Individual Differences and Impacts of Psychopathological Symptoms in Observational Reward Learning

Dongil Chung[1,2,*], HyungSeok Won[1], Yoo Joo Jeong[1], Dasom Park[1], and HeeYoung Seon[1]
[1]Ulsan National Institute of Science and Technology, Ulsan, South Korea
[2]Virginia Tech Carilion Research Institute, Roanoke, VA, USA
[*]Correspondence to: Dongil Chung (dchung@unist.ac.kr; Tel: +82-52-217-2744)

Abstract— **When making repeated decisions, individuals can learn about associations between actions and outcomes through obtained feedbacks. Such a learning process can occur based on individuals' direct experiences in the past, or simply on observed social others' actions and outcomes. Previous computational and neuroimaging studies have shown that one's learning performance is dependent on her sensitivity to reward (or punishment) and reward prediction error, the differences between experienced and expected rewards (or punishments). However, it remains unknown whether individuals' experience-based and observational learning have common or differential cognitive characteristics (e.g., value sensitivity) that affect the learning performances. Here, we use a probabilistic reward learning task, a choice task with different types of uncertainty, and computational modeling approach to quantify individuals' value sensitivity and learning performances. We further examine associations between performances in observational- and experience-based learnings with individuals' psychopathological symptoms. Particularly, depression, a most prevalent symptom in modern society and known factor that affects reward sensitivity, is used as a psychopathological measure of interest. The current study contributes to understanding how individuals' psychopathological symptoms affect their experience-based and observational reward learning.**

I. Introduction

Most human decisions are made about, among, and for social others [1]. Given the abundant opportunities to achieve information by observing social others, one does not always have to experience the consequences of choices by oneself to learn which choice is more beneficial [2-4]. Such "observational learning" of the association between choices and outcomes allows observers to update their knowledge about the environment in an indirect manner. It is well known that individuals' reward learning performances are dependent on their sensitivity to reward values and to prediction errors reflecting whether the reward was better or worse than they expected [5]. Although neuroimaging studies have shown that similar brain regions are involved in learning of information observed from social others [1, 6, 7], it still remains unknown to what extent individuals' characteristics (e.g., value sensitivity) are shared between reward learning and observational learning. For example, individuals who have high confidence about their own experience-based valuation may not use information they achieved from external sources (e.g., social others). Here, we conducted one reward learning task and one non-learning task where both of the tasks involved a series of choices between options with uncertain outcomes, and used model-based analytic approach to examine inter- and intra- task individual differences in valuation and learning. Specifically, choice patterns in a non-learning task were used to characterize individuals' preference and value sensitivities that were compared against their learning performances in a learning task where participants had chances to learn from both their own choice-outcome experiences and observation of social others' choices.

Previous studies have found that individuals' subjective valuation and learning are associated with their psychopathological symptoms. In particular, depressive symptom are known to be associated with one's diminished behavioral and neural reward responses [8-13]. A recent study, however, showed that individuals with major depression have intact value processing in a non-learning environment [14]. Because the study did not examine participants' reward learning and responses, it could be only inferred that impaired reward responses in major depression may be specific to a learning environment (c.f., [15]). In the current study, we used a within-subject design across learning and non-learning tasks, and tested the association between individuals' depressive symptoms and their estimated cognitive characteristics both in learning and non-learning environments.

II. Methods

A. Participants

32 individuals (male/female = 21/11, age = 22.41±2.33) participated in the current study. All participants provided written informed consent. The study was approved by the Institutional Review Boards of Ulsan National Institute of Science and Technology. Two participants were excluded from data analyses due to experimenter error, and eight participants were excluded due to their low performance (around 50% accuracy in one of the three conditions). After exclusion, data from 22 participants were included for the final analyses (male/female = 14/8, age = 22.55±2.24).

B. Experimental Procedures

Two behavioral decision-making tasks were used. One task involved observational and experience-based reward learning where participants had to learn which of the two given options is better ('learning task' hereafter) [3]. During the other task, participants had to make a series of choices between two certain or uncertain options ('uncertainty task' hereafter) [16]. There was no feedback provided of their choices, and therefore the task did not involve any learning processes.

