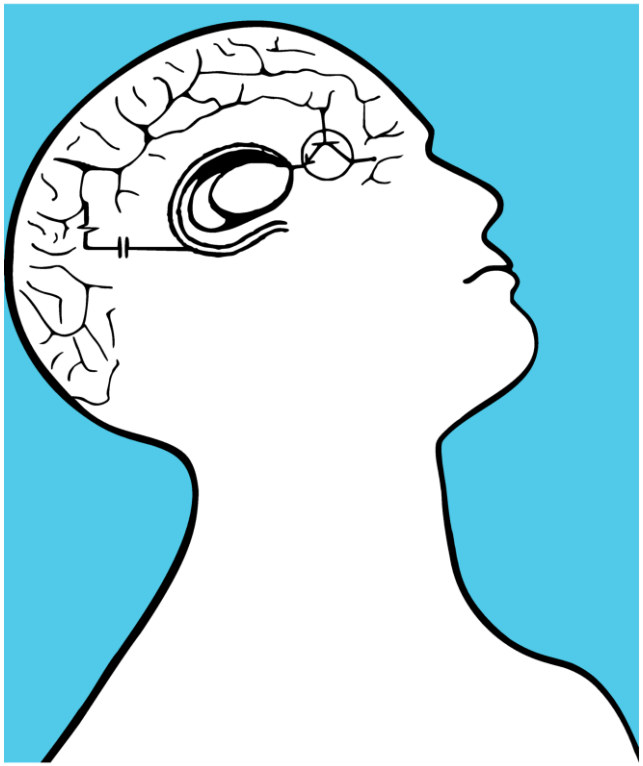
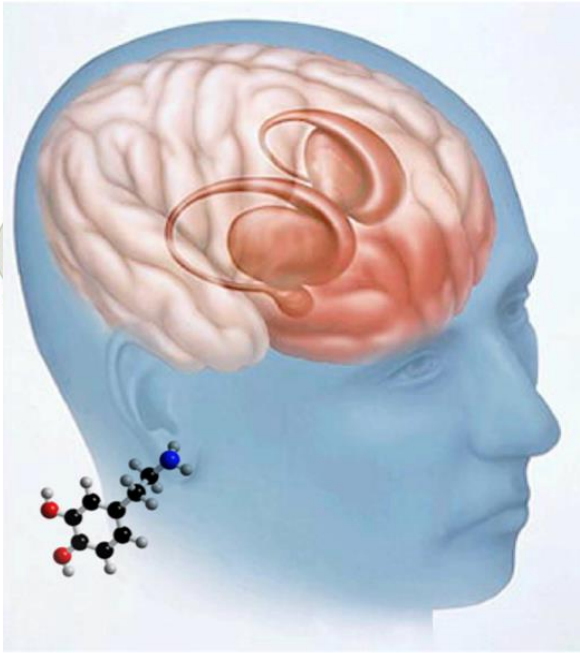
Tuning striatal dopamine signals to optimize reinforcement learning across tasks

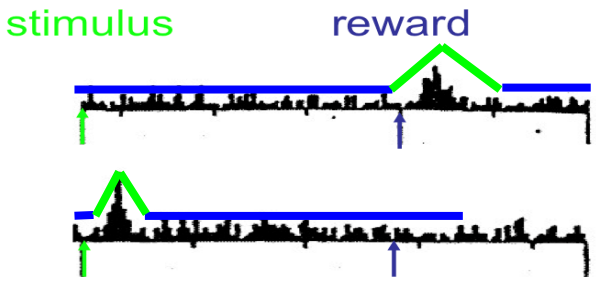


Michael J. Frank
Carney Center for Computational Brain Science
Laboratory for Neural Computation and Cognition

CARNEY INSTITUTE
FOR BRAIN SCIENCE
BROWN UNIVERSITY

[mal]Adaptive tuning of striatal dopamine signals

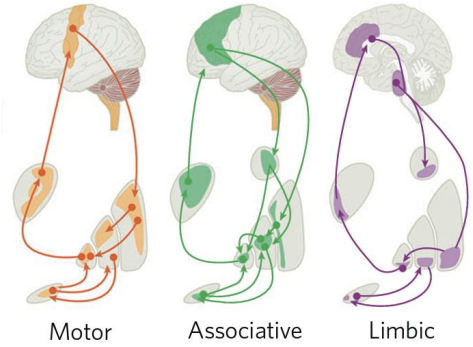




Adaptive tuning of striatal dopamine signals

- *Across* striatal regions

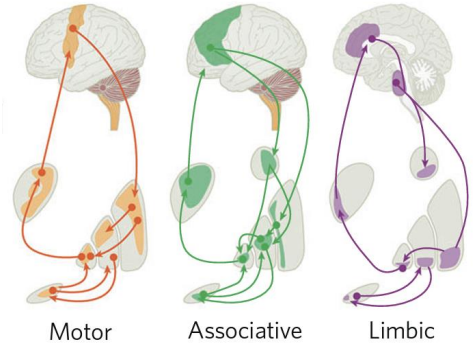
  - to credit neural circuits needed for task



Adaptive tuning of striatal dopamine signals

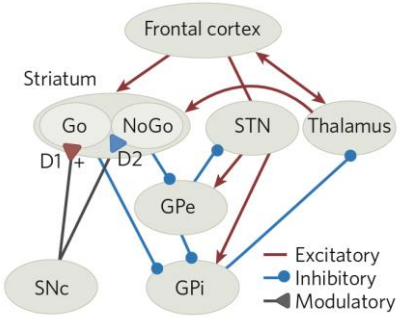
- *Across* striatal regions

  - to credit neural circuits needed for task



- *Within* striatal regions: opponency

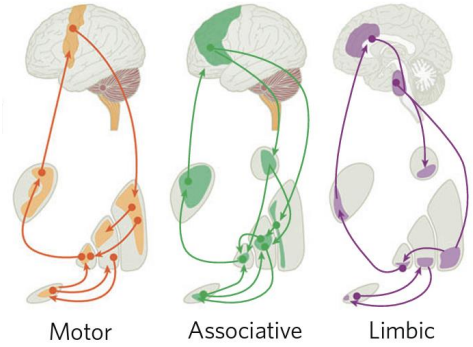
  - to optimize value-based choice



[mal]Adaptive tuning of striatal dopamine signals

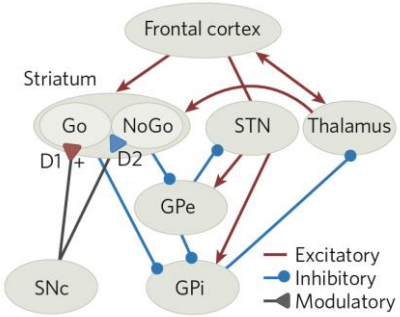
- *Across* striatal regions

  - to credit neural circuits needed for task



- *Within* striatal regions: opponency

  - to optimize value-based choice

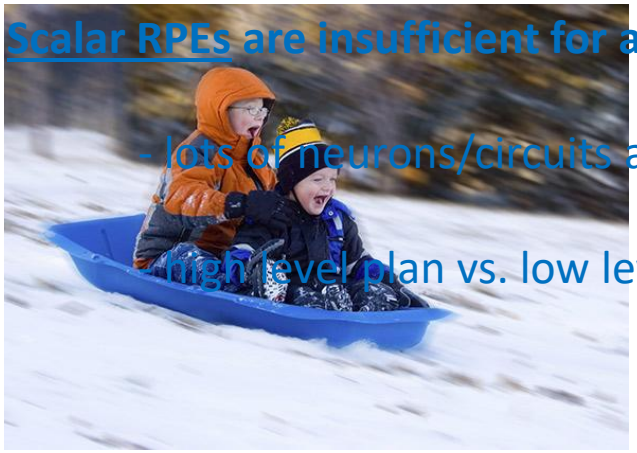


- Psychopathology as a by-product:

  - PD, gambling, ADHD, SZ, antipsychotics

Credit assignment for causal circuits

- **Scalar RPEs are insufficient for assigning credit to causal circuits**

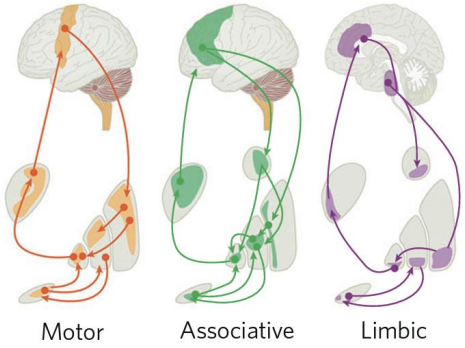


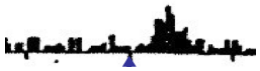
- lots of neurons/circuits active, need to reinforce the right ones

S-O
(passive)

- high level plan vs. low level motor implementation

S-A-O
(agency)



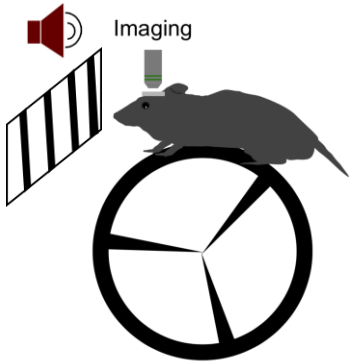


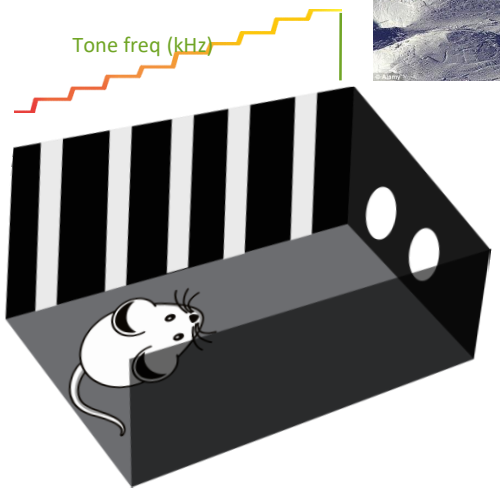
DA

DA innervates all levels of corticostriatal hierarchy…

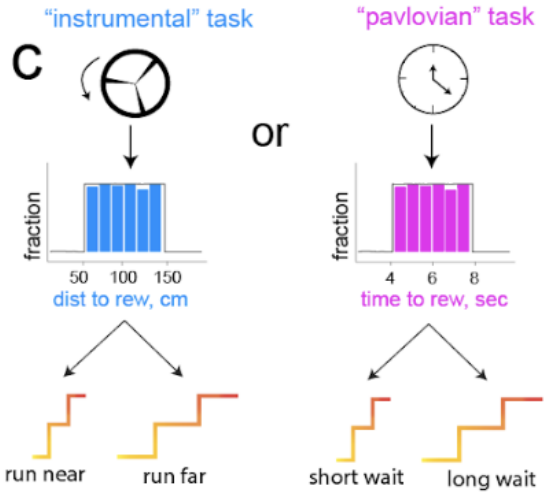
Frank & Badre 2012; Collins & Frank2013

DA and credit assignment in mice





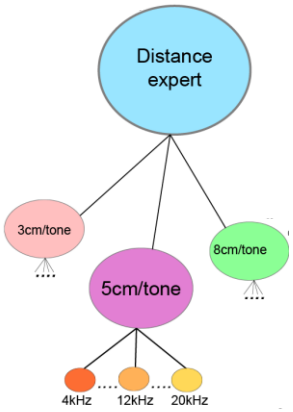
Hierarchical structure



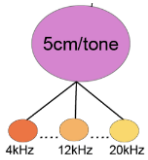
S-A-O (agency)

