# Reinforcement Learning

*Woo-Young (Young) Ahn*
*Computational Clinical Science Lab*
*Department of Psychology*
*Seoul National University*
*[ccs-lab.github.io](ccs-lab.github.io)*

# *Reinforcement Learning (RL)*

- *What is RL?*
  - *Definitions & terminology*
  - *RL in human research vs RL in AI*

- *RL models <u>(algorithms for prediction and control)</u>*
  - *Classical conditioning*
    - *Rescorla-Wagner (R-W) model*
    - *(Bayesian or non-Bayesian) extension of R-W models*
  - *Operant (instrumental) conditioning*
    - *Model-free vs Model-based learning*
    - *Pavlovian control vs Instrumental control*
- *Adaptive Design Optimization within the RL framework*
- *Naturalistic RL*
- *Limitations & Future directions*

# *Learning objectives*

*Participants will…*

- *Understand the key concepts and notations of RL (in multiple fields)*
  - *"RL is everywhere"*

- *Know (some of) popular RL models (& references)*
  - *Simple to complex models*

- *Limitations of (current) RL in CP and some new approaches*
  - *Naturalistic paradigms for RL, new (AI) algorithms in human RL research*

*What is RL?*

*Why RL?*

# *What is RL?*

*"Learning what to do" …*
*based on (an incomplete history of)*
*rewards and punishments*

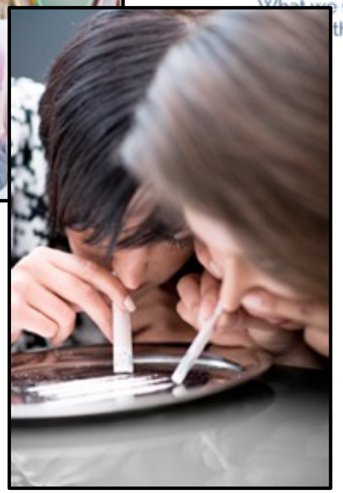
*Sutton & Barto (1998) Reinforcement Learning*
*Dayan & Labott (2000) Theoretical neuroscience*

*"Learn optimal ways to make decisions"*
*in an uncertain environment*

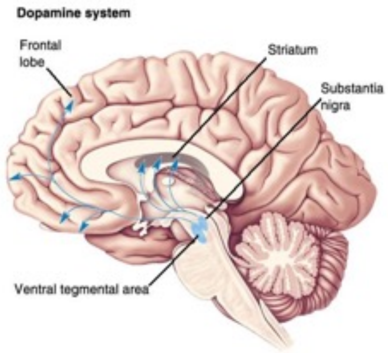
# *Why RL in CP?*

## *RL is everywhere & in multiple fields*





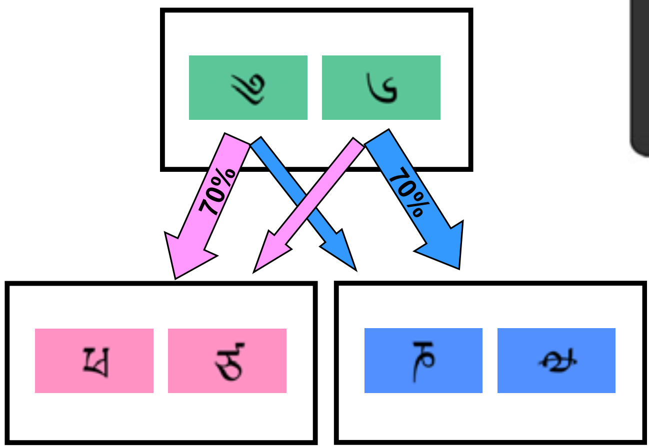
# *How do we study RL in humans?*




***"Classical laboratory tasks"***

CS+ CS- CS Neut

*e.g., O'Doherty et al (2003) Neuron*



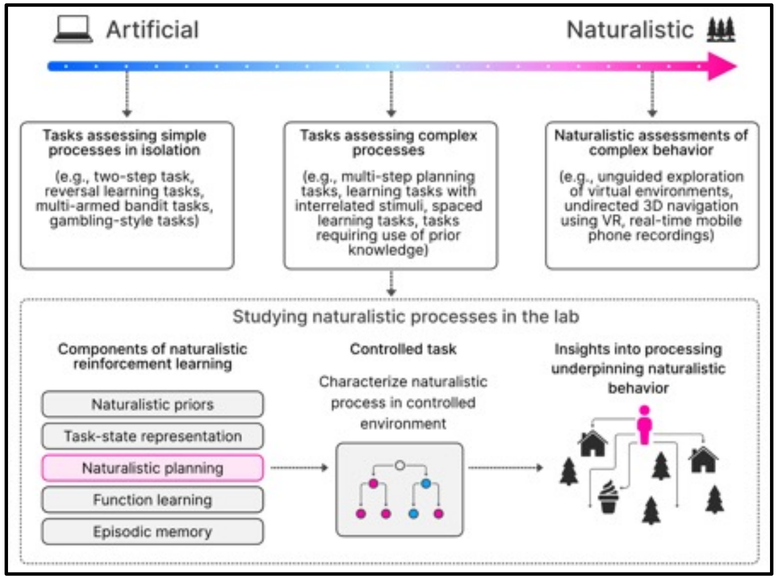

*Daw et al (2011) Neuron*

You won \$0.50, but lost \$0.75

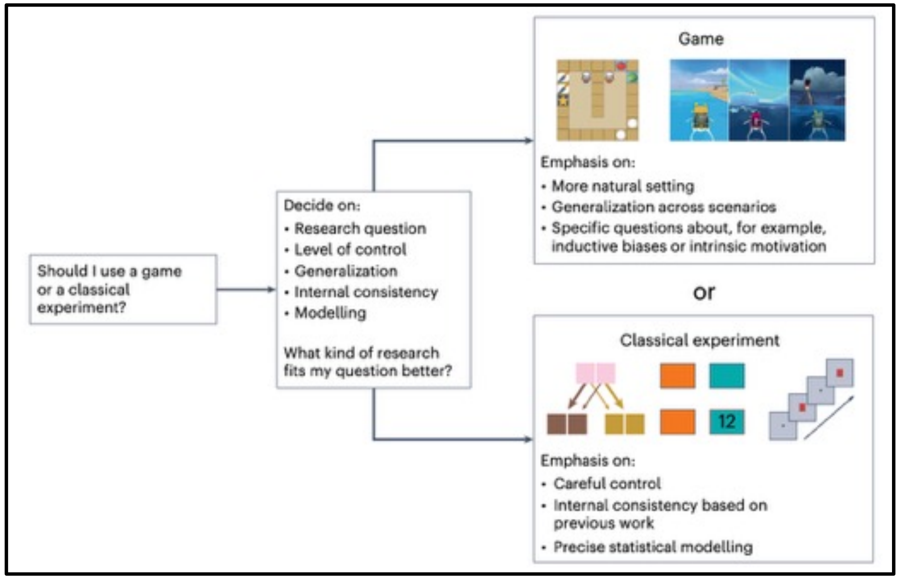

*e.g., Bechara et al (1994) Cognition*

# *How do we study RL in humans?*

## ***"Naturalistic RL tasks"***




*Wise et al (2023) Trends in Cog Sci*




*Allen et al (2024) Nature Human Behaviour*

***RL is a type of Machine Learning***

- *Supervised Learning*
- *Unsupervised Learning*
- *Reinforcement Learning*

***Q) How is RL different from other ML paradigms?***

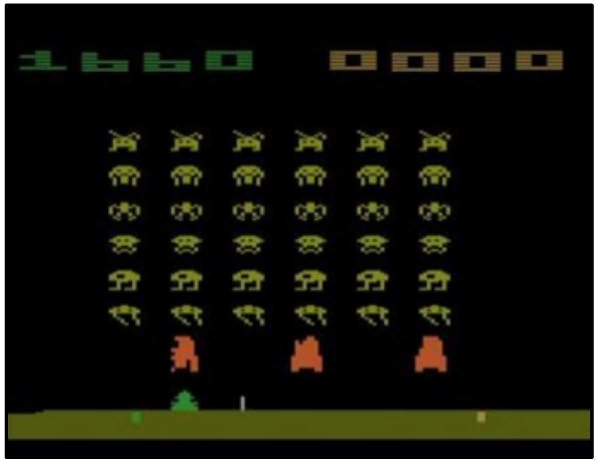
- *No external supervisor ("minimally supervised")*
- *Data is not i.i.d.! Reward signals (learn from trials and errors)*
- *Interaction with environment*
- *Goal: "all about data" vs "all about optimization"*

# What to solve with RL?

- Maximize (cumulative) reward is not the only problem
- Other topics
  - Learning reward function from examples (inverse reinforcement learning)
  - Optimize environment to maximize information gain about agent (adaptive design)
  - Transferring knowledge between domains (transfer learning, meta-learning,


Anguera et al (2013) Nature


Silver et al (2016) Nature



Mnih et al (2015) Nature

# *Machine learning / DNN in Decision Neuroscience*

Neuron

CellPress

Review

**Deep Reinforcement Learning and Its Neuroscientific Implications**

Matthew Botvinick,[1,2,*] Jane X. Wang,[1] Will Dabney,[1] Kevin J.
[1]DeepMind, London, UK
[2]University College London, London, UK
*Correspondence: botvinick@google.com
https://doi.org/10.1016/j.neuron.2020.06.014

Neuron

Article

**Using deep reinforcement learning to reveal how the brain encodes abstract state-space representations in high-dimensional environments**

Molecular Psychiatry (2019) 24:1583–1598
https://doi.org/10.1038/s41380-019-0365-9

EXPERT REVIEW

Check for updates

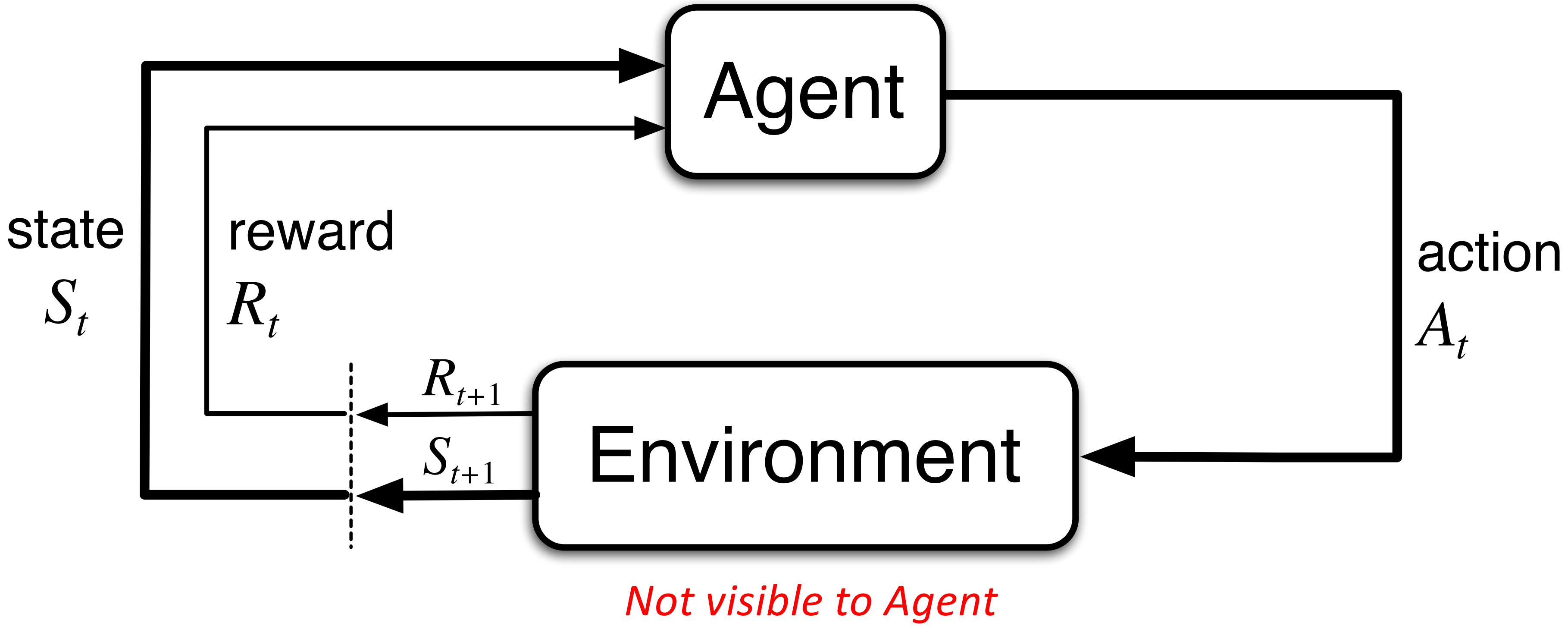
**Deep neural networks in psychiatry**

Daniel Durstewitz[1] · Georgia Koppe[1,2] · Andreas Meyer-Lindenberg[2]

*“Learn optimal ways to make decisions” in an uncertain environment*

| | *RL in human research* | *RL in AI* |
|---|---|---|
| *Goal* | *Mimic RL in real world & characterize individual differences* | *Generate optimal solution* |
| *Amount of data* | *Small* | *Very large* |
| *# parameters* | *Typically < 10* | *A lot* |
| *Parameter estimation* | *Important* | *Estimate? Often fixed* |

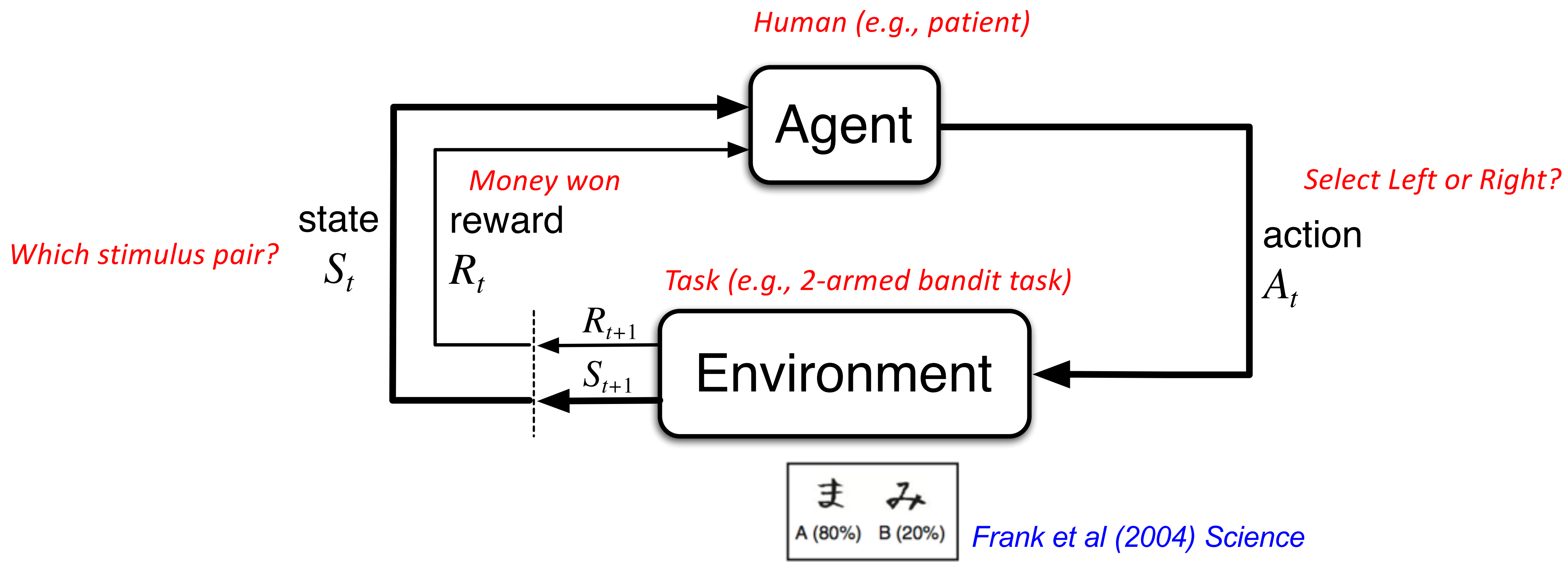
*Agent-Environment Interface*
*(check MDP lecture)*
*e.g., Maze task, Tree search, N-armed Bandit*