Uncertainty Task

Due to the potential learning effect affecting participants' behavioral patterns, all participants performed the uncertainty task on the first visit. During the task, participants were asked to make a series of choices between two gambles that have different types of uncertainty [16]. Three different types of uncertainty were introduced; risk where outcome values and corresponding probabilities were known [17], ambiguity where outcome values were known but probabilities were unknown [18], and vagueness where only the range of outcome value was known (i.e., outcome value information is partially omitted). To investigate how individuals subjectively value choices with these types of uncertainty, participants made choices between two gambles on each trial that were a combination of two of the four gamble types: certain gambles, risky gambles, ambiguous gambles, and vague gambles. There was one session where all outcome values were gains and another session where all outcome values were losses. Order of the sessions were counterbalanced and participants were paid at the end of the study, based on the outcome of a random single gamble drown from all the choices the participant made in each session. See Ref [16] for details of the task.

Learning Task

Our learning task was adapted from a previous study in observational learning [3]. Participants went through a training session on their first visit after they performed the uncertainty task. This allowed participants to get familiar with the task structure. During the task, participants were asked to learn the reward structure of various sets of fractal images. Each set of images consisted of two options where one option is always better than the other (more likely to give better outcome). In the current task, the better option had 80% chance of good outcome (+10 points in gain sessions and 0 points in loss sessions) and 20% chance of bad outcome (0 points in gain sessions and –10 points in loss sessions). On the contrary, the worse option had 80% chance of bad outcome and 20% chance of good outcome. Each trial consisted of "Observation stage" where participants get to observe one of the previous participants' behavioral choices on the same set of images, and "Action stage" where participants make a choice. As per Ref [3], there were three different conditions where each condition provided different amount of information in observation stage. The first condition ("Individual learning") did not reveal neither the other participant's action nor choice outcome. The second condition ("Action only") revealed only the choice the other participant made, and the third condition ("Action + Outcome") revealed both the choice and the outcome of the other participant experienced within the same set of images. Three types of conditions were intermixed. Note that each type of condition was associated with different set of fractal images, so that participants had to learn the outcome associations separately. In total, there were 3 gain and 3 loss sessions where each of the session had three different types of condition (individual learning, action only, and action + outcome). Gain and loss sessions alternated, and participants were notified which type of session it is at the beginning of each session.

C. Computational model

We constructed computational models for each task and estimated individual parameters that characterize each individual's valuation and learning performances.

Uncertainty Task

To formally incorporate preferences over different types of uncertainty, we drew upon expected utility theory [17] and modern portfolio theory (mean-variance framework, [19]), and constructed a subjective value function [16] that includes i) risk preference, ii) ambiguity aversion, iii) vagueness value weight (level of optimistic valuation for a range of vague outcome), iv) vagueness dispersion preference (preference for the level of vagueness), and v) sensitivity to subjective value differences between options. The Softmax choice rule was used to analyze individual subjective values of gambles with individuals' choices.

Learning Task

Learning model adapted a basic Q learning algorithm [3, 20]. Briefly, the model assumes that each participant calculates expected values of choosing option 1 or 2, which are referred as Q-values. After making a choice based on the Q-values on each trial, participants receive a feedback where they can calculate how much better (or worse) the option was than their expectation, i.e., prediction error. This error is used for updating the Q-value, so that one can have a better expectation of choosing the option. The extent to which one updates the Q-value based on the prediction error is defined as learning rate. In the current study, our learning model for participants' own experience (i.e., Action stage) follows the same logic as explained above. Note that we included two separate learning rates for gain and loss sessions, so that we can examine potential differential responses between gains and losses. We modified the suggested model from Ref [3] for Observation stage, because model-agnostic analysis results (see below) showed that participants did not improve more from only observing the other participant's action (Action only condition). With the modification, our learning model included two additional learning rates, each for learning from the observee's reward prediction error in gain and loss sessions. As in the Uncertainty Task, the Softmax choice rule was used.

