Statistical Modeling for Human Reward-based Behavioral Task Data

Xingche Guo

Department of Statistics, University of Connecticut

Presented at

The 10th International Forum on Statistics
Renmin University of China
July 6 2025

Human Behavioral Tasks

Definition: a structured experimental paradigm designed to systematically measure specific aspects of human cognition, perception, emotion, or decision-making through observable behaviors such as choices, reaction times, error rates, or movements.

Key features:

- **Controlled environment: Tasks are typically performed in a lab or digital setting where variables can be precisely manipulated.**

- **Target cognitive process: Each task is designed to probe a particular mental function (e.g., attention, memory, reward processing, inhibitory control).**

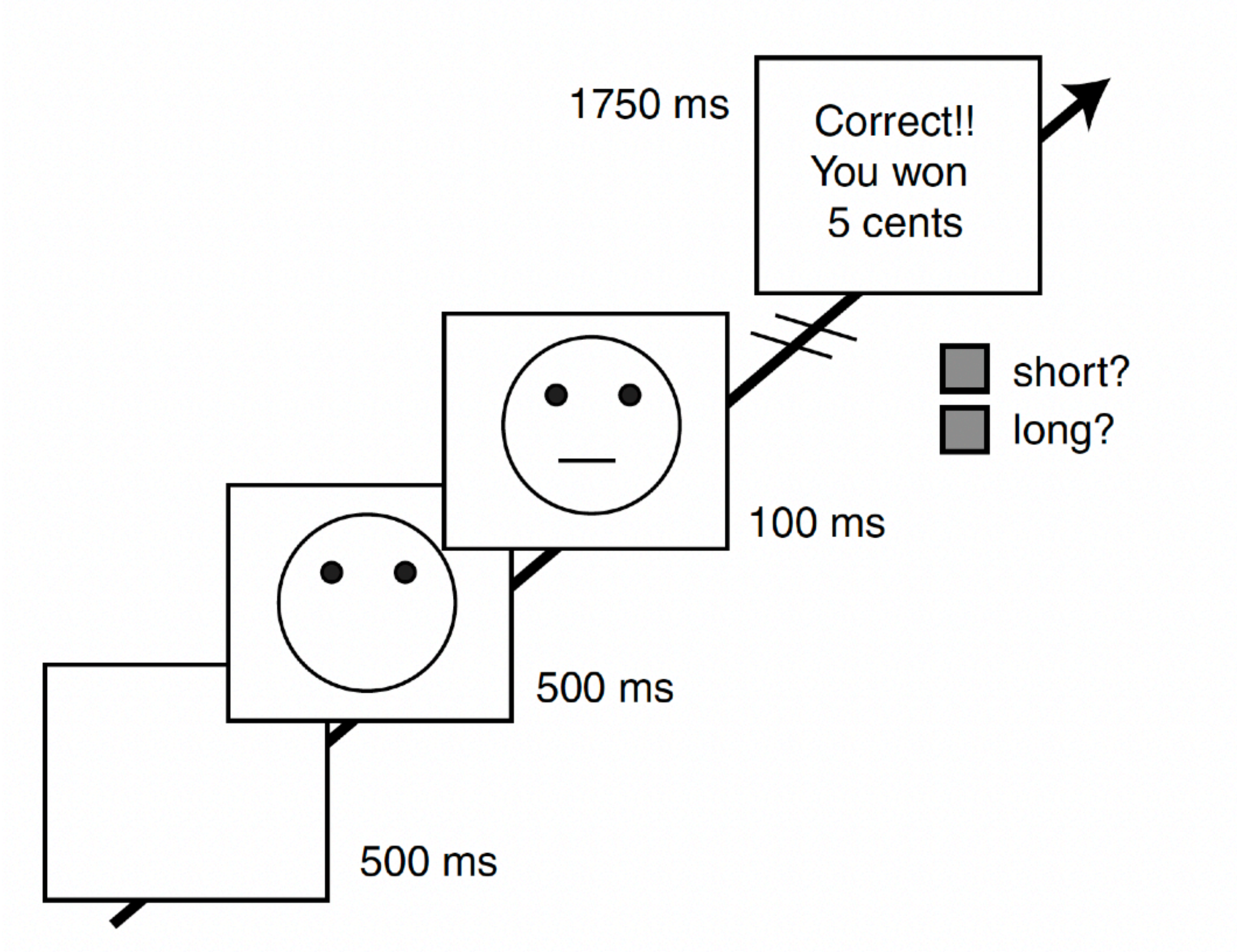
- **Outputs: Behavioral responses (e.g., response time, accuracy, choice patterns) serve as measurable data for analysis.**

- **Reproducibility: Tasks are standardized to allow replication and comparison across individuals or groups.**

Probabilistic reward task (PRT):

A computer-based behavioral experiment that measures the subject's ability to modify behavior in response to rewards.

(Pizzagalli et al., 2005)



Participant's goal: learn from the PRT system (to maximize rewards).

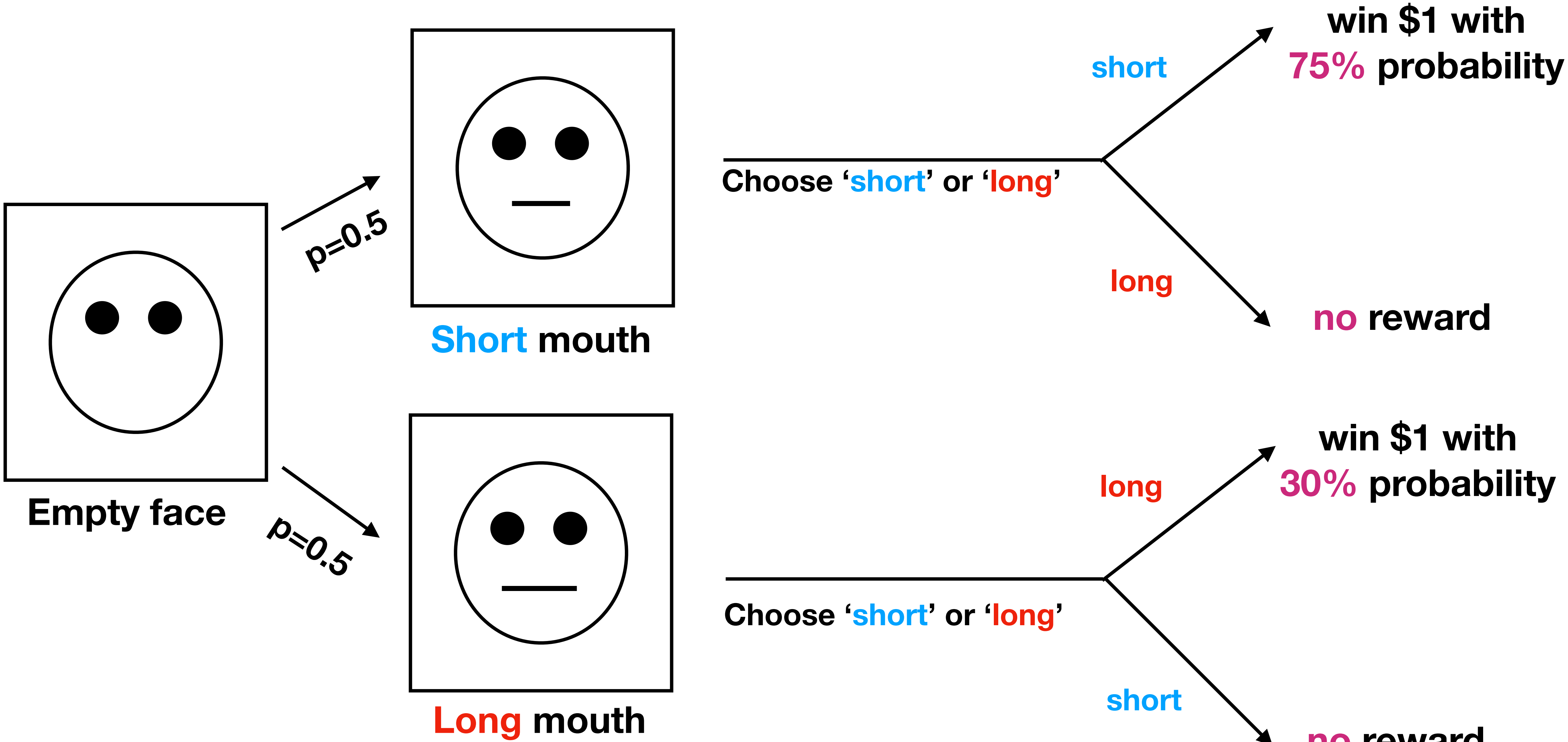
Our goal: understand how the participant learns the PRT system (not interested in PRT system).

Probabilistic reward task (PRT):

Demo (single trial)

You are told the task is to identify the correct mouth.

You don't know the reward generating mechanism.



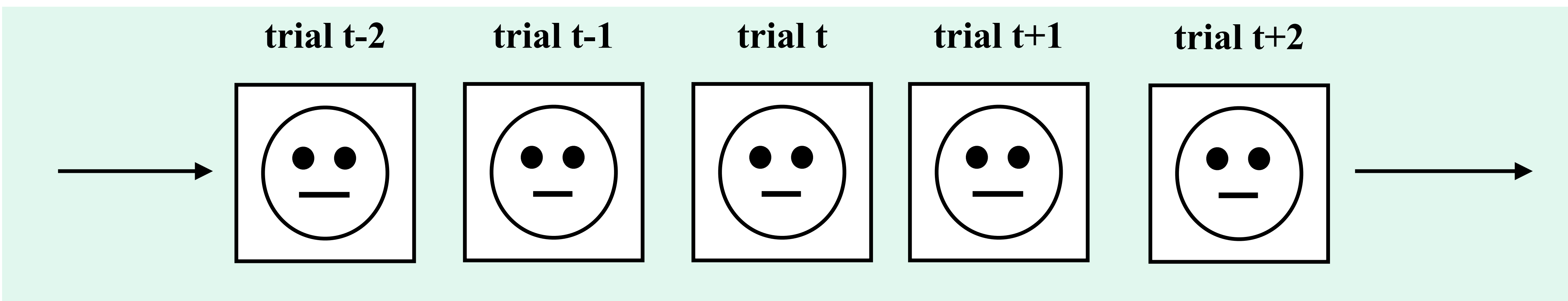
Small difference in mouth size

Rewards are imbalanced

Probabilistic reward task (PRT):

Demo (multiple trials)

A PRT session



States: (S)	long S_{t-2}	short S_{t-1}	short S_t	long S_{t+1}	short S_{t+2}	……
Actions: (A)	long A_{t-2}	short A_{t-1}	short A_t	short A_{t+1}	short A_{t+2}	
Rewards: (R)	0 R_{t-2}	1 R_{t-1}	1 R_t	0 R_{t+1}	0 R_{t+2}	
Prob. of winning:	30%	75%	75%	0%	75%	

Assess decision-making processes relevant to mental disorders via Probabilistic reward task

Observation:

An individual's learning ability and decision-making may be altered by MDD *(Pizzagalli, et al. 2005)*.

Try to Answer:

How does MDD affect the decision-making and reward learning?

- Learn slow?
- Not sensitive to reward?
- Easy to distract?
- etc…

EMBARC Study:

A clinical trial for exploring how biomarkers affect the treatment outcome for MDD *(Trivedi et al., 2016)*.

Data Types

- **Demographical and clinical data**
- **Neuroimaging data:**
  - **Task EEG/fMRI**
  - **Resting-state EEG/fMRI**
  - **etc…**
- **Human behavioral data:**
  - **Probabilistic reward task *(Pizzagalli et al., 2005)***
  - **Emotion conflict task *(Etkin et al., 2006)***
  - **etc…**

Experimental Design

- **MDD group vs Health Control group (Today's focus)**
- **In MDD group: Treatment vs Placebo**

Classical RL models *(Huys et al. 2013)*

Problem setups for PRT

Problem size: subjects ($i = 1, \ldots, n$) from a group, trials ($t = 1, \ldots, T$) for each session.

State space (S): {0, 1}: 0 = ‘long mouth’ (lean); 1 = ‘short mouth’ (rich).

Action space (A): {0, 1}: 0 = ‘long mouth’; 1 = ‘short mouth’.

Reward space (R): {0, 1}: 0 = ‘no reward’; 1 = ‘win reward’.