*Sutton & Barto (1998) Reinforcement Learning*

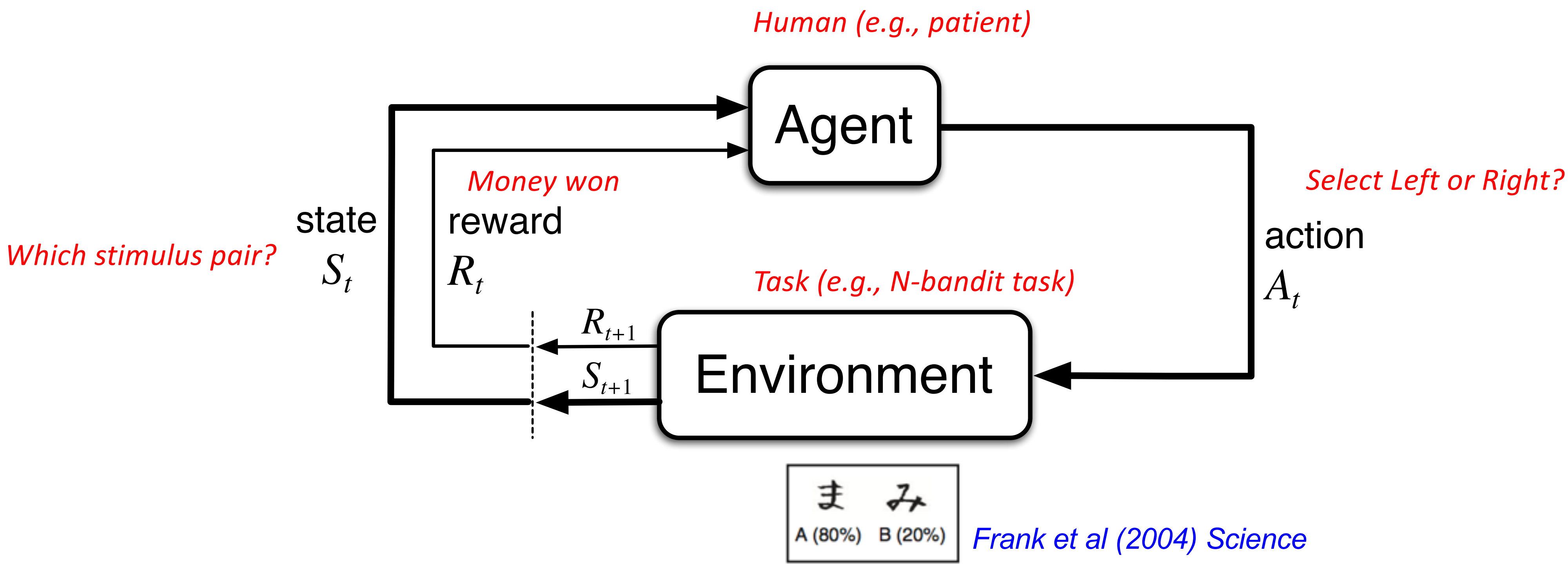
# *Typically in Computational Psychiatric research settings..*

*Model parameters → Psychologically meaningful processes/constructs*




Frank et al (2004) Science

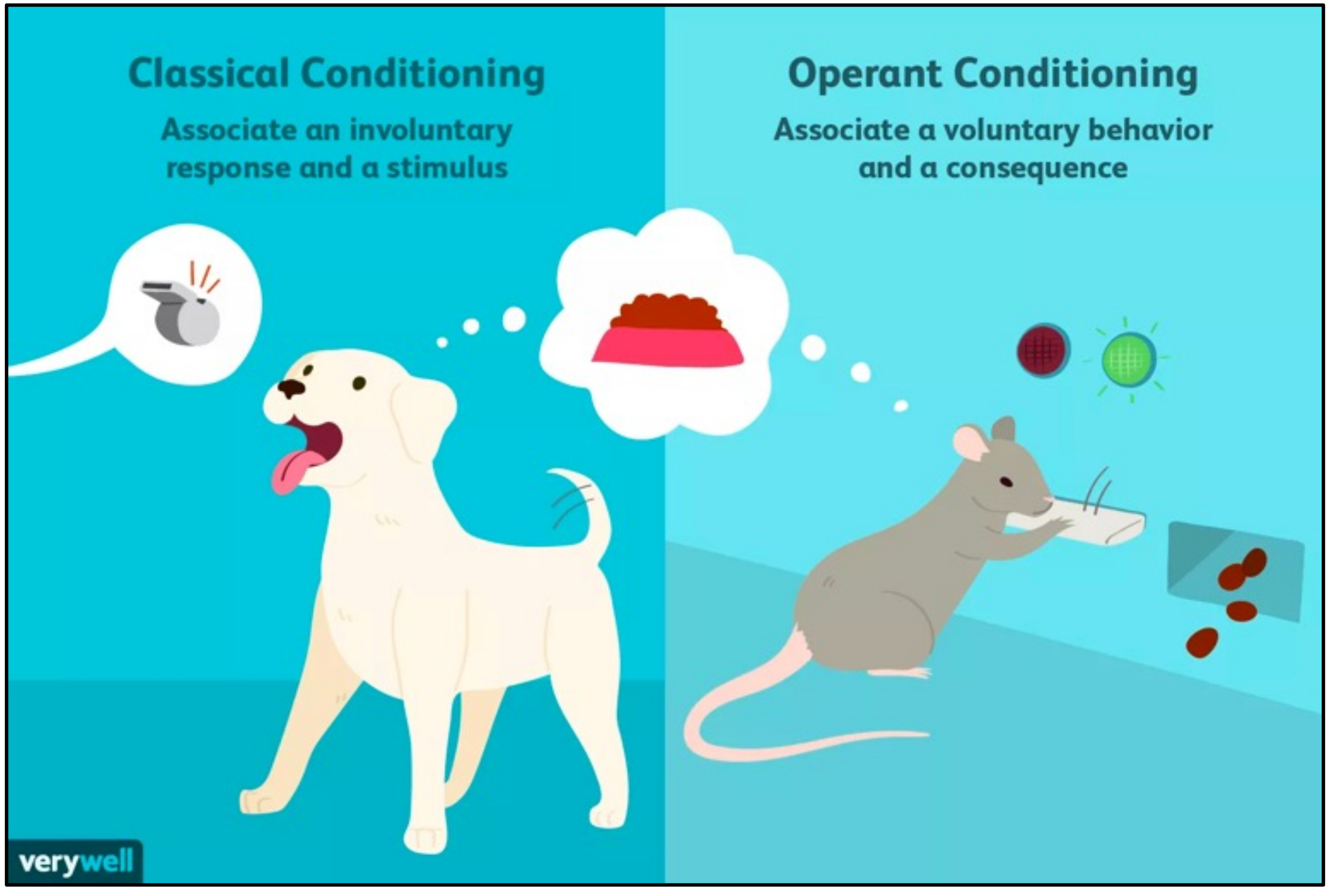
*$S_t$: State value on time (trial) t*
*$A_t$: Action value on time (trial) t*
*$R_t$: Reward on time (trial) t*
$\pi_t(a_t, s_t)$ *: Policy on time (trial) t → mapping from states to actions*




Frank et al (2004) Science

*RL models* *<u>(algorithms for prediction)</u>*

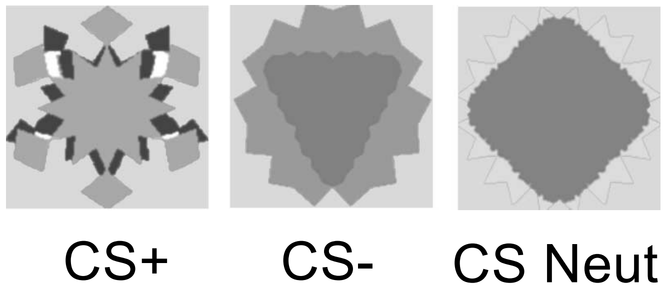
*Two experimental set-ups*
*(Not a distinction of learning mechanisms)*




https://www.verywellmind.com/classical-vs-operant-conditioning-2794861

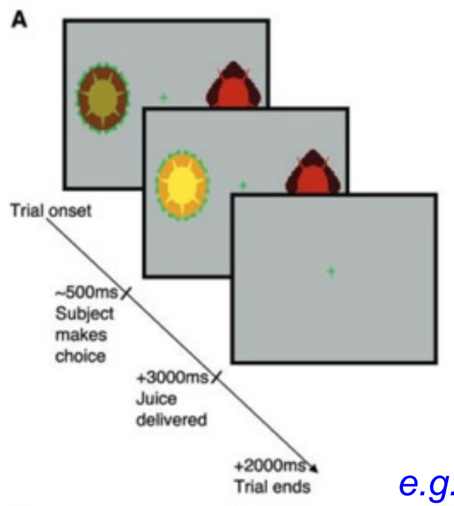
Two experimental set-ups
(Not a distinction of learning mechanisms)

Classical conditioning
(No action required)




e.g., O'Doherty et al (2003) Neuron

Operant (Instrumental)
Conditioning (Action required)




e.g., O'Doherty et al (2004) Science

# *Classical conditioning*

CS+ CS- CS Neut

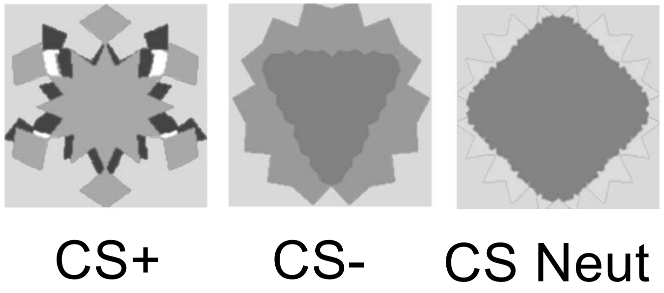
*e.g., O'Doherty et al (2003) Neuron*

*Rescorla-Wagner (R-W) model*

→ *Point estimates of $V_t$*

$$\underbrace{V_t}_{\text{Stimulus value (t)}} = \underbrace{V_{t-1}}_{\text{Stimulus value (t-1)}} + \overbrace{\alpha}^{\text{Learning rate}} \underbrace{(\underbrace{R_t}_{\text{Outcome}} - \underbrace{V_{t-1}}_{\text{Stimulus value (t-1)}})}_{\text{Prediction error}}$$

# *Classical conditioning*



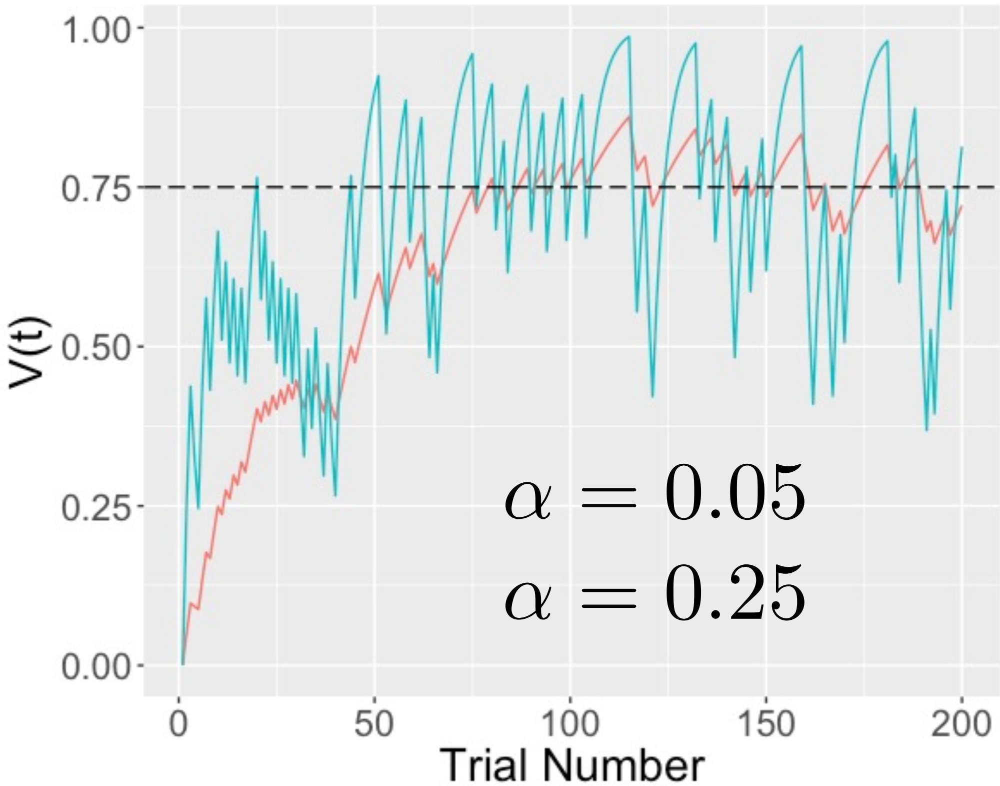

*e.g., O'Doherty et al (2003) Neuron*

** Rescorla-Wagner (R-W) model*
*→ Point estimates of $V_t$*

** Bayesian generalization of R-W*
*→ Kalman filter ➔ HGF*

*Dayan et al (2000); Kakade & Dayan (2002)*
*Daw et al (2006); Kruschke (2008); Mathys et al (2011; 2014)*

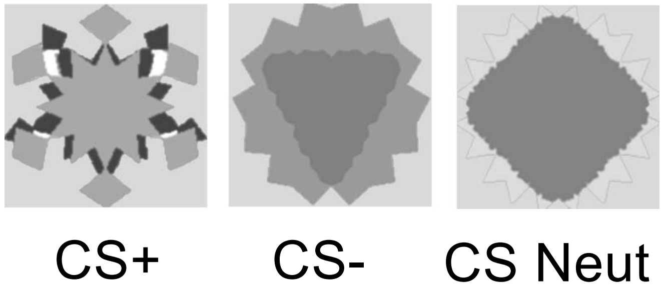
*e.g., Reward rate = 0.75*




http://haines-lab.com/post/2017-04-04-choice_rl_1/

*Also see Maaten Speekenbrink's blogs*
*https://speekenbrink-lab.github.io/blog/*

# *Classical conditioning*




*e.g., O'Doherty et al (2003) Neuron*

*Temporal Difference (TD) Learning model*

→ *Generalization of R-W (real-time model)*
→ *To account for within-trial and between-trial relationships among stimuli*

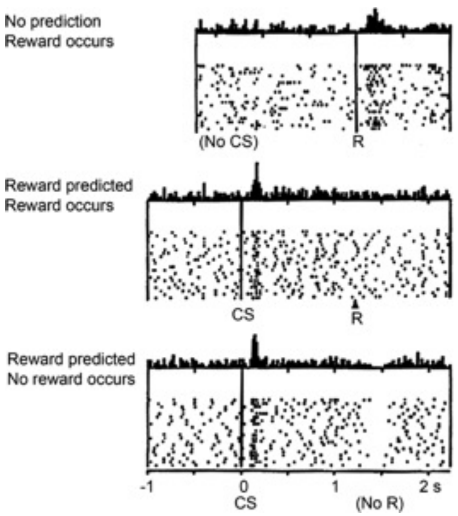
# *Reward Prediction Error*
# *TD learning model*

## Computational roles for dopamine in behavioural control

P. Read Montague[1,2], Steven E. Hyman[3] & Jonathan D. Cohen[4,5]

[1]Department of Neuroscience and [2]Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, 1 Baylor Plaza, Houston, Texas 77030, USA (e-mail: read@bcm.tmc.edu)
[3]Harvard University, Cambridge, Massachusetts 02138, USA (e-mail: seh@harvard.edu)
[4]Department of Psychiatry, University of Pittsburgh and [5]Department of Psychology, Center for the Study of Brain, Mind & Behavior, Green Hall, Princeton University, Princeton, New Jersey 08544, USA (e-mail: jdc@princeton.edu)

*Montague et al (2004) Nature*




Temporal difference (TD) learning model

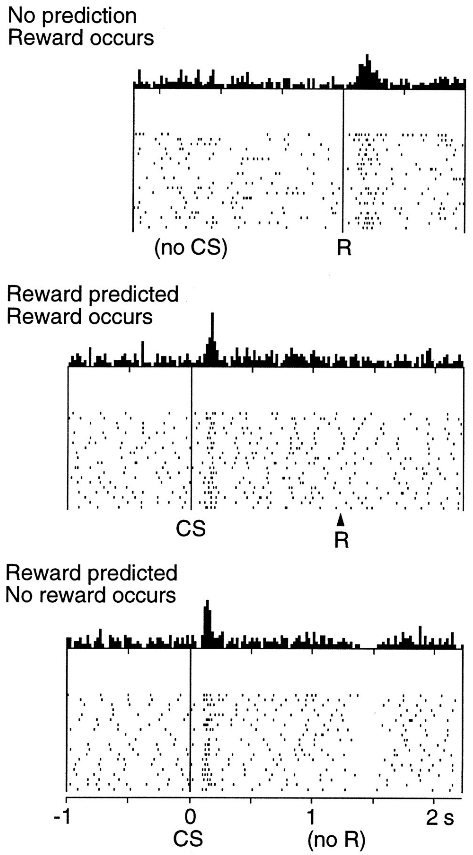
$$\delta(t) = \text{prediction error}\ (t) = E[r_t] + \boxed{\gamma \cdot \hat{V}(s_{t+1})} - \hat{V}(s_t)$$
$$\approx \text{current reward} + \gamma \cdot \text{next prediction} - \text{current prediction}$$

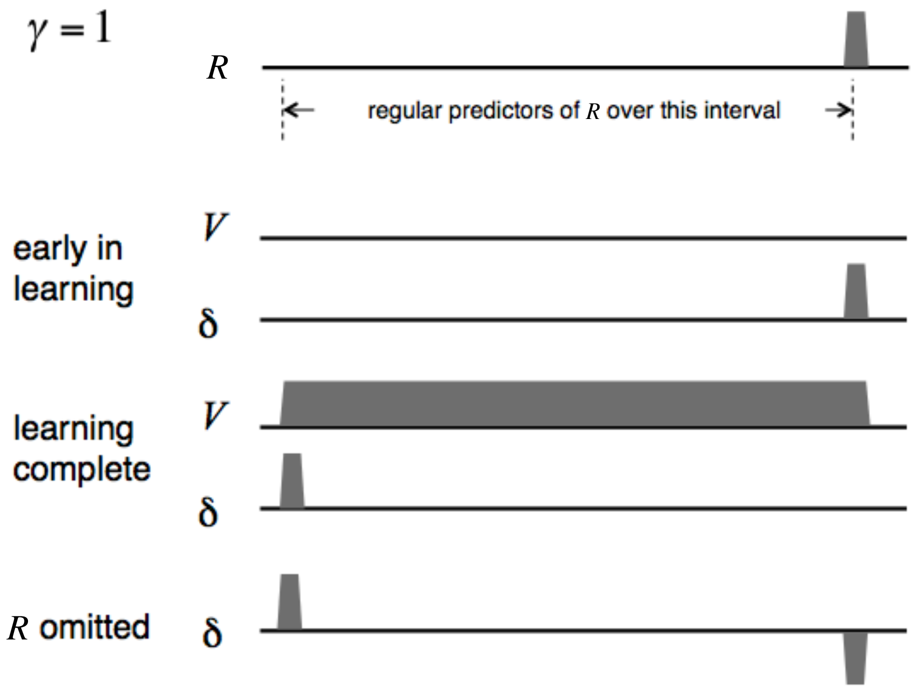
*Sutton & Barto (1998) Reinforcement Learning*

