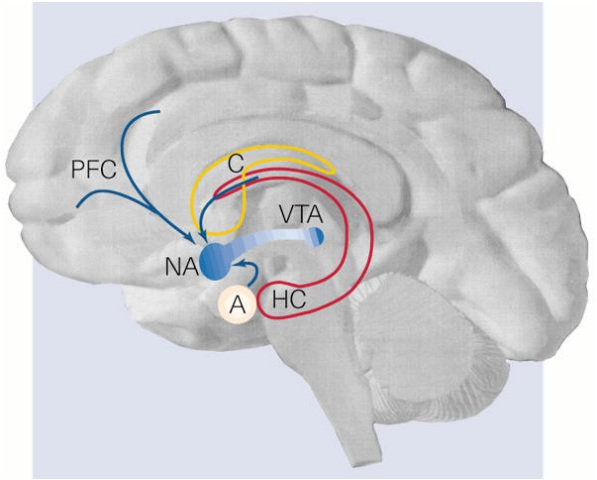


Table 1. Neurobiological Substrates for the Acute Reinforcing Effects of Drugs of Abuse

Drug of Abuse	Neurotransmitter	Sites
Cocaine and amphetamines	Dopamine Serotonin	Nucleus accumbens Amygdala
Opiates	Dopamine Opioid peptides	Ventral tegmental area Nucleus accumbens
Nicotine	Dopamine Opioid peptides?	Ventral tegmental area Nucleus accumbens Amygdala?
THC	Dopamine Opioid peptides?	Ventral tegmental area
Ethanol	Dopamine Opioid peptides Serotonin GABA Glutamate	Ventral tegmental area Nucleus accumbens Amygdala

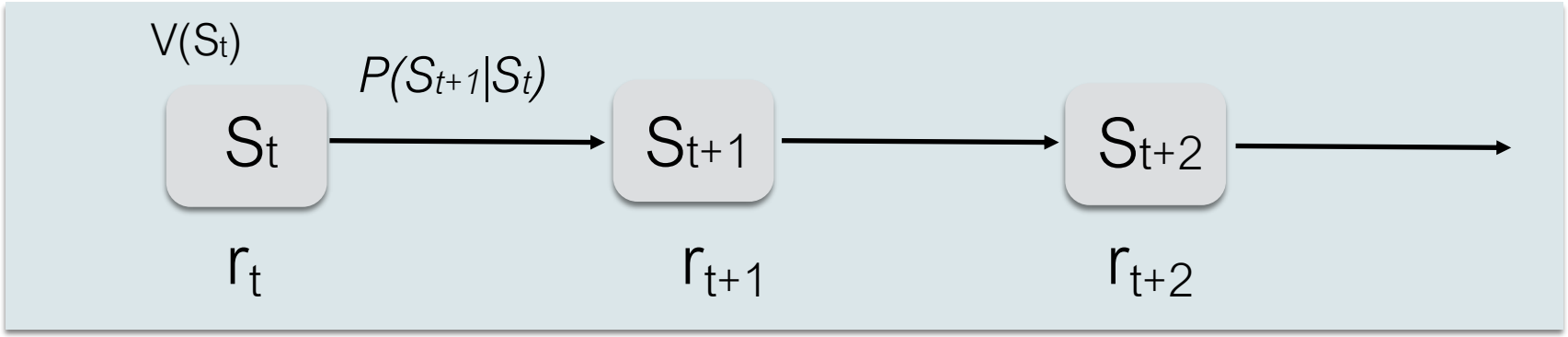
Theories of Addiction

- In the past 30 years, lots of theories, e.g.
- compulsion zone: self administration is automatically induced when, after first consumption, brain drug levels decrease from a specific range.
- set-point model (or allostasis): drugs decrease baseline level of reward sensitivity
- opponent-process theory: drug addiction = result of emotional pairing between pleasure and symptoms of withdrawal. Motivation is first related to pleasure, and then to relief from withdrawal.
- impulsivity (discounting): Incapacity to consider long-term costs, prefer immediate rewards (drugs) over larger delayed rewards (e.g., long-term health).

→ **recently, addiction as a vulnerability in the decision process;**

Inspiration from reinforcement learning

Temporal Difference (TD) learning



- Consider a succession of states S, following each other with $P(S_{t+1}|S_t)$
- Rewards observed in each state with probability $P(r|S_t)$

*(*This is a *Markov Decision Process)*

- Useful quantity to predict is the expected sum of all future rewards, given current state S_t,

= value of state S, $V(S_t)$

$$V(S_t) = E\left[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots \middle| S_t\right] = E\left[\sum_{i=t}^{\infty} \gamma^{i-t} r_i \middle| S_t\right]$$

where E denotes expected value (or mean) and gamma the discount factor

Temporal Difference (TD) learning

- Resulting learning rule:

$$V_{new}(S_t) = V_{old}(S_t) + \eta(r_t + \gamma V_{old}(S_{t+1}) - V_{old}(S_t)).$$

current reward+next prediction - current prediction

- This is TD(0) learning rule as proposed by Sutton & Barton (1990).

- reduces to Rescorla-Wagner model if only one step i.e. $V(S_{t+1})=0$.

$$V_{new}(S_t) = V_{old}(S_t) + \eta(r_t - V_{old}(S_t)).$$

Phasic dopamine signals prediction error

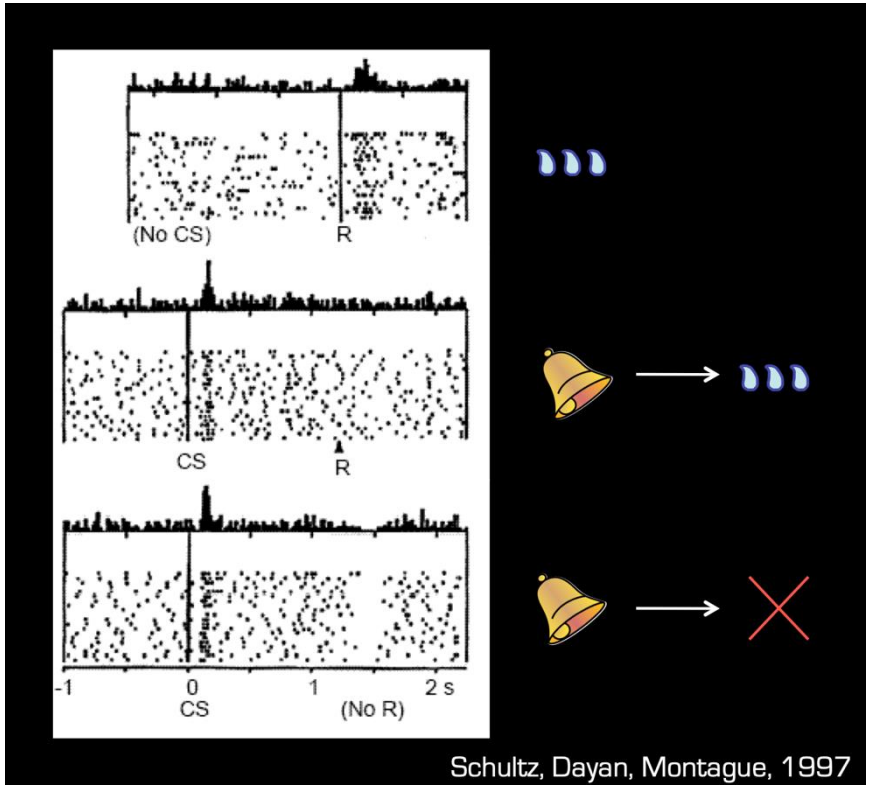
- the “largest success of computational neuroscience” [Niv]