Data for one group: $\left\{\ldots, S_{it}, A_{it}, R_{it}, \ldots\right\},\ i = 1,\ldots,n;\ \ t = 1,\ldots,T.$

Classical RL models *(Huys et al. 2013)*

Q-learning model

Expected reward (own estimate):

$$Q_{it}(a,s) = \mathbb{E}^{(\text{est})}\left(R_{it} \mid A_{it} = a, S_{it} = s\right)$$

Minimize reward prediction error: $R_{it} - Q_{it}(a,s)$

Update expected reward (gradient descent):

$$Q_{i,t+1}(a,s) = Q_{it}(a,s) + \beta_i \left(R_{it} - Q_{it}(a,s)\right)$$
$$\left(a = A_{it},\ s = S_{it}\right)$$

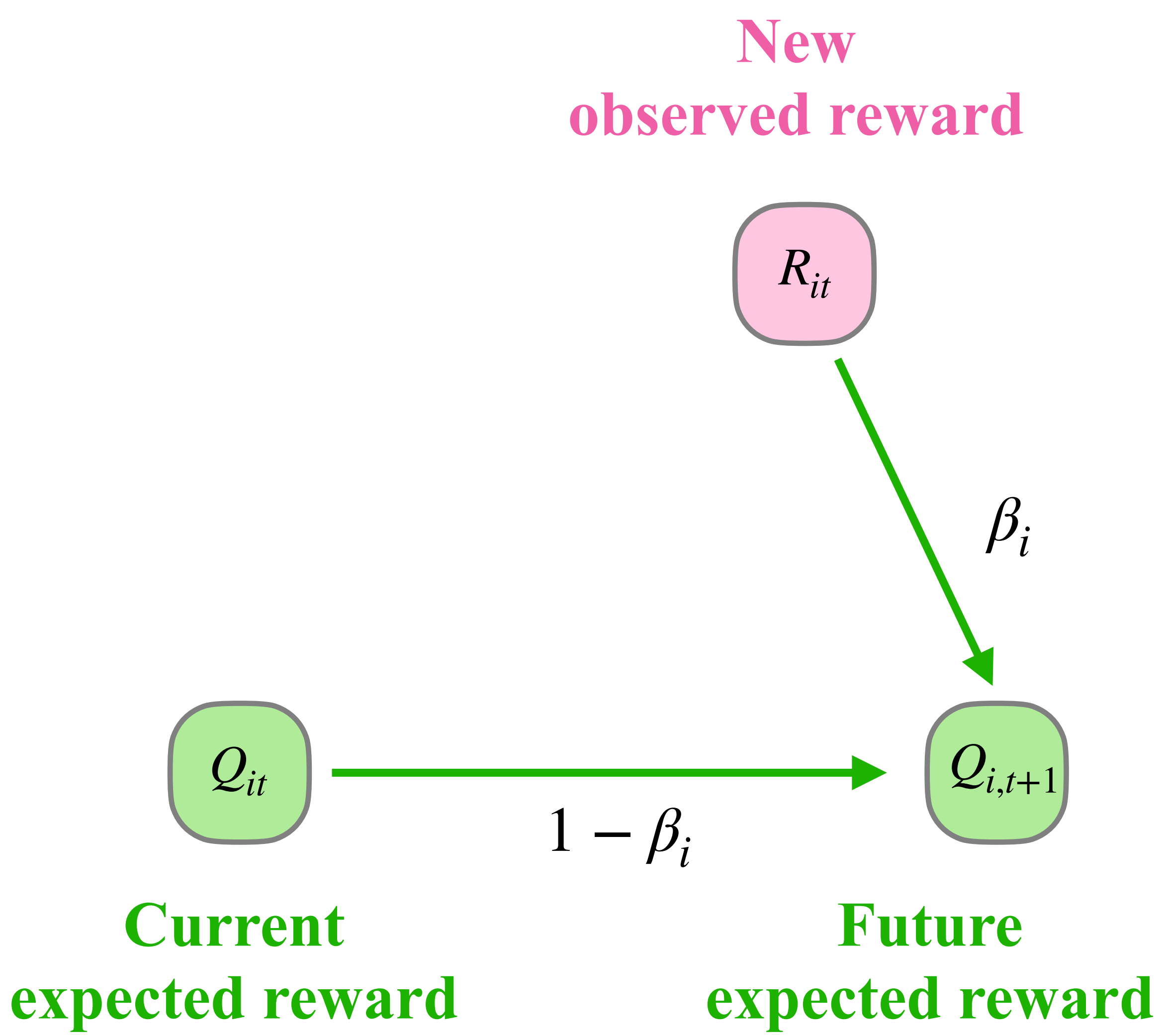
Learning rate: $\beta_i \in (0,1)$

Another view (weighted sum):

$$Q_{i,t+1}(a,s) = \left(1 - \beta_i\right) Q_{it}(a,s) + \beta_i R_{it}$$

$\beta_i \to 0$, no update,

$\beta_i \to 1$, no memory

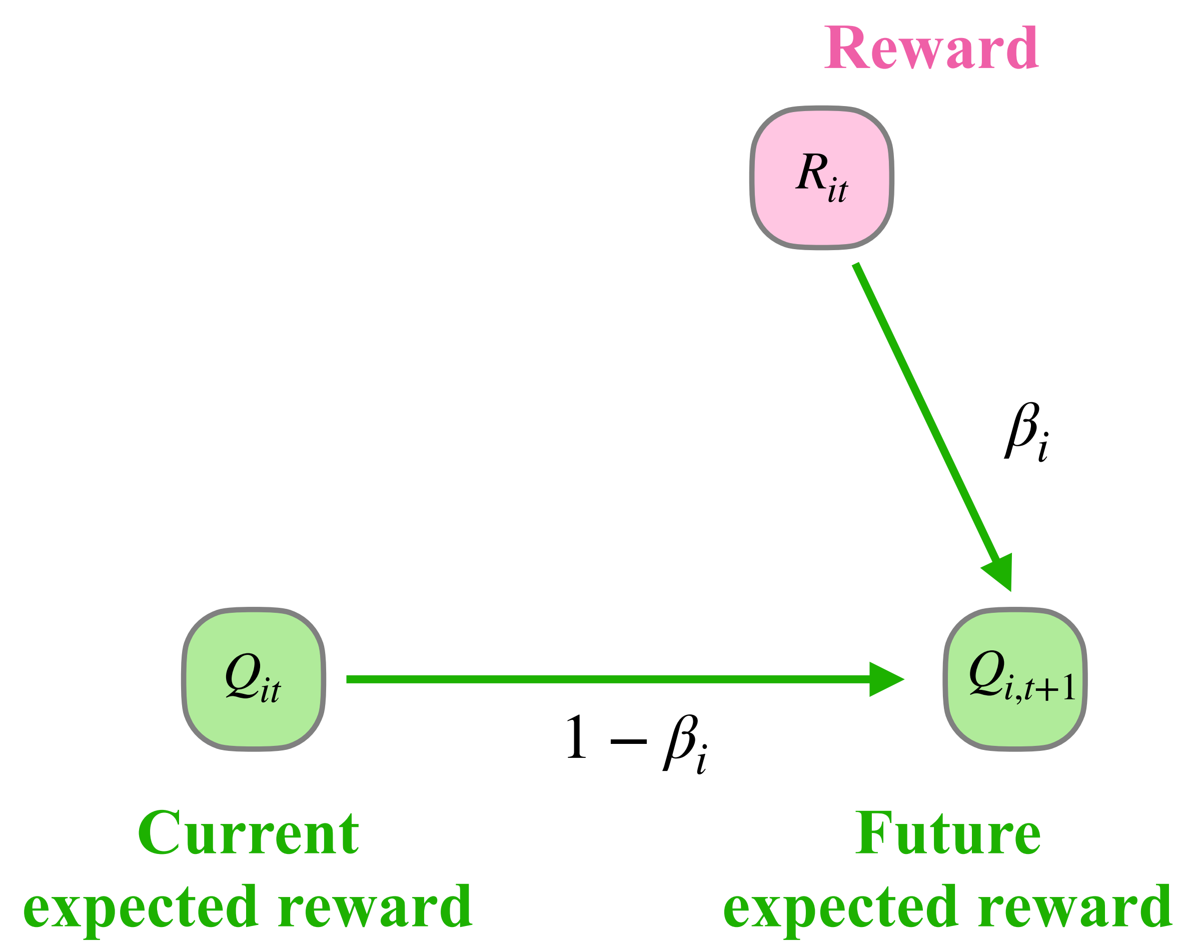


Classical RL models *(Huys et al. 2013)*

Decision making model

Contrast of expected rewards for action 1 and 0 at

the given state: $Z_{it} = Q_{it}(1,S_{it}) - Q_{it}(0,S_{it})$



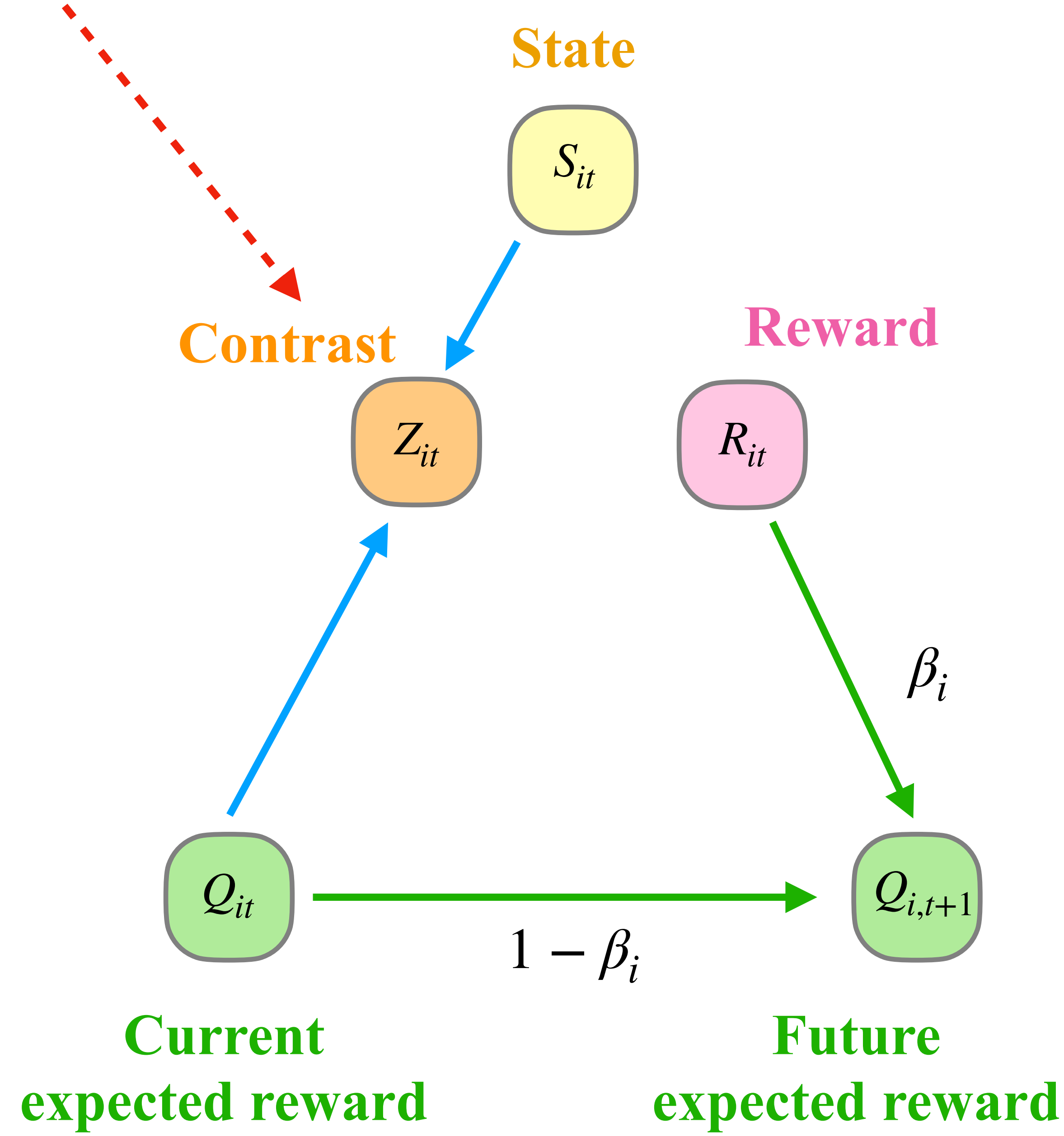
Classical RL models *(Huys et al. 2013)*

Decision making model

Contrast of expected rewards for action 1 and 0 at

the given state: $Z_{it} = Q_{it}(1,S_{it}) - Q_{it}(0,S_{it})$

weighing between two actions



Classical RL models *(Huys et al. 2013)*

Decision making model

Contrast of expected rewards for action 1 and 0 at

the given state: $Z_{it} = Q_{it}(1, S_{it}) - Q_{it}(0, S_{it})$

weighing between two actions

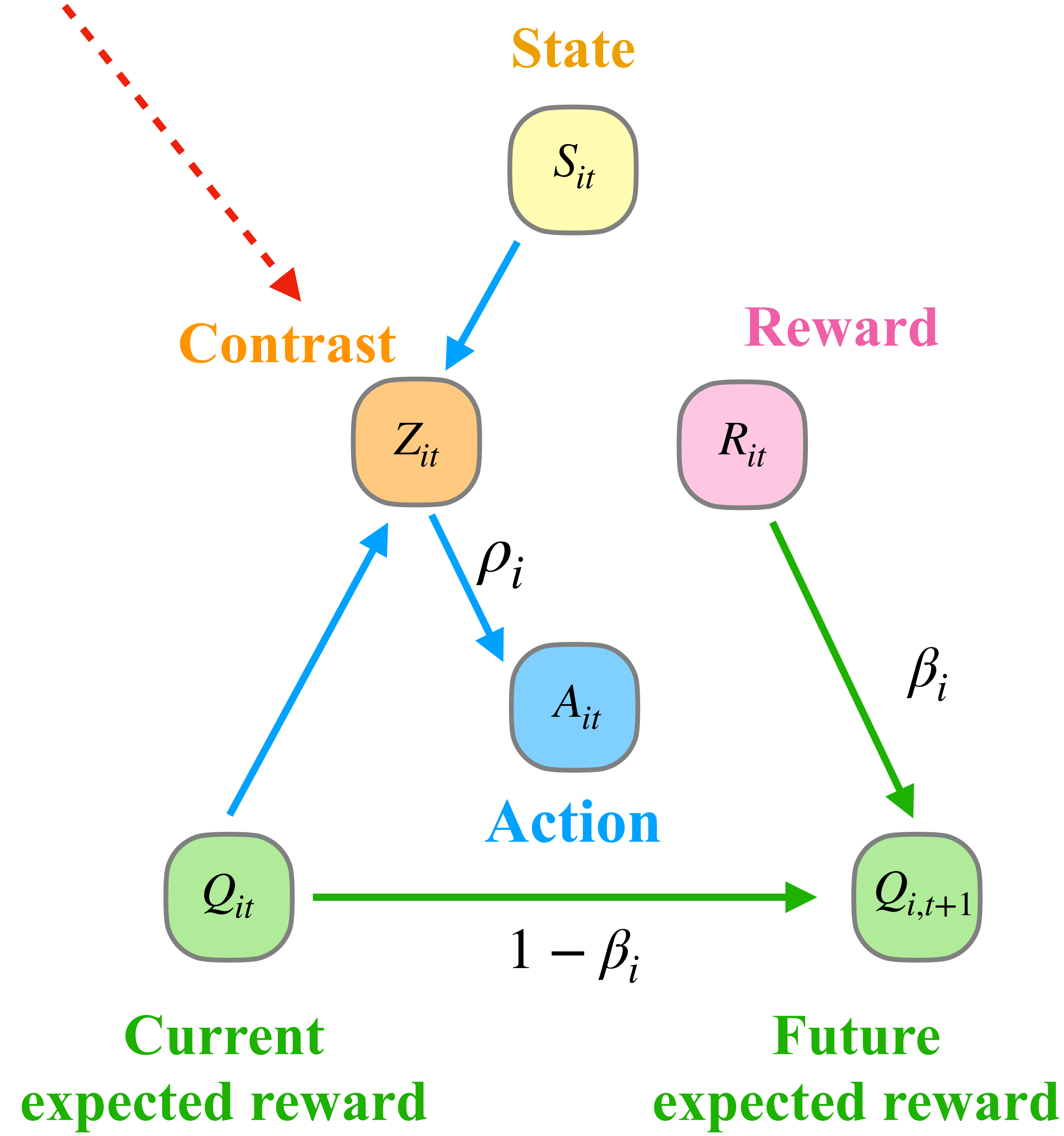
Conditional probability of taking action 1:

