Introduction to Decision Making Studies 01

Reinforcement Learning (RL) Workshop on Data Analysis Methods for Decision Research

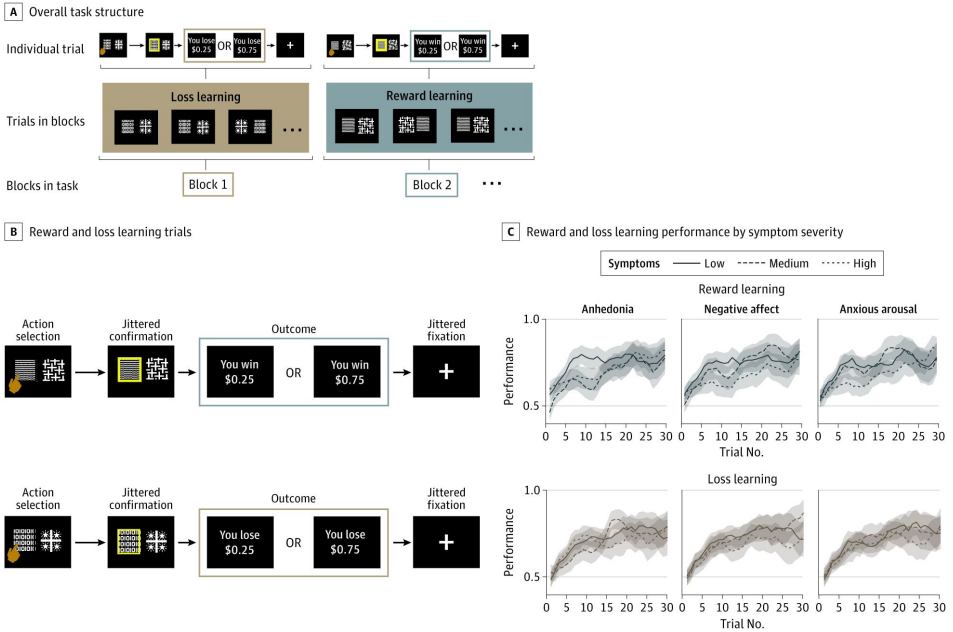
Presenter: Qiuyu Yu

Probabilistic Learning Task

Reward & Punishment

stimulus: two choices, participant does not know which one is better.

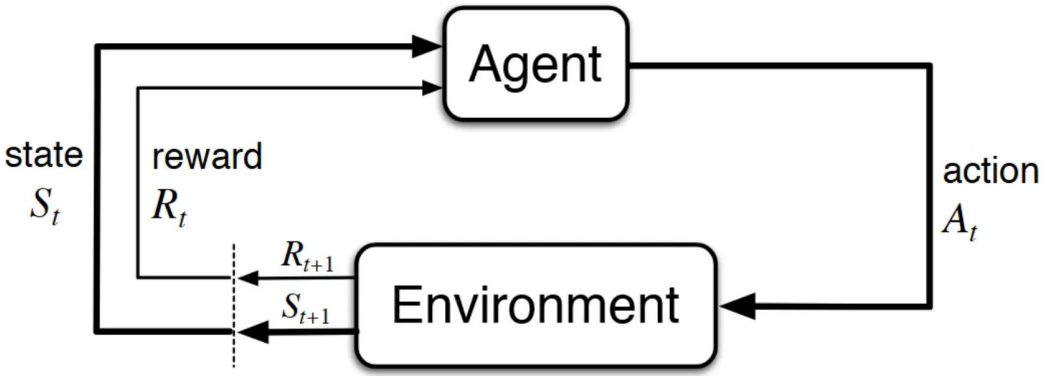
outcome:
One stimulus had a higher (75%) probability of leading to a better monetary outcome and a lower probability (25%) of leading to a worse monetary outcome, while the probabilities for the other stimulus were reversed (i.e., smaller probability of better outcome and larger probability of worse outcome).



RL Framework

Actions can be any decisions we want to learn how to make, and the **states** can be anything we can know that might be useful in making them.

That all of what we mean by goals and purposes can be well thought of as the **maximization of the expected value of the cumulative sum** of a received scalar signal (called **reward**).



Estimated parameters: RL Model

α (learning rate): Indicates how quickly you update your expectations (Q) after receiving feedback each time. How well does a person grasp the current learning patterns?

ρ (outcome sensitivity): Perception of outcomes, such as some individuals being more sensitive to punishment. Multiplicatively scaled more extreme outcome values, resulting in differential valuation of large vs small outcome values.

τ (outcome shift): Baseline shift. Overall more optimistic/pessimistic, viewing all outcomes as better/worse than they actually are. Linearly shifted all outcome values, resulting in an overall positive or negative valuation bias.

RL Model: Q Value

NewEstimate ← OldEstimate + StepSize X [Target - OldEstimate]

Q1=0, R=reward or feedback from each trial, alpha=[0,1]

$$Q_{n+1} = Q_n + \alpha[R_n - Q_n] = (1-\alpha)^n Q_1 + \sum_{i=1}^{n} \alpha(1-\alpha)^{n-i} R_i$$

$$Q_{t+1}(a) = Q_t(a) + \alpha(\rho R_t + \tau - Q_t(a))$$

softmax

The PE is weighted with a learning rate parameter η, such that larger learning rates close to 1 lead to fast adaptation of reward expectations, and small learning rates near 0 lead to slow adaptation. The process of choosing between options can be described by the softmax choice rule (Luce, 1959). This choice rule models the probability $p_i(t)$ that a decision maker will choose one option i among all options j:

$$p_i(t) = \frac{e^{(\beta(t) \times V_i(t))}}{\sum_{j=1}^{n} e^{[\beta(t) \times V_j(t)]}}. \quad (2)$$

Estimated Parameters: Decision making

β (inverse temperature): Does the individual greedily always choose the option with the highest Q value? (This can indicate whether the individual is conducting the experiment diligently.)