Phasic DA = Prediction Error

A Neural Substrate of Prediction and Reward

Wolfram Schultz, Peter Dayan, P. Read Montague*

The capacity to predict future events permits a creature to detect, model, and manipulate the causal structure of its interactions with its environment. Behavioral experiments suggest that learning is driven by changes in the expectations about future salient events such as rewards and punishments. Physiological work has recently complemented these studies by identifying dopaminergic neurons in the primate whose fluctuating output apparently signals changes or errors in the predictions of future salient and rewarding events. Taken together, these findings can be understood through quantitative theories of adaptive optimizing control.



Schultz, Dayan, Montague, 1997

http://www.sciencemag.org • SCIENCE • VOL. 275 • 14 MARCH 1997

Redish's (*Science*, 2004) model: cocaine messes up with prediction error

REPORTS 1944 10 DECEMBER 2004 VOL 306 SCIENCE www.sciencemag.org

Addiction as a Computational Process Gone Awry

A. David Redish

Addictive drugs have been hypothesized to access the same neurophysiological mechanisms as natural learning systems. These natural learning systems can be modeled through temporal-difference reinforcement learning (TDRL), which requires a reward-error signal that has been hypothesized to be carried by dopamine. TDRL learns to predict reward by driving that reward-error signal to zero. By adding a noncompensable drug-induced dopamine increase to a TDRL model, a computational model of addiction is constructed that overselects actions leading to drug receipt. The model provides an explanation for important aspects of the addiction literature and provides a theoretic viewpoint with which to address other aspects.

$$\delta'_t = \delta_t + D_{\text{drug}}$$

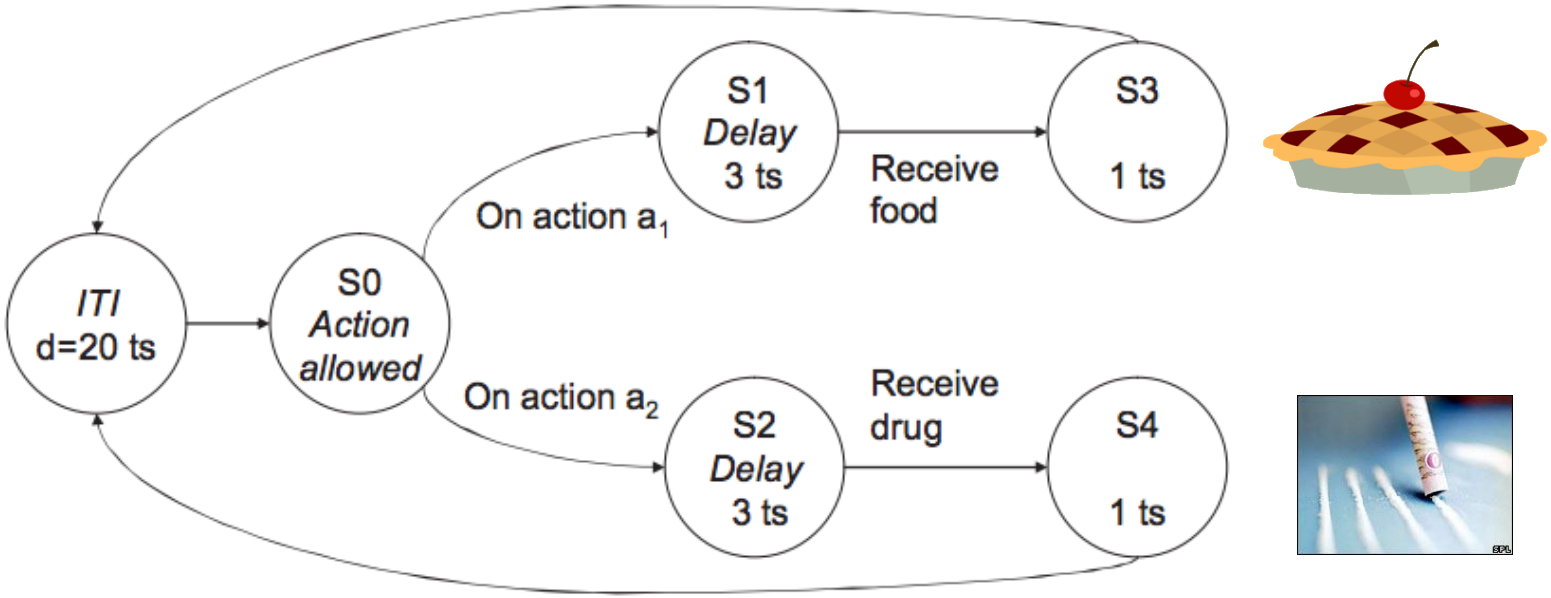
- Cocaine and other drugs produce a transient increase in dopamine
- idea: this dopamine surge induce an increase in prediction error δ that can't be compensated by changes in values

$$\delta = \max\{\underbrace{\gamma^d[R(S_l) + V(S_l)] - V(S_k) + D(S_l)}_{\text{normal TD error + drug bump}}, \quad \underbrace{D(S_l)}_{\text{drug bump alone}} \}.$$

where $D(S_l)$ indicates a dopamine surge occurring on entry into S_l.

Consequence: values of states leading to the drug increase without bound.