$$\text{logit } P\left(A_{it} = 1 \mid Z_{it}\right) = \rho_i Z_{it}$$

Reward sensitivity: $\rho_i > 0$:

if $\rho_i \to \infty$, $P(A_{it} = 1 \mid Z_{it} = 1) \to 1$,

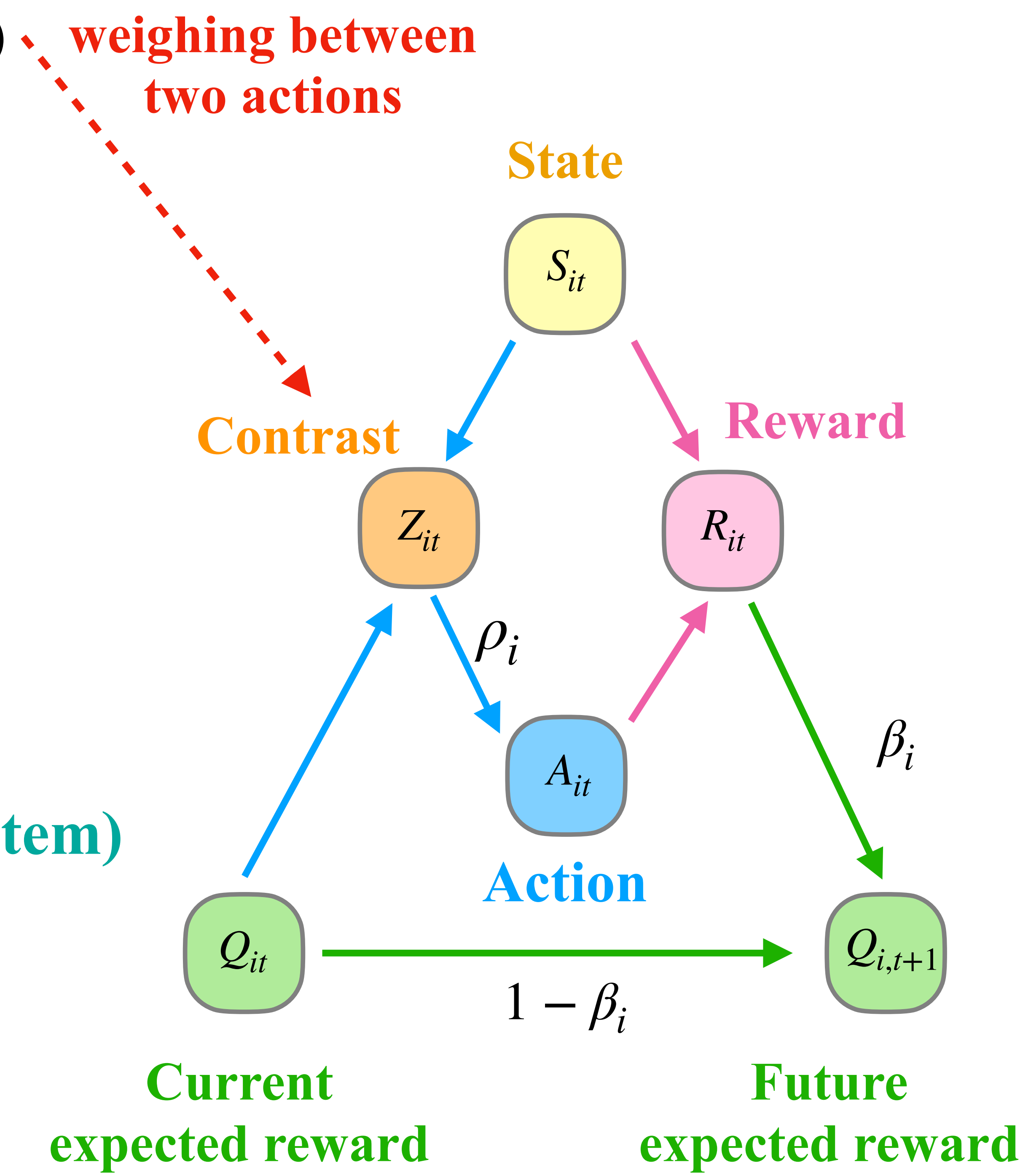
if $\rho_i \to 0$, $P(A_{it} = 1 \mid Z_{it} = 1) \to 0.5$.



Classical RL models *(Huys et al. 2013)*

Decision making model

Contrast of expected rewards for action 1 and 0 at the given state: $Z_{it} = Q_{it}(1, S_{it}) - Q_{it}(0, S_{it})$



Conditional probability of taking action 1:

$$\text{logit } P\left(A_{it} = 1 \mid Z_{it}\right) = \rho_i Z_{it}$$

Reward sensitivity: $\rho_i > 0$:

if $\rho_i \to \infty$, $P(A_{it} = 1 \mid Z_{it} = 1) \to 1$,

if $\rho_i \to 0$, $P(A_{it} = 1 \mid Z_{it} = 1) \to 0.5$.

Reward generating model (from PRT system)

$$P(R_{it} = 1 \mid S_{it} = A_{it} = 1) = 0.75$$

$$P(R_{it} = 1 \mid S_{it} = A_{it} = 0) = 0.3$$

$$P(R_{it} = 1 \mid S_{it} \neq A_{it}) = 0$$

Semiparametric RL model

Guo, X., Zeng, D., Wang, Y. (2024). A Semiparametric Inverse Reinforcement Learning Approach to Characterize Decision Making for Mental Disorders. *Journal of the American Statistical Association.*

Semiparametric RL model

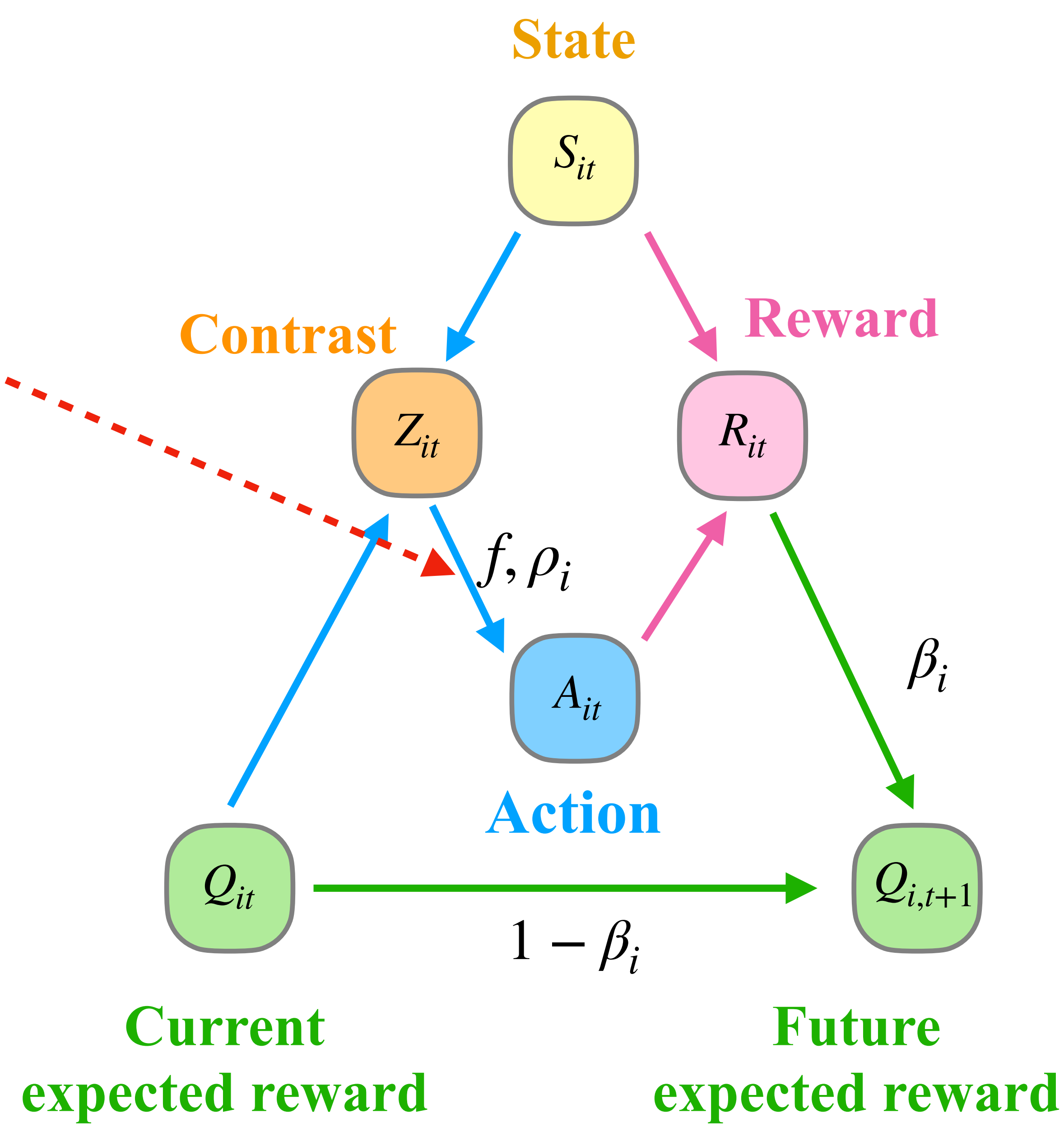
Decision making model (Our contribution)

Contrast of expected rewards for action 1 and 0 at the given state: $Z_{it} = Q_{it}(1, S_{it}) - Q_{it}(0, S_{it})$

Conditional probability of taking action 1:

$$\text{logit } P\left(A_{it} = 1 \mid Z_{it}\right) \quad = f(\, \rho_i Z_{it} \,)$$

Reward sensitivity function: $f(\,\cdot\,)$



Semiparametric RL model

Decision making model (Our contribution)

Contrast of expected rewards for action 1 and 0 at the given state: $Z_{it} = Q_{it}(1, S_{it}) - Q_{it}(0, S_{it})$

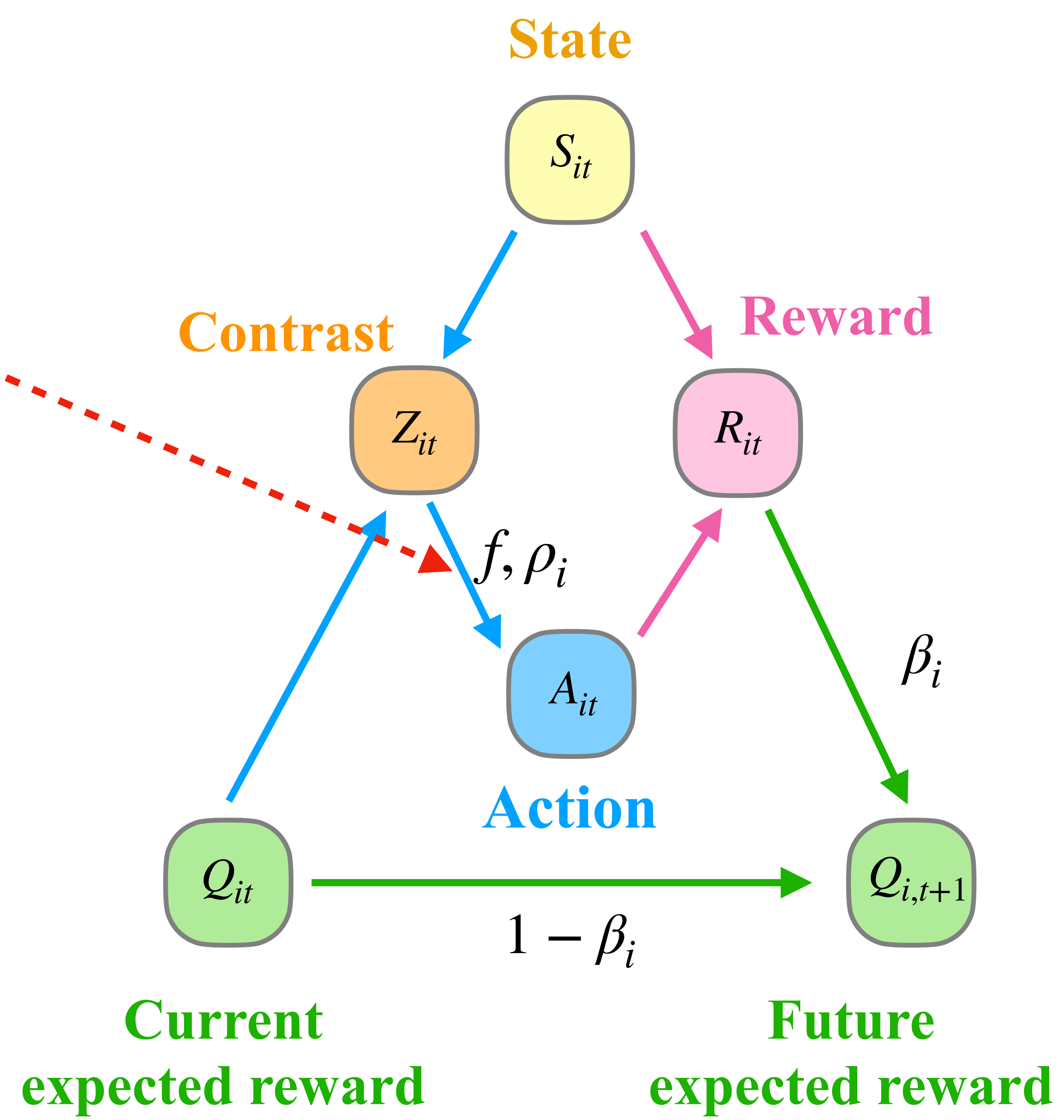
Conditional probability of taking action 1:

$$\text{logit } P\left(A_{it} = 1 \mid Z_{it}\right) = f(\rho_i Z_{it})$$

Reward sensitivity function: $f(\cdot)$

We further assume:

(i). $f(\cdot)$ **non-decreasing; (ii)** $f(0) = 0$



Semiparametric RL model

Decision making model (Our contribution)

Contrast of expected rewards for action 1 and 0 at the given state: $Z_{it} = Q_{it}(1, S_{it}) - Q_{it}(0, S_{it})$

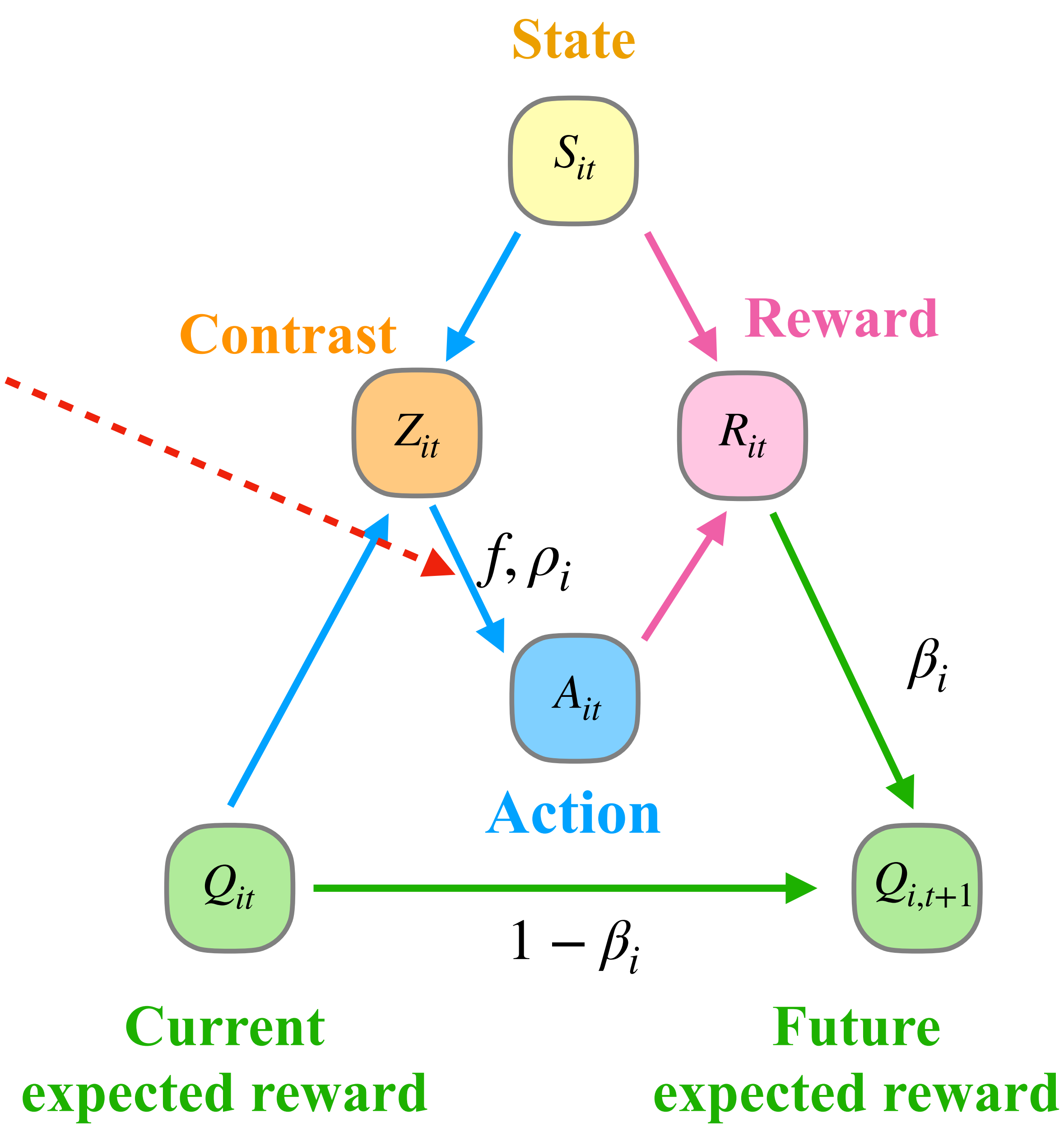
Conditional probability of taking action 1:

$$\text{logit } P\left(A_{it} = 1 \mid Z_{it}\right) = f(\rho_i Z_{it})$$

Reward sensitivity function: $f(\cdot)$

We further assume:

(i). $f(\cdot)$ **non-decreasing; (ii)** $f(0) = 0$



Properties:

(i). $P(A_{it} = 1 \mid Z_1) \geq P(A_{it} = 1 \mid Z_2)$, if $Z_1 \geq Z_2$

(ii). $P(A_{it} = 1 \mid Z_{it} = 0) = 0.5$

Application to EMBARC Study

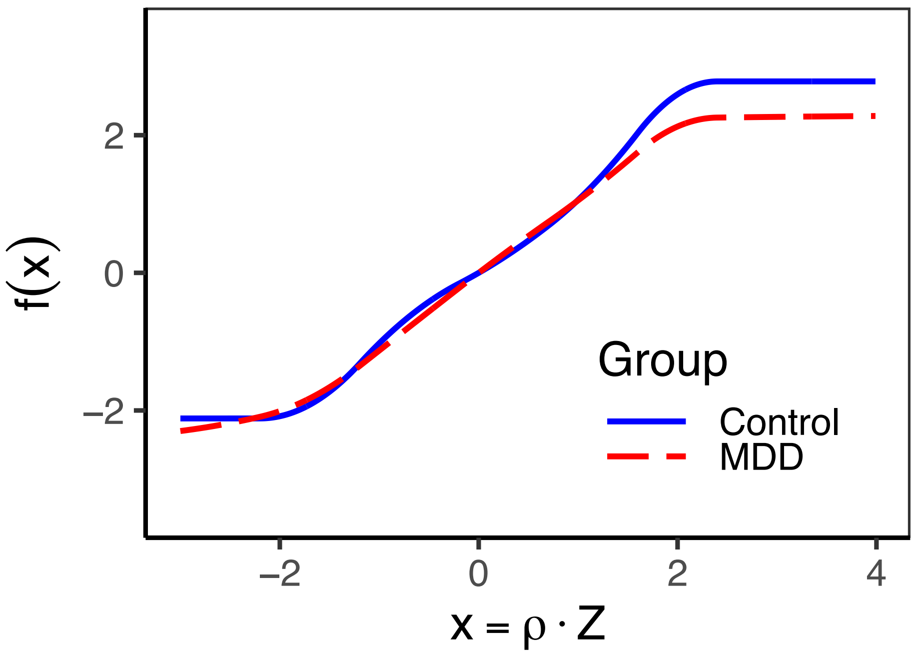
Results: MDD vs Control

Learning Rate:

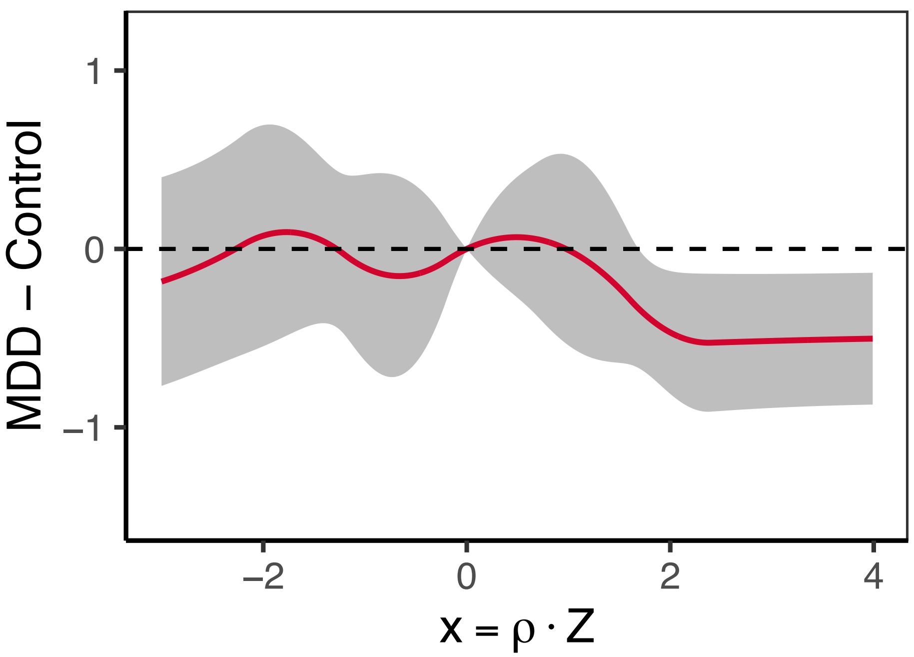
The difference of learning rate between MDD group and Control group is not significant.

Reward sensitivity function $f(\cdot)$:

$f(\cdot)$ for MDD and Control



$f(\cdot)$ difference between MDD and Control



- **Nonlinear (a floor and ceiling effect).**
- **The Control group has a larger reward sensitivity function compared to the MDD group when the contrast is a large positive value.**

What does the floor and ceiling effect of f(.) tell us?

Consider 3 decision-making models:

Visualize $P(A = 1 \mid Z)$

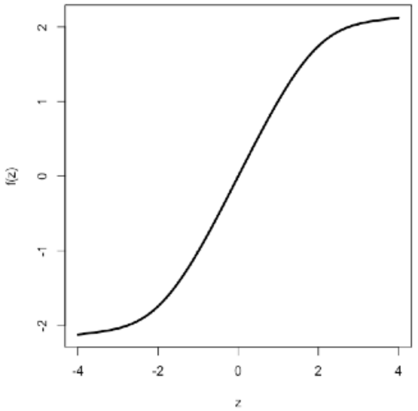
Classical RL:

$$P(A = 1 \mid Z) = \frac{1}{1 + \exp(-Z)}$$

Semiparametric RL:

$$P(A = 1 \mid Z) = \frac{1}{1 + \exp(-f(Z))}$$

$f(z)$:



Mixture (Classical RL and random):

$$P(A = 1 \mid Z, U = 1) = \frac{1}{1 + \exp(-Z)}$$

$$P(A = 1 \mid Z, U = 0) = 0.5, \quad P(U = 1) = 0.8$$

What does the floor and ceiling effect of f(.) tell us?