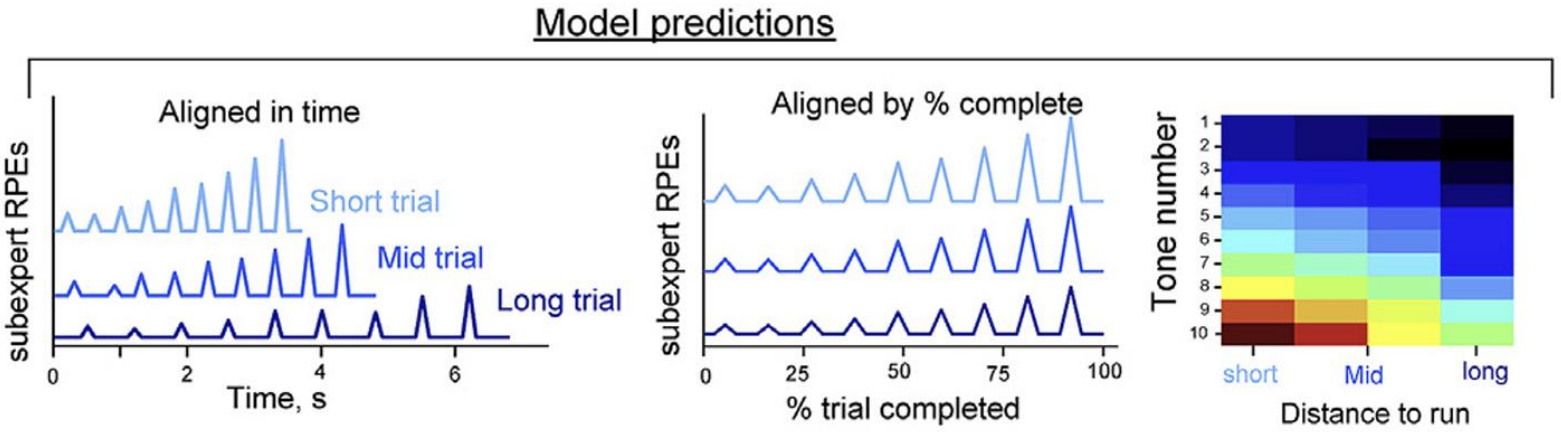
S-O (passive)

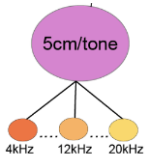
Multi-agent RL

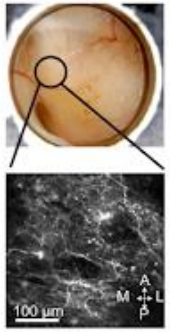


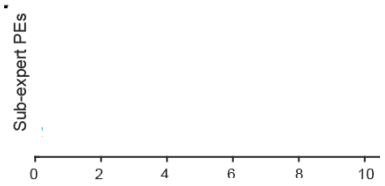
Sequential RPEs with progress to reward



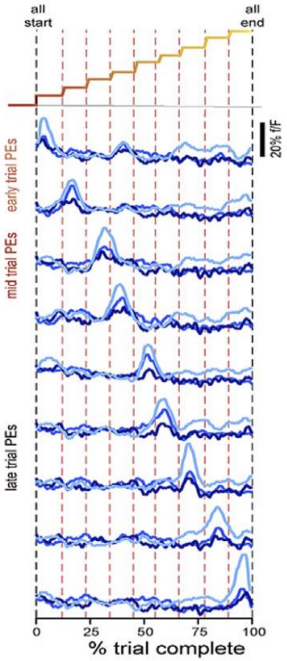


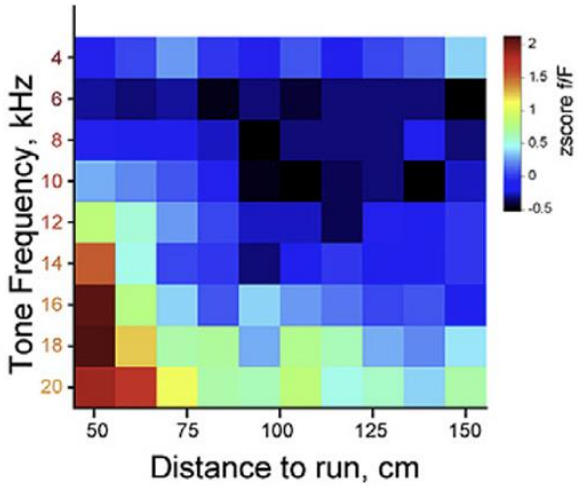






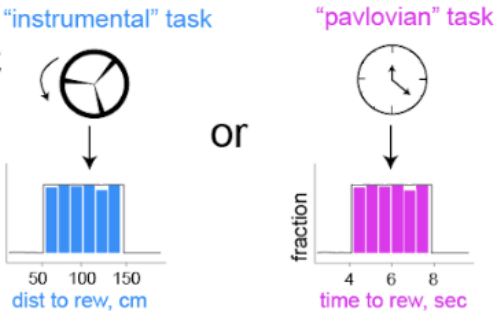
Sub-expert RPEs





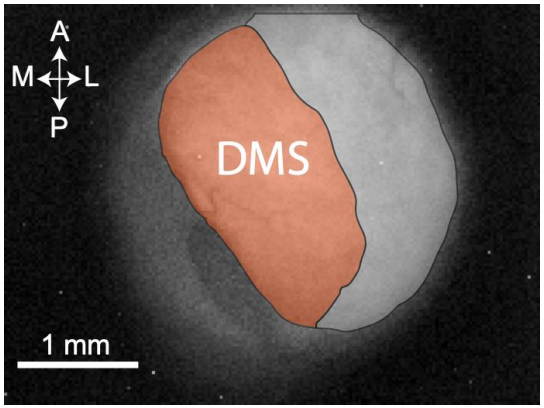
Hamid, Frank & Moore 2021 *Cell*

But how to assign credit to causal circuit?

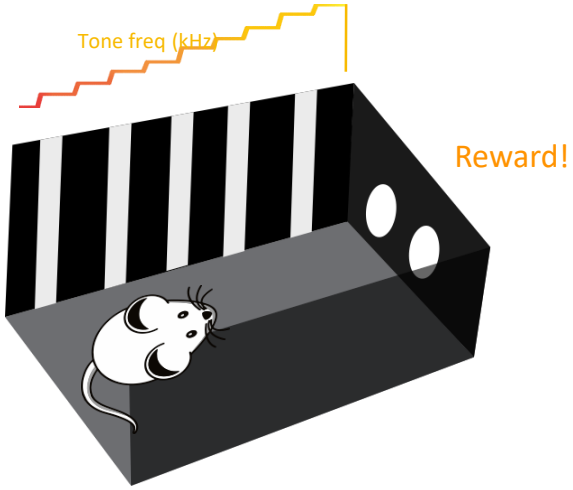


DMS:

Goal-directed, "action-outcome" *(agency)*



Are my actions responsible for attaining rewards?

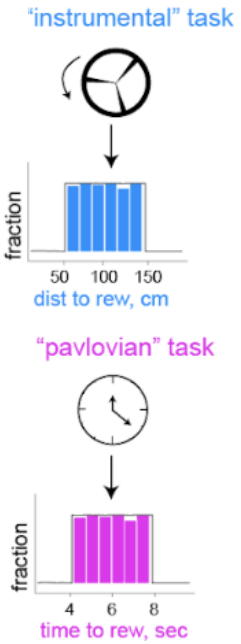


Actions affect world?

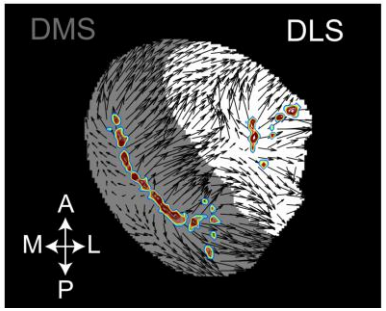
instrumental task ↑ DMS reinforcement

pavlovian task ↓ DMS reinforcement

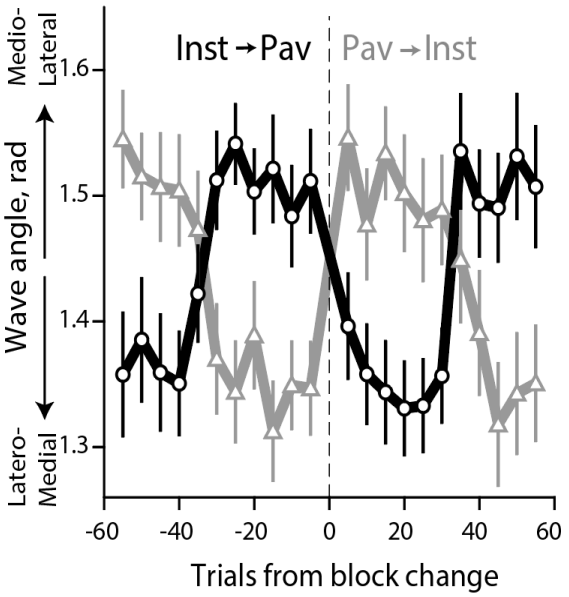
Credit assignment at rewards: directional DA *waves*



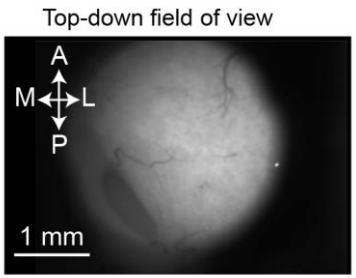
Instrumental task



Medio-lateral wave



Task reversals



ajectory is

o task structure (vector RPE)

- **can be used for credit assignment…**

but HOW?

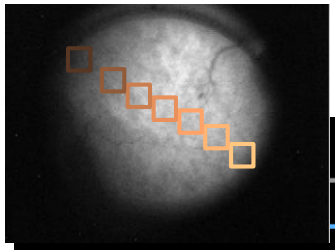
Cell

Article

Wave-like dopamine dynamics as a mechanism for spatiotemporal credit assignment

Arif A. Hamid,[1,3,6,*] Michael J. Frank,[2,3,5,4,*] and Christopher I. Moore[1,3,5,4,*]

Waves can support credit assignment via timing delays



Waves induce DA offsets…

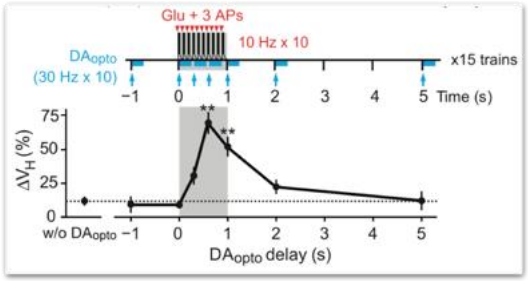
Peak normalized

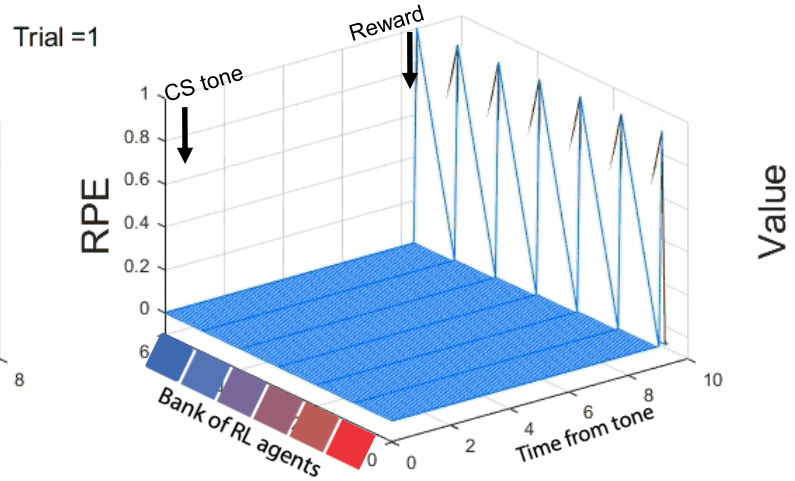
… which impact striatal plasticity

A silent eligibility trace enables dopamine-dependent synaptic plasticity for reinforcement learning in the mouse striatum