# Q) How TD learning accounts for the phasic response of a dopamine neuron?

Sutton & Barto (2017) Reinforcement Learning, 2nd Ed., Chapter 15

$$\boxed{\delta_t = R_t + \gamma V(s_t) - V(s_{t-1})}$$




Reward onset

$$\delta_t = R_t + V_t - V_{t-1} = R_t + 0 - 0 = R_t$$

Cue onset

$$\delta_t = R_t + V_t - V_{t-1} = 0 + R_t - 0 = R_t$$

Reward onset

$$\delta_t = R_t + V_t - V_{t-1} = 0 + 0 - R_t = -R_t$$

Recent works: Kim et al (2020) Cell; Gershman et al (2024) Nature Neuro. Masset et al (2025) Nature

*Instrumental learning*

*Model-based vs Model-free*

# *Model-based vs Model-free*

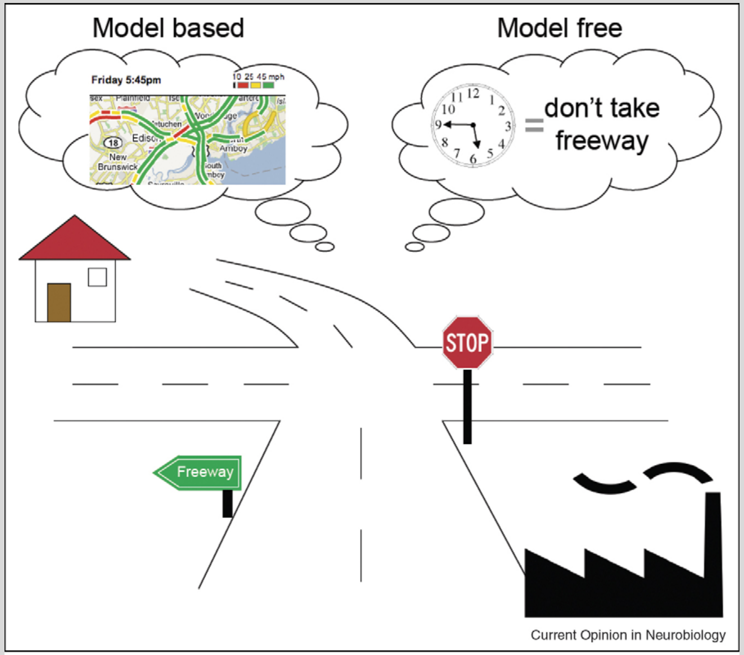


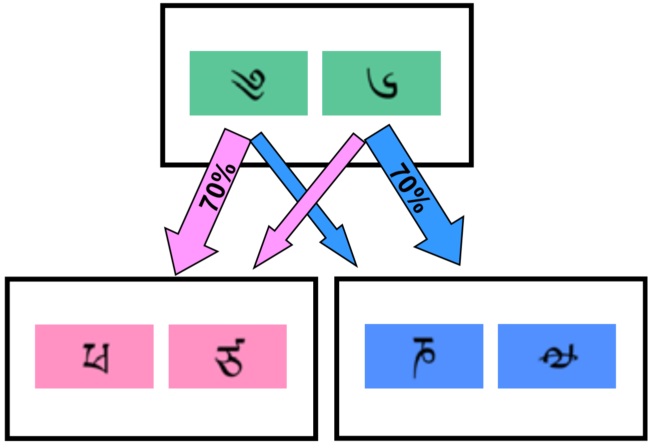

Figure 1: Two ways to choose which route to take when traveling home from work on friday evening.

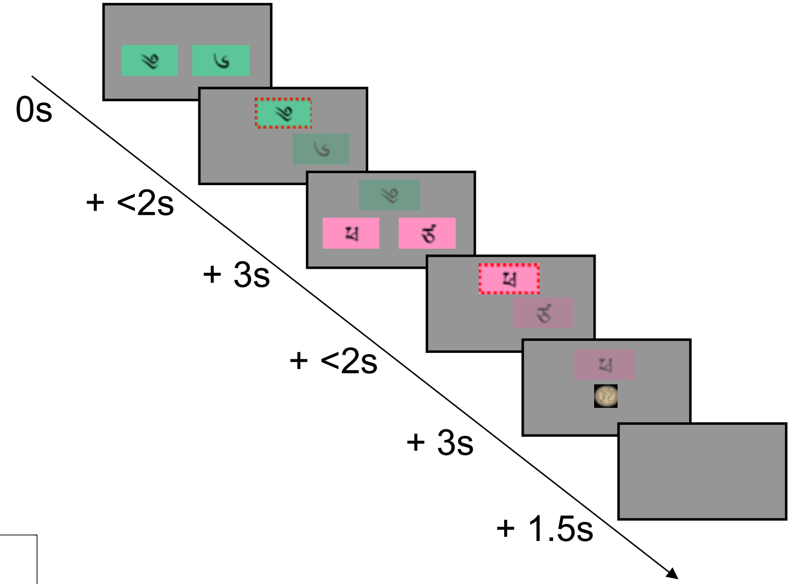
Dayan & Niv (2008)

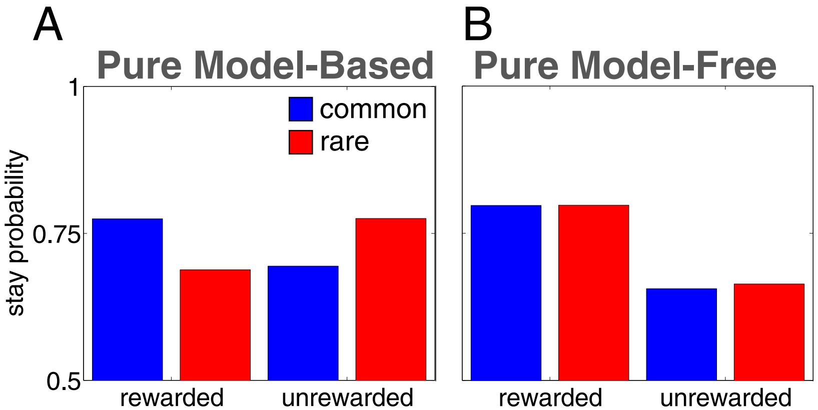
- *Model-based (goal-directed) learning: build a model of an environment. Effortful but flexible.*
- *Model-free (habitual) learning: relies on trials-and-errors. Efficient but inflexible.*
- *(Clinical) examples: compulsive behaviors, etc.*

# *Two-Step task*

*Daw et al (2011) Neuron*



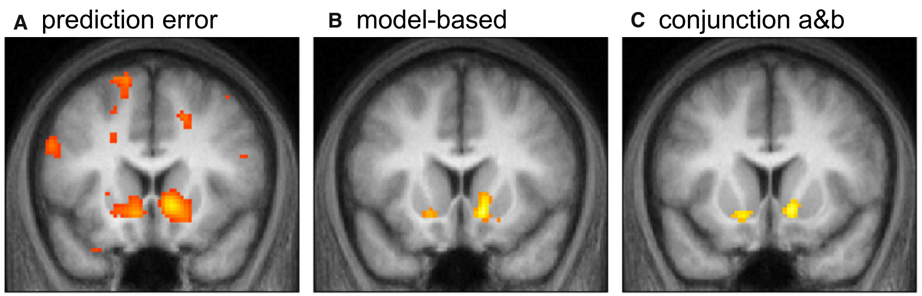
# *Computational model*

*Daw et al (2011) Neuron*
*Wunderich et al (2012) Neuron*

- *Separately calculate $V^{MF}$ and $V^{MB}$ (assuming full knowledge of the environment).*

- *Omega (ω): weight for model-based (MB)*
  - *0 (completely model-free) ≤ $\omega_{MB}$ ≤ 1 (completely model-based)*

$$V^{Hybrid} = \omega \cdot V^{MB} + (1-\omega) \cdot V^{MF}$$




*Daw et al (2011) Neuron*

PLOS | COMPUTATIONAL BIOLOGY

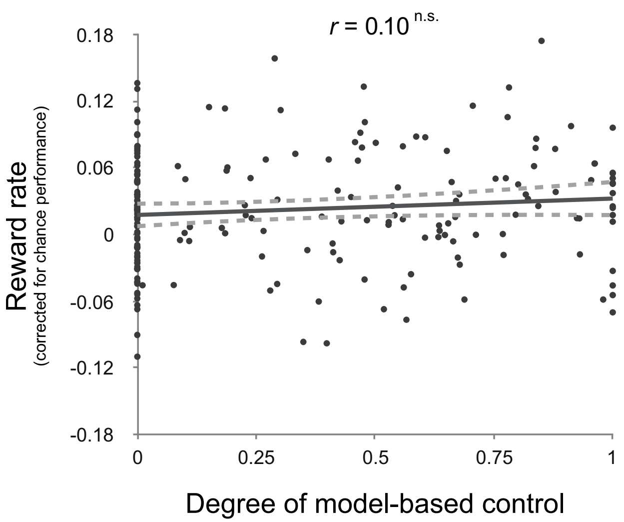
RESEARCH ARTICLE

# When Does Model-Based Control Pay Off?

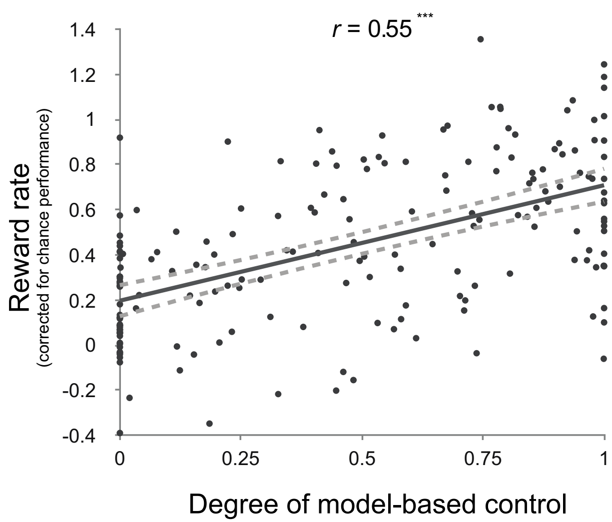
**Wouter Kool[1]*, Fiery A. Cushman[1☯], Samuel J. Gershman[1,2☯]**

**1** Department of Psychology, Harvard University, Cambridge, Massachusetts, United States of America,
**2** Center for Brain Science, Harvard University, Cambridge, Massachusetts, United States of America

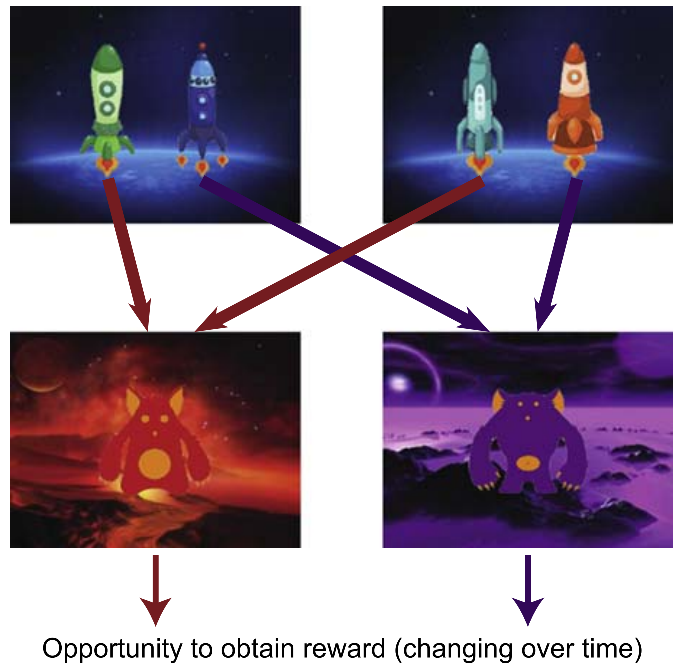
☯ These authors contributed equally to this work.
* wkool@fas.harvard.edu




*Daw Two-Step Task*

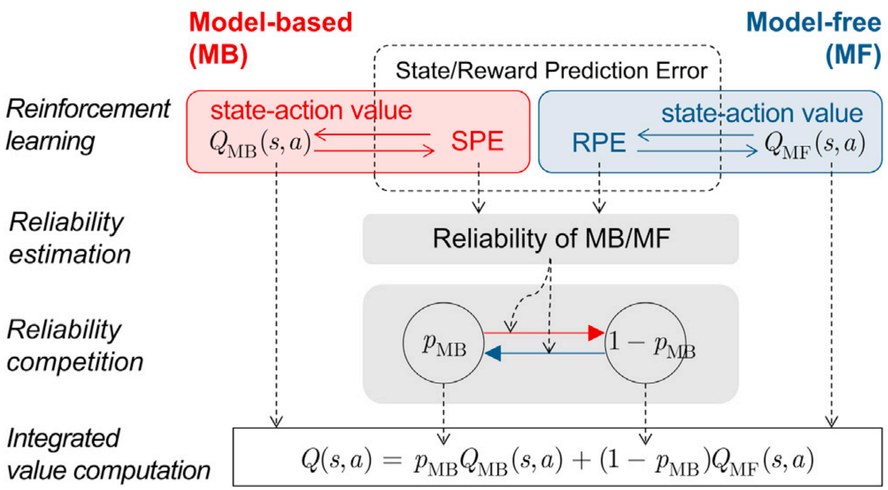


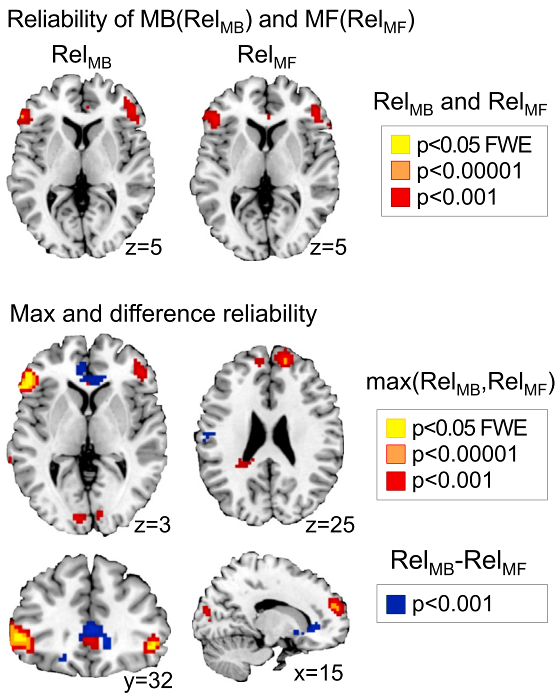

*Kool Two-Step Task*




*Kool et al (2016) PloS Comput Biol*

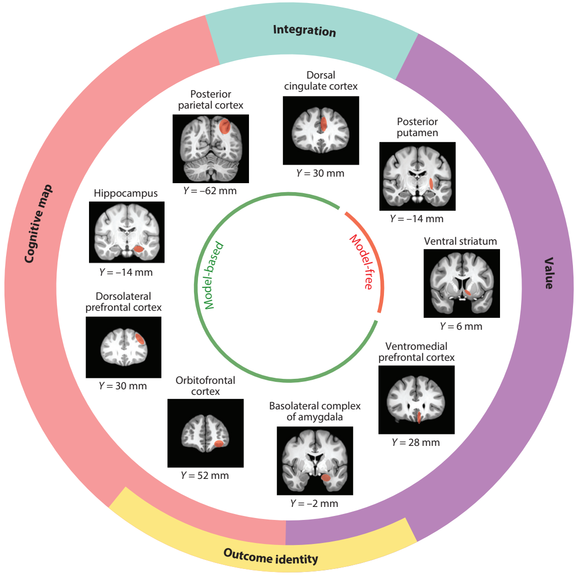
# *Reliability-based arbitration between model-based and model-free*



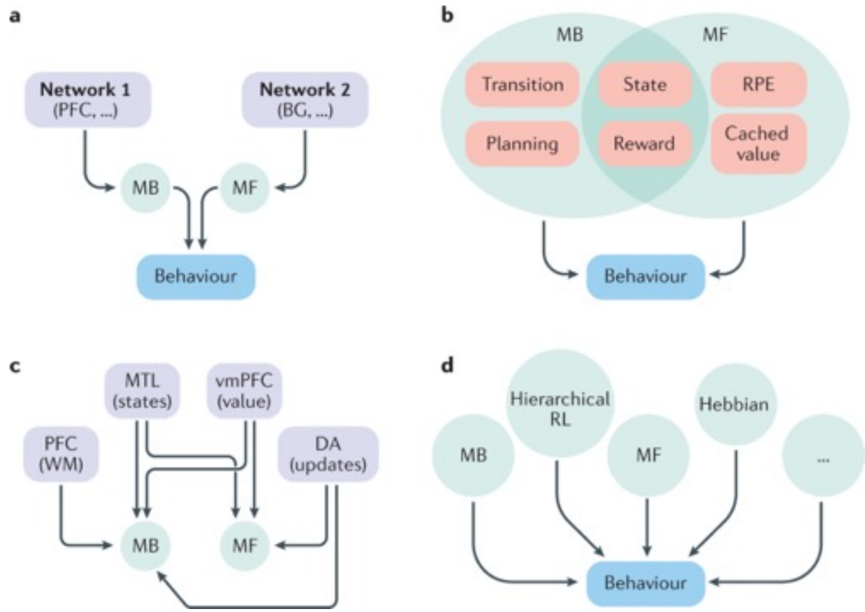




*Lee et al (2014) Neuron*
*Daw et al (2005) Nature Neuroscience*
*Wang et al (2018) Brain & Neuro. Advances*