Redish's (*Science*, 2004) model



$$V(S_k) \leftarrow V(S_k) + \eta_V \delta$$

$$\delta = \max\{\gamma^d [R(S_l) + V(S_l)] - V(S_k) + D(S_l),\ D(S_l)\}$$

- **Drug is hijacking the learning pathways, creating a prediction error where there should be none.**

Redish's (2004) model

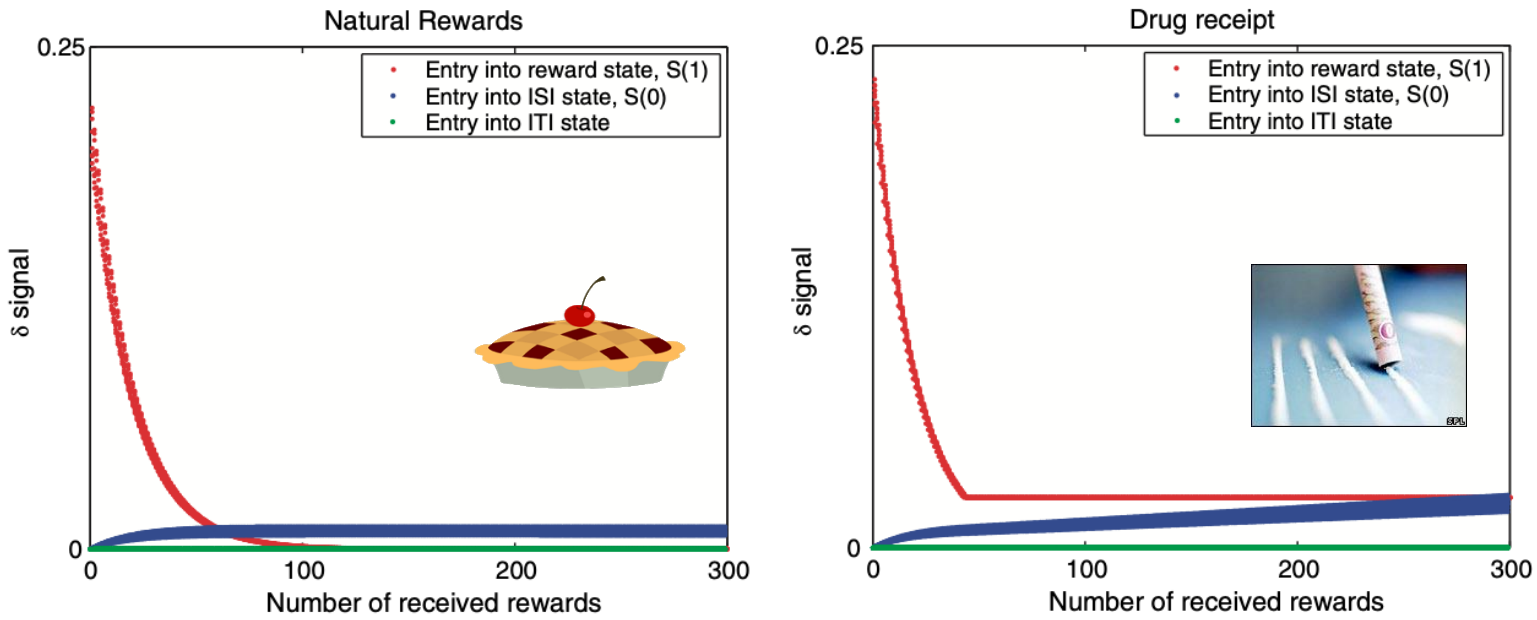


Fig. 3. Dopamine signals. (**Left**) With natural rewards, dopamine initially occurs primarily at reward receipt (on entry into reward state S_1) and shifts to the conditioned stimulus [on entry into interstimulus-interval (ISI) state S_0] with experience. (State space is shown in fig. S7.) (**Right**) With drugs that produce a dopamine signal neuropharmacologically, dopamine continues to occur at the drug receipt (on entry into reward state S_1) even after experience, as well as shifting to the conditioned stimulus (on entry into ISI state S_0), thus producing a double dopamine signal.

- **Drug is hijacking the learning pathways, creating a prediction error where there should be none.**

Redish's (2004) model: Predictions

- With repeated experience, drug choice becomes:

1) less sensitive to alternative non-drug reinforcers [some evidence];

2) more inelastic to costs [confirmed]

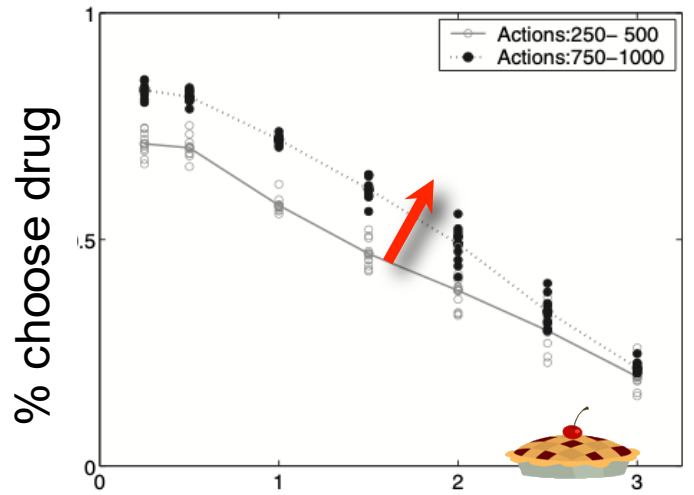


Fig. 1. Probability of selecting a drug-receipt pathway depends on an interaction between drug level, experience, and contrasting reward. Each line shows the average probability of selecting the drug-receipt pathway, $S_0 \xrightarrow{a_2} S_2$, over the contrasting reward pathway, $S_0 \xrightarrow{a_1} S_1$, as a function of the size of the contrasting reward $R(S_3)$. (State space is shown in fig. S1.) Drug receipt on entering state S_4 was $R(S_4) = 1.0$ and $D(S_4) = 0.025$. Individual simulations are shown by dots. Additional details provided in (*14*).

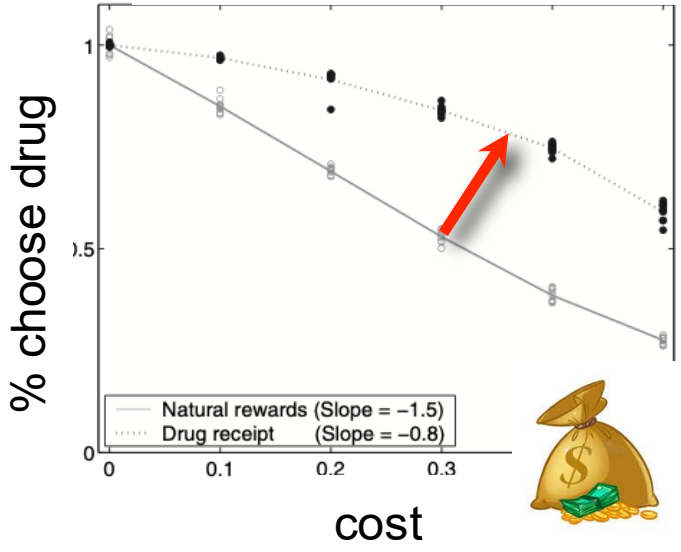


Fig. 2. Elasticity of drug receipt and natural rewards. Both drug receipt and natural rewards are sensitive to costs, but natural rewards are more elastic. Each dot indicates the number of choices made within a session. Sessions were limited by simulated time. The curves have been normalized to the mean number of choices made at zero cost.