Tomomi Shindou, Mayumi Shindou, Sakurako Watanabe, Jeffery Wickens

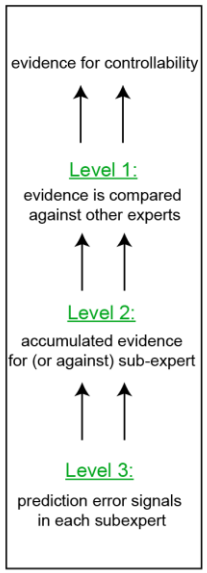
Yagishita et al 2014

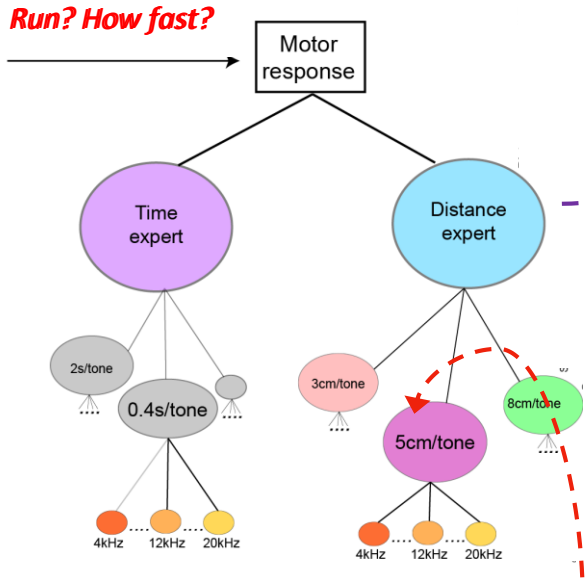


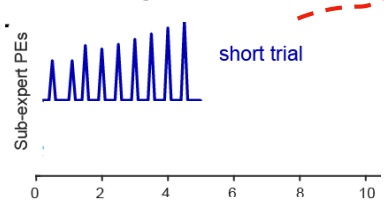


But how to control waves in the first place?

"Mixture of striatal experts" to direct, and learn from, DA

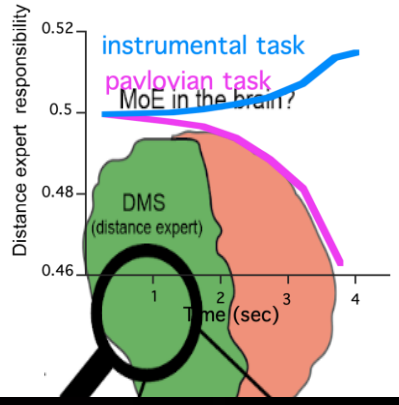






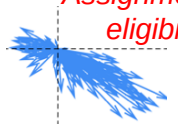
Sub-expert RPEs

DMS in control!

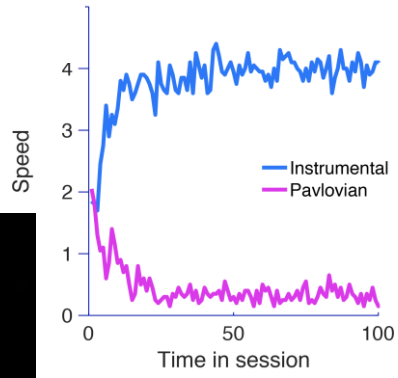


'Local' RPEs signal which experts are wrong!

Credit Assignment via eligibility



Model 'running'



Hamid Frank & Moore 2021; after Frank & Badre 2012

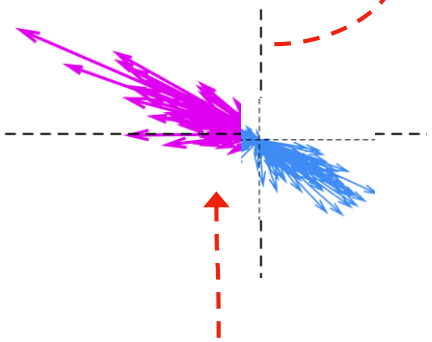
see Barbera et al '16, Matamales et al '20 for MSN 'expert-like' clusters

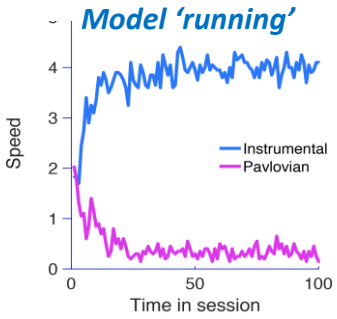
Predictions

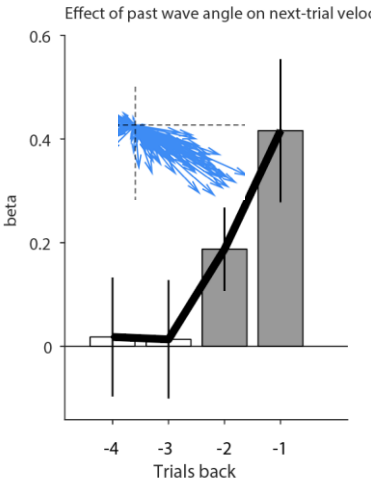
Data

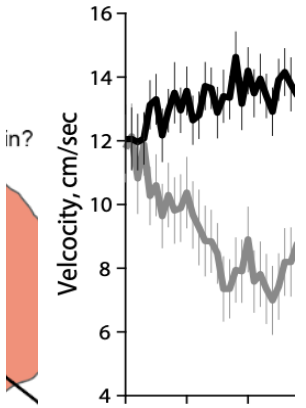
DA Waves at Reward

Credit assignment



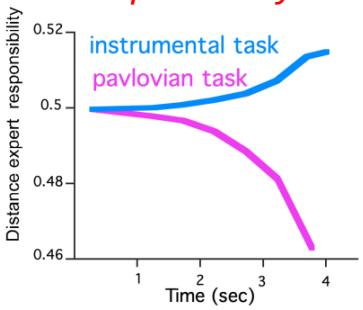


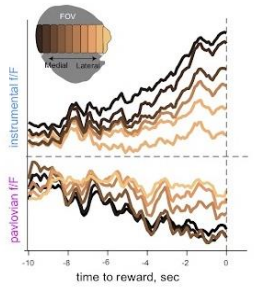


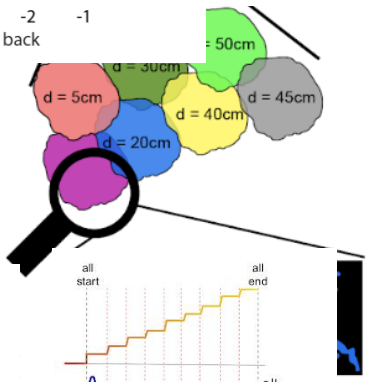


Expert responsibility

Task-dependent DA Ramps

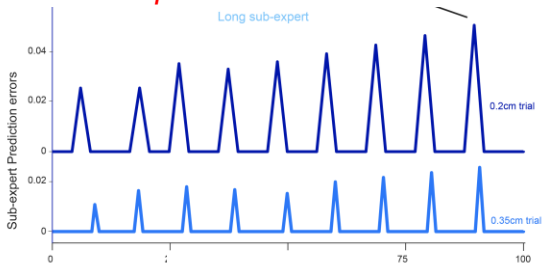




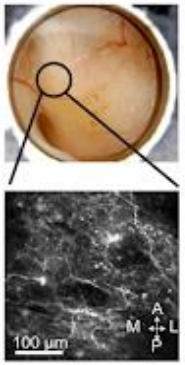


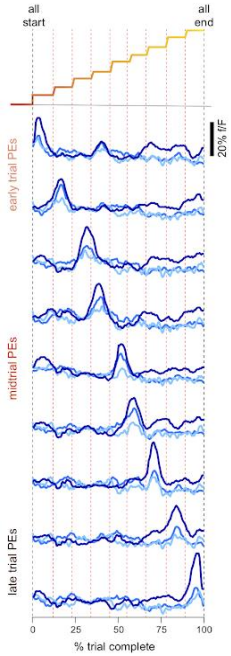
Sub-expert task RPEs

DA RPE's (at tones)



% complete





[mal]Adaptive tuning of striatal dopamine signals

- *Across* striatal regions

  - to credit circuits responsible for task reward

- *Dopamine waves directed to striatal regions depending on task demands!*
- *Implications for psychopathology (e.g., schizophrenia; Maia & Frank 2017)*

- *Within* striatal regions: opponency

  - to optimize value-based choice

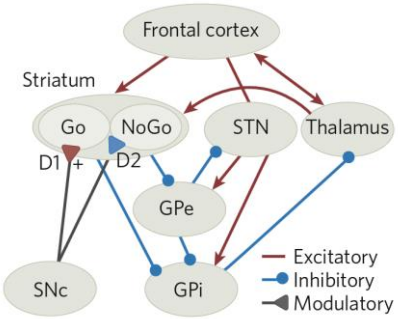
[mal]Adaptive tuning of striatal dopamine signals

- *Across* striatal regions

  - to credit circuits responsible for task reward

- *Within* striatal regions: <u>opponency</u>

  - to optimize value-based choice



- psychopathology as a byproduct: PD, gambling, ADHD, antipsychotics

Dopamine and Opponent Pathways

Distinct roles for direct and indirect pathway striatal neurons in reinforcement

Alexxai V Kravitz[1,4], Lynne D Tye[1,2,4] & Anatol C Kreitzer[1–3]

Distinct Roles of Synaptic Transmission in Direct and Indirect Striatal Pathways to Reward and Aversive Behavior

Takatoshi Hikida,[1,2] Kensuke Kimura,[1,3] Norio Wada,[1] Kazuo Funabiki,[1] and Shigetada Nakanishi[1,*]

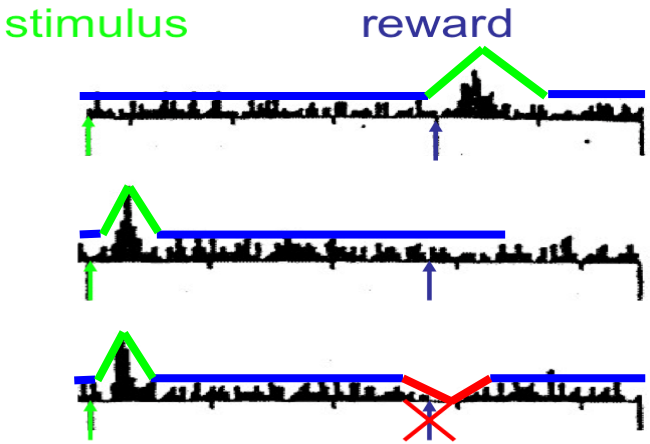
mbling...