Consider 3 decision-making models:

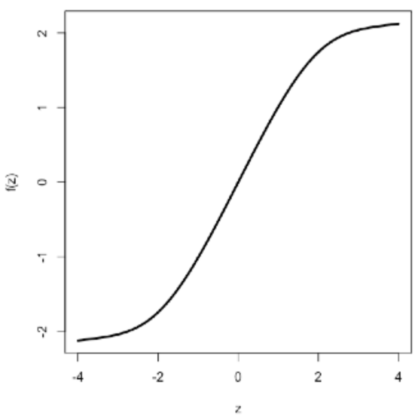
Classical RL:

$$P(A = 1 \mid Z) = \frac{1}{1 + \exp(-Z)}$$

Semiparametric RL:

$$P(A = 1 \mid Z) = \frac{1}{1 + \exp(-f(Z))}$$

$f(z)$:

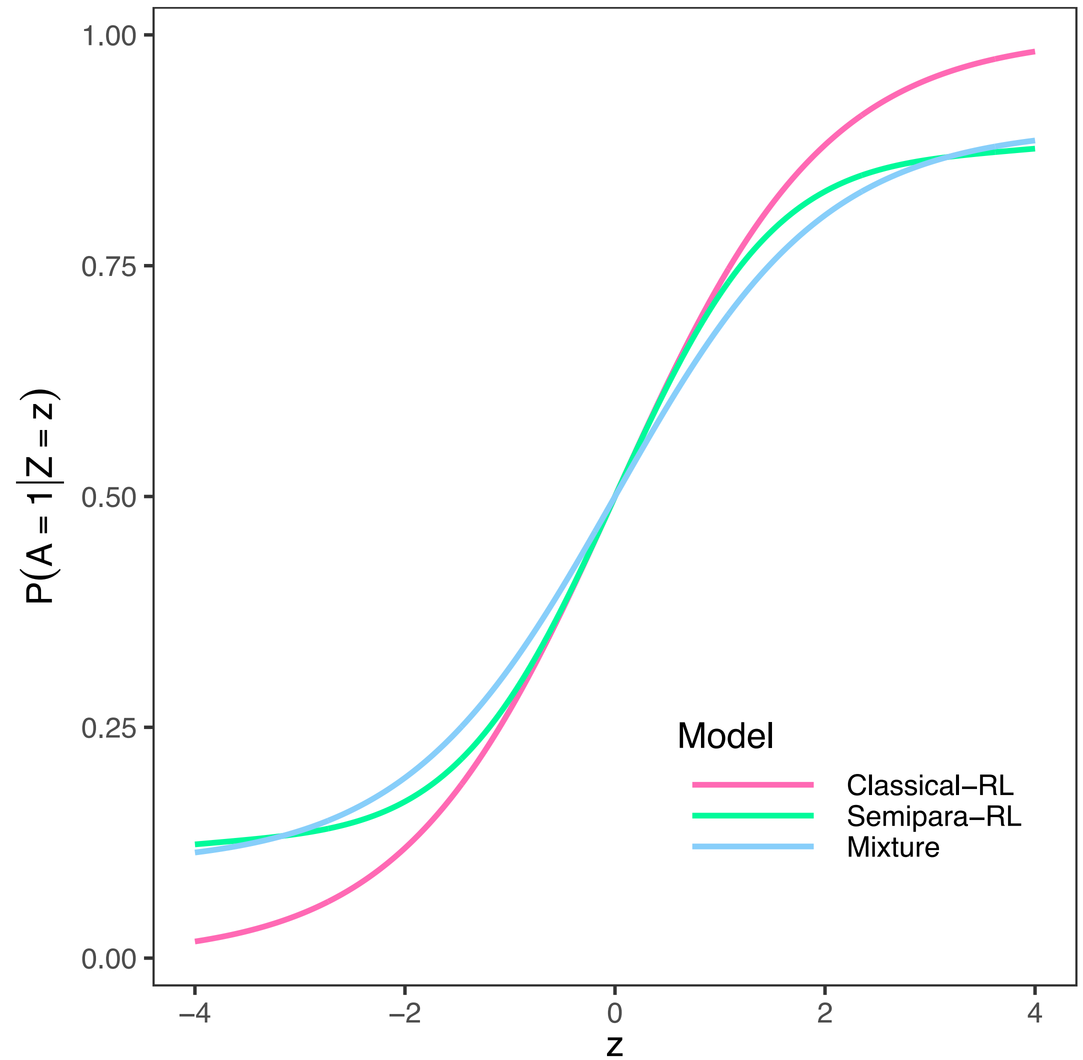


Mixture (Classical RL and random):

$$P(A = 1 \mid Z, U = 1) = \frac{1}{1 + \exp(-Z)}$$

$$P(A = 1 \mid Z, U = 0) = 0.5, \quad P(U = 1) = 0.8$$

Visualize $P(A = 1 \mid Z)$



What does the floor and ceiling effect of f(.) tell us?

Consider 3 decision-making models:

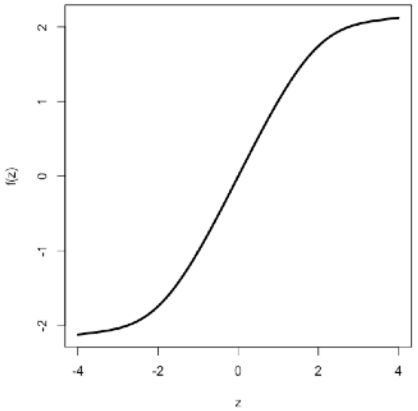
Classical RL:

$$P(A = 1 \mid Z) = \frac{1}{1 + \exp(-Z)}$$

Semiparametric RL:

$$P(A = 1 \mid Z) = \frac{1}{1 + \exp(-f(Z))}$$

$f(z)$:

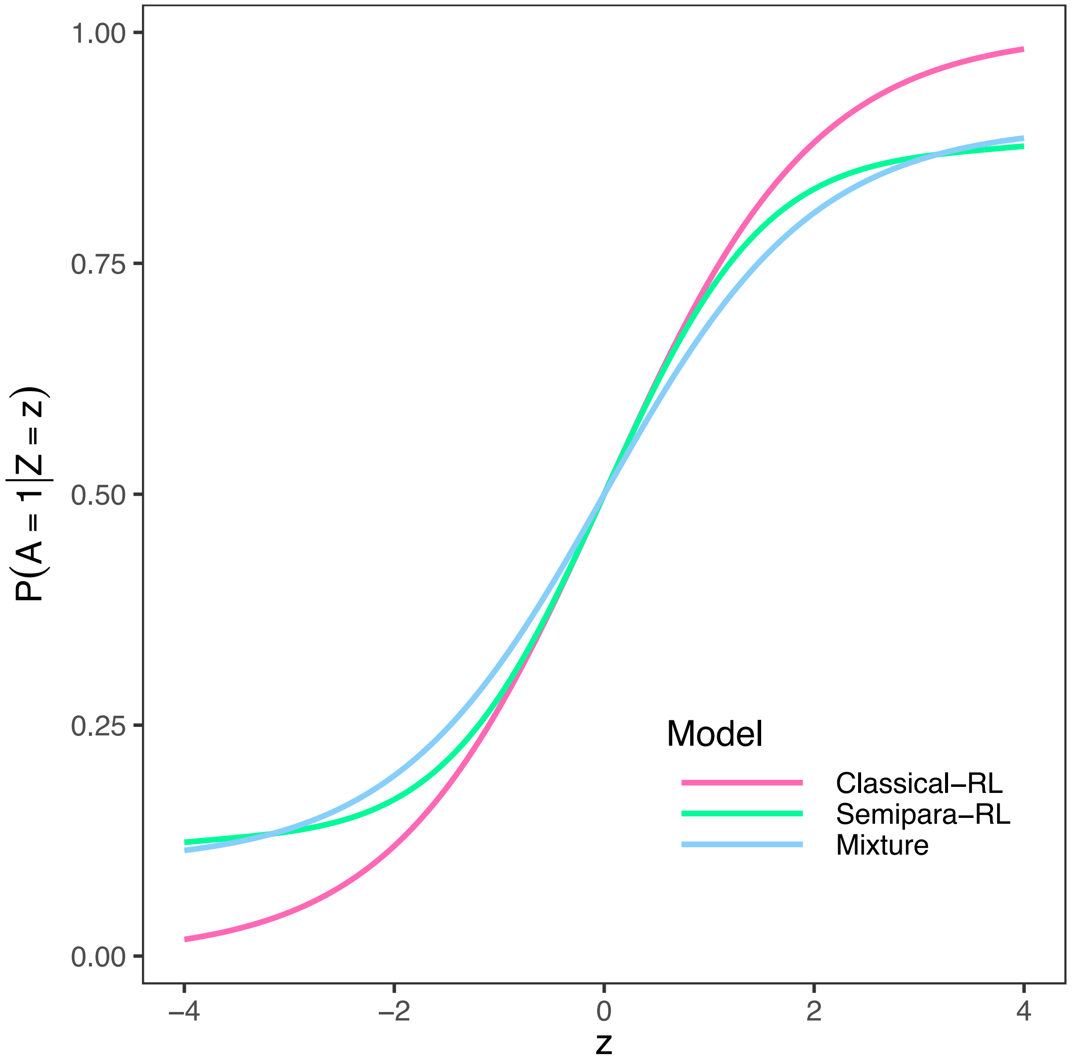


Mixture (Classical RL and random):

$$P(A = 1 \mid Z, U = 1) = \frac{1}{1 + \exp(-Z)}$$

$$P(A = 1 \mid Z, U = 0) = 0.5, \quad P(U = 1) = 0.8$$

Visualize $P(A = 1 \mid Z)$



Question: Is decision-making more complex than a single RL model?

(*Iigaya et al., 2018*; *Ashwood et al., 2022*) provide evidence that subjects employ multiple decision-making strategies for decision-making.

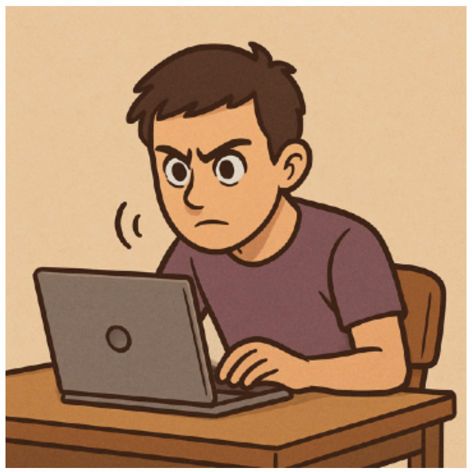
RL-HMM framework

Guo, X., Zeng, D., Wang, Y. (2025). HMM for Discovering Decision-Making Dynamics Using Reinforcement Learning Experiments. *Biostatistics*

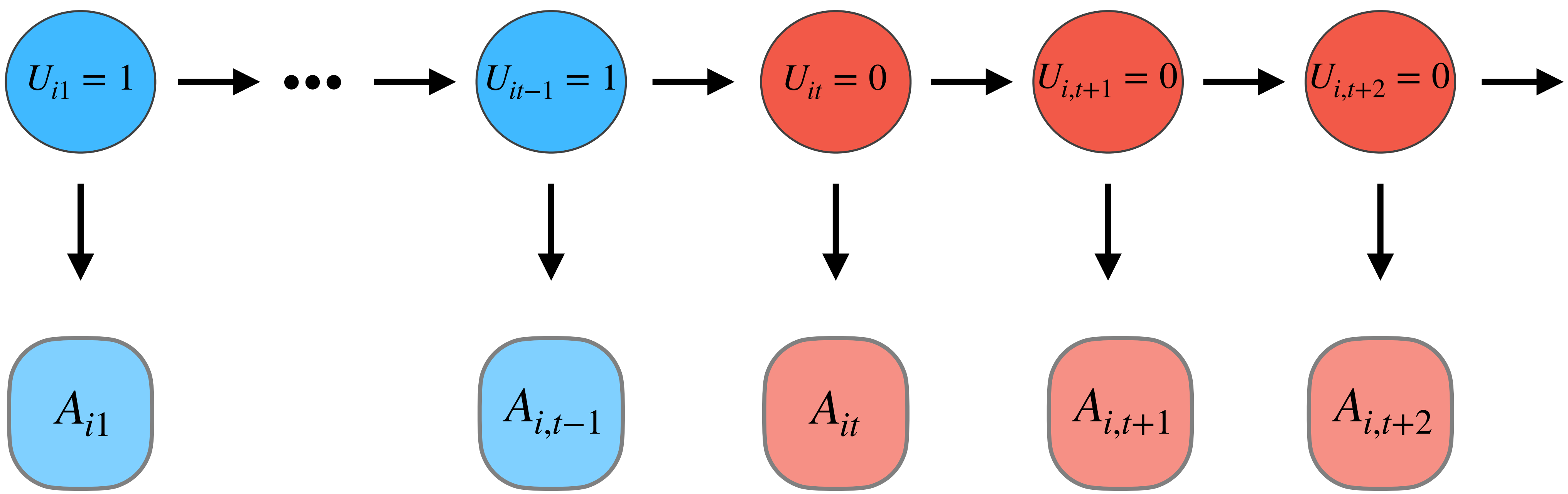
RL-HMM framework

The hidden Markov model structure for two decision-making phases

engaged



lapse



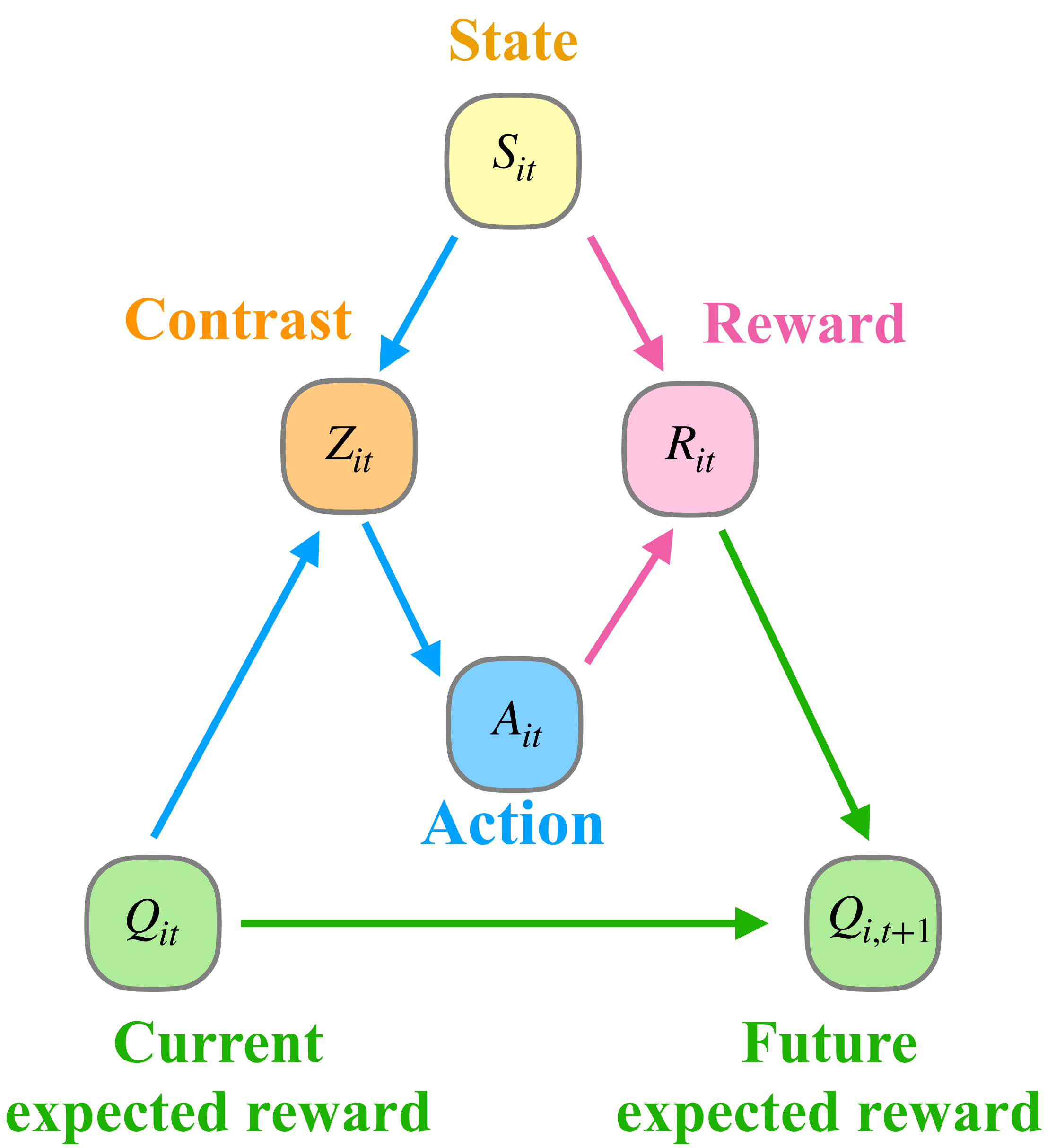
Decisions via expected reward Q.