Inferior lateral prefrontal and frontopolar cortex




*O'Doherty et al (2017) Annu. Rev. Psychology*



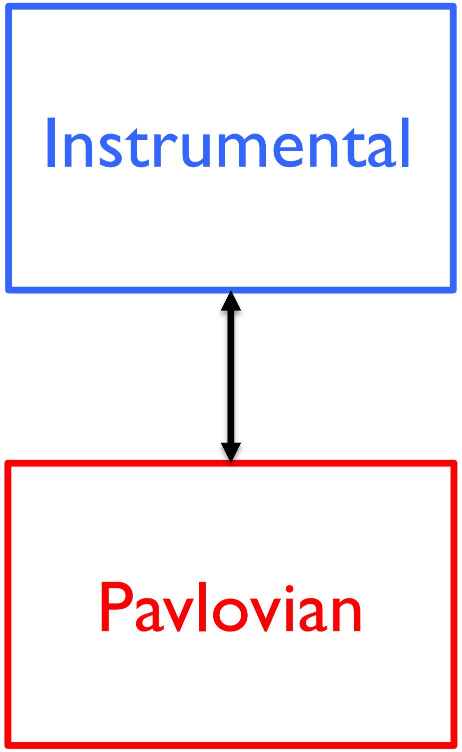

*Collins & Cockburn (2020) Nature Reviews Neuroscience*

Current Biology
Review

# The Role of Hippocampal Replay in Memory and Planning

H. Freyja Ólafsdóttir[1], Daniel Bush[2], and Caswell Barry[1]

# *Pavlovian vs Instrumental control*




"Learn what's best"

Approach for (+)
Withdraw from (-)

**"Pavlovian bias"**

Opinion

CellPress

**Action versus valence in decision making**

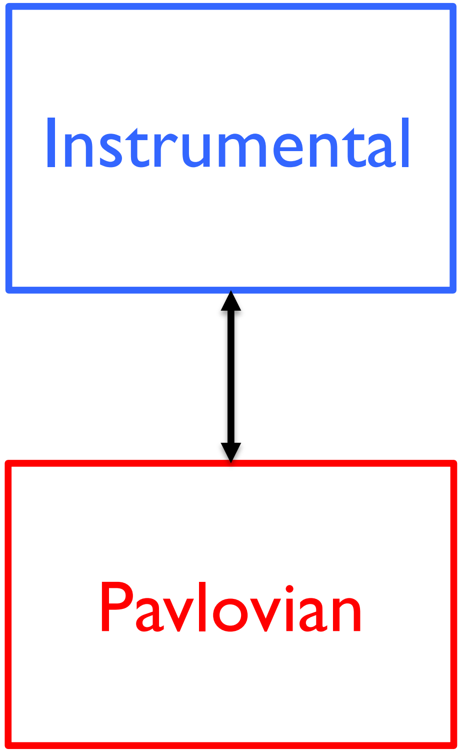
**Marc Guitart-Masip[1,2], Emrah Duzel[3,4,5], Ray Dolan[2], and Peter Dayan[6]**

[1] Aging Research Centre, Karolinska Institute, SE-11330 Stockholm, Sweden
[2] Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London WC1N 3BG, UK
[3] Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, UK
[4] Otto von Guericke University Magdeburg, Institute of Cognitive Neurology and Dementia Research, D-39120 Magdeburg, Germany
[5] German Center for Neurodegenerative Diseases, D-39120 Magdeburg, Germany
[6] Gatsby Computational Neuroscience Unit, University College London, London W1CN 3AR, UK

Balleine & O'Doherty (2010); Dayan et al (2006); Dayan (2013); Dayan & Niv (2008); Dolan & Dayan (2013); Dayan & Berridge (2014); Rangel et al (2008)

# *Orthogonalized Go/Nogo task*
# *Pavlovian-Instrumental competition*

Guitart-Masip et al (2012) Neuroimage
Also, see Huys et al (2011) Plos Comp Biology




"Learn what's best"

Approach for (+)
Withdraw from (-)

**"Pavlovian bias"**

Balleine & O'Doherty (2010); Dayan et al (2006); Dayan (2013); Dayan & Niv (2008); Dolan & Dayan (2013); Dayan & Berridge (2014); Rangel et al (2008)

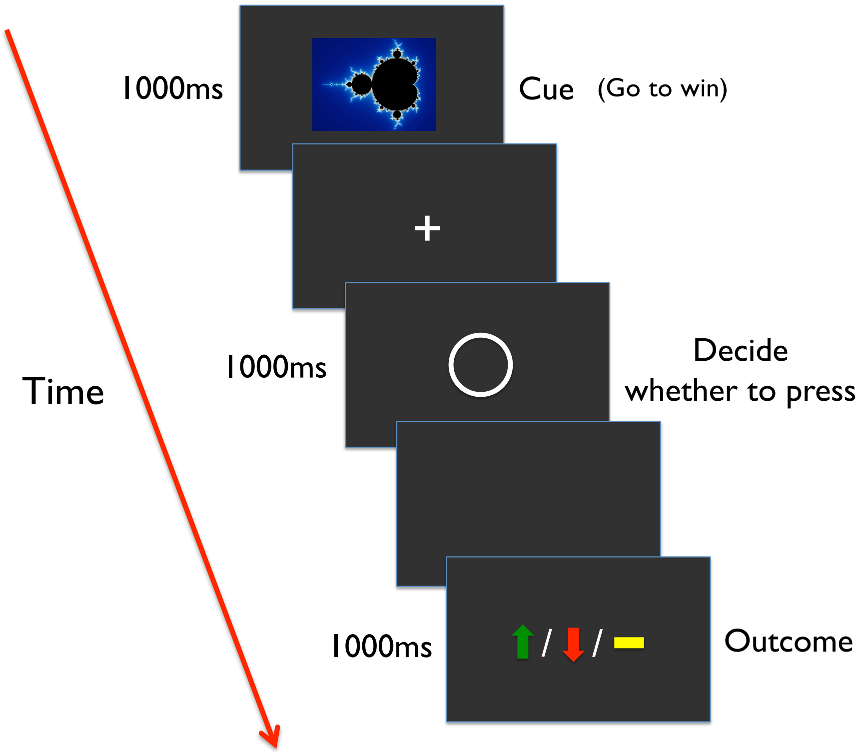
# *Orthogonalized Go/Nogo task*

| | Loss | Gain |
|---|---|---|
| Go | Go to avoid | Go to win |
| Nogo | Nogo to avoid | Nogo to win |

Pavlovian congruent
Pavlovian incongruent

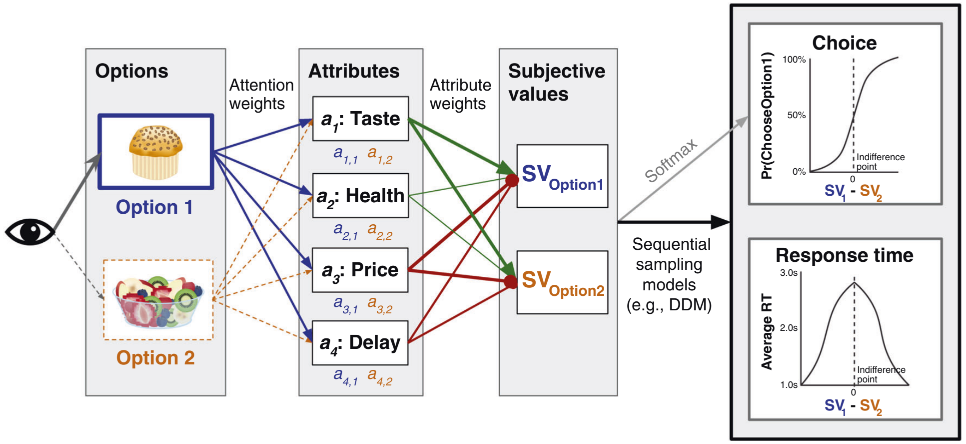
- *4 cues (conditions)*

  *2 actions (Go / Nogo) x*
  *2 valence (Gain / Loss)*



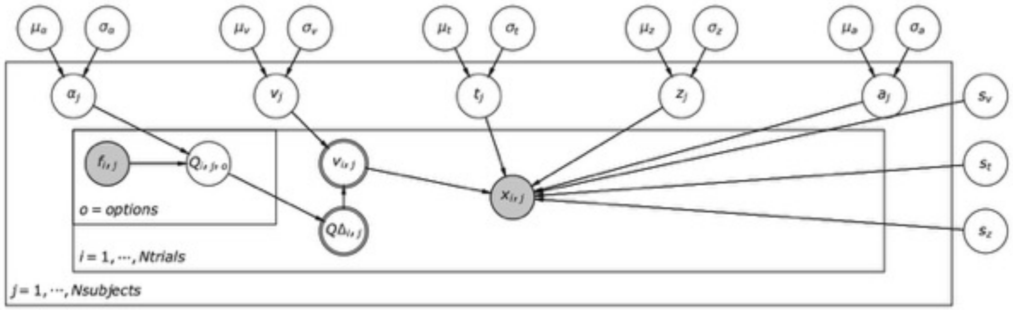

Guitart-Masip et al (2012) Neuroimage
Cavanagh et al (2013) J Neuro

# More…




Collins & Shenhav et al (2021) Neuropsychopharmacology

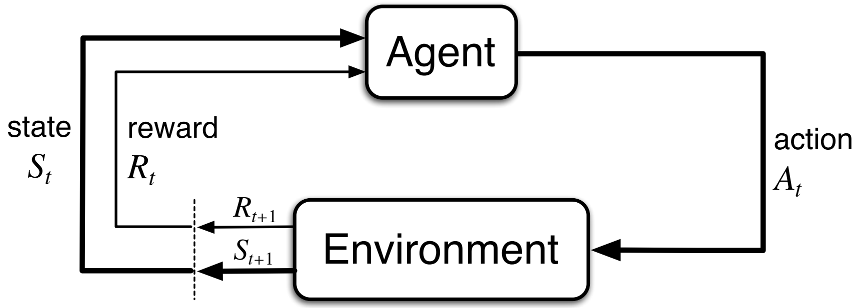
*RL + sequential sampling models*

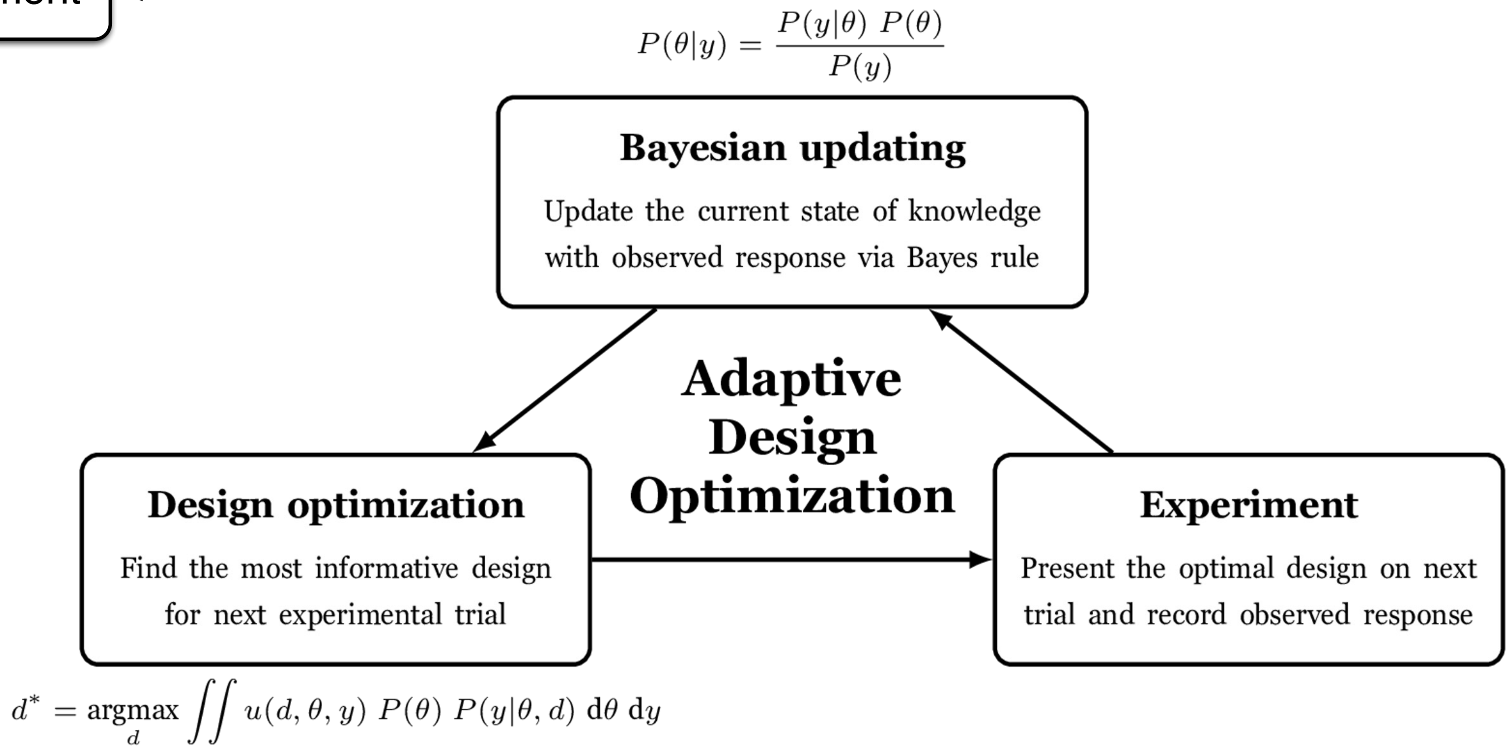



Pedersen & Frank (2020) Computational Brain & Behavior

*Adaptive Design Optimization within the RL framework*

*Optimize experiments on the fly!*


Agent
state
$S_t$
reward
$R_t$
action
$A_t$
$R_{t+1}$
$S_{t+1}$
Environment


$P(\theta|y) = \frac{P(y|\theta)\ P(\theta)}{P(y)}$
Bayesian updating
Update the current state of knowledge with observed response via Bayes rule
Adaptive Design Optimization
Design optimization
Find the most informative design for next experimental trial
Experiment
Present the optimal design on next trial and record observed response
$d^* = \underset{d}{\operatorname{argmax}} \iint u(d, \theta, y)\ P(\theta)\ P(y|\theta, d)\ \mathrm{d}\theta\ \mathrm{d}y$

$$p(\theta|y,d) = \frac{p(y|\theta,d)p(\theta)}{p(y|d)}$$

*Bayesian updating of model parameters ($\theta$)*

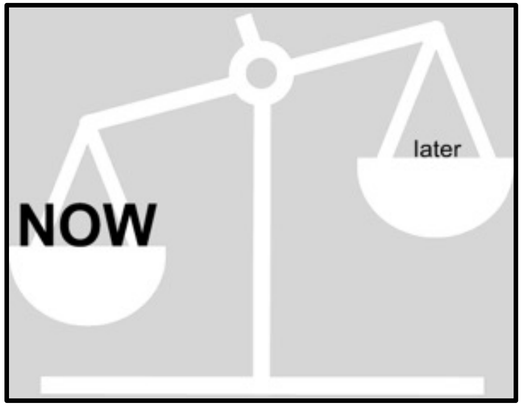
Find optimal design ($d^*$)

*Design & Conduct a mini-experiment on every trial*
*“What’s the most informative design ($d^*$) we should use?”*

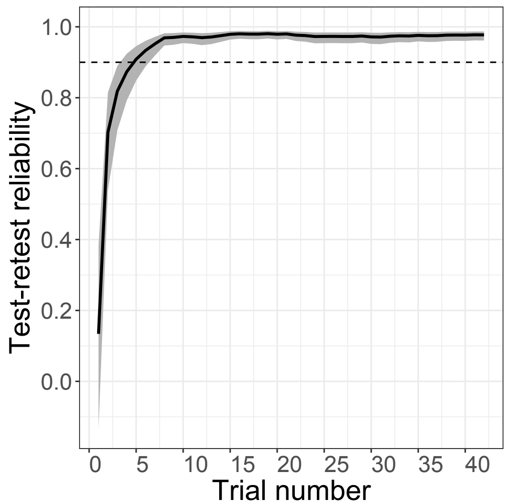
e.g., <u>$320</u> now vs $800 in <u>3 years</u>

Observe Data ($y$)

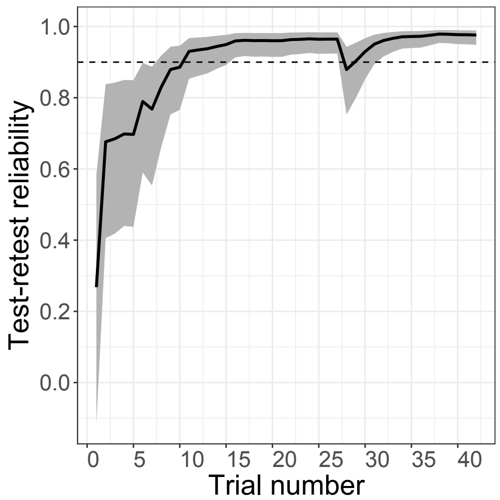
$$d^* = \underset{d}{\operatorname{argmax}} \int\int u(d,\theta,y)p(y|\theta,d)p(\theta)dyd\theta$$