idance learning

progression

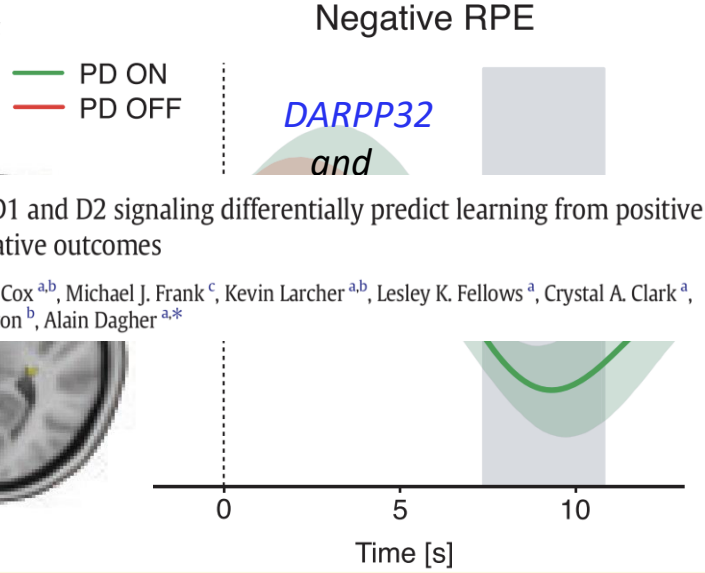
Frank et al 04-07

D1/D2 effects on choice **and** learning; Frank 04-05



Individual differences

PET imaging genetics



Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes

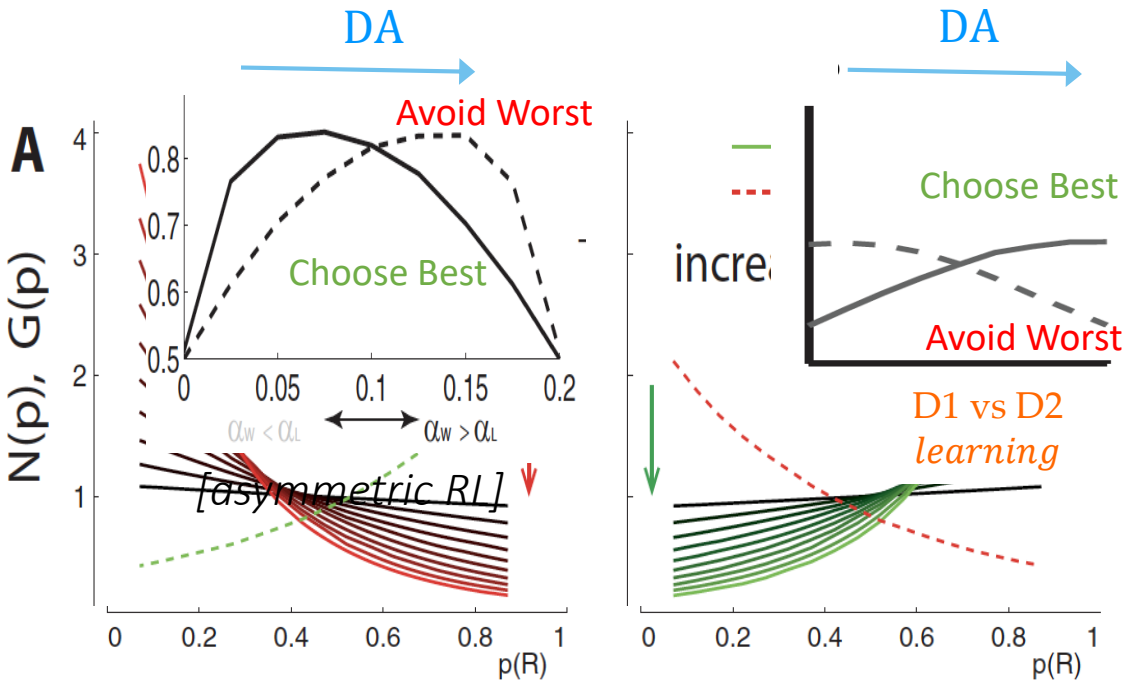
Sylvia M.L. Cox [a,b], Michael J. Frank [c], Kevin Larcher [a,b], Lesley K. Fellows [a], Crystal A. Clark [a], Marco Leyton [b], Alain Dagher [a,*]

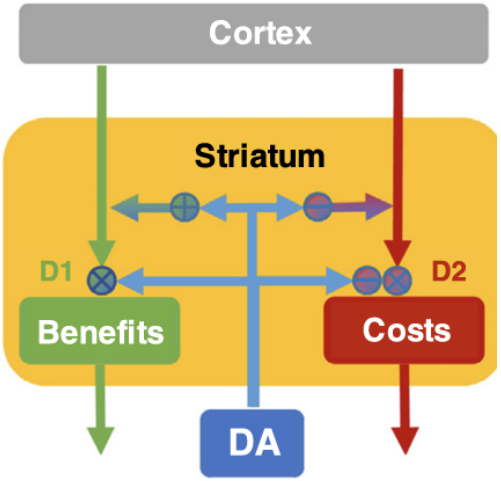
Schultz et al 1997; Montague et al 1996; Frank et al., 2004; 2007; 2009; Collins & Frank 2014; Cox et al 2015
Cools et al 2006; Bronach et al 2019; Pessiglione et al 2006; Palminteri et al 2009, Doll et al 2011; Cockburn et al 2014

Anne Collins

Opponent Actor Learning (OpAL):

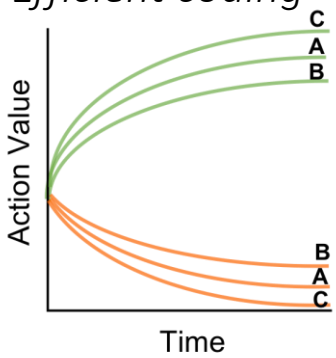
D1 vs. D2 experts





D1 / D2 actors for "Efficient coding"

[without Hebbian plasticity]



Actor Learning

$$G_a(t+1) = G_a(t) + \alpha_G G_a(t) \times \delta(t)$$
$$N_a(t+1) = N_a(t) + \alpha_N N_a(t) \times -\delta(t)$$

Critic generates RPE

$$\delta(t) = R - V(t)$$

[3-factor Hebbian plasticity]

Collins, A.G.E. & Frank, M.J. 2014

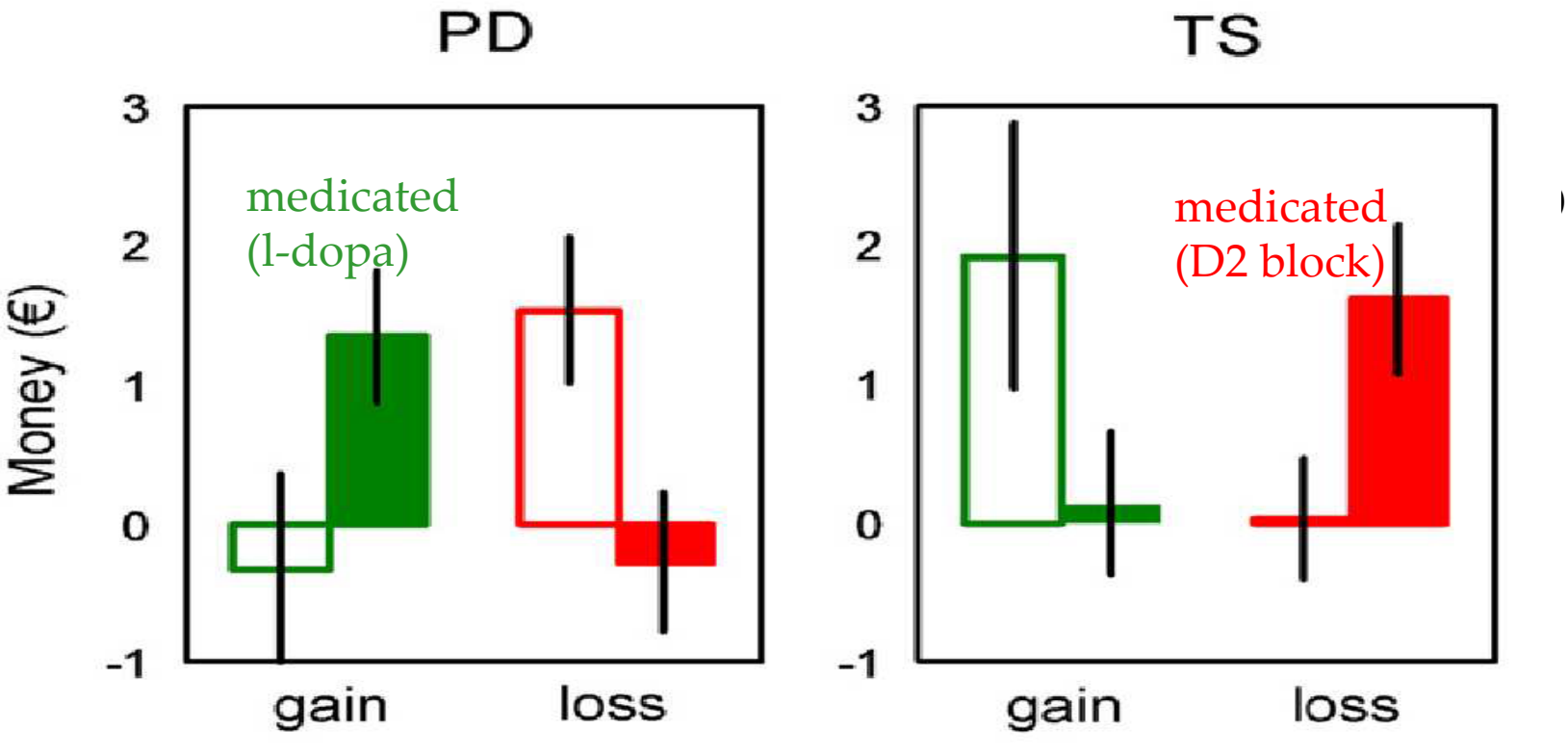
Dopamine-RL and psychopathology

REVIEW COMPUTATION AND SYSTEMS

nature neuroscience

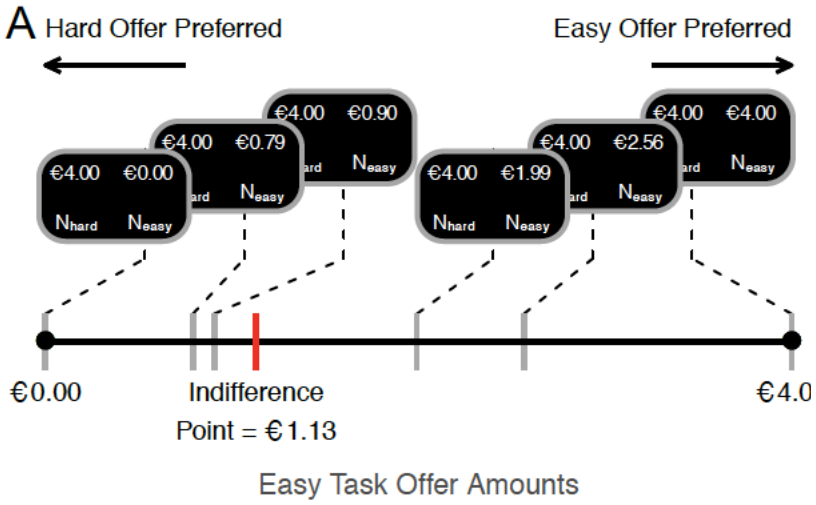
From reinforcement learning models to psychiatric and neurological disorders

Tiago V Maia[1,2] & Michael J Frank[3,4]



Palminteri et al 2009

Dopamine and the cost/benefit of cognitive effort: "smart drugs" and ADHD



Everyone wants the most they can possibly get
For the least they can possibly do
-Todd Snider, "Easy Money"

