Neurocognitive mechanism of monetary and nicotine reward processing among smokers: A preliminary study

Jeung-Hyun Lee (jhleeangel@snu.ac.kr)
Department of Psychology, Seoul National University
Seoul, South Korea

Eunhwi Lee (lehlsy0904@snu.ac.kr)
Department of Psychology, Seoul National University
Seoul, South Korea

Joshua W. Brown (jwmbrown@indiana.edu)
Department of Psychological and Brain Sciences, Indiana University
Bloomington, IN, USA

Woo-Young Ahn (wahn55@snu.ac.kr)
Department of Psychology / Department of Brain and Cognitive Sciences, Seoul National University
Seoul, South Korea

Abstract:

Optimal decision-making often requires incorporating contextual information, which is often impaired in drug addiction. While it is crucial to utilize actual drug rewards in understanding disrupted reward processing in drug addiction, little attempts have been made to use actual drug reward as rewards. To address this gap, we used a variant of the two-armed bandit reinforcement learning task paired with computational models and a real-time vaping device to deliver nicotine reward to smokers. Our study provides preliminary evidence that smokers exhibited a reduced tendency to consider contextual information, specifically when anticipating nicotine rewards, as opposed to monetary ones. We expect that these findings will contribute to a deeper understanding of disrupted reward processing in drug addiction.

Keywords: reinforcement learning; context; range adaptation; reward frequency; drug addiction

Introduction

Context-dependent evaluation of reward is a fundamental feature of adaptive decision-making. During reinforcement learning (RL), subjective value of each given outcome is altered depending on the context, in which options are presented.

Drug addiction is characterized by a failure to adapt to the diminishing rewards that drugs offer. Inflexible and compulsive decision-making patterns are often referred to as disrupted range adaptation (Gueguen et al., 2023). Individuals persist in using drugs even though the rewards from drug intake become less frequent than before, indicating decreased sensitivity to reward frequency (Luijten et al., 2017). Yet, it remains unknown whether nicotine use disorder is associated with altered range adaptation or the altered influence of reward frequency.

Here, we aim to address this gap by using a variant of 2-bandit task and computational modeling that can account for both range adaptation and reward frequency. A range adaptation model of context-dependent RL posits that rewards are evaluated within a range of potential outcomes. A series of recent work has shown how range adaptation applies to systematic errors made during choices in new contexts (Palminteri & Lebreton, 2021; Bavard et al, 2021). Reward frequency is another important contextual variable that affects value-based decision making, as preferences toward an option could be determined by the relative frequency of rewards granted (Steward et al, 2006; Hayes & Wedell. 2023).

Another critical gap in the field is that monetary rewards have been predominantly used to study substance use disorders, rather than actual drug rewards. Compared to secondary rewards (e.g., money), drug rewards produce immediate and direct physiological effects, thereby directly affecting dopamine pathways and disrupting reward processing (Modak et al., 2021). In the current study, we address this issue by using a real-time vaping device to test the hypothesis that active smokers would exhibit different reward processing patterns across different types of rewards (nicotine vs. money).

Methods

Participants

In this preliminary study, we have recruited a total of 18 participants (Non-smokers N=14 with no history of diagnosis to any psychiatric disorder, Smokers N=4 with a diagnosis of nicotine use disorder). Participants who completed two visits and were included in the analysis. We used the Structured Clinical Interview for DSM-5 Disorders: Clinical version (SCID-5) for diagnosis (First et al., 2106).

Experimental procedure and task

During each visit, each group engaged in a series of reinforcement learning (RL) tasks that differed in reward probability of each stimulus (non-smokers) and reward type (smokers). Non-smokers performed either the original or the new task with modified task parameters on each visit, with the order counterbalanced. They were instructed that they would receive a monetary incentive based on their task performance. Smokers, on the other hand, performed the original task with the money and the nicotine condition. During the nicotine condition, they were made to vape instead of money as a reward, using a customized vaping device in the laboratory.

The details of the original task (**Figure 1A**) can be found in Bavard et al, (2022). The primary focus of the task is transfer error, where participants tend to choose the option that was rewarding in the previous learning context, even though it has a smaller expected value (EV) compared to the alternative option in the new transfer context. The new task design (**Figure 1B**) aimed to control for the potential effect of reward frequency on transfer effect, by setting the probability of rewarding to 0.5.

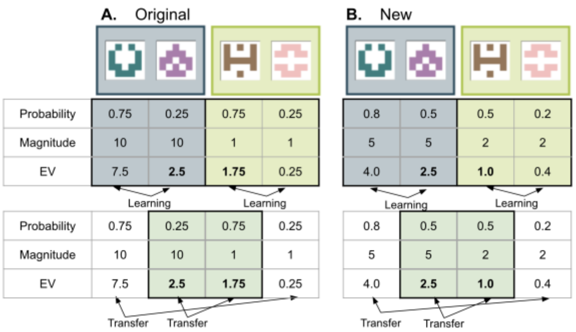


Figure 1: Original and New task design

Computational models

We considered RL models that differ in how each encodes choice outcomes and implemented three RL models with hierarchical Bayesian modeling (Kruschke, 2015) using Rstan (Stan Development Team, 2024).

The ABSOLUTE model We used a baseline model based on the delta rule and assumes that outcomes are encoded in an absolute fashion. Here, the Q-value is updated on each trial, based on the prediction error term (1). On trial *t*, chosen (c) and unchosen (u) option values of the current context *s* are updated, where α_c and α_u are the learning rates for each option and δ_t is prediction error.