Random decisions.

RL-HMM framework

engaged vs lapse

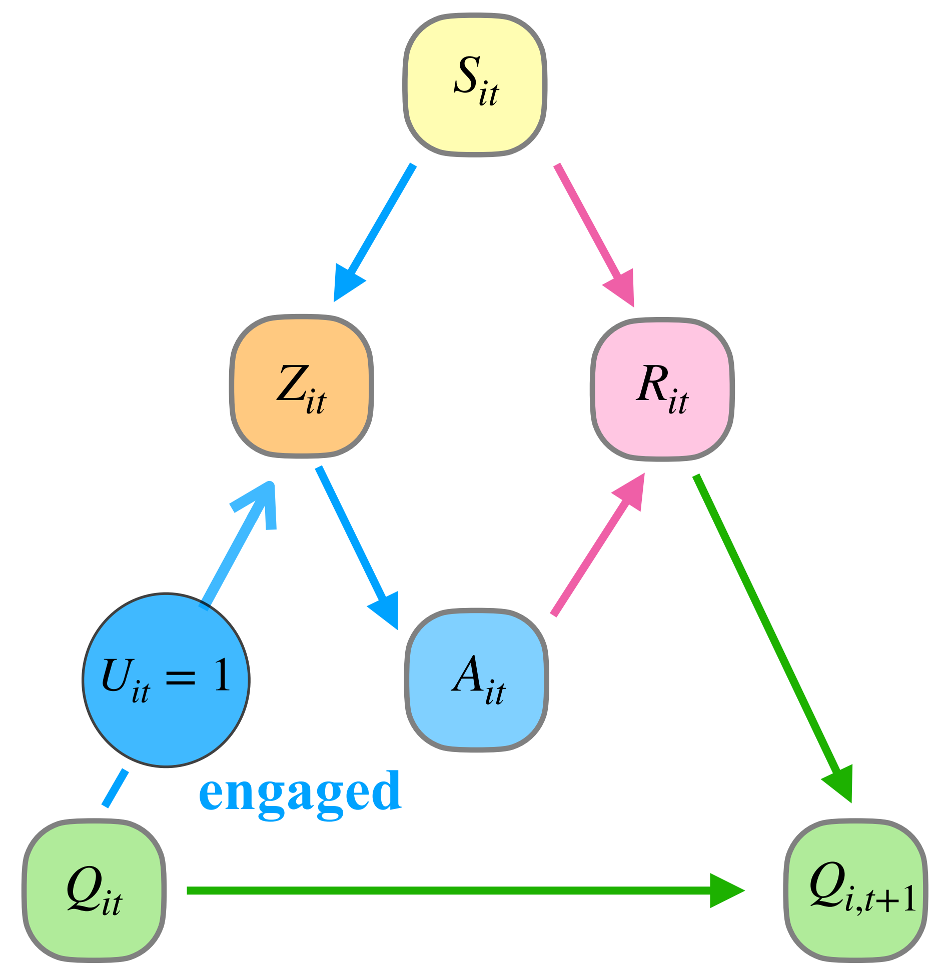
RL framework



RL-HMM framework

Learning strategy: engaged

$U_{it} = 1$



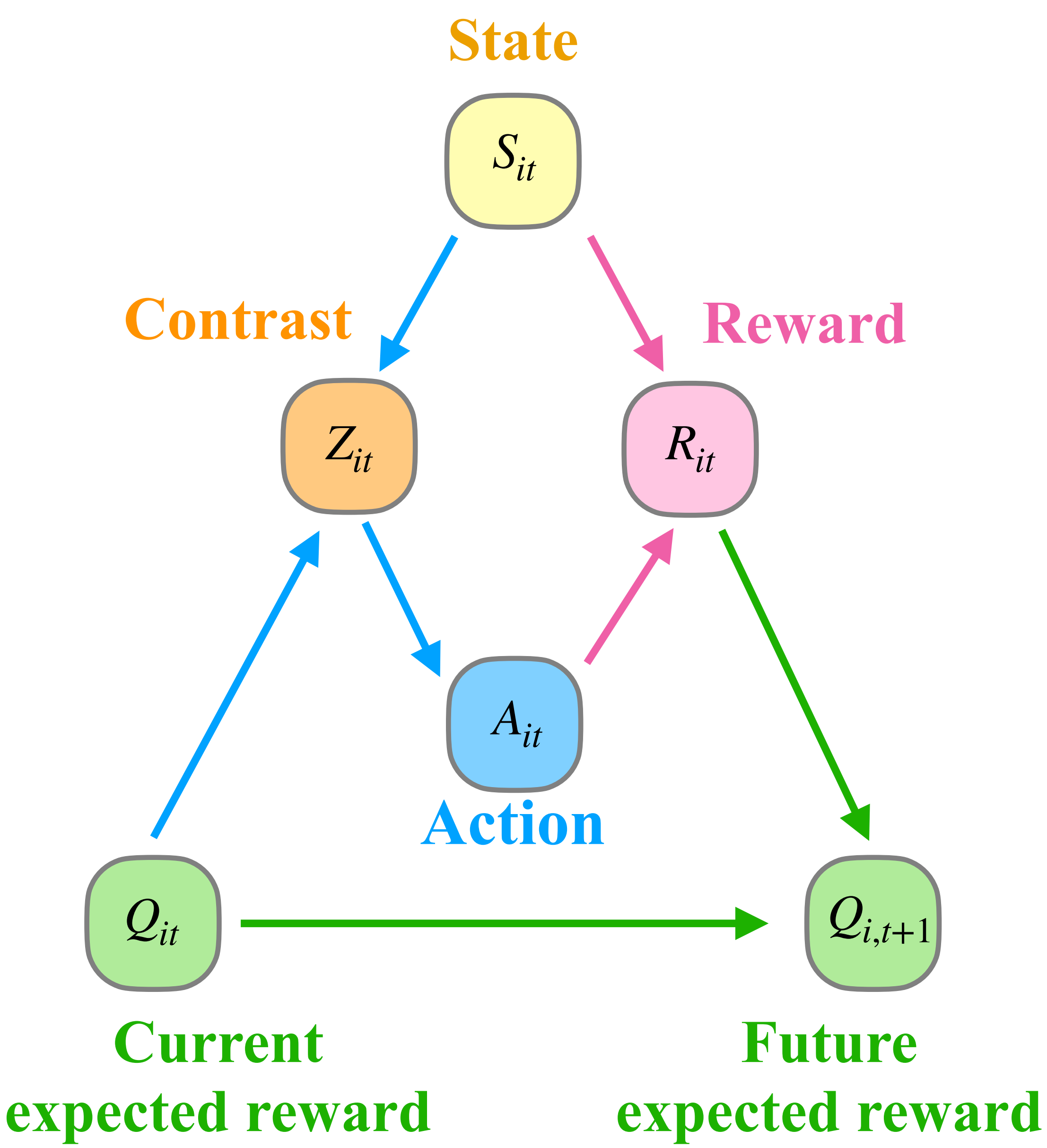
The same decision-making model as the RL framework.

$$\text{logit } P\left(A_{it} = 1 \mid U_{it} = 1,\, Z_{it}\right) = \rho Z_{it}$$

RL-HMM framework

engaged vs lapse

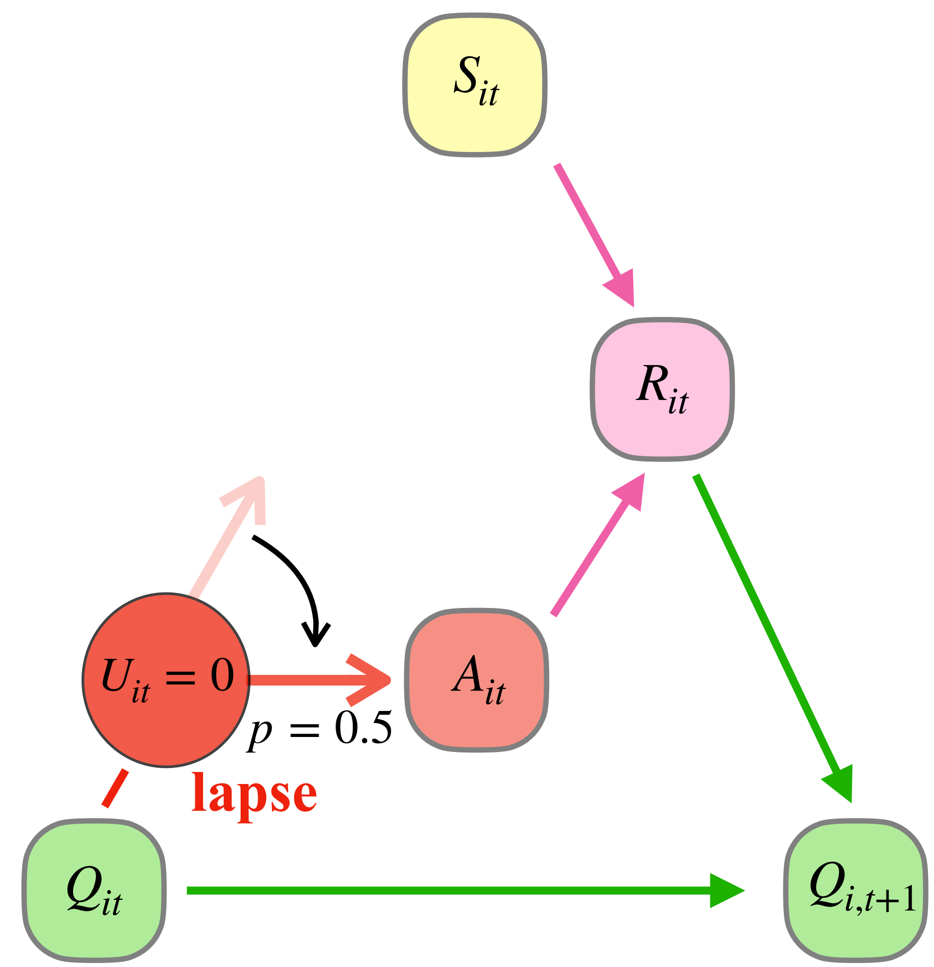
RL framework



RL-HMM framework

Learning strategy: lapse

$U_{it} = 0$



Random decisions.

$$P\left(A_{it} = 1 \mid U_{it} = 0\right) = 0.5$$

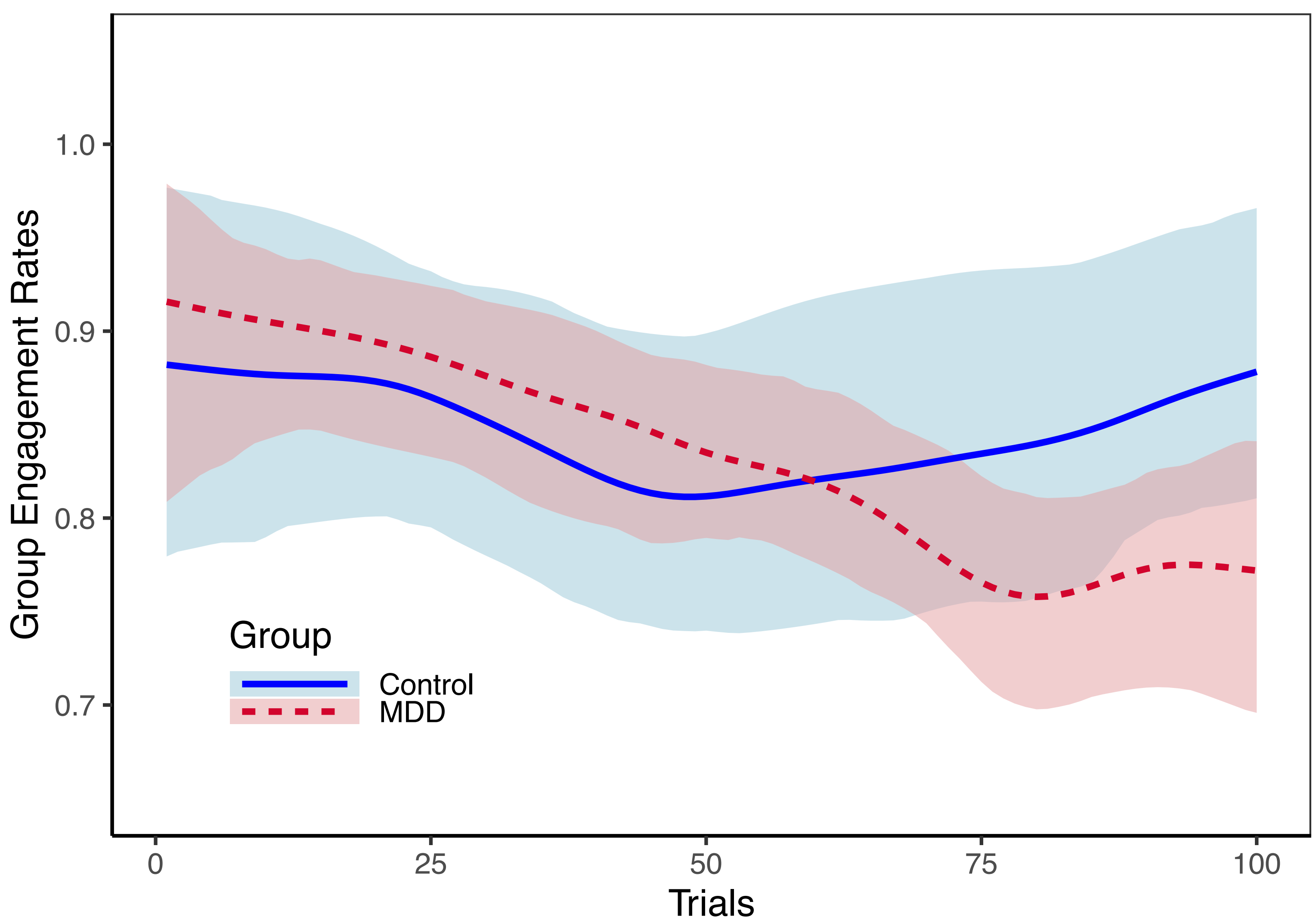
Application to EMBARC Study

Results: MDD vs Control

Individual engaged probability at trial *t*: $H_i(t) = P\left(U_{it} = 1 \mid \boldsymbol{A}_{i[1:T]}\right)$