Redish's (2004) model: testing the predictions

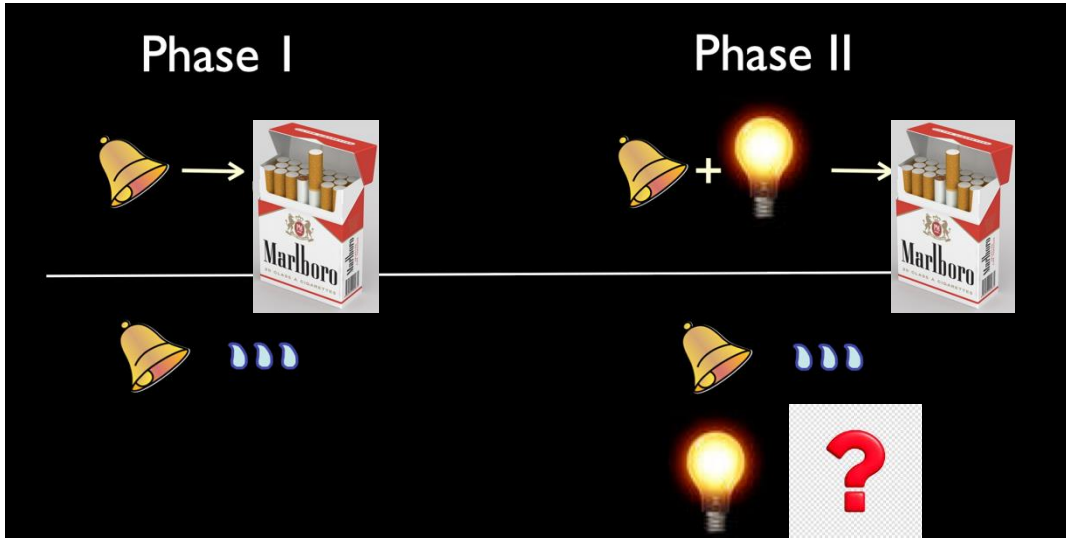
- Double surge of dopamine in drug experiments

> validated but in different structures of the nucleus accumbens [Aragona et al 2009]

- Drugs would not show Blocking (when drugs are the reward)

[Panilio et al 2007, Jaffe et al 2014]

> a subset of animals don't show blocking with nicotine.



Redish's (2004) model: testing the predictions

- A lever delivers high dose of cocaine, then reduced to lower dose :
  Does the rat adapt how he values the lever (lower their reward expectation)?
- Redish's model predicts that he shouldn't.
  > Theory not validated. But maybe a subset problem again?

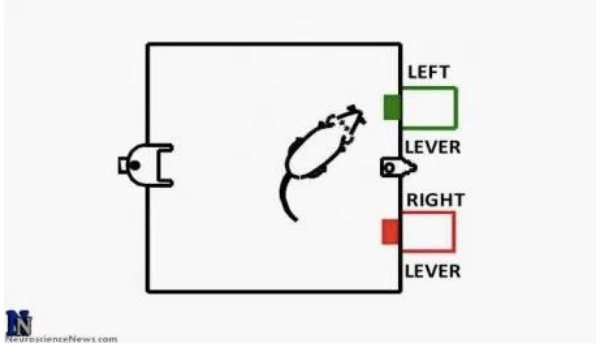
Published in final edited form as:
Behav Brain Res. 2010 October 15; 212(2): 204–207. doi:10.1016/j.bbr.2010.03.053.

Learning That a Cocaine Reward is Smaller Than Expected: A Test of Redish's Computational Model of Addiction

Katherine R. Marks, **David N. Kearns**, **Chesley J. Christensen**, **Alan Silberberg**, and **Stanley J. Weiss**
Psychology Department American University

Abstract

The present experiment tested the prediction of Redish's [7] computational model of addiction that drug reward expectation continues to grow even when the received drug reward is smaller than expected. Initially, rats were trained to press two levers, each associated with a large dose of cocaine. Then, the dose associated with one of the levers was substantially reduced. Thus, when rats first pressed the reduced-dose lever, they expected a large cocaine reward, but received a small one. On subsequent choice tests, preference for the reduced-dose lever was reduced, demonstrating that rats learned to devalue the reduced-dose lever. The finding that rats learned to lower reward expectation when they received a smaller-than-expected cocaine reward is in opposition to the hypothesis that drug reinforcers produce a perpetual and non-correctable positive prediction error that causes the learned value of drug rewards to continually grow. Instead, the present results suggest that standard error-correction learning rules apply even to drug reinforcers.



Questions and extensions

- Redish's model, extensions and RL framework

--> a new generation of models and model-driven experiments.

Lots of remaining challenges:

- addiction to ordinary rewards such as fatty foods, which unlike cocaine produce a dopamine signal that can be accommodated
- addiction to non-stimulant substances which depend less on mesolimbic dopamine (e.g. alcohol)
- describing withdrawal symptoms -- opponent mechanisms
- why do people want to get sober?
- why do people relapse?; accounting for effect of stress.
- vulnerability: only a minority of people become addicted -- while other people can enjoy casual use, why? (drug use and drug addition are two different things !).

Multi-systems theories: Model-based vs Model-Free

- TD learning models are called "model-free" because the structure of the environment is not learnt explicitly (i.e. transition probabilities., reward probabilities).
- Debated how much human learning/ decision-making is "model-free" vs "model-based"

- Model-based correspond to planning, deliberative
- Model-free corresponds to habitual, inflexible, procedural