Bolkan et al 2022 Nature Neuroscience

Westbrook et al. (2020) *Science*

[mal]Adaptive tuning of striatal dopamine signals

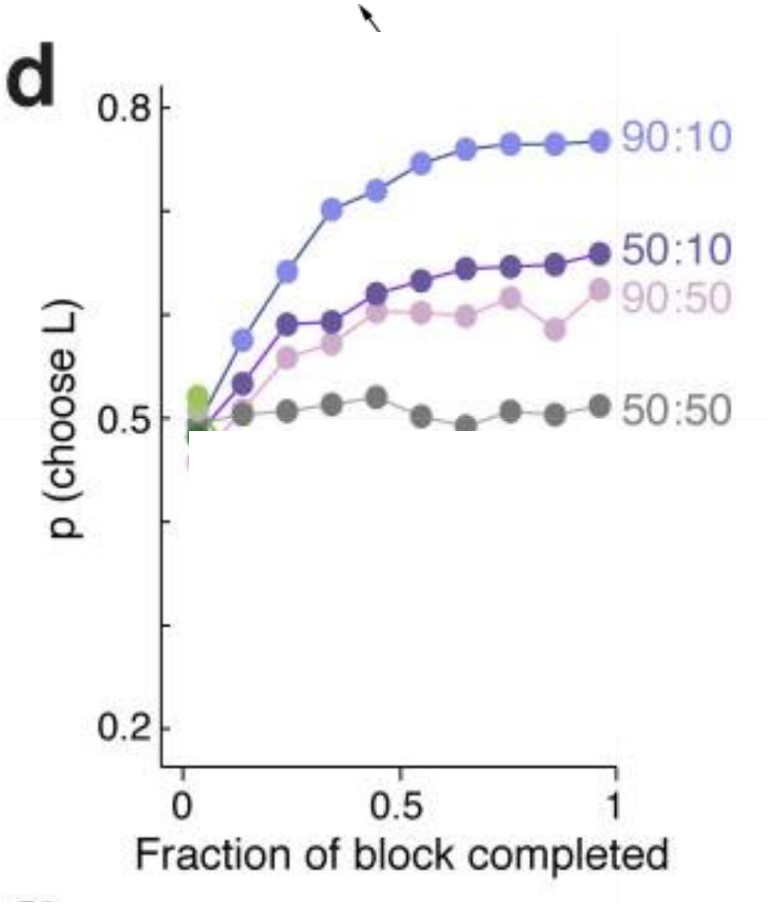
- *Across* striatal regions

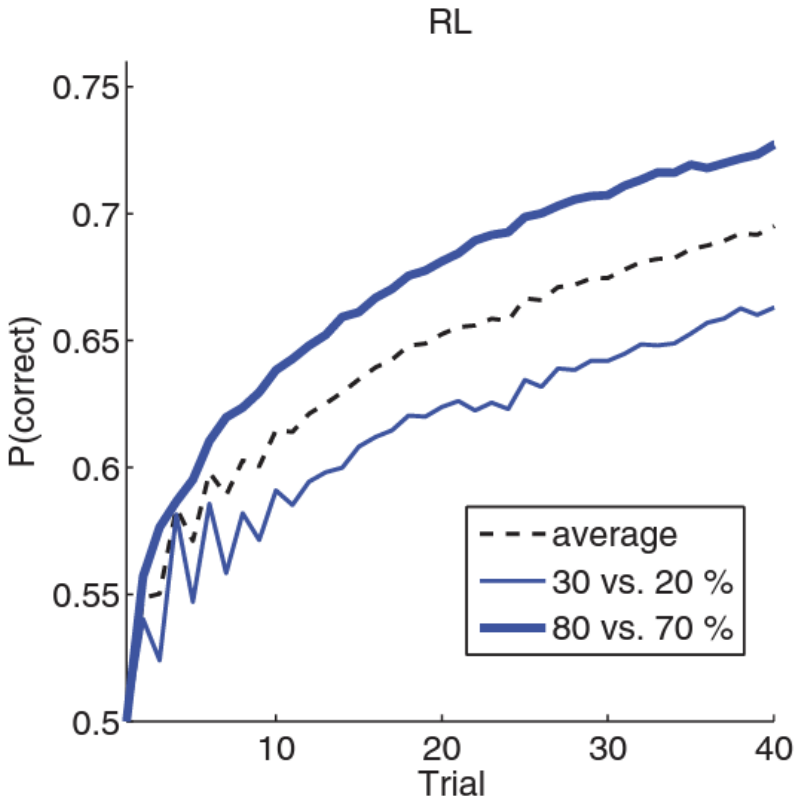
  - to credit circuits responsible for task reward

- *Within* striatal regions: opponency *(why?)*

  - to optimize value-based choice!

Opponency: What for?

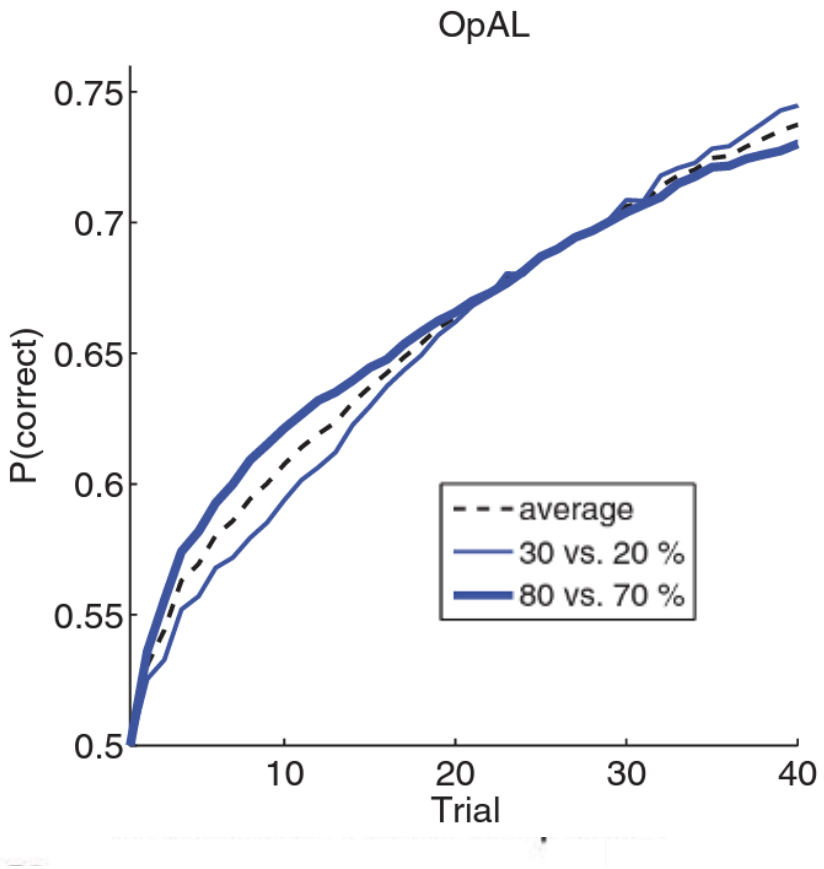




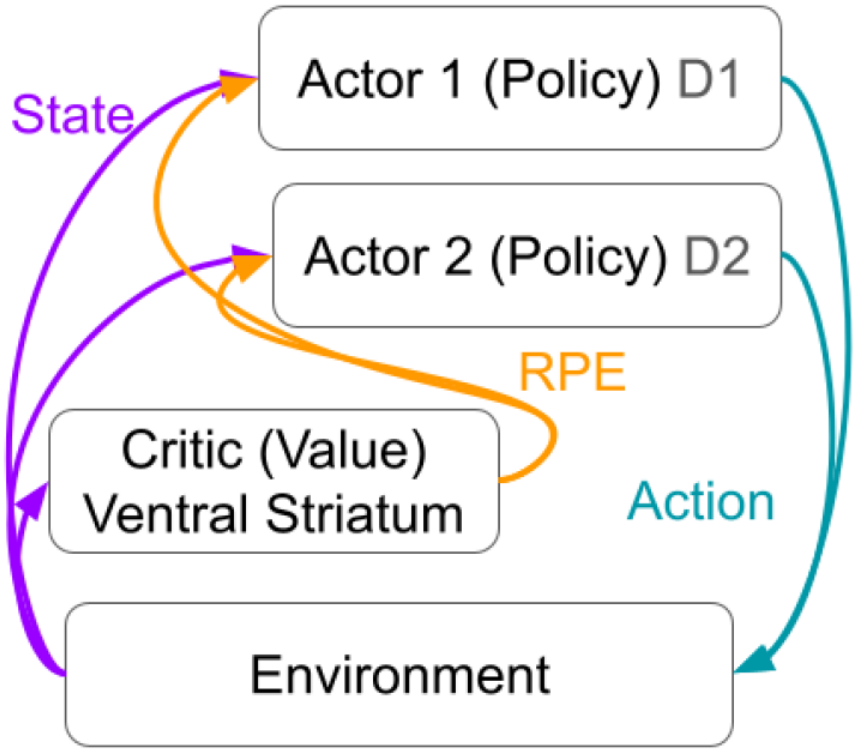
Rodent RL data: no such lean deficit!

Hamid et al 2016

Opponency: What for? specialized policies in D1 vs D2 actors



This is for fixed and balanced DA levels: can we do better?



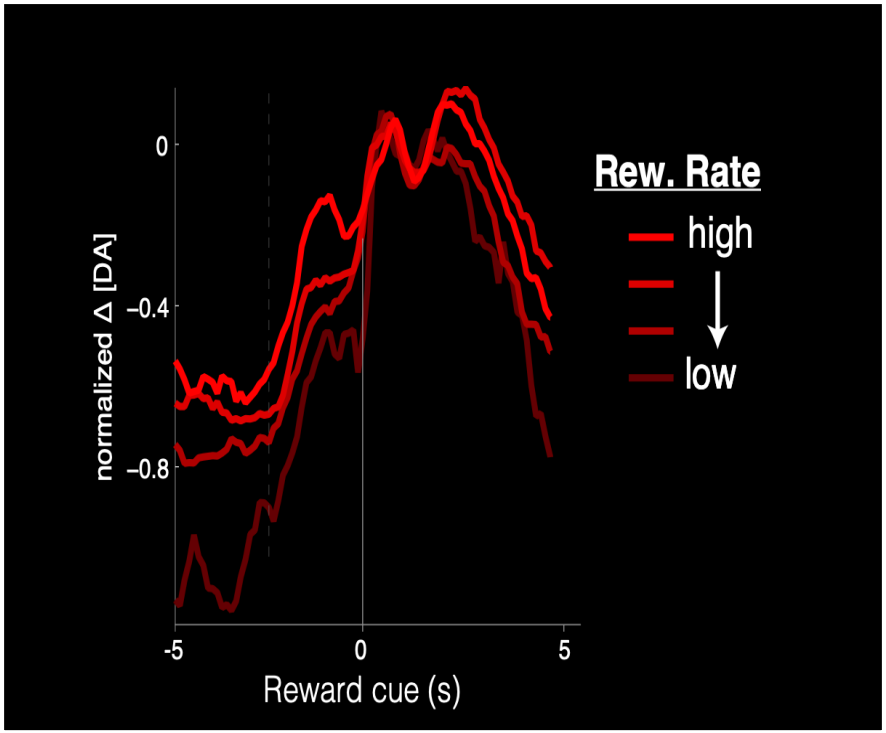
Opponent Actor Learning (OpAL) Model
Collins & Frank 2014