Group engaged rate at trial *t*: $\bar{H}(t) = n^{-1} \sum_{i=1}^{n} H_i(t)$

Group engaged rates (MDD vs Control)



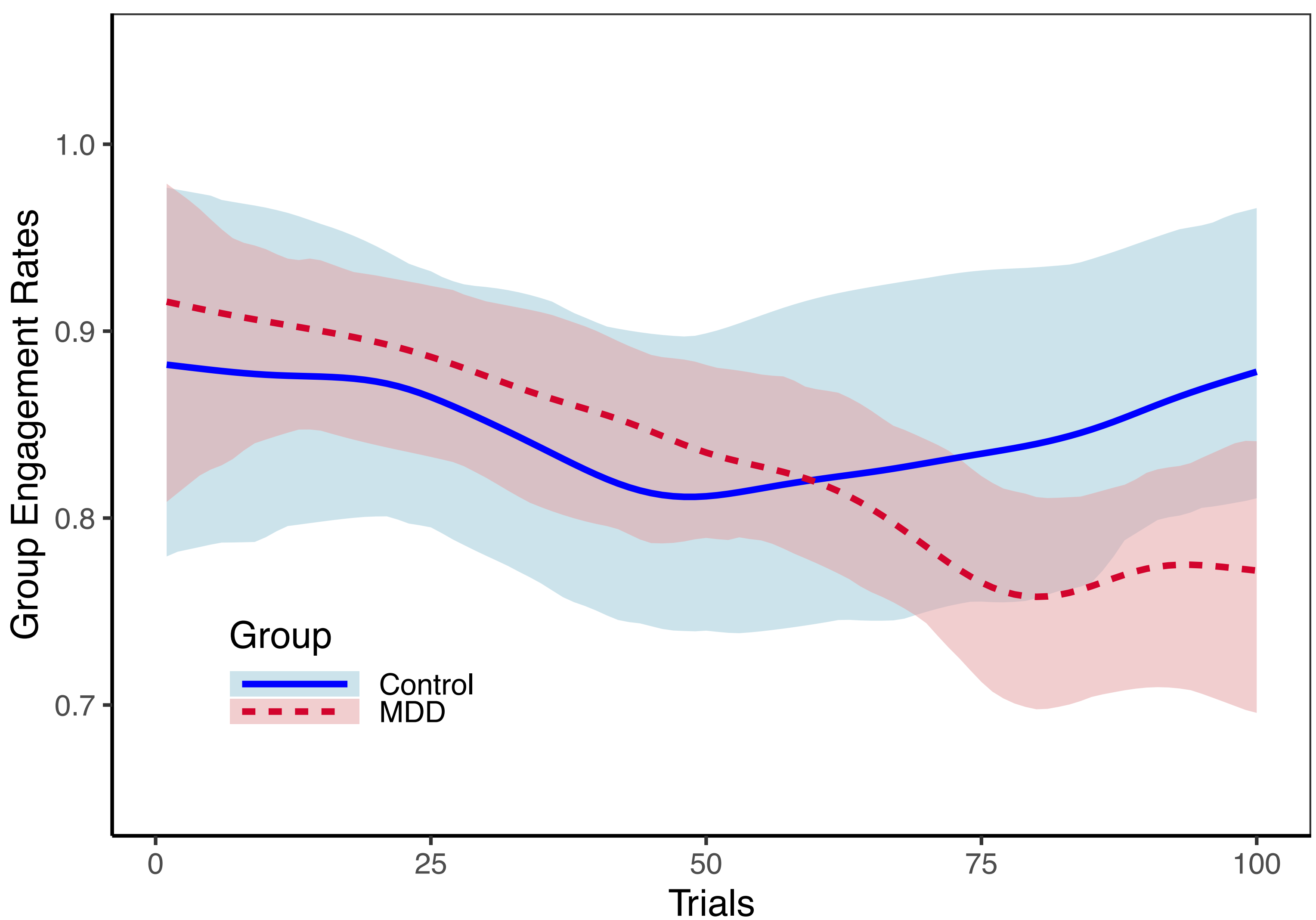
Application to EMBARC Study

Results: MDD vs Control

Individual engaged probability at trial t: $H_i(t) = P\left(U_{it} = 1 \mid A_{i[1:T]}\right)$

Group engaged rate at trial t: $\bar{H}(t) = n^{-1} \sum_{i=1}^{n} H_i(t)$

Group engaged rates (MDD vs Control)



MDD group potentially experiences greater difficulty in concentration compared to the control group at the second half of the task.

Application to EMBARC Study

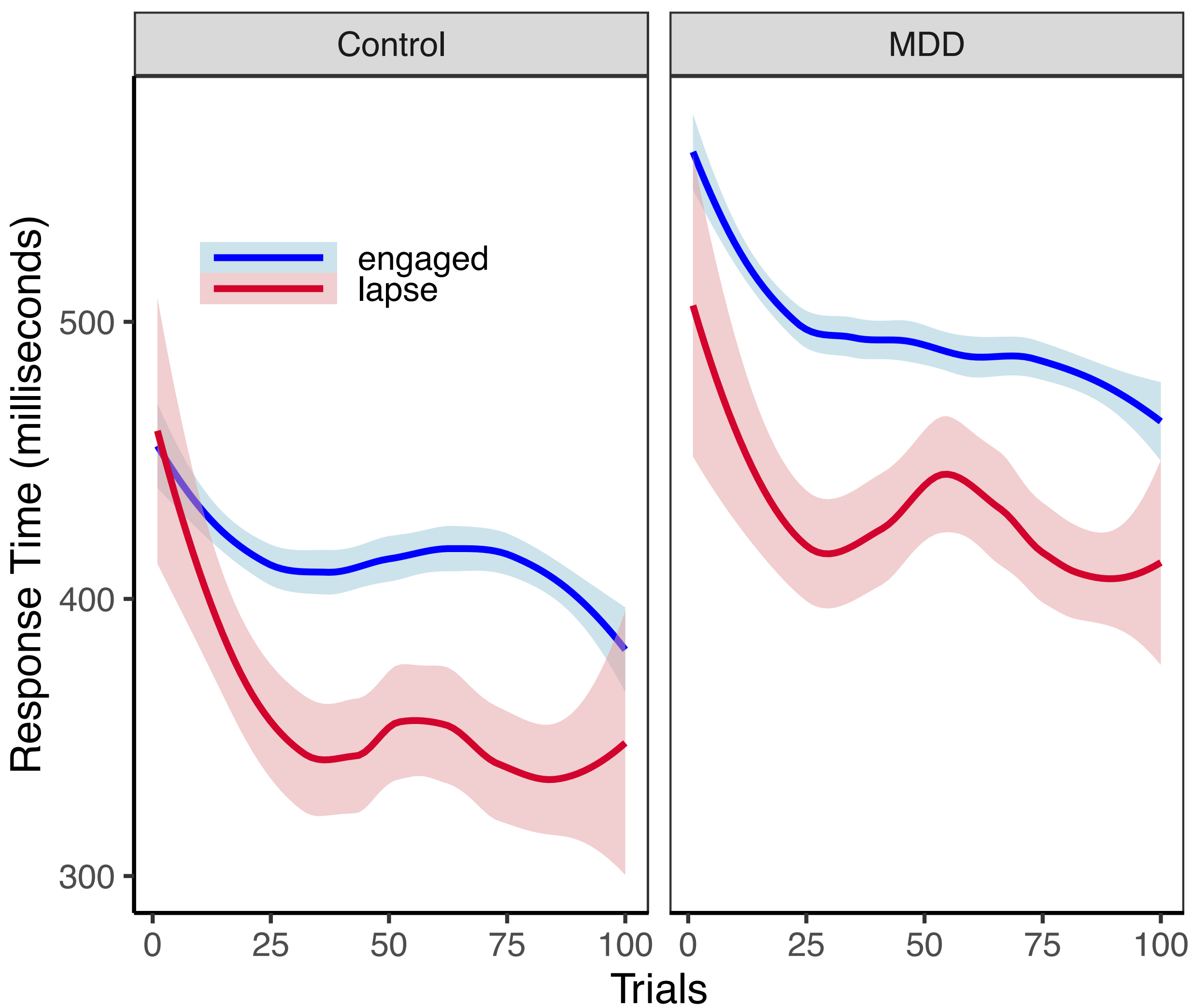
Results: MDD vs Control

Individual engaged probability at trial t: $H_i(t) = P\left(U_{it} = 1 \mid \boldsymbol{A}_{i[1:T]}\right)$

Predict the decision-making strategies: **engaged**, if $H_i(t) \geq 0.5$ **lapse**, if $H_i(t) < 0.5$

Response time (decision making time): time between state-showing and action-taking.

Response time vs Trials



Application to EMBARC Study

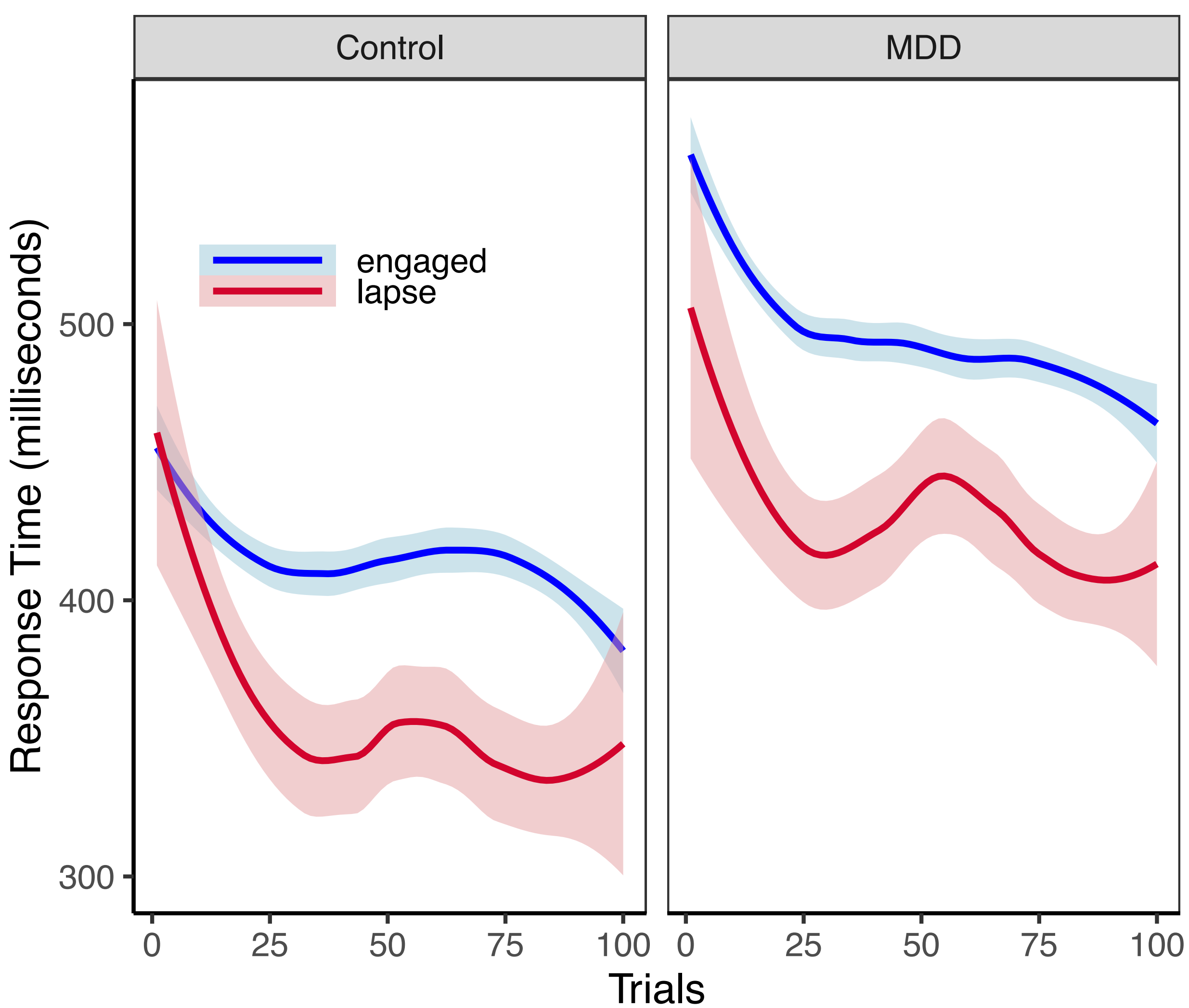
Results: MDD vs Control

Individual engaged probability at trial t: $H_i(t) = P\left(U_{it} = 1 \mid A_{i[1:T]}\right)$

Predict the decision-making strategies: **engaged**, if $H_i(t) \geq 0.5$ **lapse**, if $H_i(t) < 0.5$

Response time (decision making time): time between **state-showing** and **action-taking**.