III. RESULTS

A. *Model-agnostic analysis*

Percentage of correct choice (i.e., better option) was calculated to first depict their learning across the task (**Fig. 1**). From the early stage of the task, in both gain and loss sessions, individuals showed higher accuracy in Action + Outcome condition (red) compared with that in Individual learning condition (black). In contrast to the previous study [3], individuals did not show improved accuracy in Action only condition (blue) (**Fig. 1**). In gain sessions, total points individuals earned were significantly higher in Action + Outcome condition than other two conditions. Although trending, loss sessions showed a comparable pattern. These results indicate that individuals successfully used social others' experience (observed choices and results of social others) to learn expected values of the options.

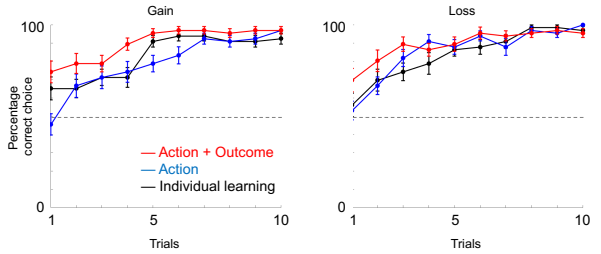

Figure 1. Model-agnostic results of learning performances. Individuals showed relatively faster learning performance in Action + Outcome condition compared with other conditions. Error bars represent standard error of the mean; Dotted lines represent 50% accuracy.

B. *Model-based analysis*

Individual parameters included in the suggested computational model were estimated using hierarchical Bayesian estimation [21, 22]. As described in the methods section, four learning rate parameters per individual were estimated, separately defining for gain and loss sessions, and for Individual learning and Observed learning (Action + Outcome) conditions. On average, individuals showed larger learning rates for loss sessions than for gain sessions (**Fig. 2**), indicating that they were more responsive to losses [23]. Between conditions, participants showed larger learning rates where they observed others (**Fig. 2**). This pattern was more apparent in loss sessions. These results suggested that on average, individuals were using others' positive and negative experiences more than they use their own experiences.

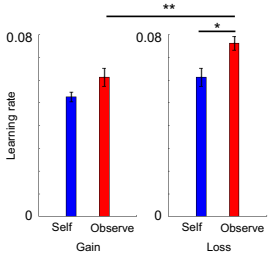

Figure 2. Model-based learning rates for individual and observation-based learning. Individuals showed higher learning rates for observation-based learning (observe) compared with experience-based learning (self) (repeated measures ANOVA, $F(3, 63)=7.61$, $P = 0.00020$). Particularly, the trend was more significant in loss sessions (post-hoc paired t-test, $t(21)=-2.46$, $P = 0.023$). Regardless of how individuals learned (self or observe), they showed larger learning rates in loss than gain sessions (post-hoc paired t-test; self: $t(21) = -2.04$, $P = 0.055$; observe: $t(21)=-3.65$, $P = 0.0015$). Error bars represent standard error of the mean; *P<0.05, **P<0.01.

To examine individual differences in observational learning, we calculated correlation coefficients between individual parameters. Individuals who showed higher value sensitivity (i.e., larger inverse temperature) showed larger learning rates in Individual learning condition, regardless of gain or loss sessions (**Fig. 3, upper**). Those who learned the fastest from their own experiences (high learning rate in Individual learning) showed the slowest learning rate for observed experiences (**Fig. 3, lower**). These results are consistent with our expectation that individuals who put larger weight on the information they experienced themselves place smaller weight on the information achieved from social others.