Alana Jaskir

Dynamic DA to leverage opponency: OpAL*

Reward Rich Environment

Reward Lean Environment



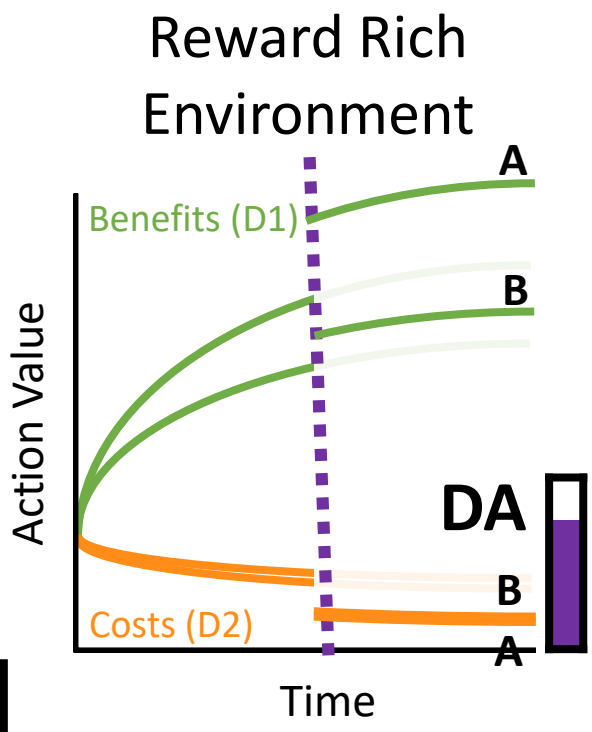
Background [DA] grows with reward history

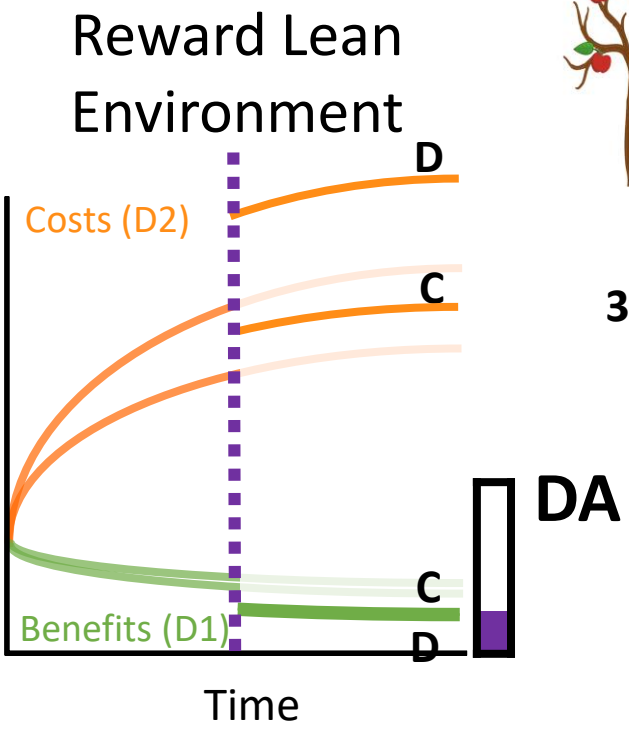
Hamid et al 2015

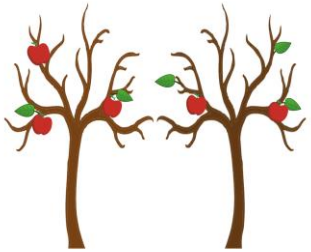
➔ *Use ongoing estimate of reward history to modulate DA*

A > B

80% 70%

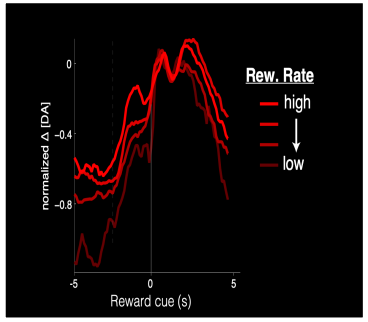






C > D

30% 20%

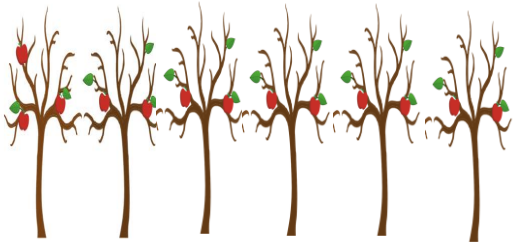


Jaskir, A. & Frank, M.J. 2022 *bioRxiv*

Striatal opponency supports adaptive RL across environments

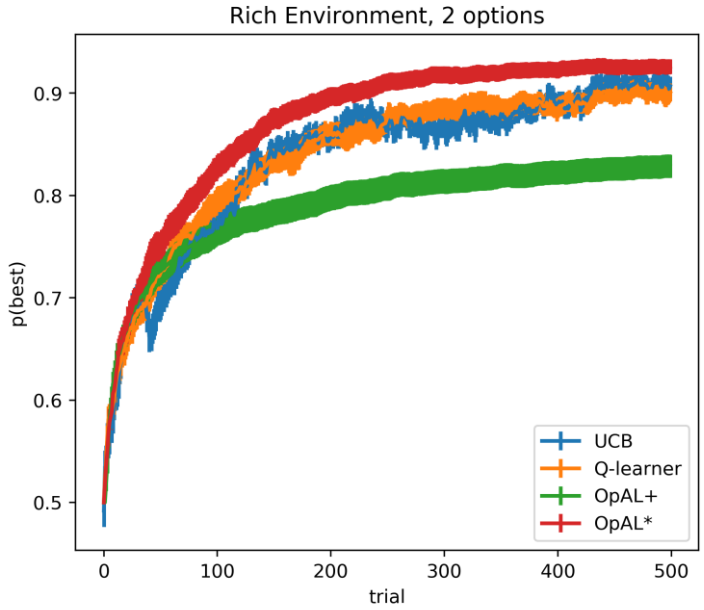
A B

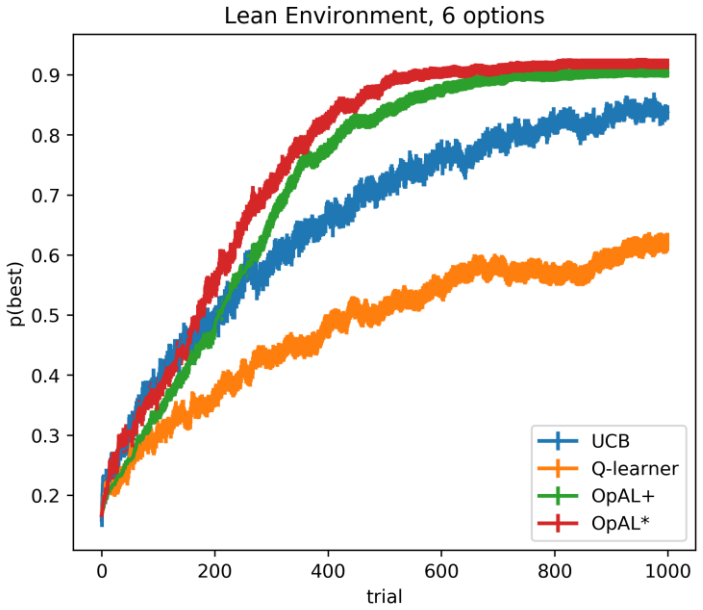
80% 70%



C D E F G H

30% 20 20 20 20 20%



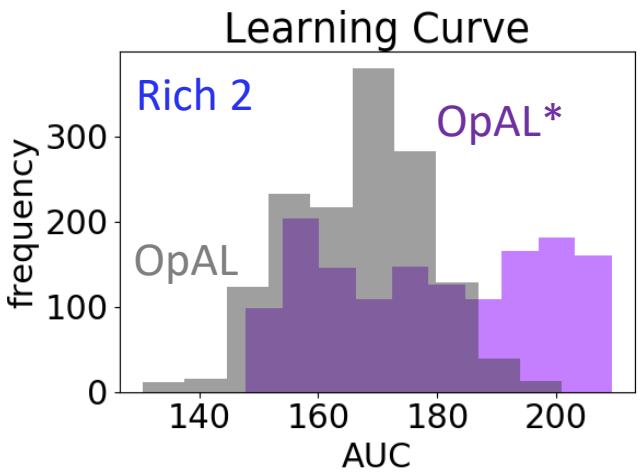


Each model is given its "best possible shot" (parameters optimized)

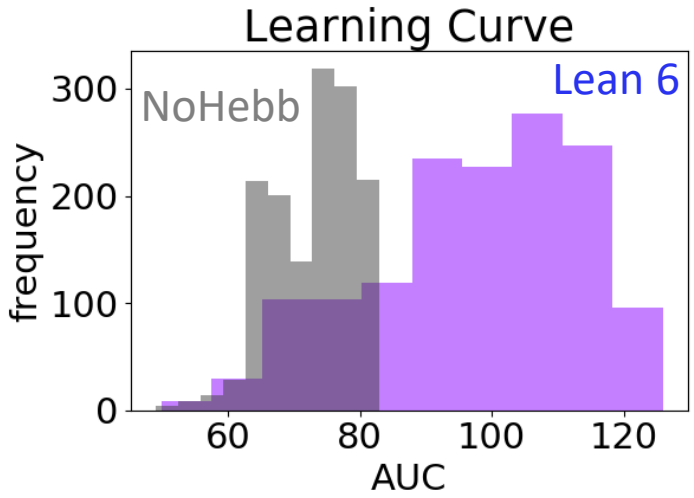
Jaskir, A. & Frank, M.J. 2022 *bioRxiv, in revision*

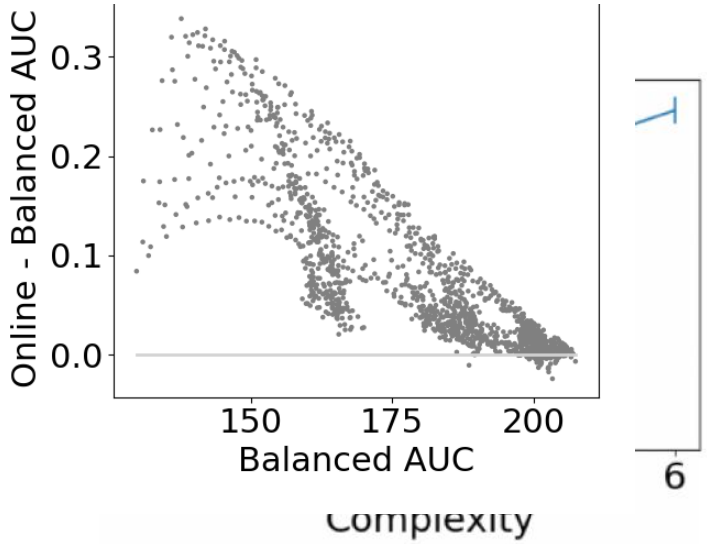
Robust advantages across parameters

Dynamic DA contributions



Hebbian contributions





OpAL advantages require:*
dynamic DA and Hebbian opponency

OpAL also improves adaptive risk-taking*

Advantages grow with complexity and sparse reward

Jaskir, A. & Frank, M.J. 2022 *bioRxiv, in revision*

Summary:

DA dynamics optimize learning and choice

- **DA dynamics** *across* striatum
  - **Transients**: reward prediction errors tailored to local "expert"
  - **Waves**: support RL credit assignment to subregions
  - '**Tone':** reward history

- **Opponency *within*** striatal subregion
  - Normative advantage, especially in:
    sparse reward environments with multiple choice options
  - adaptive risk-taking

DA effects in psychopathology as a by-product

Many open questions at circuit and computation level….

Howard Hughes Medical Institute

National Institute of Mental Health

Thanks!