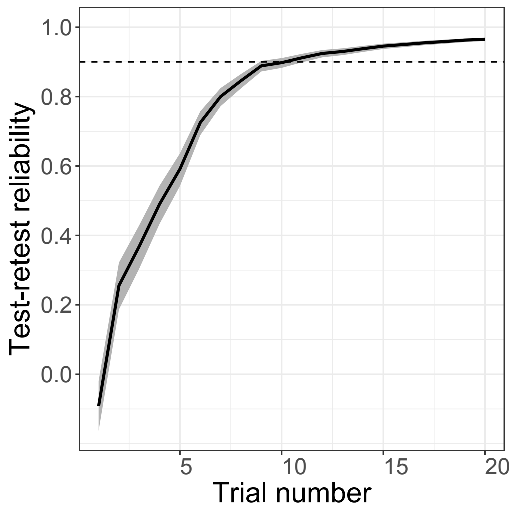
# *Up to 0.98 test-retest reliability within ~10 trials*




*College students*

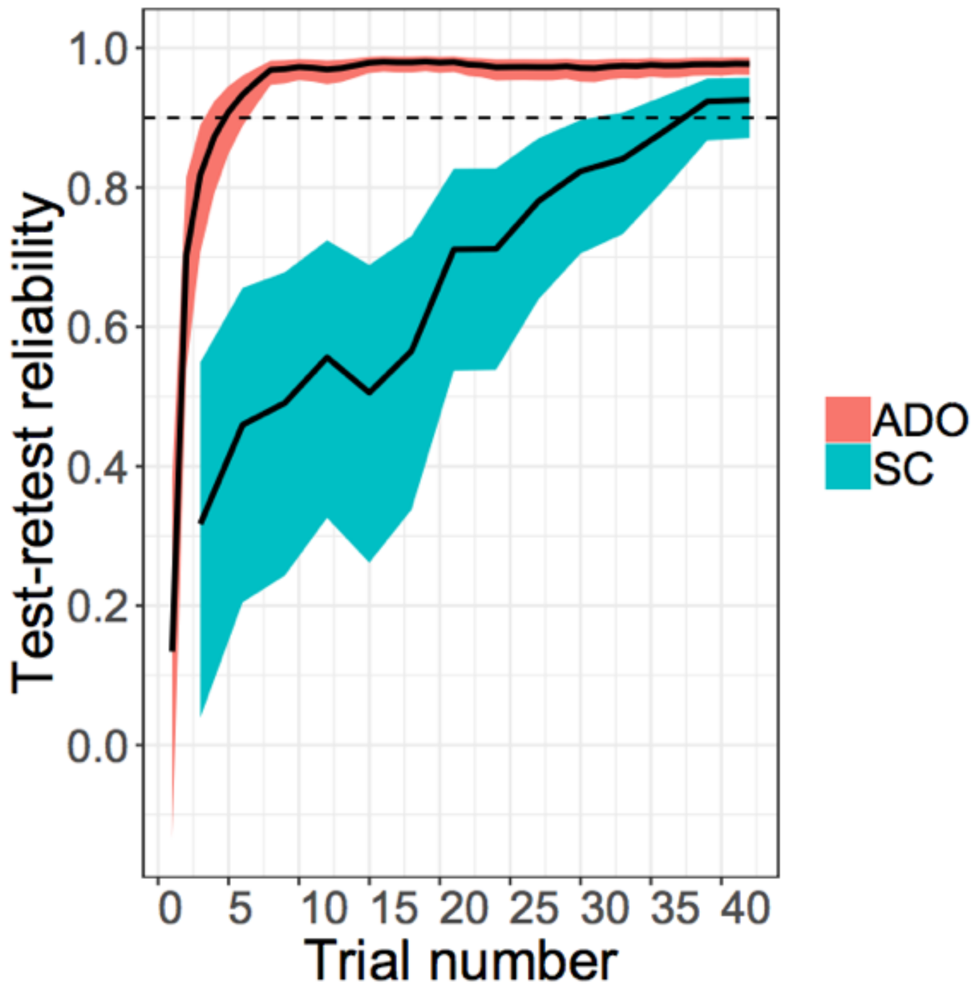



*Patients with SUDs*




*Online Amazon MTurk*

*Ahn et al (2020) Scientific Reports*

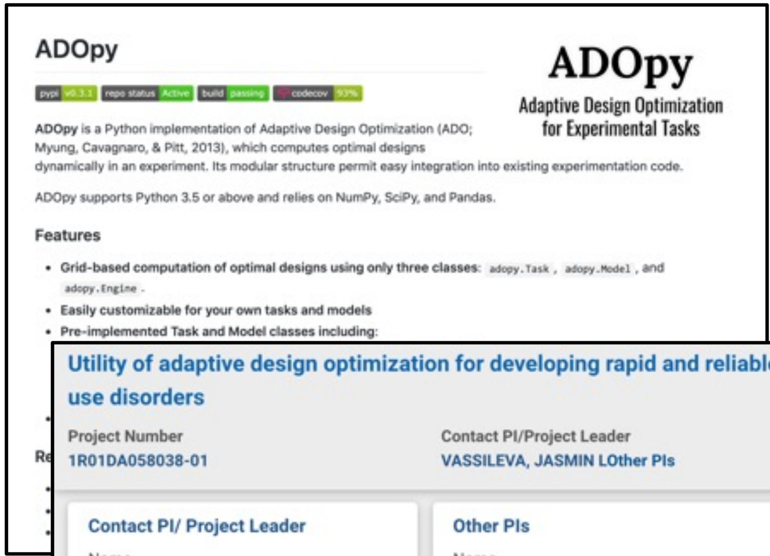



*3-5 times more precise*

*3-8 times more efficient*

*Ahn et al (2020) Scientific Reports*

https://github.com/adopy/adopy



ADOpy

ADOpy
Adaptive Design Optimization
for Experimental Tasks

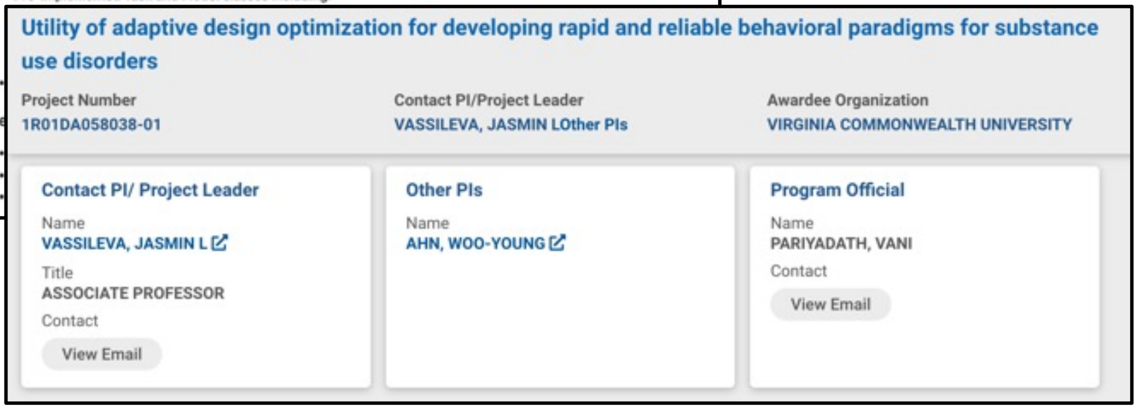
pypi v0.3.1 | repo status Active | build passing | codecov 93%

ADOpy is a Python implementation of Adaptive Design Optimization (ADO; Myung, Cavagnaro, & Pitt, 2013), which computes optimal designs dynamically in an experiment. Its modular structure permit easy integration into existing experimentation code.

ADOpy supports Python 3.5 or above and relies on NumPy, SciPy, and Pandas.

Features

- Grid-based computation of optimal designs using only three classes: `adopy.Task`, `adopy.Model`, and `adopy.Engine`.
- Easily customizable for your own tasks and models
- Pre-implemented Task and Model classes including:

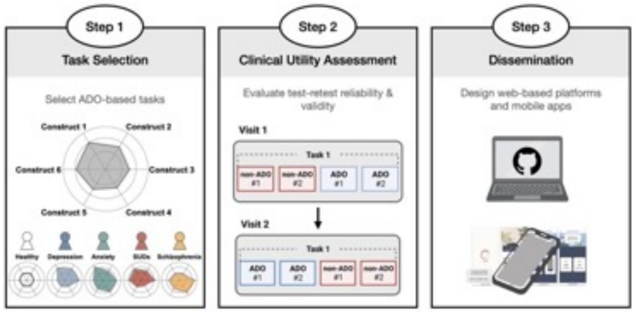


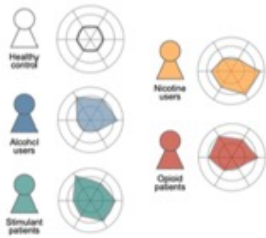
Utility of adaptive design optimization for developing rapid and reliable behavioral paradigms for substance use disorders

| Project Number | Contact PI/Project Leader | Awardee Organization |
|---|---|---|
| 1R01DA058038-01 | VASSILEVA, JASMIN LOther PIs | VIRGINIA COMMONWEALTH UNIVERSITY |

| Contact PI/ Project Leader | Other PIs | Program Official |
|---|---|---|
| Name<br>VASSILEVA, JASMIN L<br>Title<br>ASSOCIATE PROFESSOR<br>Contact<br>View Email | Name<br>AHN, WOO-YOUNG | Name<br>PARIYADATH, VANI<br>Contact<br>View Email |

*NIH / NIDA R01 (2023-2028)*
*MPI: Vassileva & **Ahn***



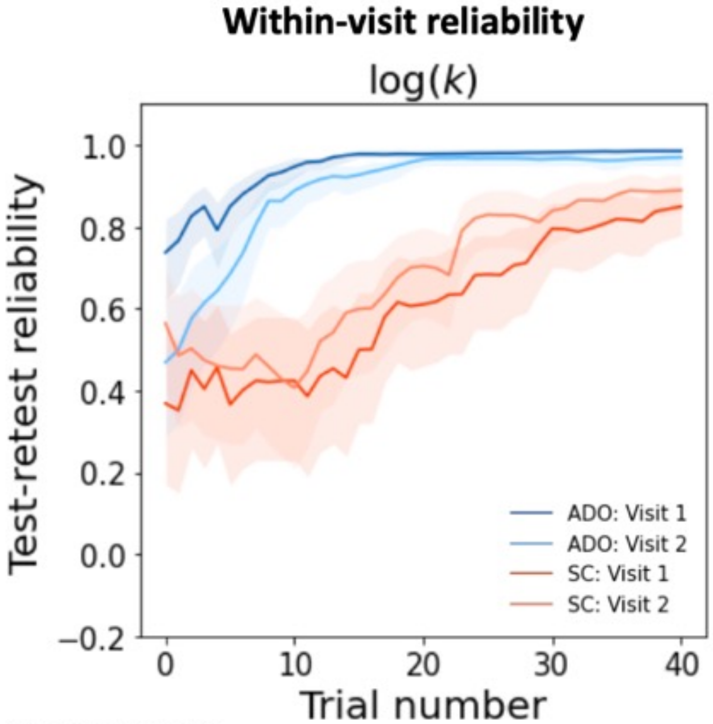

*Jeongyeon Shin*

*<u>ADO-powered fMRI Experiments</u>*

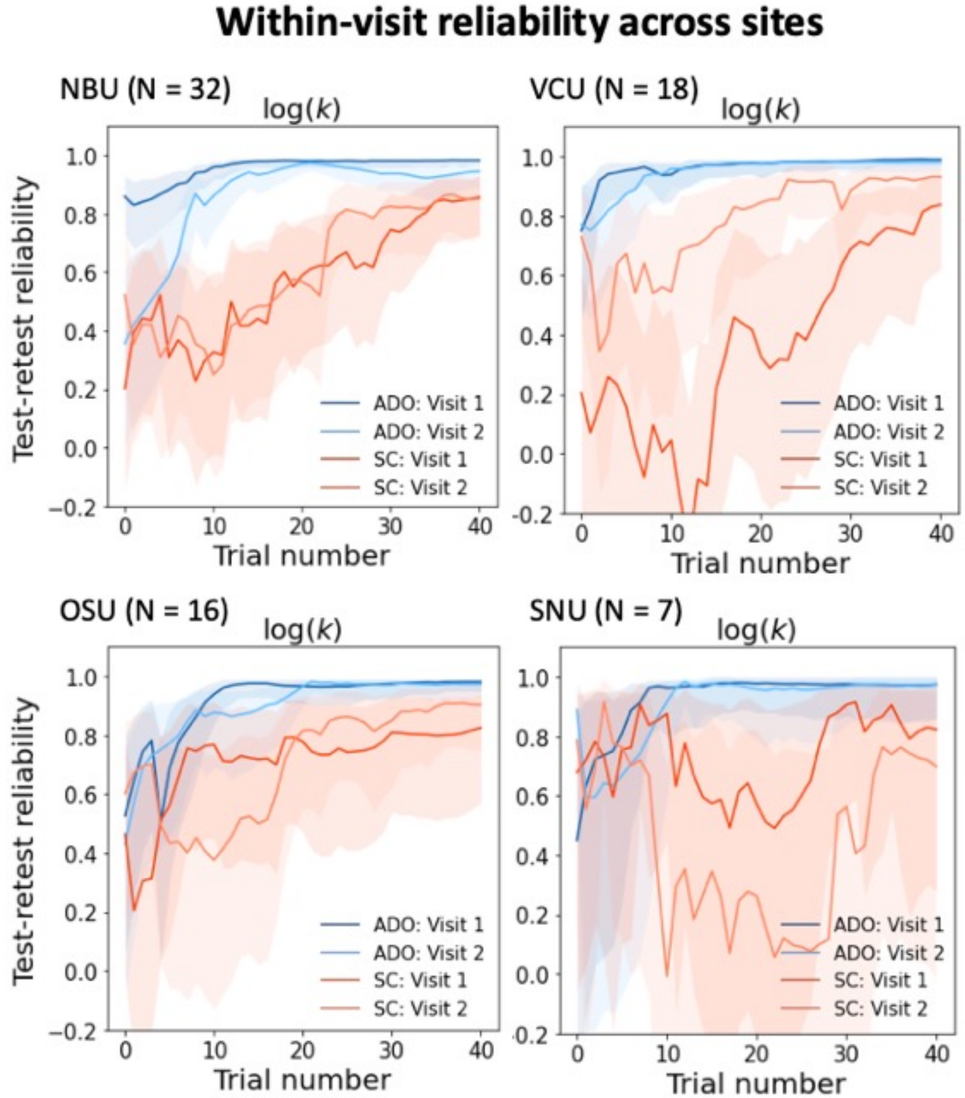
*Shin et al (in preparation)*

*Yang, Pitt, Ahn, Myung (2020) Behavior Research Methods*
*Kwon, Lee, & Ahn (2023) Biol. Psychiatry: CNNI*
*Vassileva, Lee, Psederska & Ahn (2023) Computational Neuroscience*

| Country | Recruitment site | N |
|---|---|---|
| Bulgaria | New Bulgarian Univ. (NBU) | 50 |
| USA | Ohio State Univ. (OSU) | 33 |
| | Virginia Commonwealth Univ. (VCU) | 23 |
| South Korea | Seoul National Univ. (SNU) | 13 |

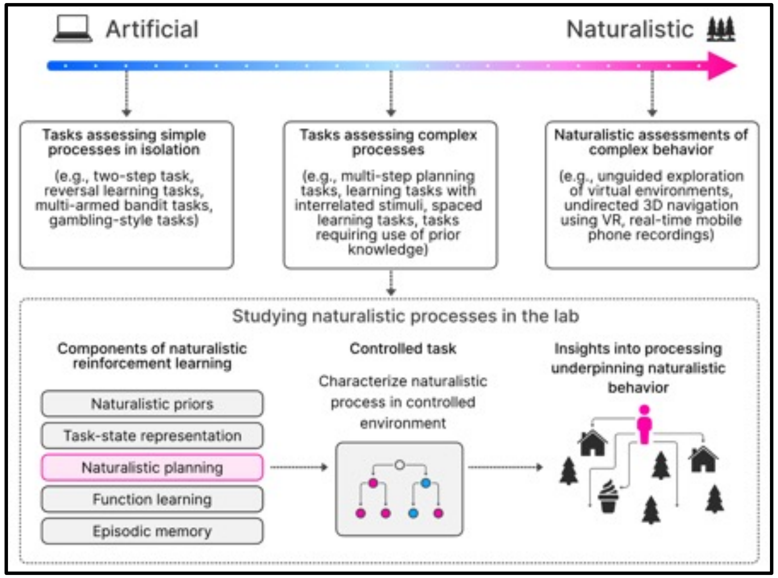



* Shaded area: 95%CI

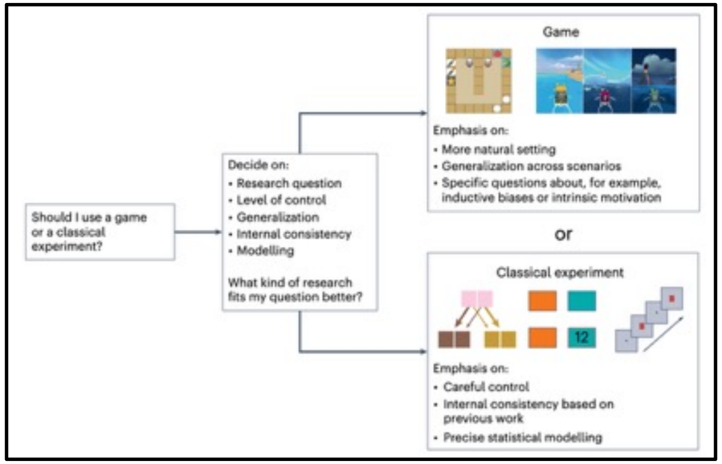


# *Last but not the least major limitation*

*Overly simplified "toy" problems/tasks*

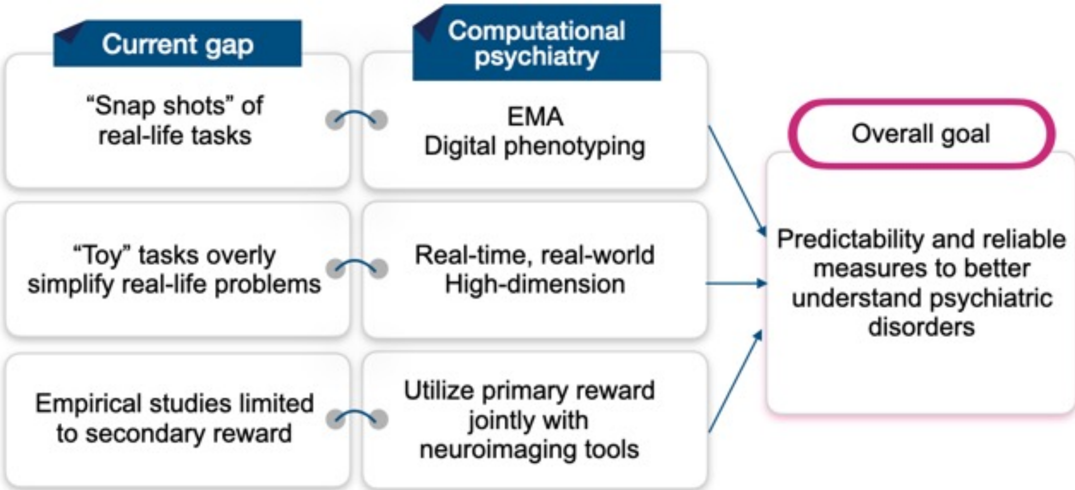



*Wise et al (2023) Trends in Cog Sci*




*Allen et al (2024) Nature Human Behaviour*

# *Overly simplified "toy" problems/tasks*

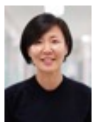



*Ahn, Lee, & Kim (invited review) Current Directions in Psychological Science*
*Lee, Lee, Im, O'Doherty, & Ahn (invited review) Nature Reviews Psychology*