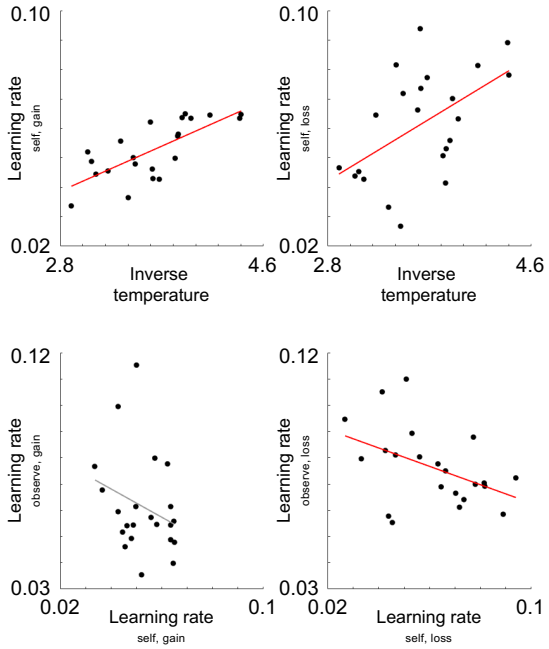

Figure 3. Inter-parameter associations. Participants who had higher value sensitivity (larger inverse temperature) showed larger learning rates in both gain (Pearson's correlation $r = 0.73$, $P = 1.16e-04$) in loss (Pearson's correlation $r = 0.52$, $P = 0.013$) sessions. In loss sessions, individuals who had larger learning rates in Individual learning had smaller learning rate for observation-based learning (Pearson's correlation $r = -0.45$, $P = 0.037$). Each point represents an individual participant. Red lines show significant regression results, and a grey line shows a non-significant regression result.

To examine whether individual characteristics in cognitive process for non-learning environment explain individuals' learning performances, we examined associations between risk preference (signed sensitivity to reward variance) and

learning rates. Individuals who were most averse to risk during Uncertainty task showed the largest learning rate in gain sessions (**Fig. 4**). However, individuals' behavioral patterns in loss session learning task was not associated with their risk preference for gambles with potential losses. These results suggest that there may be additional process that modulates individuals' learning pattern above and beyond their general decision-making tendency (risk preference in non-learning environment).

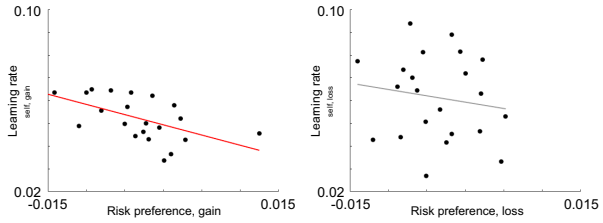

Figure 4. Relationship between individuals' risk preference and learning rates. In gain sessions, individuals who are risk averse (negative value on x-axis) showed larger learning rate (Pearson's correlation r = -0.59, P = 0.0041). No such relationship was found in loss sessions. Each point represents an individual participant. Red and grey lines show significant and non-significant regression result, respectively.

Individuals' learning rates, preference for different types of uncertainty, and value sensitivity were correlated with self-reported questionnaire data, including Barrett impulsivity [24], Beck Depression Inventory (BDI-II) [25], and Mood and Anxiety Symptom Questionnaire (MASQ) [26]. Consistent with previous studies [14, 15], we did not find any evidence showing that depression symptom impairs individuals' valuation, reward learning, nor observational learning performances. The current within-subject design shows that, in a behavioral level, individuals' symptom severity in depression is not associated with their subjective valuation nor with reward learning performances.

IV. Discussions & Conclusions

The current study examined individual differences in observational learning and valuation during decision-making under uncertainty. Our results replicated that individuals tend to learn in faster speed when they use information achieved from social others than their experiences. Individual differences in non-learning task within gain domain were associated with individuals' learning performances in reward learning task. However, this partial association between tasks were not observed between individuals' risk preference and learning rate for observational reward learning, which provides evidence for differential processes between learning from different sources: social (observational) and nonsocial (experience-based) information. We did not find any evidence showing that mild depression symptom impairing one's valuation nor learning performances. This null finding consistent with previous studies again showcases the strength of model-based analytics in dissociating multiple factors of cognitive process. Future study should follow to understand why social information bias exists in learning and from where such individual differences originate.

Acknowledgment

This work was supported in part by UNIST (1.180031.01 and 1.180073.01 to DC) and National Research Foundation of Korea (NRF-2018M3C1B8013691 and NRF-2018R1D1A1B07043582 to DC). We thank C. Burke for kindly sharing the original source code for the task.