Response time vs Trials



- **'Engaged' strategy takes more time to make decisions compared to the 'lapse' strategy.**
- **Control group takes less time to make decisions than the MDD group.**

How to incorporate the information of response time?

- **The decision-making state (lapse or engaged) is unknown. Simulation studies show a high false positive rate in identifying it.**

- **Response time may help predict decision-making phases — lapses tend to have shorter response times than engaged states.**

Question:

Can we jointly modeling decision-making processes and response times?

RL-HMM-DDM framework

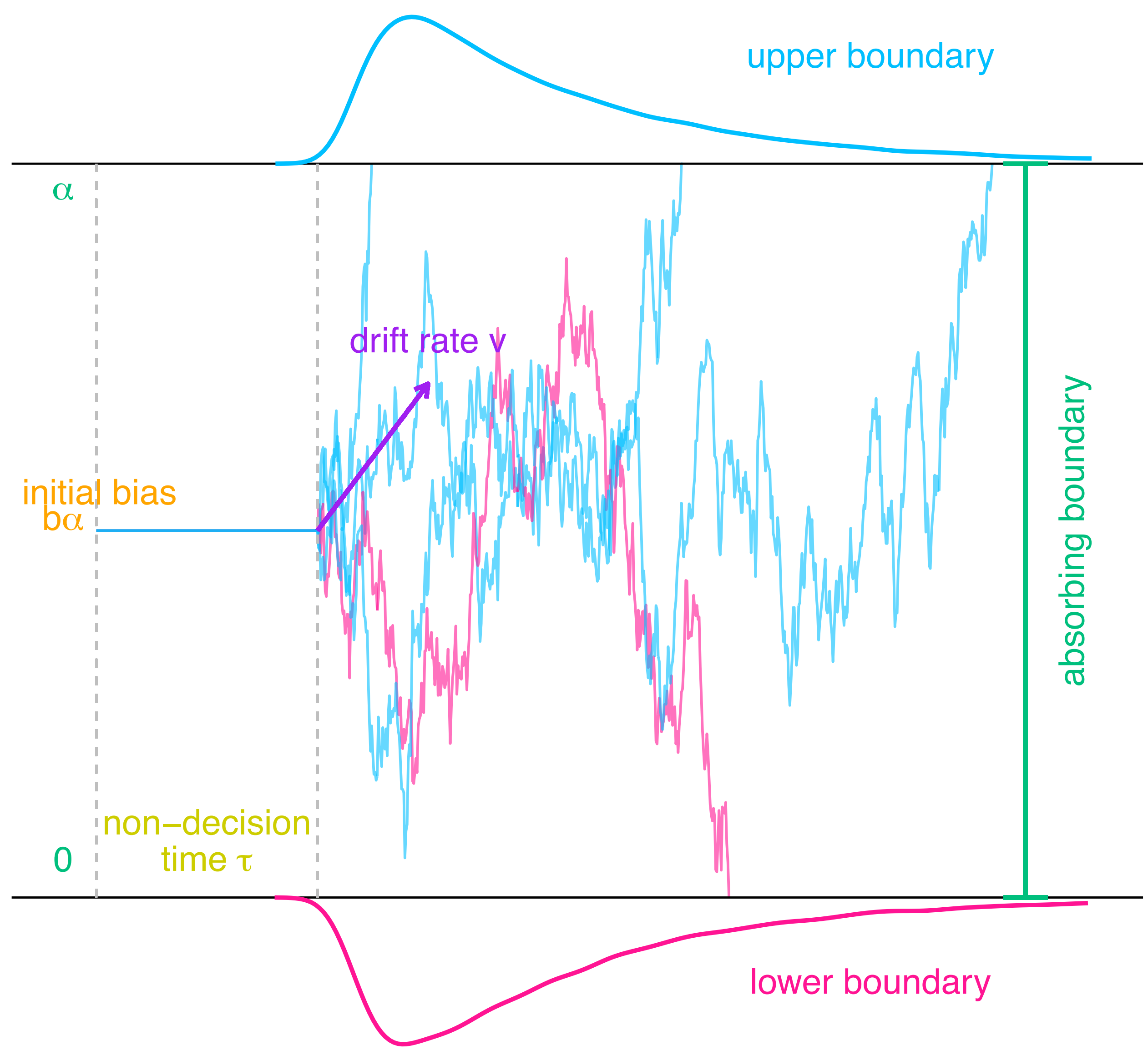
Bian, Y., **Guo, X.**, Wang, Y. (2025+). Joint modeling for learning decision-making dynamics in behavioral experiments. *Revision at the Annals of Applied Statistics.*

Model response time with Drift Diffusion Model (DDM) *(Ratcliff, 1978)*

- **DDM conceptualizes decision-making as a continuous evidence accumulation process.**
- **DDM is a Wiener process with drift**

$$\frac{dX(t)}{dt} \sim \text{Normal}\left(v, \sigma^2\right), \quad X(0) = b\alpha$$

- **The decision is made until he evidence reaches one of two absorbing boundaries corresponding to the decision choices.**
- **The response time is the total time needed to hit one of the absorbing boundaries.**



Reinforcement Learning-Drift-Diffusion Model (RL-DDM) *(Pederson et. al., 2017)*

- **For any DDM the joint distribution of decision and response time follows a Wiener first-passage time (WFPT) distribution.**

$$\left(T_{i,j}, A_{i,j}\right) \sim \text{WFPT}\left(\alpha, b, v_{i,j}, \tau\right)$$

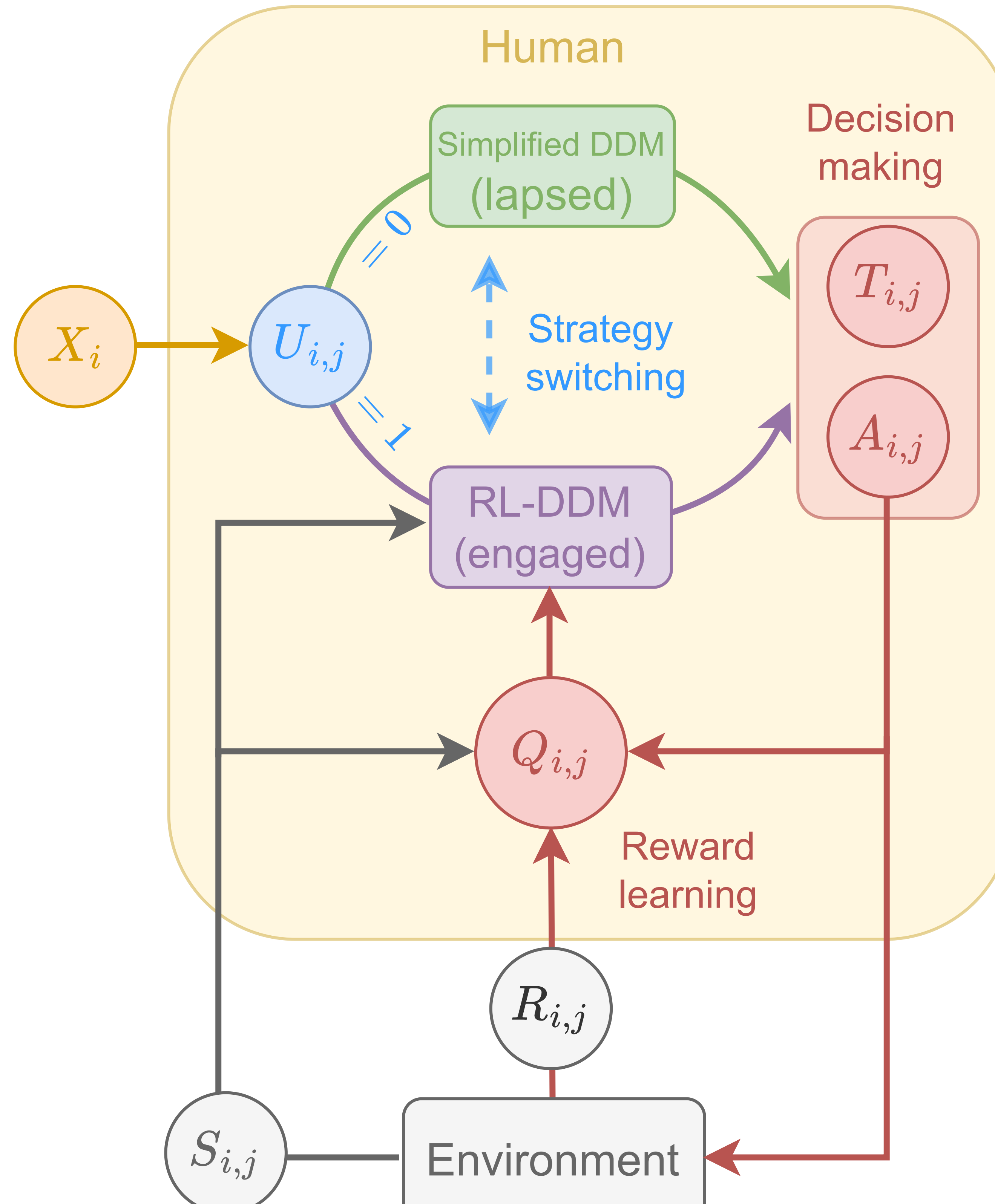
- **For any DDM, the drift rate v characterizes the speed of evidence accumulation. v > 0 favors decision "1", v < 0 favors decision "0".**

- **In reward tasks, RL-HMM links the drift rate to the contrast of the two expected rewards.**

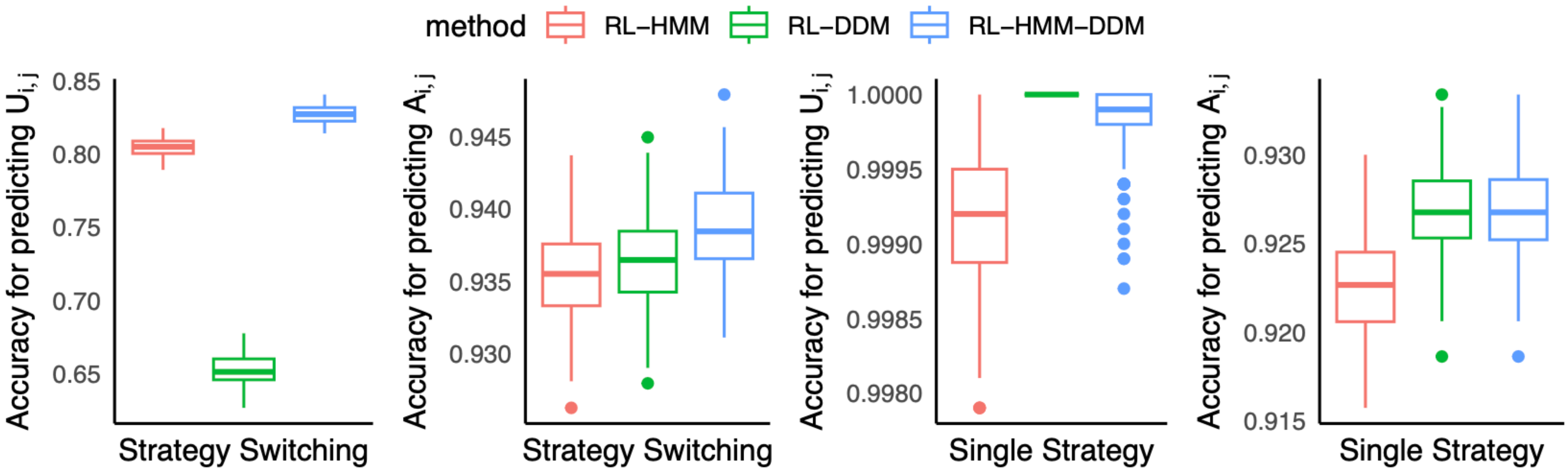
$$v_{i,j} = cZ_{i,j} = c\left\{Q_{i,j}\left(1, S_{i,j}\right) - Q_{i,j}\left(0, S_{i,j}\right)\right\}$$

RL-HMM-DDM Framework

- '**Engaged**' decisions are based on an **RL-DDM**.
- '**Lapse**' decisions are based on an **DDM** with **no initial bias** and **zero drift**, mimicking random guessing.
- **Latent state switching is captured by an HMM.**



Simulation Results



- RL-DDM performs unsatisfactorily in scenarios involving strategy switching.
- RL-HMM poorly predicts $U_{i,j}$ and $A_{i,j}$.
- RL-HMM-DDM reduces bias and achieves the best accuracy.

Take Home Message

- **Enhance reward-based decision modeling by considering multiple strategies and integrating response times.**

- **MDD patients are less sensitive to reward, harder to concentrate in tasks, and need longer time to make the decisions.**

- **A promising research direction for statisticians is studying brain-behavior association by jointly analyzing the behavioral data and brain imaging.**

Thank you