*Sang-Ho Lee*


*Min-hwan Oh*

*Explore vs Exploit w/ Minecraft*

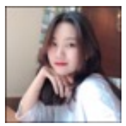

*Wonmok Shim*


*Hyeonmin Lee*

*"Real" reward (nicotine)*


*Josh Brown*


*Jeung-Hyun Lee*


*Eunhwi Lee*

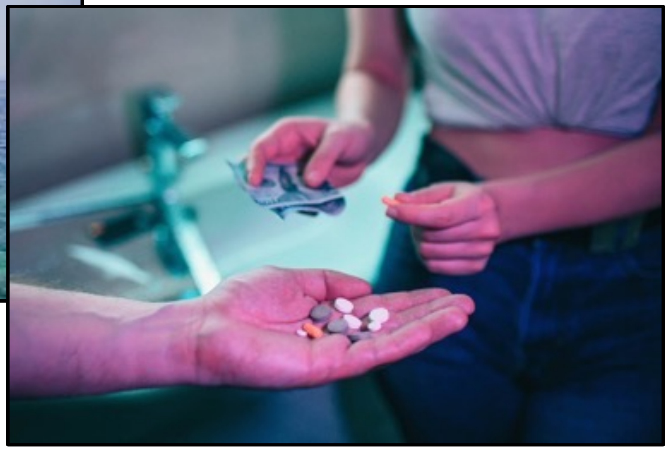
*Naturalistic (movie) paradigm*

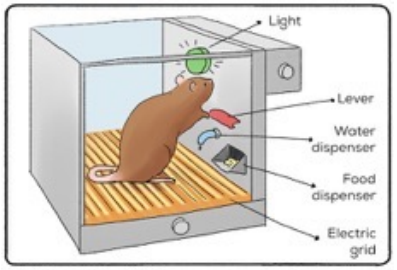

*Monica Rosenberg*


*Mina Kwon*

*Overly simplified "toy" problems/tasks*
*→ Real-time/real-world tasks*

*Sang-Ho Lee*






Lee, Song, Oh, & Ahn (2024) Psychological Science

# *Measures of Impulsivity*

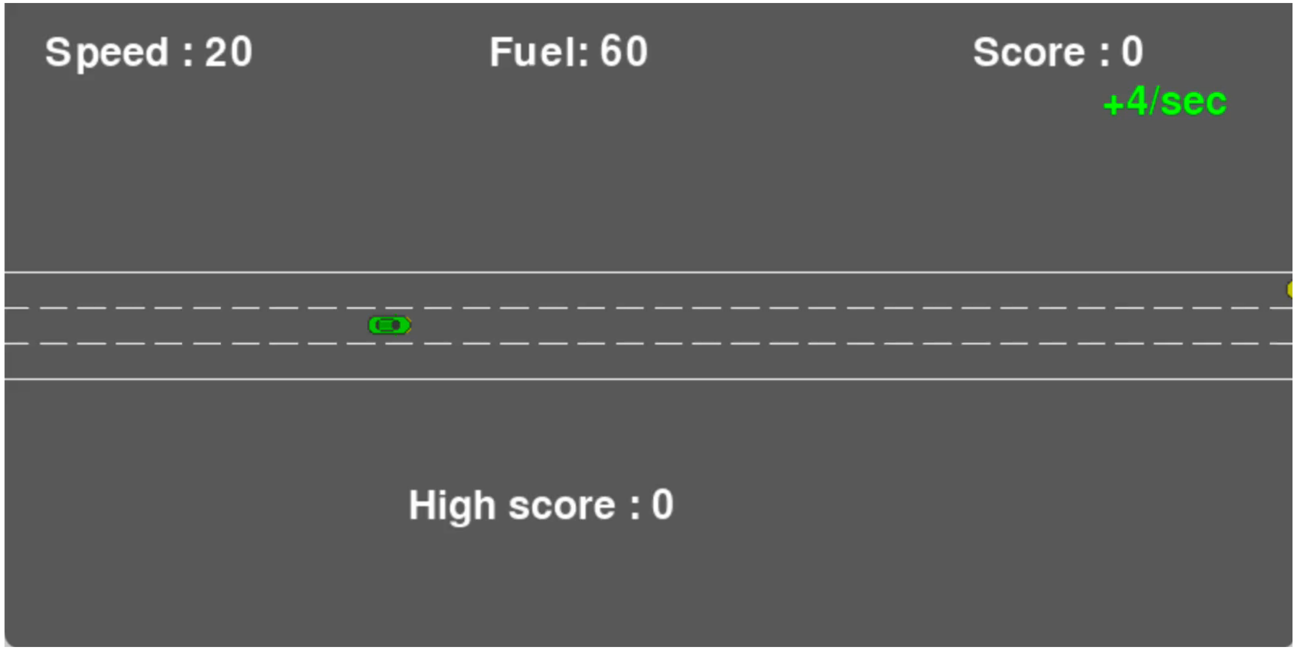
- ***Self-report measures*** *and* ***Behavioral measures*** *are the two most widely used methods to assess impulsivity.*


- ***A Challenge:*** *Self-report measures and behavioral measures often show weak correlations with each other.*

*Sharma et al (2014), Saunders et al. (2018), Murphy et al (2016), Frey et al. (2017), Dang et al (2020)*

# (Real-time) “Highway Task”

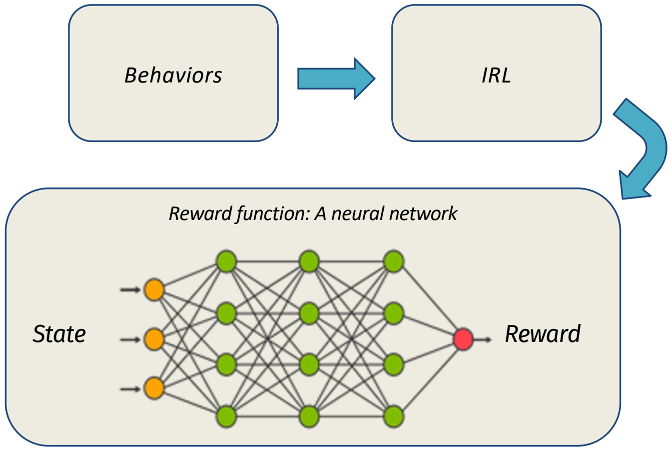
Sang-Ho Lee (KAIST)




*Lee, Song, Oh, & Ahn (2024) Psychological Science*

# *Inverse Reinforcement Learning*

*Arora, Doshi (2021)*
*Fu et al (2017)*




*To model complex behaviors in a real-time task, we harness deep neural networks + inverse reinforcement learning (IRL).*

- *IRL infers reward structures underlying observed behaviors.*

*Min-hwan Oh*

**Algorithm 1** Adversarial inverse reinforcement learning

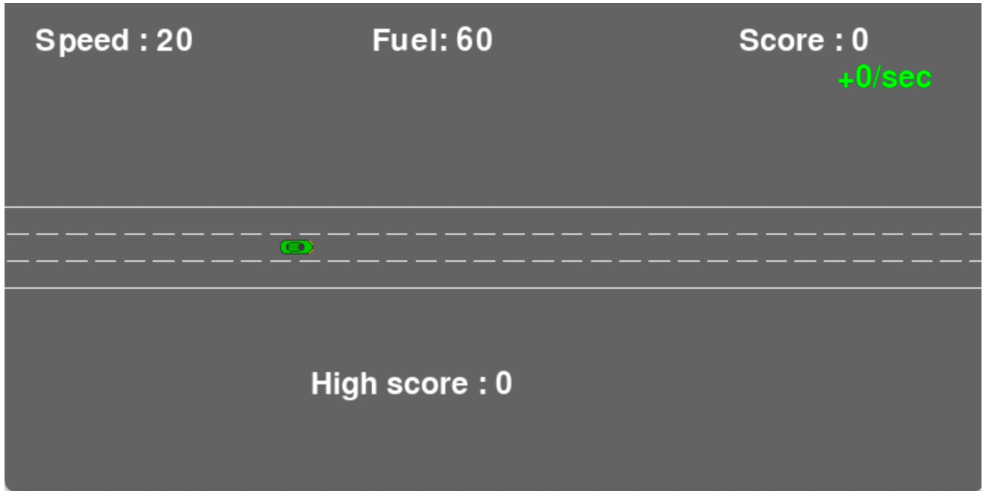
1: **for** iteration $i$ in {1,..., N} **do**
2: Obtain observed trajectories $\tau_i^O$ from the data.
3: Initialize discriminator $D_{\theta,\phi}(s, a, s')$ and policy $\pi$.
4: **for** step $t$ in {1,..., T} **do**
5: Sample trajectories $\tau_i$ using the policy function $\pi$.
6: Train $D_{\theta,\phi}$ to classify the data $\tau_i^O$ from samples $\tau_i$.
7: Update reward $r(s, a, s') \leftarrow \log\left(D_{\theta,\phi}(s, a, s')\right) - \log\left(1 - D_{\theta,\phi}(s, a, s')\right)$.
8: Update policy $\pi$ with respect to $r(s, a, s')$.
9: **end for**
10: **end for**

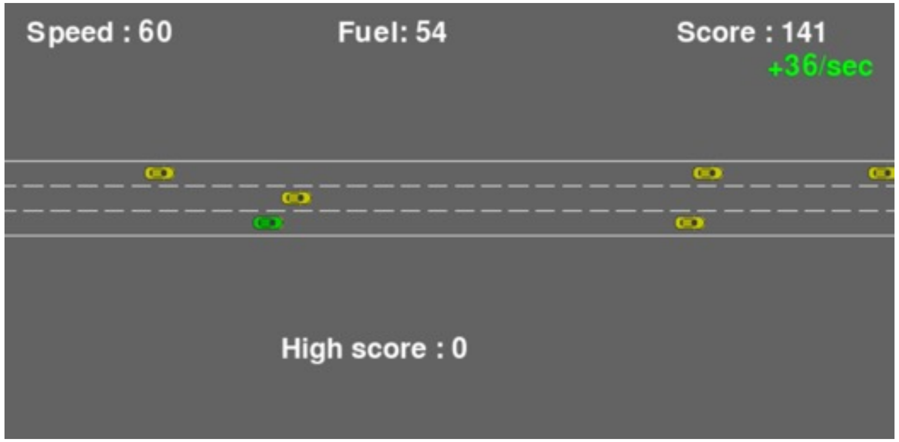
*We aim to explore individual differences in impulsivity by comparing individual reward structures inferred by IRL.*

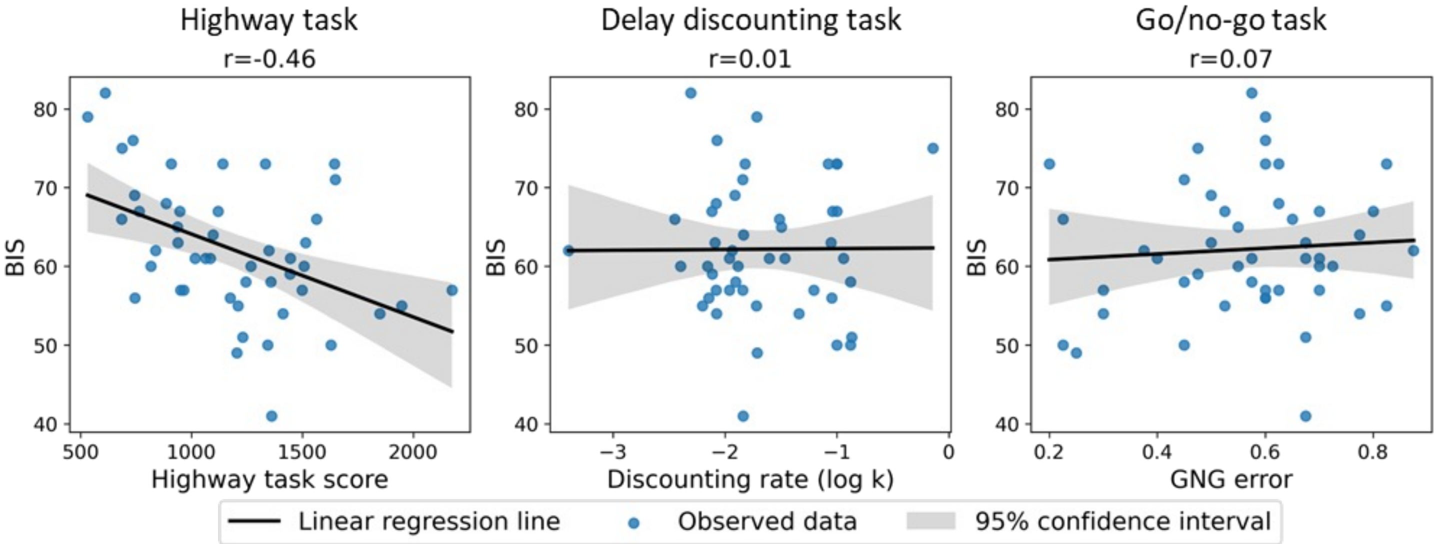
*Lee, Song, Oh, & Ahn (2024) Psychological Science*