References

1. Ruff, C.C. and E. Fehr, *The neurobiology of rewards and values in social decision making.* Nature Reviews Neuroscience, 2014. **15**(8): p. 549.
2. Christopoulos, G.I. and B. King-Casas, *With you or against you: social orientation dependent learning signals guide actions made for others.* Neuroimage, 2015. **104**: p. 326-335.
3. Burke, C.J., et al., *Neural mechanisms of observational learning.* Proceedings of the National Academy of Sciences, 2010. **107**(32): p. 14431-14436.
4. Apps, M.A., M.F. Rushworth, and S.W. Chang, *The anterior cingulate gyrus and social cognition: tracking the motivation of others.* Neuron, 2016. **90**(4): p. 692-707.
5. Sutton, R.S. and A.G. Barto, *Reinforcement learning: An introduction.* 2011.
6. Hackel, L.M., B.B. Doll, and D.M. Amodio, *Instrumental learning of traits versus rewards: dissociable neural correlates and effects on choice.* Nature Neuroscience, 2015. **18**(9): p. 1233.
7. Suzuki, S., et al., *Learning to simulate others' decisions.* Neuron, 2012. **74**(6): p. 1125-1137.
8. Huys, Q.J., et al., *Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis.* Biology of mood & anxiety disorders, 2013. **3**(1): p. 12.
9. Association, A.P. and A.P. Association, *Diagnostic and statistical manual of mental disorders (DSM).* Washington, DC: American psychiatric association, 1994: p. 143-7.
10. Nelson, B.D., et al., *Blunted Neural Response to Rewards as a Prospective Predictor of the Development of Depression in Adolescent Girls.* American Journal of Psychiatry, 2016: p. appi. ajp. 2016.15121524.
11. Pizzagalli, D.A., et al., *Reduced hedonic capacity in major depressive disorder: evidence from a probabilistic reward task.* Journal of psychiatric research, 2008. **43**(1): p. 76-87.
12. Treadway, M.T. and D.H. Zald, *Reconsidering anhedonia in depression: lessons from translational neuroscience.* Neuroscience & Biobehavioral Reviews, 2011. **35**(3): p. 537-555.
13. Weinberg, A., et al., *Blunted neural response to rewards as a vulnerability factor for depression: Results from a family study.* 2015.
14. Chung, D., et al., *Valuation in major depression is intact and stable in a non-learning environment.* Scientific reports, 2017. **7**: p. 44374.
15. Rutledge, R.B., et al., *Association of neural and emotional impacts of reward prediction errors with major depression.* JAMA psychiatry, 2017. **74**(8): p. 790-797.
16. Chung, D., et al., *Evidence for preference consistency across risky, ambiguous, and vague gambles.* 2017.

17. Bernoulli, D., *Exposition of a new theory on the measurement of risk.* Econometrica: Journal of the Econometric Society, 1954: p. 23-36.
18. Ellsberg, D., *Risk, ambiguity, and the Savage axioms.* The quarterly journal of economics, 1961: p. 643-669.
19. Markowitz, H.M., *Portfolio selection: efficient diversification of investments*. Vol. 16. 1959: Yale University Press, New Haven.
20. Pessiglione, M., et al., *Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans.* Nature, 2006. **442**(7106): p. 1042.
21. Ahn, W.-Y., N. Haines, and L. Zhang, *Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package.* Computational Psychiatry, 2017.
22. Daw, N.D., *Trial-by-trial data analysis using computational models.* Decision making, affect, and learning: Attention and performance XXIII, 2011. **23**: p. 3-38.
23. Tom, S.M., et al., *The neural basis of loss aversion in decision-making under risk.* Science, 2007. **315**(5811): p. 515-518.
24. Patton, J.H., M.S. Stanford, and E.S. Barratt, *Factor structure of the Barratt impulsiveness scale.* Journal of clinical psychology, 1995. **51**(6): p. 768-774.
25. Beck, A.T., R.A. Steer, and G.K. Brown, *Beck depression inventory-II.* San Antonio, 1996. **78**(2): p. 490-8.
26. Watson, D., et al., *Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales.* Journal of abnormal psychology, 1995. **104**(1): p. 3.