- Possibly relevant to pathology

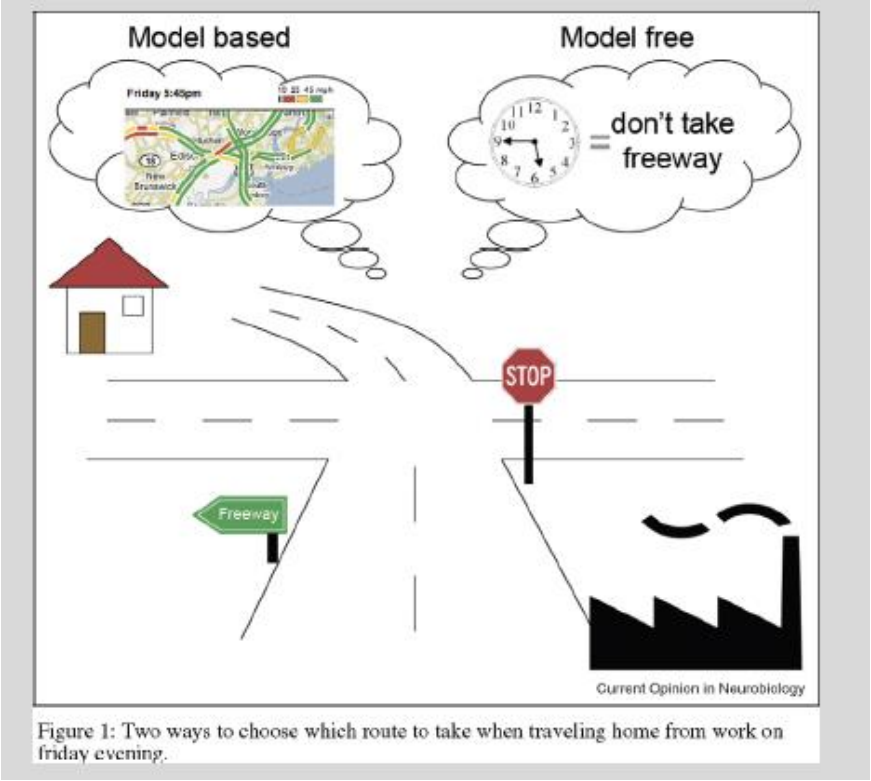


Figure 1: Two ways to choose which route to take when traveling home from work on friday evening.

Multi-systems theories: Model-based vs Model-Free

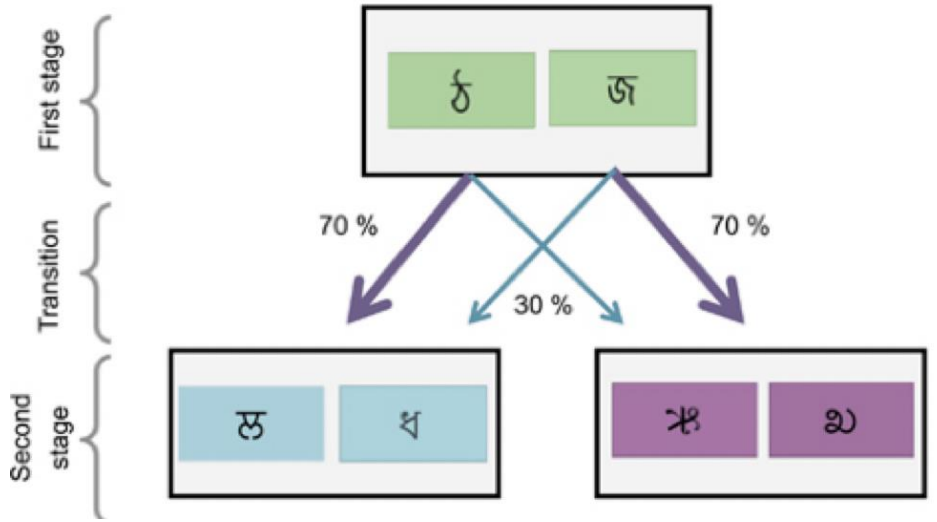
- It might be possible to execute the same task with one or the other of the systems.
- Damage to one system can drive behaviour to be controlled by the other
- There are multiple failure models in each system and interaction.
- Drug addiction could correspond to a disruption of the model-based system and shift to model-free/ habitual system.

habitual system

goal-directed system

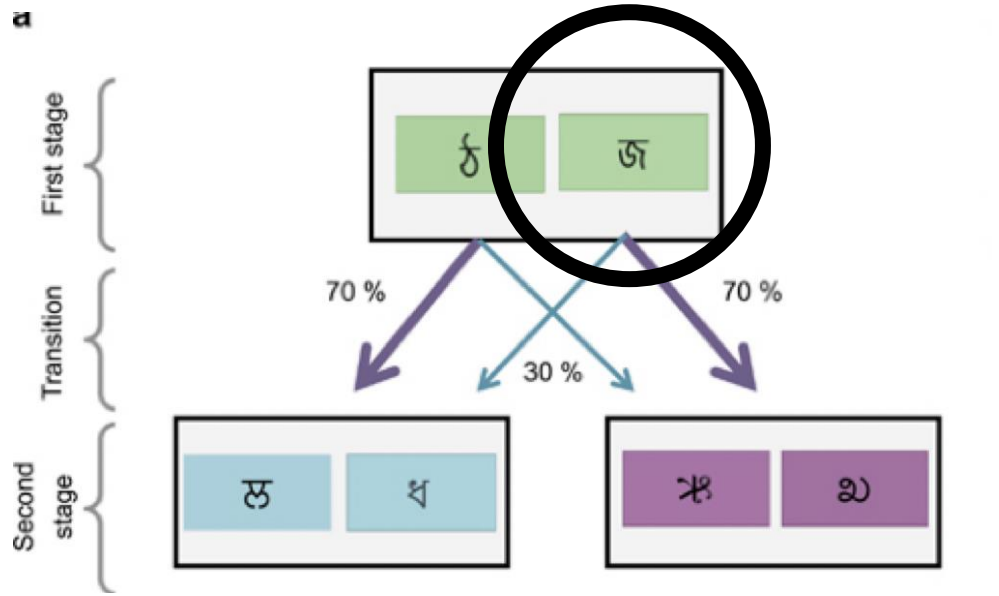
Two-Steps Decision Task

- How can we assess the relative involvement of both systems in humans? (Daw et al. 2011)
- On each trial, choosing between 2 stimuli leads with fixed probabilities to one of 2 pairs of stimuli in stage 2.
- Each of the four 2nd-stage stimuli is associated with a probabilistic outcome (money $$).
- Those probabilities change slowly and independently across the trials.