ω (choice perseveration): Inertia, where individuals persist with previous choices or switch rapidly.

split-half reliability; test-retest reliability (longitudal)

Probabilistic Decision Making Modeling

use w as last choice history, code choice $_{t-1}$ as 0 and 1

$$P(A)_t = \frac{e^{\beta * Q(A)_t}}{\left(e^{\beta * Q(A)_t} + e^{\beta * Q(B)_t}\right)}$$

$$P(A)_t = \frac{e^{\beta * Q(A)_t + choice_{t-1} * \omega}}{\left(e^{\beta * Q(A)_t + choice_{t-1} * \omega} + e^{\beta * Q(B)_t + |(1 - choice_{t-1})| * \omega}\right)}$$

Data Analysis and Modeling

Build models based on different variance (rewarding and punishment).

Model the above parameters and compare models based on BIC.

Test the independence of model parameters.

To test which combination of parameters best represented participants' behavior on the task, models with learning rate α plus 1) one valuation parameter, outcome sensitivity ρ (similar to [20]; model $\alpha + \rho$; 2 free parameters for reward and loss learning, respectively); 2) both valuation parameters (model $\alpha + \rho + \tau$; 3 free parameters per valence); 3) one decision parameter, inverse temperature β (model $\alpha + \beta$; 2 free parameters per valence); 4) both decision parameters (model $\alpha + \beta + \omega$; 3 free parameters per valence) were tested. Models without inverse temperature β as a free parameter fixed this parameter based on its estimated value ($\beta \approx 7$).

Integrating with fMRI data

Add RL parameters as parametric regressors and then compute beta values. GLM

First level imaging analyses used parametric regressors of prediction error δ or outcome value Rt at the time of outcome and expected value Qt of the chosen option at the time of onset.

All RL parameters have done z-transformed.

Regressors were separated by valence (reward or loss) and all regressors were modeled as stick functions.

Add other symptoms and time series in Model

$$\alpha_{total} = \alpha_{intercept} + time*\alpha_{time} + baseline\_anhedonia*\alpha_{anhedonia} + time*\Delta anhedonia*\alpha_{time*symptom} + \varepsilon$$

$$\alpha_{total} = \alpha_{intercept} + time*\alpha_{time} + baseline\_anhedonia*\alpha_{anhedonia} + treatment*\alpha_{treatment} + time*\Delta anhedonia*\alpha_{time*symptom} + time*treatment*\alpha_{time*treatment} + baseline\_anhedonia*treatment*\alpha_{symptom*treatment} + \Delta anhedonia*treatment*time*\alpha_{symptom*time*treatment} + \varepsilon$$

Stan

data: input preprocessed data .

transformed data: Secondary processing of preprocessed data (as needed).

parameters: parameters to be sampled by HMC.

transformed parameters: preprocess parameters (as needed).

model: Insert into the model according to the mathematical formula.

generated quantities: postprocess the model

Stan - Data and Parameters

```
data{
  int nTrials;
  int<lower=1,upper=2> choice[nTrials];
  int<lower=-1,upper=1> reward[nTrials];
}
parameters {
  real<lower=0,upper=1> alpha; // learning rate
  real<lower=0,upper=20> tau;  // softmax inverse tem
}
```

Stan - Model

```
model {
  vector[2] v;
  real pe;
  v = rep_vector(0, 2);
  for (t in 1:nTrials) {
    choice[t] ~ categorical_logit(tau*v);
    pe = reward[t] - v[choice[t]]; // prediction error
    v[choice[t]] = v[choice[t]] + alpha * pe; // value update
  }
```

Reference & Resource

- Brown, V. M., Zhu, L., Solway, A., Wang, J. M., McCurry, K. L., King-Casas, B., & Chiu, P. H. (2021). Reinforcement Learning Disruptions in Individuals With Depression and Sensitivity to Symptom Change Following Cognitive Behavioral Therapy. *JAMA psychiatry*, *78*(10), 1113–1122. https://doi.org/10.1001/jamapsychiatry.2021.1844 **(Supplement Files)**
- Mukherjee, D., Filipowicz, A. L. S., Vo, K., Satterthwaite, T. D., & Kable, J. W. (2020). Reward and punishment reversal-learning in major depressive disorder. *Journal of Abnormal Psychology, 129*(8), 810–823. https://doi.org/10.1037/abn0000641
- Sutton, R. S., & Barto, A. (2014). *Reinforcement learning: An introduction* (Nachdruck). The MIT Press. Chapter 1, 2, 3
- Pedersen, M.L., Frank, M.J. & Biele, G. The drift diffusion model as the choice rule in reinforcement learning. Psychon Bull Rev 24, 1234–1251 (2017). https://doi.org/10.3758/s13423-016-1199-y
- Luce, R. D. (1959). Individual choice behavior: A theoretical analysis. New York; NY: Wiley.
- https://youtu.be/oI138zRMQUs?si=_Jyx__sQuQomYwII
- https://github.com/lei-zhang/talks_and_workshops/tree/main