# Task Performance Example

- *This is an agent trained by deep Q-learning.*

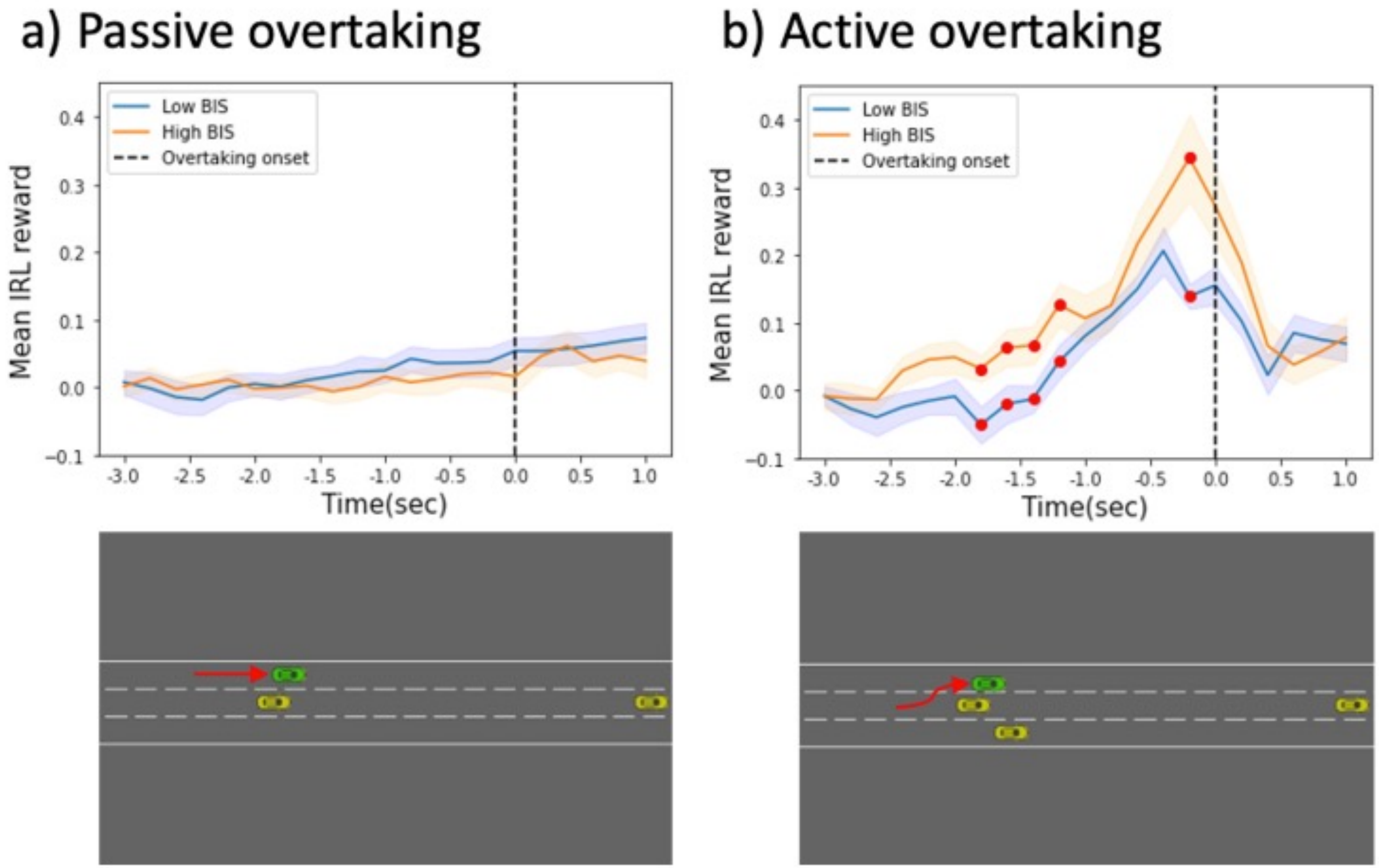




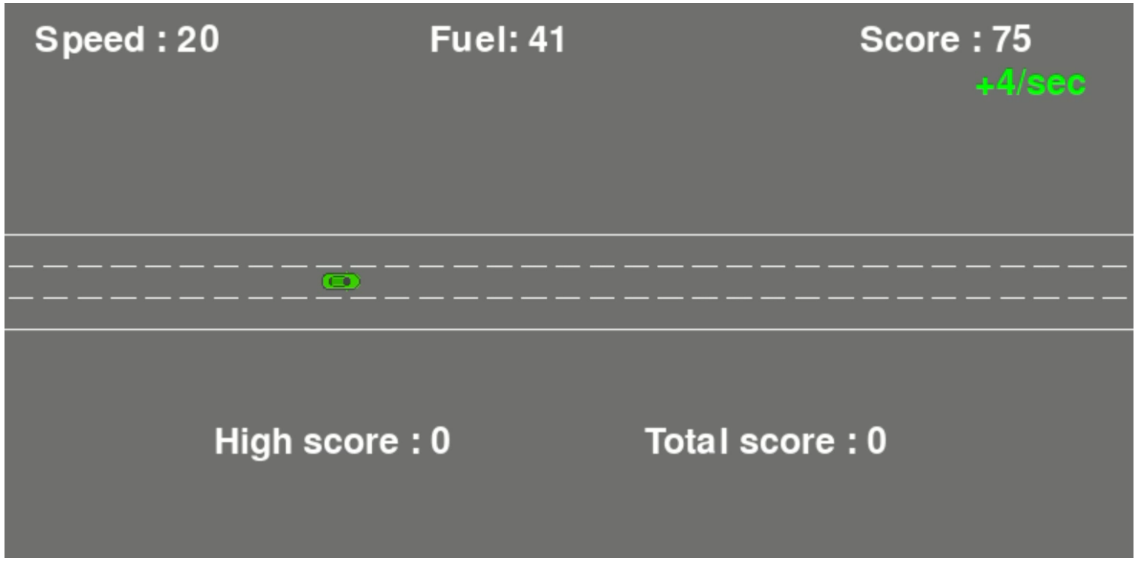

Lee, Song, Oh, & Ahn (2024) Psychological Science

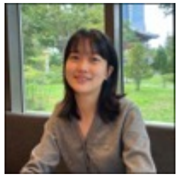
*Lee, Song, Oh, & Ahn (2024) Psychological Science*



# *Next Questions*

- *IRL reward ≈ reward signals in the brain?*
- *More realistic tasks?*
  - *3D version of the Highway task using Unity (w/ Robert Whelan, TCD)*
  - *"Cutting-in"*

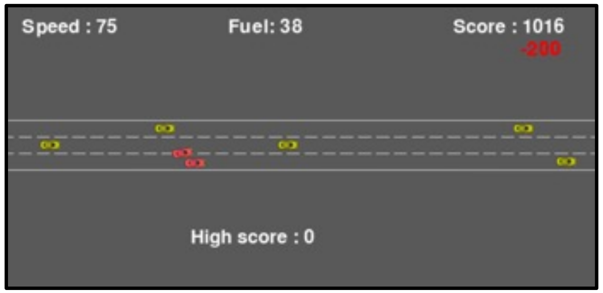



*Chaeyoun Chung*

*Joonha Kim*

## *(Preliminary) fMRI Study w/ Highway task*

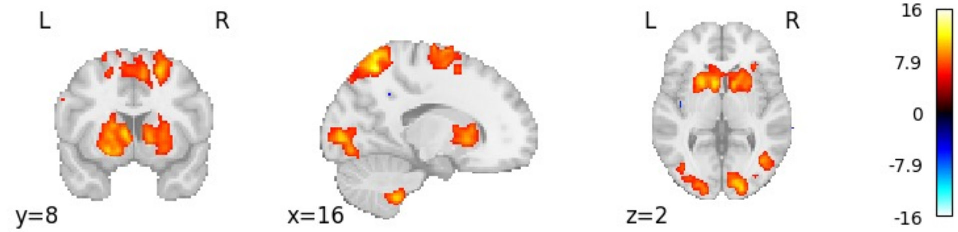
- *fMRI Study w/ healthy controls (N=45)*
- *Focus on two events*
  - *"Overtaking"*
  - *"Crash"*

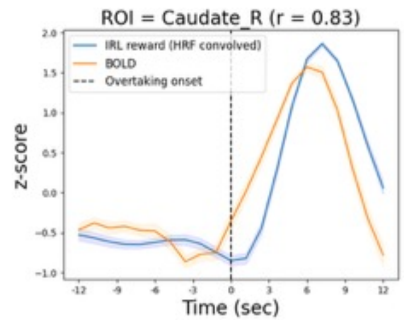



*Lee, Oh, & Ahn (2024) CCN2024*
*Lee, Chung, Oh, & Ahn (in preparation)*

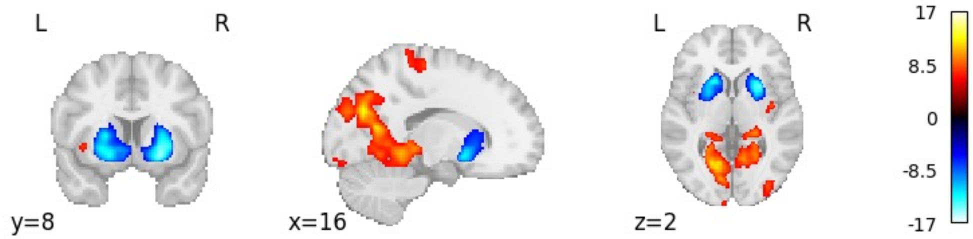
# GLM analysis


## Overtake onset



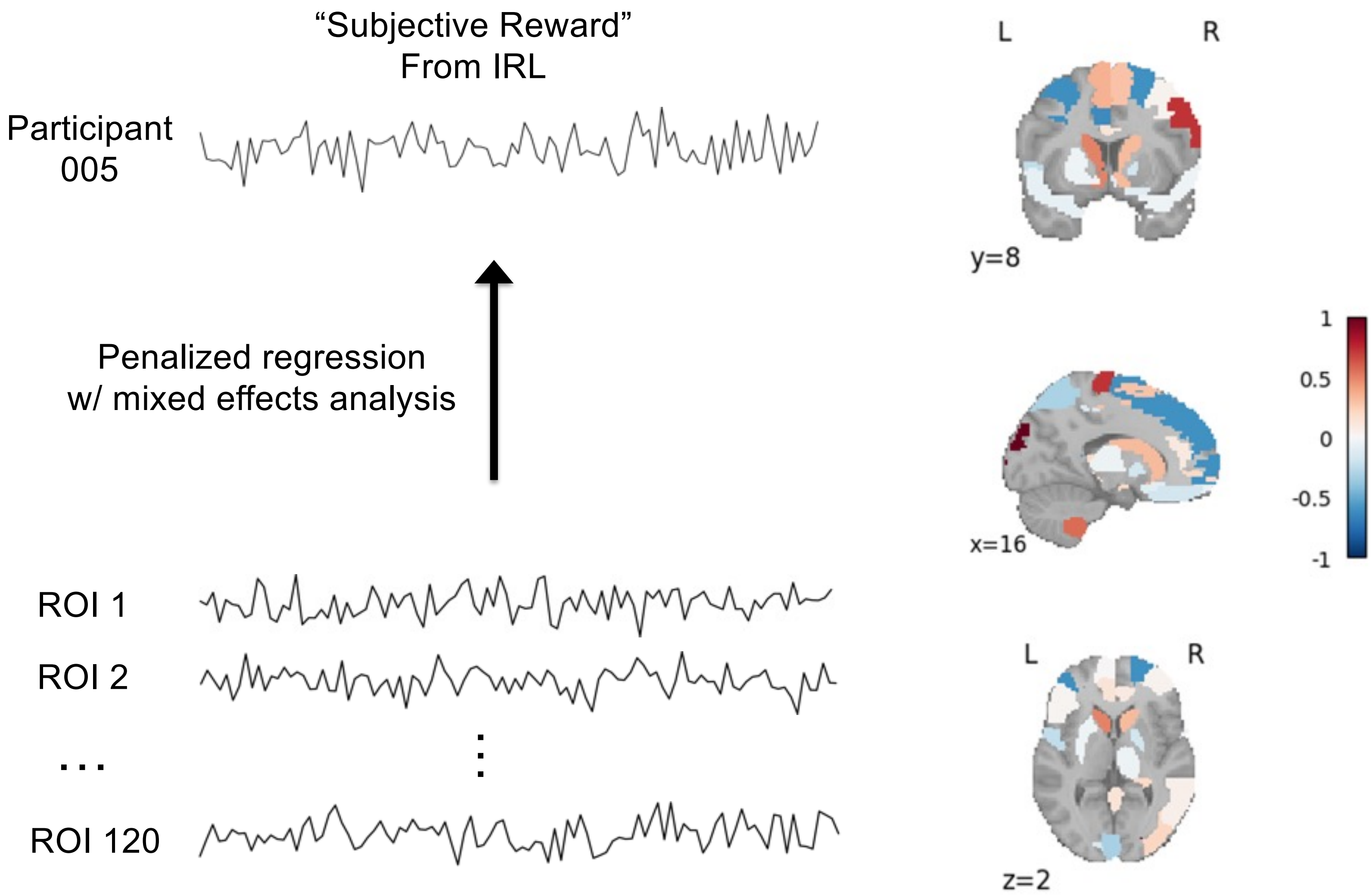




## Crash onset



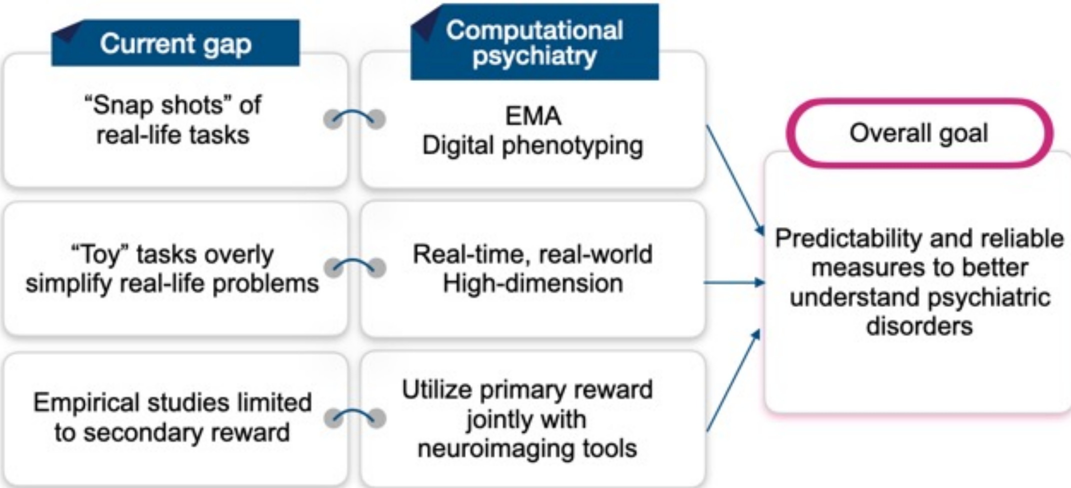

*Lee, Oh, & Ahn (2024) CCN2024*
*Lee, Chung, Oh, & Ahn (in preparation)*

*P value threshold = 0.001 (corrected p< 0.05, FDR corrected)*



"Subjective Reward"
From IRL
Participant
005
Penalized regression
w/ mixed effects analysis
ROI 1
ROI 2
...
ROI 120
L
R
y=8
x=16
z=2
1
0.5
0
-0.5
-1

# *Overly simplified "toy" problems/tasks*

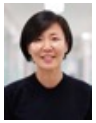



*Ahn, Lee, & Kim (invited review) Current Directions in Psychological Science*
*Lee, Lee, Im, O'Doherty, & Ahn (invited review) Nature Reviews Psychology*

*Real-time Driving task*


*Sang-Ho Lee* *Min-hwan Oh*

*Explore vs Exploit w/ Minecraft*




*Wonmok Shim* *Hyeonmin Lee*

*"Real" reward (nicotine)*



*Josh Brown* *Jeung-Hyun Lee* *Eunhwi Lee*

*Naturalistic (movie) paradigm*

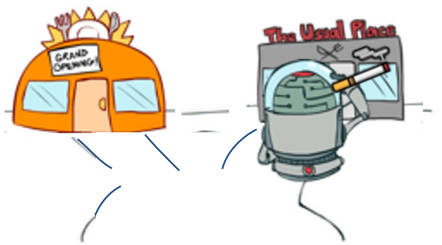

*Monica Rosenberg* *Mina Kwon*

# *Exploration/exploitation bias in addiction?*

***Exploration Bias***

*"Failure to properly use prior information and excessive focus on irrelevant information"*

- ***Smokers*** *in reward* *(Morris et al., 2016)*
- ***Binge eating disorder*** *in loss* *(Morris et al., 2016)*

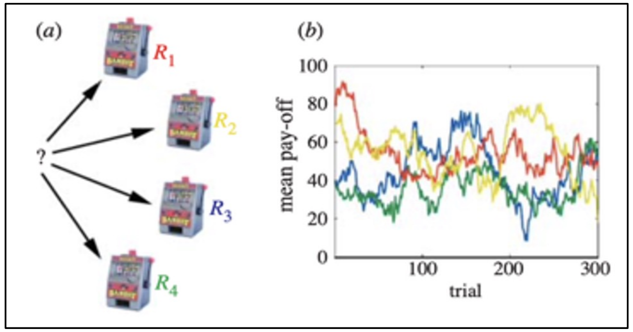



***Exploitation Bias***

*"Repetition of habitual behaviors without acquiring new information"*

- ***Methamphetamine users*** *(Harle et al., 2015; Addicott et al., 2017)*
- ***Alcohol use disorder*** *(Morris et al., 2016)*
- ***Habitual smokers*** *(Addicott et al., 2014a)*

# *Multi Armed-Bandit Task*

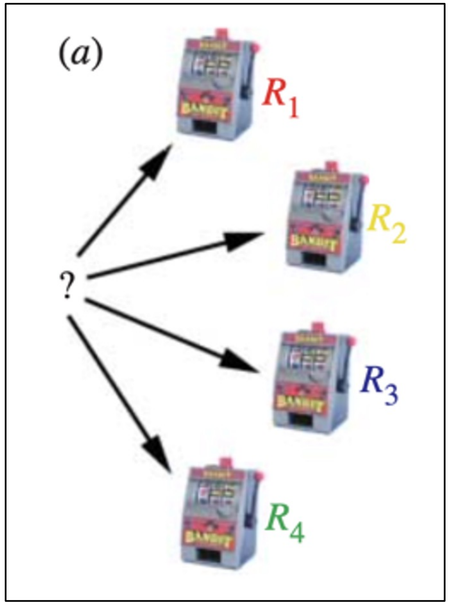



*Daw et al (2006) Nature*


*Dining options!*

<u>*Limitations?*</u>

- *Low ecological validity*
- *Cost (time & effort) for navigation*
- *Locomotion & visual exploration*