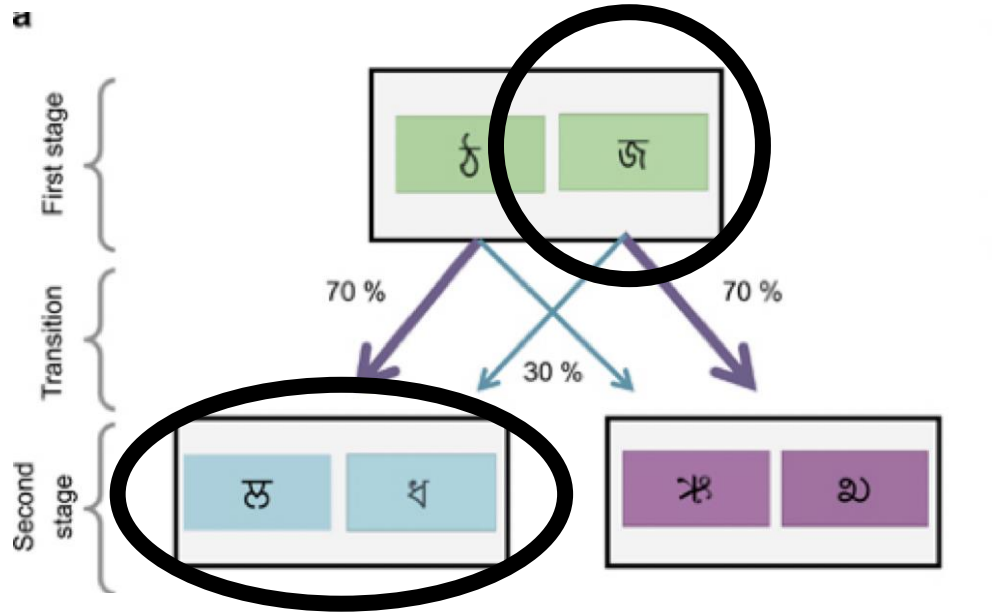
Two-Steps Decision Task

- Model-based and model-free strategies make different predictions about the influence of the outcome obtained after the second stage onto subsequent first-stage choices.



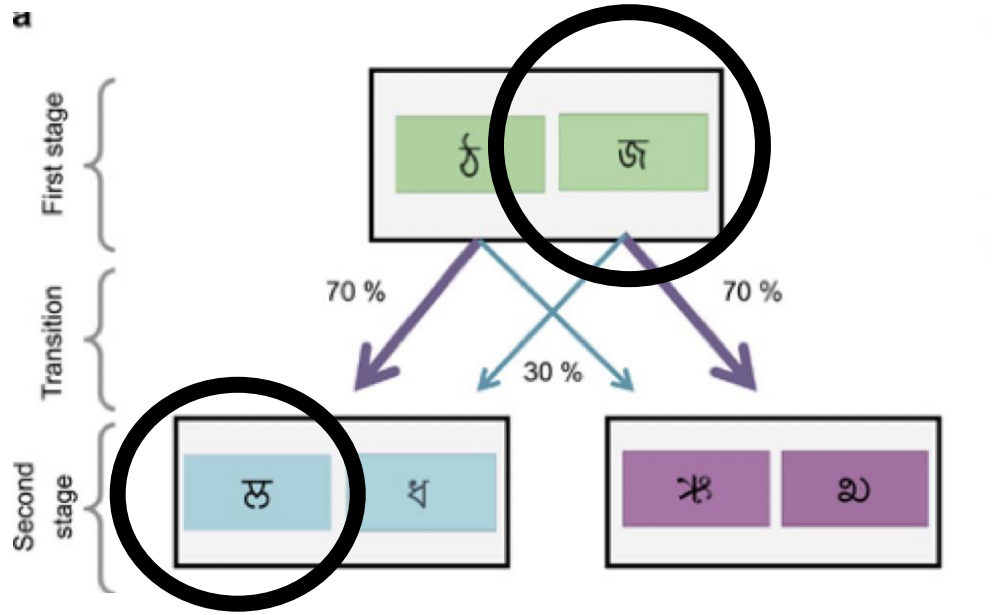
Two-Steps Decision Task

- Model-based and model-free strategies make different predictions about the influence of the outcome obtained after the second stage onto subsequent first-stage choices.



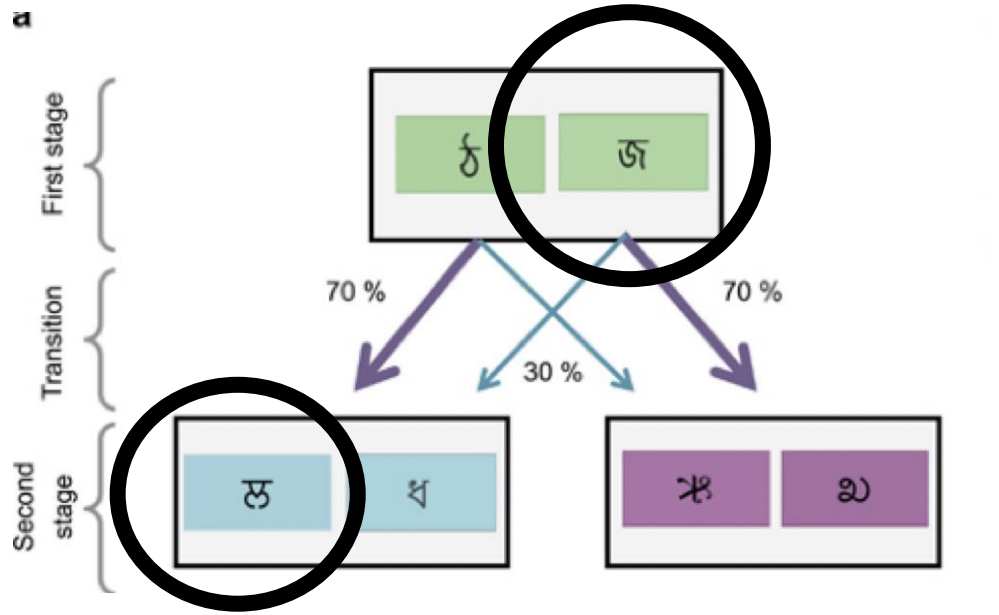
Two-Steps Decision Task

- Model-based and model-free strategies make different predictions about the influence of the outcome obtained after the second stage onto subsequent first-stage choices.



Two-Steps Decision Task

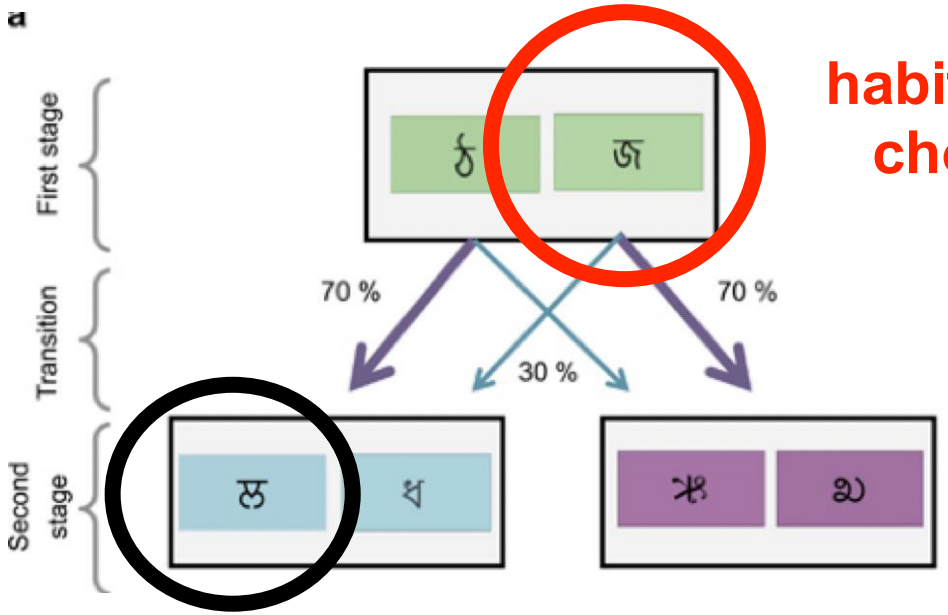
- Model-based and model-free strategies make different predictions about the influence of the outcome obtained after the second stage onto subsequent first-stage choices.