CARNEY INSTITUTE FOR BRAIN SCIENCE
BROWN UNIVERSITY

Alana Jaskir

Arif Hamid

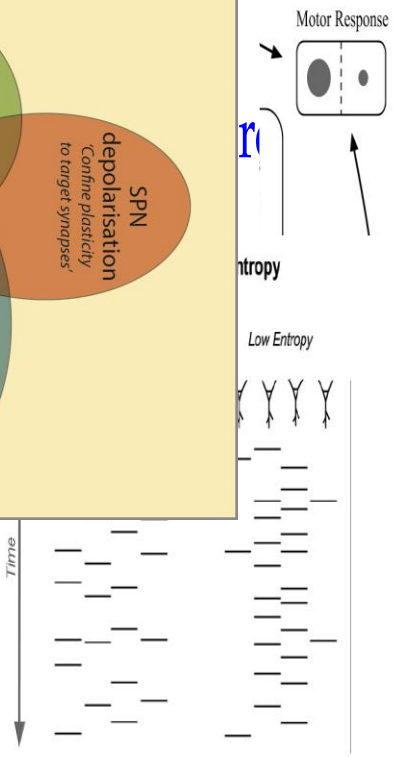
Andrew Westbrook

Roshan Cools
Chris Moore
Ines Belghiti
Aneri Soni
David Badre

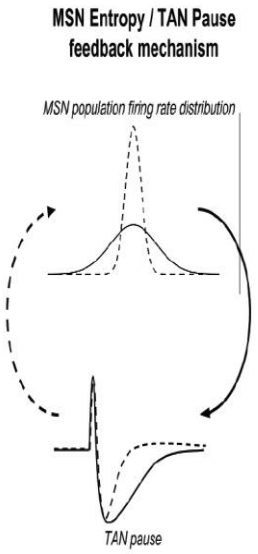
Anne Collins (Berkeley)

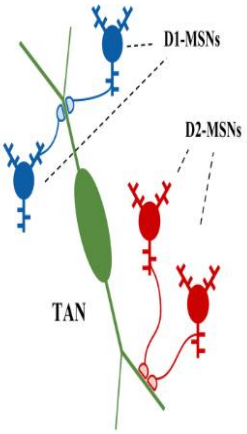
Lab for Neural Computation & Cognition

Reynolds et al., 2022
Nature Communications



ve learning rates

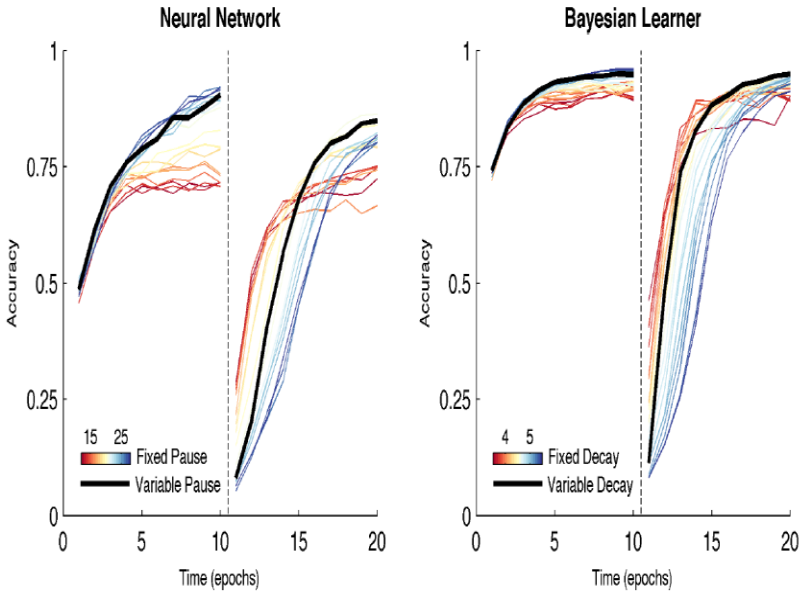




MSN-TAN collaterals: Bolam et al '86; Chuhma et al 11; Gonzalez et al 13

MSN entropy → longer TAN pauses

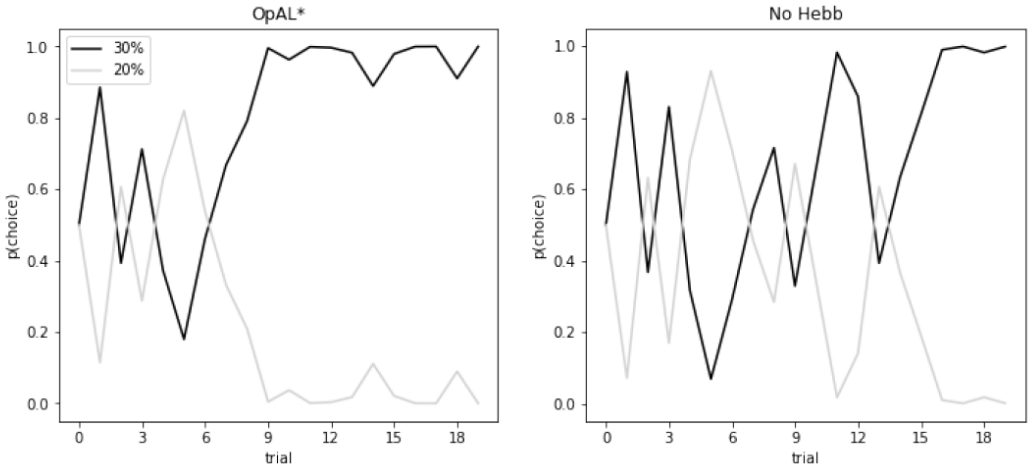
BG-TAN net is analogous to Bayesian learner

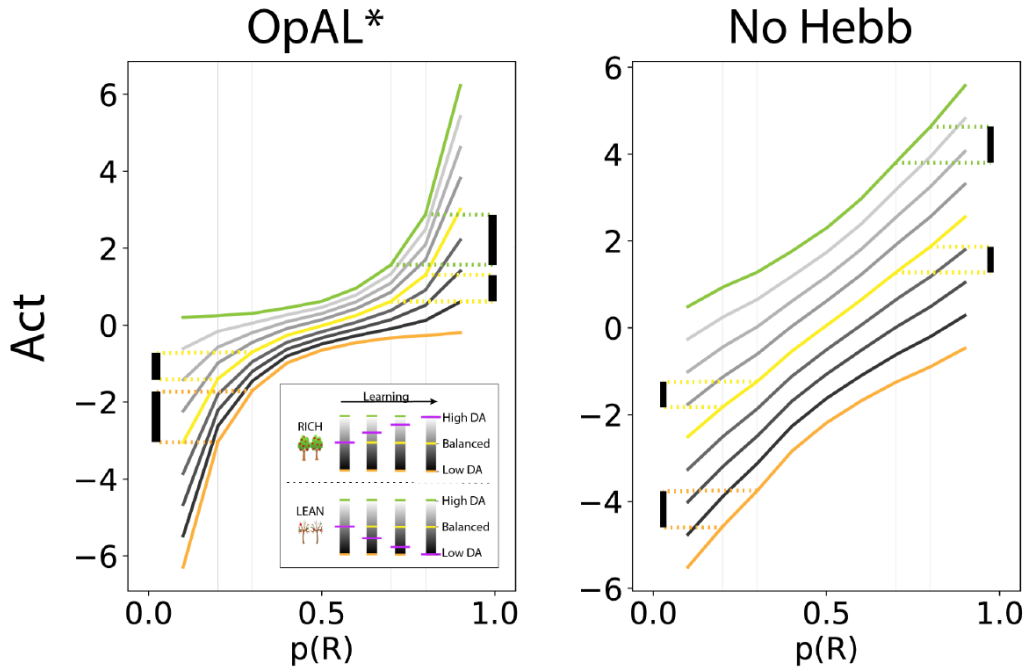


TAN feedback mechanism promotes robust learning when uncertain but prevents overlearning thereafter

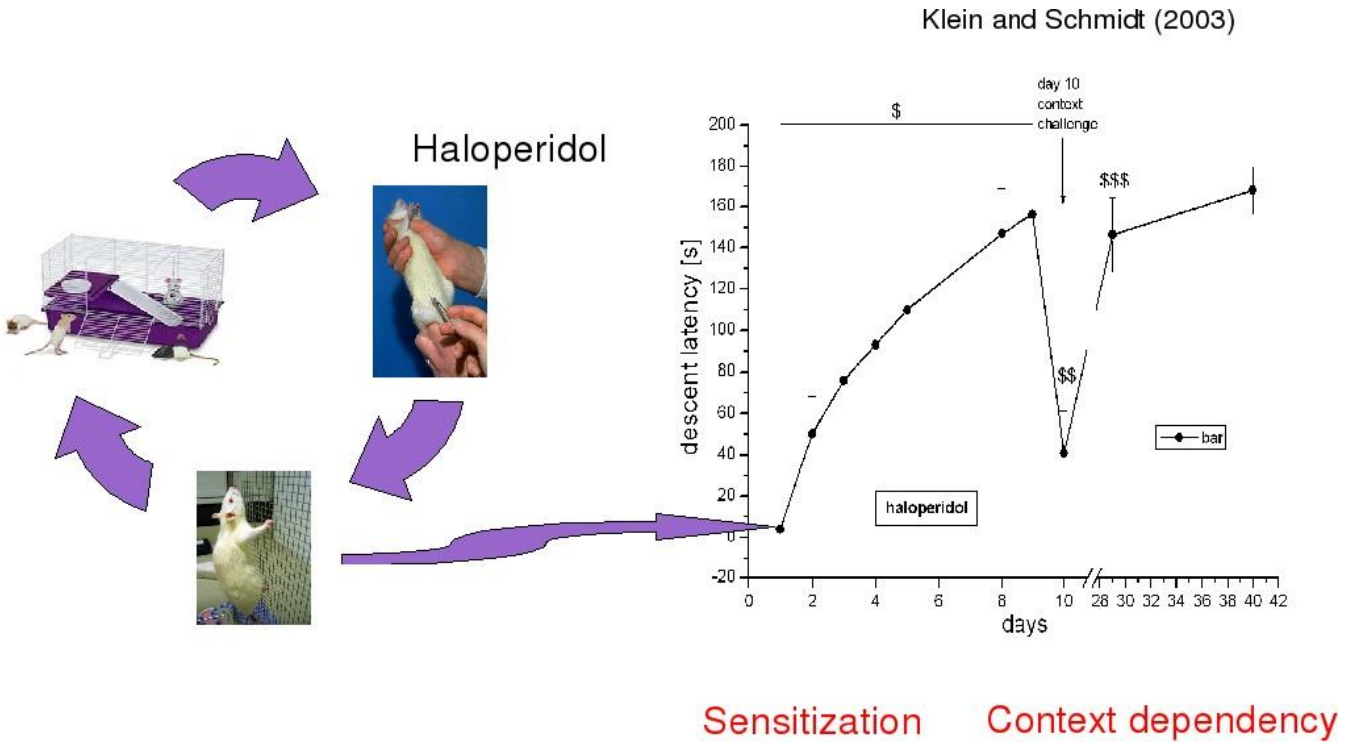
Franklin & Frank, 2015, *eLife*

OpAL* Mechanism



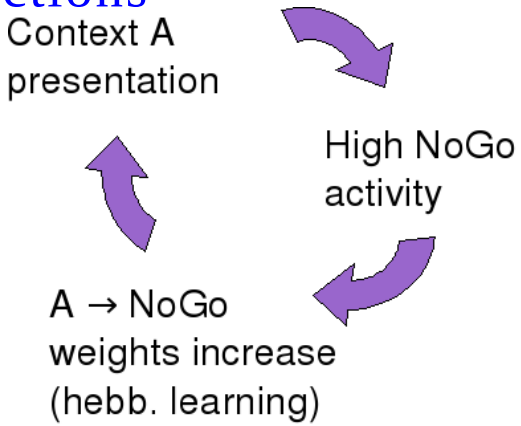


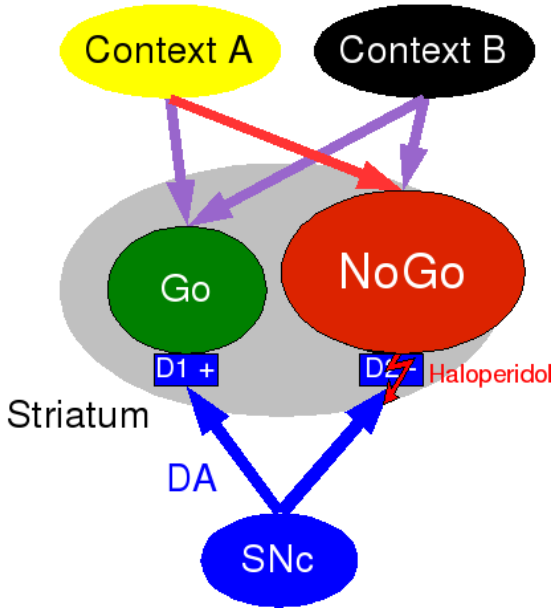
Can PD be learned?
Catalepsy sensitization by DA depletion or haloperidol



A case of exaggerated NoGo learning??