*2D MAB*

MINECRAFT

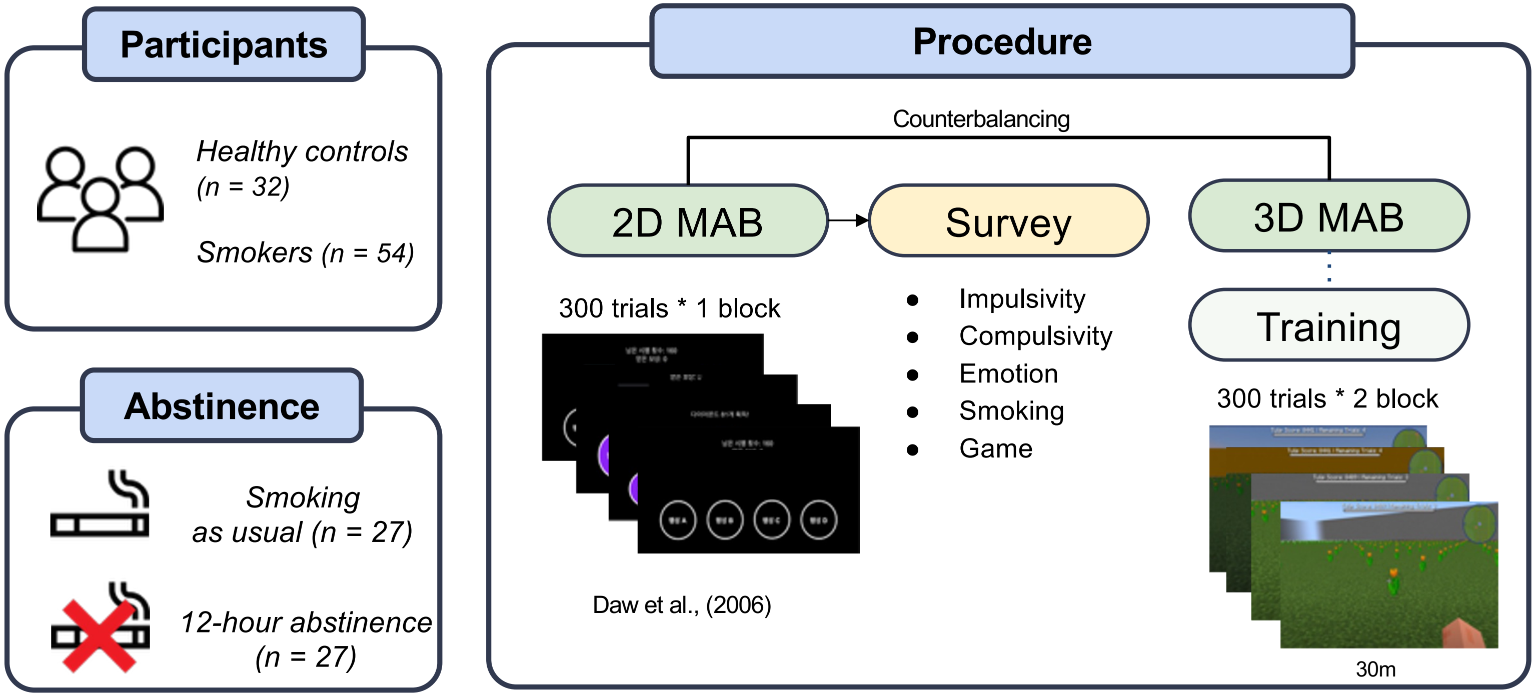
*3D MAB*

*Hyeonmin Lee*

*Lee et al (in preparation)*


Tulip Score: 8441 | Remaining Trials: 4

# *Participants and Procedure*



# Computational Modeling

- *Hierarchical Bayesian parameter estimation (MCMC) w/ Stan*

- *Model Comparison (leave-one-out CV information criterion; LOOIC)*
  - *Model 1: [2D/3D] Kalman filter + Softmax* *(Daw et al., 2006)*
  - *Model 2: [3D] Kalman Filter + Softmax with the navigation cost*
    - *In 3D MAB, model 2 is the best-fitting model in all three groups*

$$P(\text{choice}_{i,t} = j) = \frac{\exp(\boxed{\beta_i} Q_{i,j,t})}{\sum_{k=1}^{Narm} \exp(\boxed{\beta_i} Q_{i,k,t})}$$

$$Q'_{i,j,t} = \mu_{\text{ev},i,j,t} - \boxed{\gamma_i} \cdot \text{cost}_{i,j,t}$$

$\gamma_i$ *- cost sensitivity*

$\beta_i$ *- exploitation vs exploration parameter*

# Locomotion & visual exploration measures

**Locomotion**

- Coordinate Variance
- Enclosing Diameter
- Field Entropy
- Tortuosity
- Total Path Length
- Turning Angle Standard Deviation
- Sum of Distances from Origin
- Maximum Distance from Origin

**Visual Exploration**

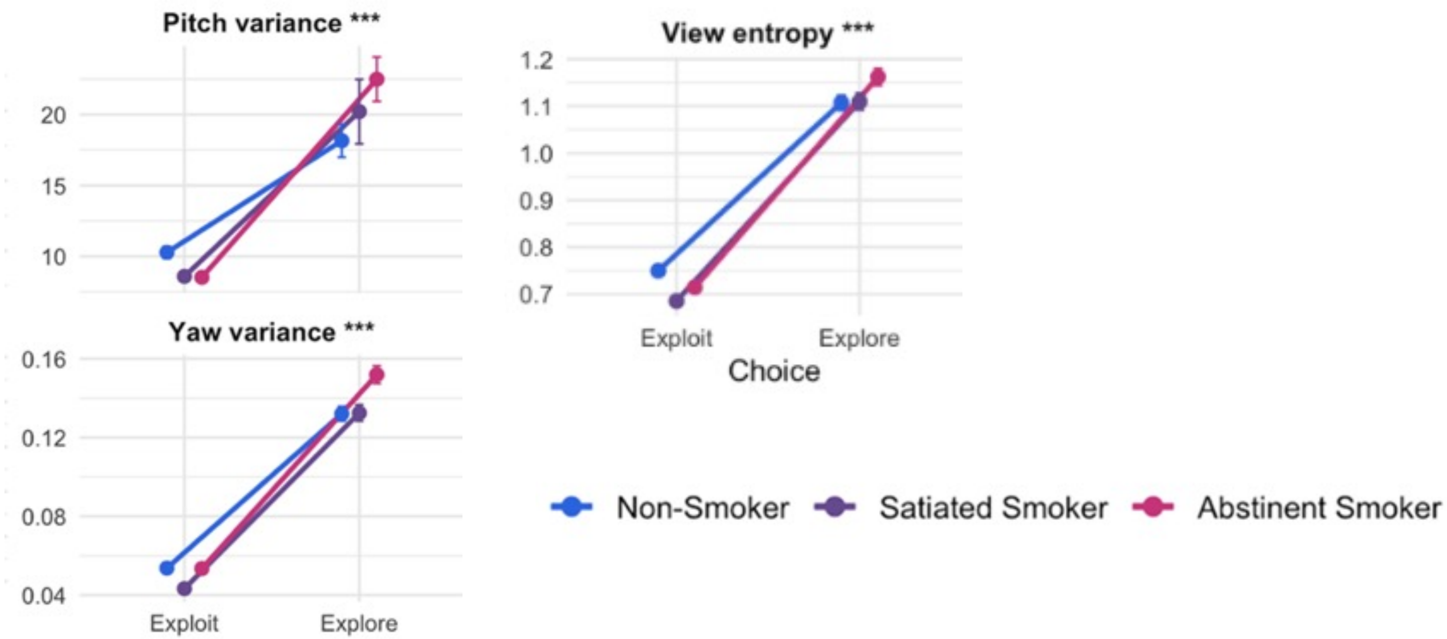
- Yaw/Pitch Variance
- View Entropy
- Total Angular Displacement
- Mean/Max Angular Speed

# *Modeling Results*

- *In 2D MAB, no group difference*
- *In 3D MAB, higher exploitation and reduced cost sensitivity among abstinent smokers*


$\beta_i$ *exploitation vs exploration* $\gamma_i$ *- cost sensitivity*

# *Visual exploration results*




*Abstinent smokers show greater (task-irrelevant) visual exploration during exploration*

***Abstinent smokers***
***= more distracted / disorganized?***

# “Real” rewards in the magnet


vs

Jeung-Hyun Lee

Eun-Hwi Lee

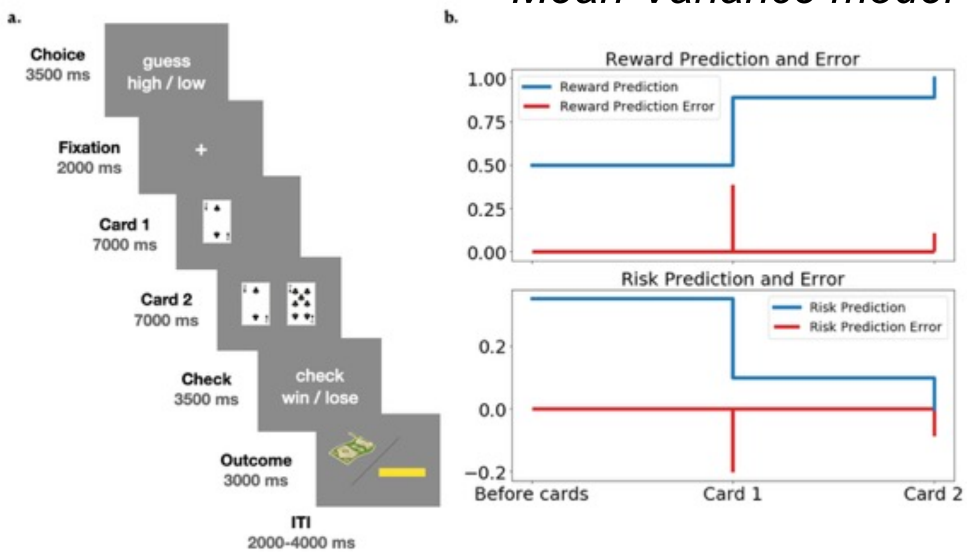
Joshua Brown, Indiana Univ.

Modak et al. (2021) Neuroimage: Clinical
Lim et al (2020) BP: CNNI

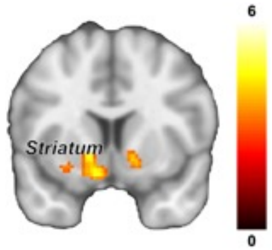

*Joshua Brown, Indiana Univ.*

## *Neural correlates of reward and risk w/ monetary and nicotine rewards*

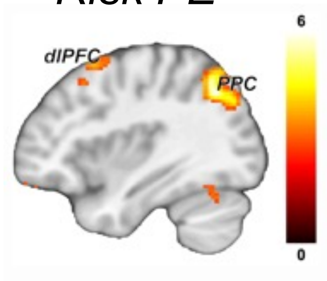
*Mean-Variance model*




*Reward PE*




*Risk PE*



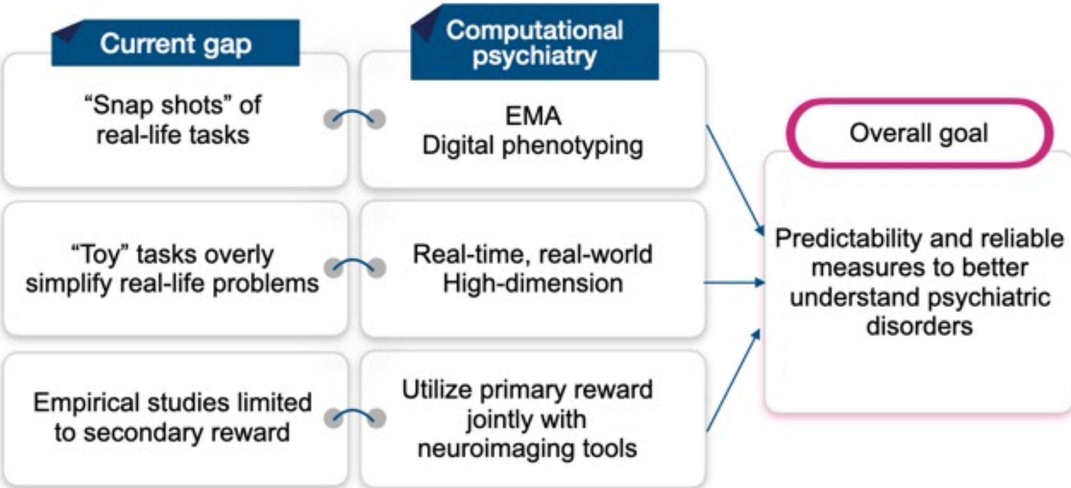

*Jeung-Hyun Lee*

*Eun-Hwi Lee*

*Preuschoff et al (2008) Journal of Neuroscience*
*Preuschoff et al (2006) Neuron*

*Lee, Lee, Brown, & Ahn (2024) CCN2024*

# *Time for naturalistic paradigms!?*



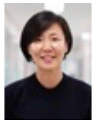

*Ahn, Lee, & Kim (invited review) Current Directions in Psychological Science*
*Lee, Lee, Im, O'Doherty, & Ahn (invited review) Nature Reviews Psychology*

*Real-time Driving task*


*Sang-Ho Lee*


*Min-hwan Oh*

*Explore vs Exploit w/ Minecraft*


*Wonmok Shim*


*Hyeonmin Lee*

*"Real" reward (nicotine)*


*Josh Brown*


*Jeung-Hyun Lee*


*Eunhwi Lee*

*Naturalistic (movie) paradigm*


*Monica Rosenberg*


*Mina Kwon*

# *Challenges ahead of us*

- *(Simple) computational modeling is hard*
- *Phenotypes/(bio)markers are not reliable*
- *Dynamic nature of the human mind*
- *Overly simplified "toy" problems/tasks*
- *"Good" mathematical models for "the mind"*

Article

# A foundation model to predict and capture human cognition

https://doi.org/10.1038/s41586-025-09215-4

Received: 26 October 2024

Accepted: 29 May 2025

Published online: 02 July 2025

Open access

Check for updates

Marcel Binz[1✉], Elif Akata[1], Matthias Bethge[2],
Julian Coda-Forno[1], Peter Dayan[2,4], Can Dem
Thomas L. Griffiths[7], Susanne Haridi[1,8], Aksh
Sreejan Kumar[7], Tobias Ludwig[2,4], Marvin M
Surabhi S. Nath[2,4,8], Joshua C. Peterson[10], Mi
Johannes A. Schubert[4], Luca M. Schulze Bu
Mirko Thalmann[1], Fabian J. Theis[12,13,14], Vuon
Konstantinos Voudouris[1], Robert Wilson[16], K
Huadong Xiong[16] & Eric Schulz[1]

Article

# Discovering cognitive strategies with tiny recurrent neural networks

https://doi.org/10.1038/s41586-025-09142-4

Received: 15 May 2023

Accepted: 12 May 2025

Published online: 02 July 2025

Open access

Check for updates

Li Ji-An[1], Marcus K. Benna[1] & Marcelo G. Mattar[2,3✉]

Understanding how animals and humans learn from experience to make adaptive decisions is a fundamental goal of neuroscience and psychology. Normative modelling frameworks such as Bayesian inference[1] and reinforcement learning[2] provide valuable insights into the principles governing adaptive behaviour. However, the simplicity of these frameworks often limits their ability to capture realistic biological behaviour, leading to cycles of handcrafted adjustments that are prone to researcher subjectivity.

*“Build a model that can predict and simulate human behavior in any domain”*

Tutorial C

## Reinforcement Learning using the hBayesDM Package

### AM Session

8:15 am - 11:45 am

 Ahn

 Lee

 Im

### Info

In this tutorial, participants will learn how to use a Bayesian package called hBayesDM (supporting R and Python) for modeling various reinforcement learning and decision making (RLDM) tasks. A short overview of (hierarchical) Bayesian modeling will be also provided. Participants will also learn important steps and issues to check when reporting modeling results in publications.

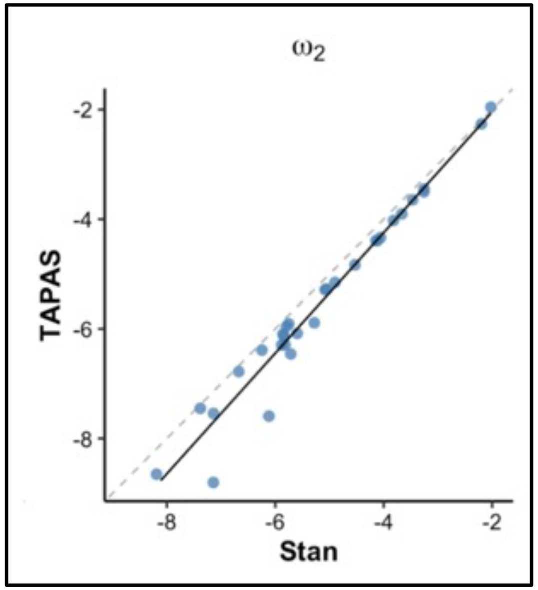
hBayesDM 1.3.0 | Reference | Articles | Changelog

# Hierarchical Bayesian Analysis on Hierarchical Gaussian Filter

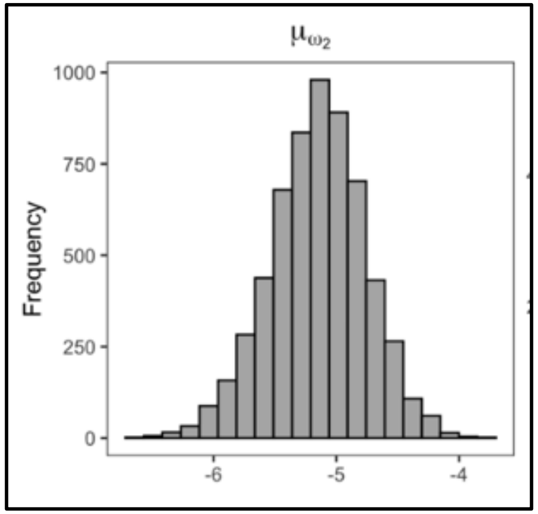
Source: `vignettes/hgf_tutorial.Rmd`

By Jinwoo Jeong, Juha Lee, Yusom Jo, Woo-Young Ahn | September 2, 2025

https://ccs-lab.github.io/hBayesDM/articles/hgf_tutorial.html




## *(Hierarchical) Bayesian analysis of HGF*




*Jinwoo Jeong*

*Juha Lee*

*Yusom Jo*

Computational Clinical Science Laboratory


*Thanks!*

ccs-lab.github.io /
happylaboratory.org