$$$$

Two-Steps Decision Task

- Model-based and model-free strategies make different predictions about the influence of the outcome obtained after the second stage onto subsequent first-stage choices.

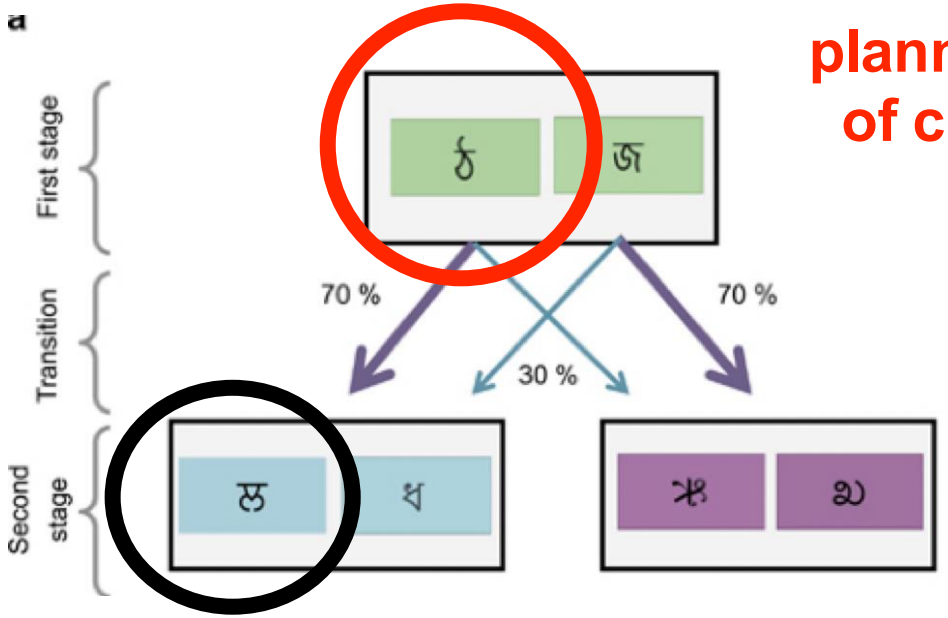


habit system: increase value of choice that led to the action

$$$$

Two-Steps Decision Task

- Model-based and model-free strategies make different predictions about the influence of the outcome obtained after the second stage onto subsequent first-stage choices.

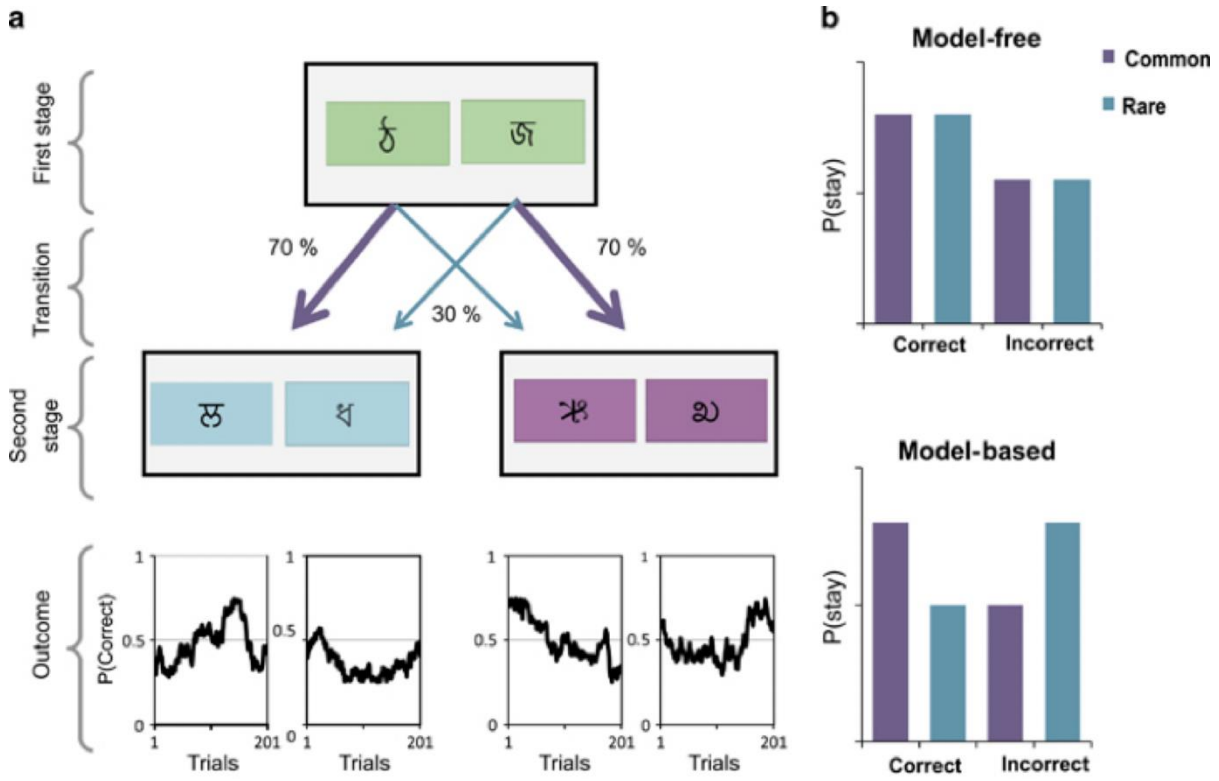


planning system: increase value of choice that is most likely to allow that action