Dopamine depletion: performance ↔ learning interactions

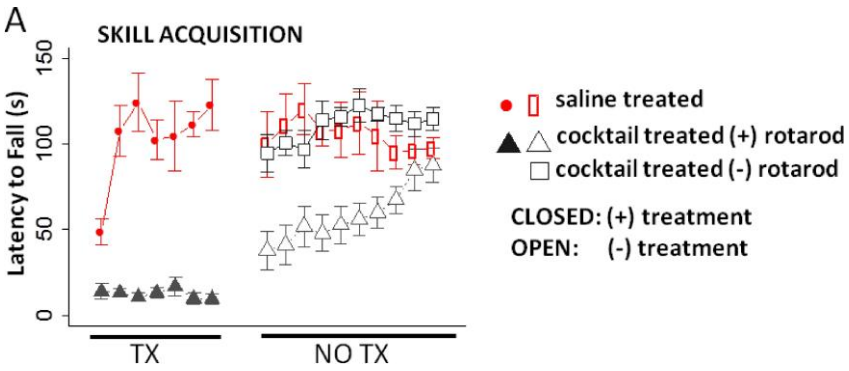




cf context-dependent catalepsy sensitization; Wiecki et al09

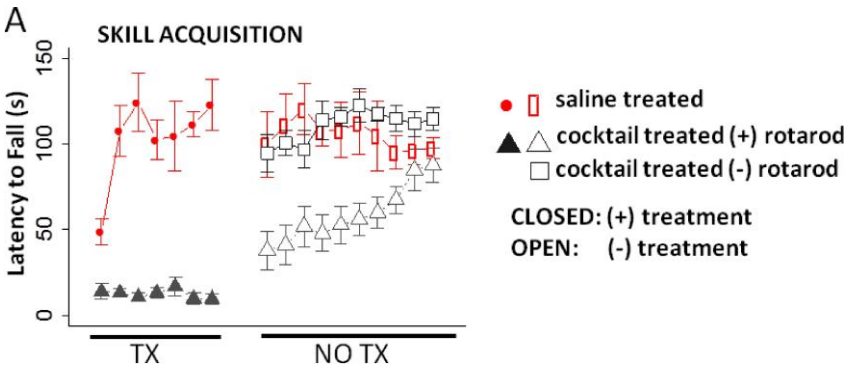
Striatal-dependent motor task: Accelerating Rotarod

Data

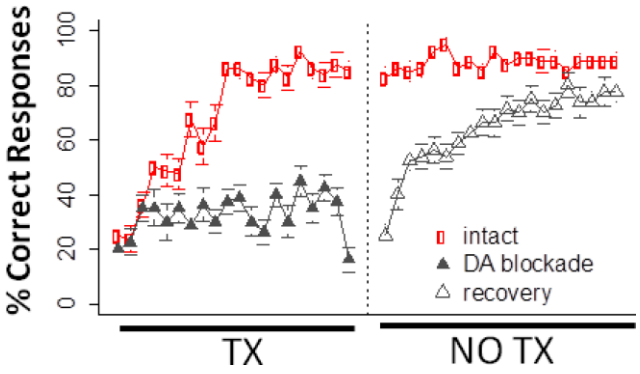


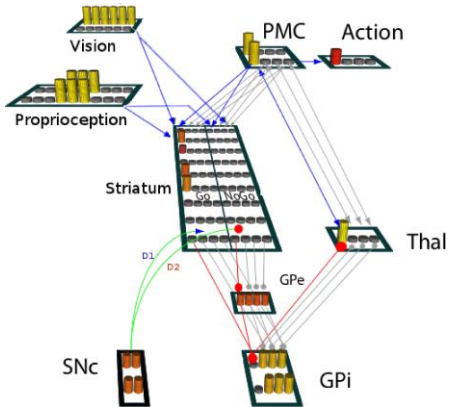
Striatal-dependent motor task: Accelerating Rotarod

Data



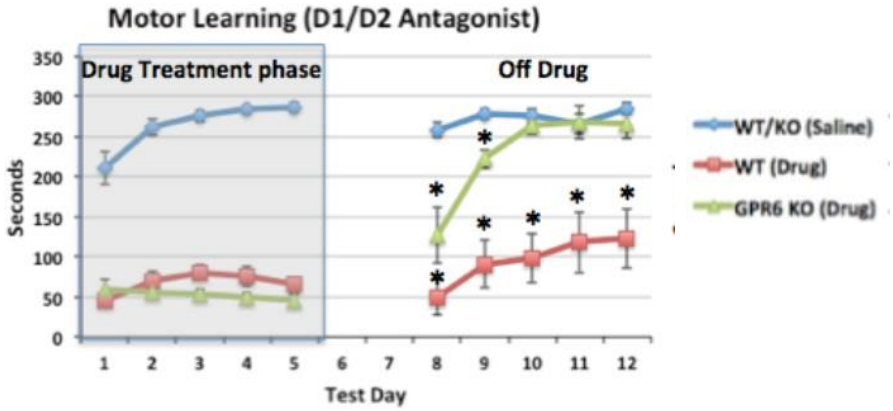
Model





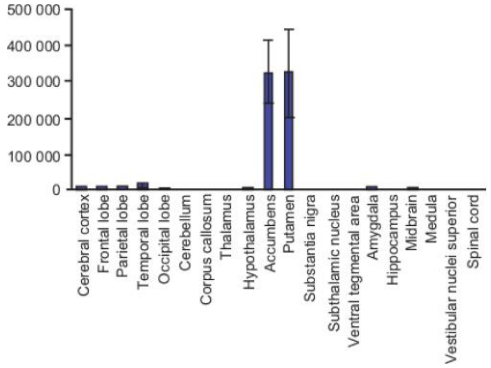
Beeler et al, 2012

GPR6 knockout protects against aberrant D2 learning

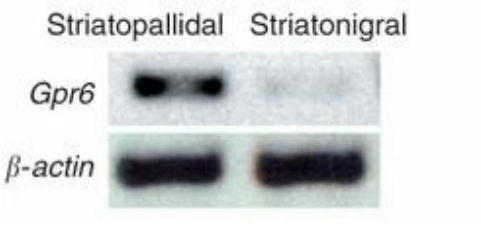


preliminary data, with Kevin Bath, Anuj Patel

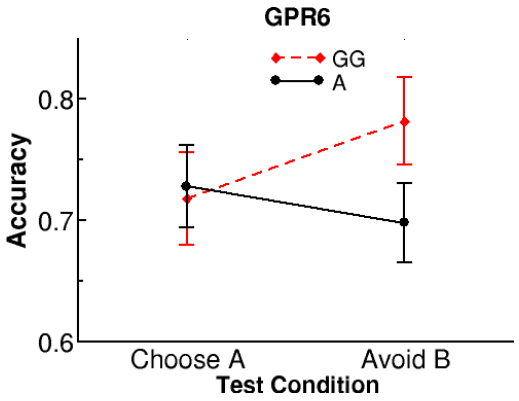
GPR6: highly specific to indirect (NoGo) pathway



postmortem human (Roth et al 06)

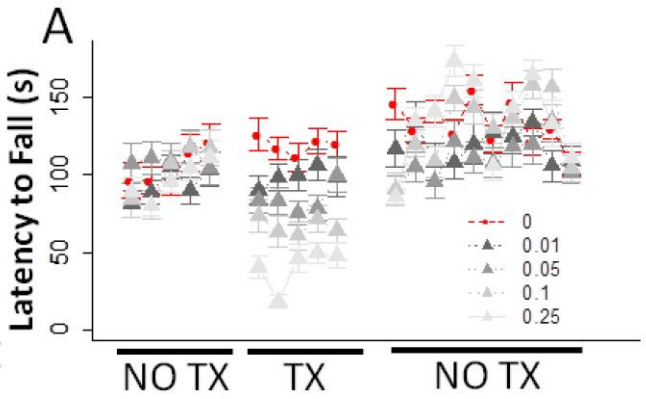


mouse (Lobo et al07)

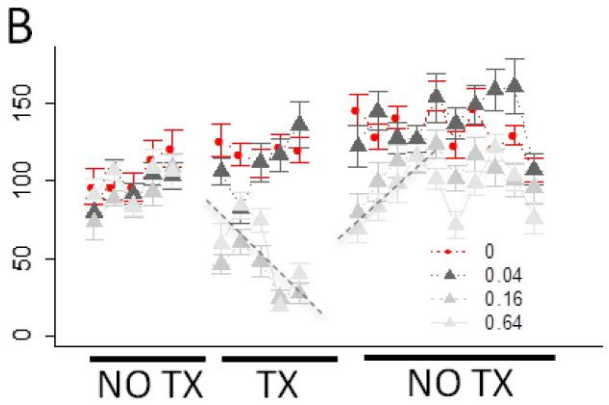


Selective D1 / D2 blockade during established skill

D1 block

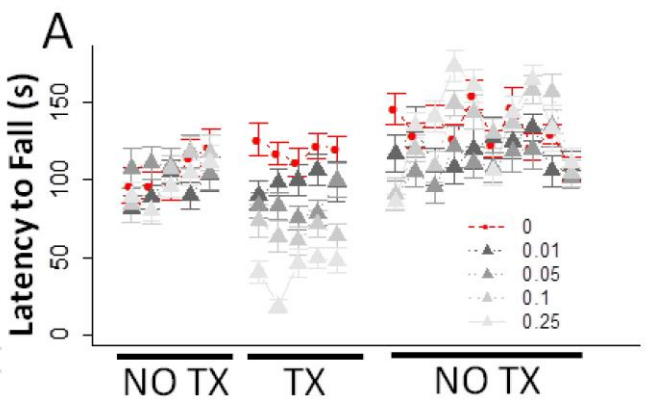


D2 block

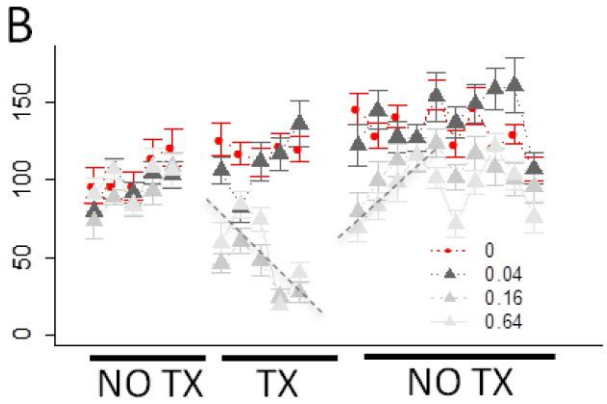


Selective D1 / D2 blockade during established skill

D1 block



D2 block



Model