(1)

$$Q_{t+1}(s,c) = Q_t(s,c) + \alpha_c * \delta_{c,t}$$
$$Q_{t+1}(s,u) = Q_t(s,u) + \alpha_u * \delta_{u,t}$$

The RANGE model We adapted a model from the previous study and implemented in a hierarchical Bayesian way. A context-dependent outcome $R_{RAN,t}$ is computed and with R_{MAX} and R_{MIN} are updated on each trial *t* if the given outcome is greater than their current values.

(2)

$$R_{RAN,t} = \frac{R_{OBJ,t} - R_{MIN,t}(s)}{R_{MAX,t}(s) - R_{MIN,t}(s) + 1}$$

The ORL model Next, we newly adapted the Outcome Representation Learning (ORL) model (Haines et al., 2018) that includes reward frequency term (ω) that updates expected frequency to win (EF)(3). In the model, expected value (EV) and EF are updated by two separate prediction error terms ($\delta_{val,t}$, $\delta_{freq,t}$).

(3)

$$V_{t+1}(s,c) = EV_{t+1} + EF_{t+1} * \omega$$
$$V_{t+1}(s,u) = EV_{t+1} + EF_{t+1} * (1-\omega)$$

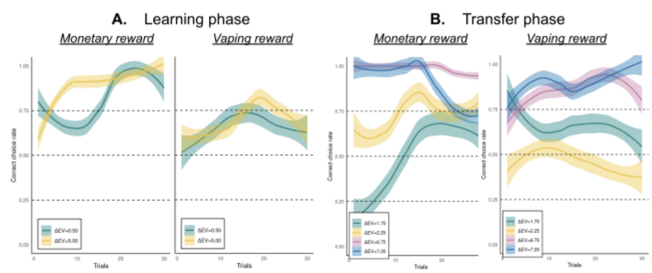
Results

Behavioral results

The results showed that both the original and the new task designs captured the systematic pattern of range adaptation. Thus, we replicated the transfer error in previous studies (e.g., Bavard et al., 2021) as we compared the ratio of optimal choices made during the transfer phase among non-smokers. When we compared two types of rewards (nicotine vs. money) among smokers, there was a marginally significant lower transfer error for nicotine rewards in the smallest EV difference condition (Nicotine=0.67, Money=0.47, p=0.086, t (6)=-0.6, marked as green in **Figure 2B**). This preliminary result may support the hypothesis that smokers adapt less to contextual information when learning about drug than monetary rewards.

Figure 2: Correct choice rate for original and new tasks

Modeling results



We fitted all participants' data using three models. When we compared the model performance using the leave-one-out information criterion (LOOIC), the ORL model outperformed the other two models (**Table 1**), suggesting that reward frequency could have played a critical role in outcome encoding.

Table 1: Model comparison.

Model	LOOIC	SE
(1) ABSOLUTE	1500.9	141.7
(2) RANGE	2744.6	342.1
(3) ORL (ω)	1259.8	148.8

Conclusions

In the current study, we provide preliminary results suggesting that smokers may have distinct neurocognitive mechanisms for processing nicotine and monetary rewards. Smokers showed a distinct pattern of learning nicotine outcome comparably slower than monetary reward, and thus making less transfer errors for nicotine reward, which could be explained with increased sensitivity to drug reward frequency. We expect to further provide more robust evidence on this possibility.

Acknowledgments

This work was supported by National Research Foundation of Korea Grant No. 2021M3E5D2A0102249311 funded by the Ministry of Science, Information and Communication Technologies and Future Planning (to W.-Y.A.); BK21 FOUR Program Grant No. 5199990314123 (to W.-Y.A.); funding from the Seoul National University Creative Pioneering Researchers Program (to W.-Y.A.); Seoul National University Artificial Intelligence Graduate School Program Grant No. 2021-0-01343 (to W.-Y.A.)

References

Bavard, S., Rustichini, A., & Palminteri, S. (2021). Two sides of the same coin: Beneficial and detrimental consequences of range adaptation in human reinforcement learning. *Science Advances*, 7(14).

First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2016). User's guide for the SCID-5-CV Structured Clinical Interview for DSM-5® disorders: Clinical version. *American Psychiatric Publishing, Inc.*

Gueguen, M., C.M., Anlló, H., Bonagura, D., Kong, J., Hafezi, S., Palminteri, S., & Konova, A. B. (2023). Recent Opioid Use Impedes Range Adaptation in Reinforcement Learning in Human Addiction. *Biological Psychiatry*.

Haines, N., Vassileva, J., & Ahn, W. Y. (2018). The Outcome-Representation Learning Model: A Novel Reinforcement Learning Model of the Iowa Gambling Task. *Cognitive Science*, 1-28.

Hayes, W., M. & Wedell, D. H. (2023). Testing models of context-dependent outcome encoding in reinforcement learning. *Cognition*, 230.

Kruschke, J. (2015). Doing Bayesian Data Analysis. *Academic Press.*

Luuijten, M., Schellekens, A. F., Kühn, S., Machielse, M. W. J., & Sescousse, G. (2017). Disruption of Reward Processing in Addiction: An Image-Based Meta-analysis of Functional Magnetic Resonance Imaging Studies, *Jama Psychiatry*,

Palminteri, S. & Lebreton, M. (2021). Context-dependent outcome encoding in human reinforcement learning. *Current Opinion in Behavioral Sciences*, 41, 144-151.

Priyamvada M., Hutslar, C., Polk, R., Atkinson, E., Fisher, L., Macy, J., Chassin, L., Presson, C., Finn, P. R., & Brown, J. W. (2021). Neural bases of risky decisions involving nicotine vapor versus monetary reward. *Neuroimage: Clinical*, 32

Stan Development Team (2024). “RStan: the R interface to Stan.” R package version 2.32. 6, https://mc-stan.org/.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, 53 (1), 1-26