$$$$

Two-Steps Decision Task

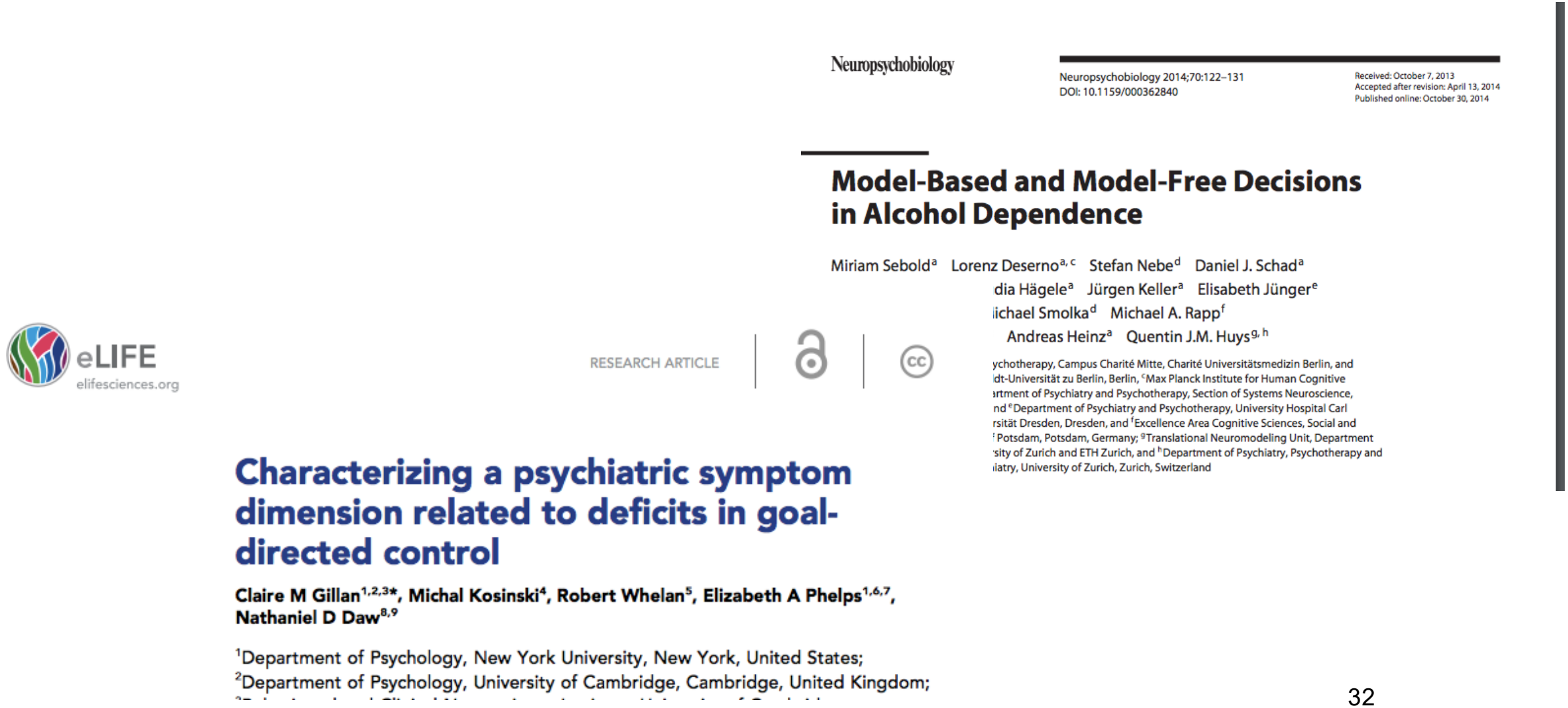
- Model-based and model-free strategies make different predictions about the patterns of responses
- Comparison to those patterns and fitting models to participants responses can be used to quantify the contribution of each system in humans.



$$Q_{\text{hyb}}(s_1, a_1) = \omega\, Q_{\text{MB}}(s_1, a_1) + (1 - \omega)\, Q_{\text{MF}}(s_1, a_1)$$

- The MB model uses the true transition probabilities in the first stage.
- omega quantifies how much learning is MB vs MF in individual participants.

- Deficits in goal-directed decision-making have been linked to compulsive behaviour and intrusive thoughts (Gillian et al 2016).



Neuropsychobiology

Neuropsychobiology 2014;70:122–131
DOI: 10.1159/000362840

Received: October 7, 2013
Accepted after revision: April 13, 2014
Published online: October 30, 2014

Model-Based and Model-Free Decisions in Alcohol Dependence

Miriam Sebold[a] Lorenz Deserno[a, c] Stefan Nebe[d] Daniel J. Schad[a]
dia Hägele[a] Jürgen Keller[a] Elisabeth Jünger[e]
ichael Smolka[d] Michael A. Rapp[f]
Andreas Heinz[a] Quentin J.M. Huys[g, h]

ychotherapy, Campus Charité Mitte, Charité Universitätsmedizin Berlin, and
dt-Universität zu Berlin, Berlin, [c]Max Planck Institute for Human Cognitive
rtment of Psychiatry and Psychotherapy, Section of Systems Neuroscience,
nd [e]Department of Psychiatry and Psychotherapy, University Hospital Carl
rsität Dresden, Dresden, and [f]Excellence Area Cognitive Sciences, Social and
Potsdam, Potsdam, Germany; [g]Translational Neuromodeling Unit, Department
sity of Zurich and ETH Zurich, and [h]Department of Psychiatry, Psychotherapy and
iatry, University of Zurich, Zurich, Switzerland

eLIFE
elifesciences.org

RESEARCH ARTICLE

Characterizing a psychiatric symptom dimension related to deficits in goal-directed control

Claire M Gillan[1,2,3*], Michal Kosinski[4], Robert Whelan[5], Elizabeth A Phelps[1,6,7], Nathaniel D Daw[8,9]

[1]Department of Psychology, New York University, New York, United States;
[2]Department of Psychology, University of Cambridge, Cambridge, United Kingdom;

Conclusions

- Psychiatric disorders are increasingly viewed as deficits in learning and decision-making

- This makes RL tasks and modelling relevant to their study.

- Prominent models of addiction suggest that drug intake hijacks the learning processes (because dopamine surges interferes with the representation of prediction errors), hence leading to aberrant valuation of states leading to the drug.

- Decision-making depends on multiple systems acting concurrently.

Drug addiction could correspond to a disruption of the model-based system and shift to model-free/ habitual system.

- Current computational addiction work emphasizes multiple systems, heterogeneity, and task/parameter identifiability.

Readings

- Redish, A. D. (2004). Addiction as a Computational Process Gone Awry. Science, 306(5703), 1944–1947.

- Chapter 9 (Addiction) by Redish, in CP textbook.

- Drummond & Niv (2020). Model-based decision making and model-free learning. In: Current Biology. 2020 ; Vol. 30, No. 15. pp. R860-R865