Inverse reinforcement learning captures value representations in the reward circuit in a real-time driving task: a preliminary study

Sang Ho Lee[1] (sh.lee.11@snu.ac.kr)
Min-hwan Oh[2] (minoh@snu.ac.kr)
Woo-Young Ahn[1,3] (wahn55@snu.ac.kr)

[1]Department of Psychology, Seoul National University
Seoul, Korea
[2]Graduate School of Data Science, Seoul National University
Seoul, Korea
[3]Department of Brain and Cognitive Sciences, Seoul National University
Seoul, Korea

Abstract:

A challenge in using naturalistic tasks is to describe complex data beyond simple summary of behaviors. Lee et al. (2024) showed that an inverse reinforcement learning (IRL) algorithm combined with deep neural networks is a practical framework for modeling real-time behaviors in a naturalistic task. However, it remains unknown whether the reward function inferred by IRL reflects value representations in the reward circuit. In this preliminary study (N=10), we investigate the neural correlates of the reward inferred by IRL. Human participants were scanned using fMRI while performing a real-time driving task (i.e., highway task). We show that the trajectory of IRL reward during the task strongly correlates with the trajectory of BOLD signals in the reward circuit including the prefrontal cortex, the striatum, and the insula. The results demonstrate the validity of the IRL as a modeling framework that explains both behaviors and the brain activity in a real-time task.

Keywords: naturalistic task, inverse reinforcement learning, fMRI, deep neural networks

Recent technological advances in computational power and data acquisition methods (e.g., virtual reality, mobile devices) have increased the use of naturalistic tasks and data in neurocognitive studies (Parsons, 2015). A challenge in using naturalistic tasks is to describe the observed data beyond simple summary of behaviors. Data from naturalistic tasks often encompass a multitude of dimensions that define an immense number of possible states and actions, posing challenges in modelling the behaviors in the tasks (Thompson et al., 2019; Wise et al., 2023).

Lee et al. (2024) used an inverse reinforcement learning (IRL) algorithm combined with deep neural networks (Fu et al., 2017) to model the behaviors in a real-time driving task (i.e., highway task), in which participants controlled a car in a simulated highway. The objective of IRL is opposite to that of "forward" reinforcement learning (RL; Sutton & Barto, 2018). IRL learns the reward function underlying observed behaviors, whereas RL learns behavioral policy by observing rewards. Lee et al. (2024) showed that real-time trajectories of IRL reward in the highway task provide indicators of impulsivity, with impulsive participants showing higher rewards in risky situations.

The reward inferred by IRL is interpreted as a participant's subjective reward, but it is unclear whether the IRL reward indeed reflects value representations in the brain, especially the reward circuit. To evaluate the validity of IRL as a modeling framework for understanding reward processing in the brain during a real-time task, we conducted an experiment in which participants performed the highway task while undergoing fMRI scanning. We hypothesized that IRL reward trajectories would be similar to the trajectories of BOLD signals in the reward circuit including the prefrontal cortex, the striatum, and the amygdala (O'Doherty et al., 2017). We hypothesized that brain regions in the reward circuit would show heightened activity during overtaking (rewarding event) and reduced activity during crashing (aversive event).

Results

We first analyzed the data using a generalized linear model (GLM) to investigate whether the salient events in the task modulate the reward circuit. The independent variables in the GLM were the onsets of the two events of interest (overtaking and crashing; Lee et al., 2024). Six head movement regressors and button-press regressors were also included as covariates. **Figure 1** shows the neural correlates of overtaking and crashing. As hypothesized, the striatum exhibited increased and decreased activity during overtaking and crashing, respectively. The results

suggest that IRL is a valid framework for investigating reward processing in the brain.

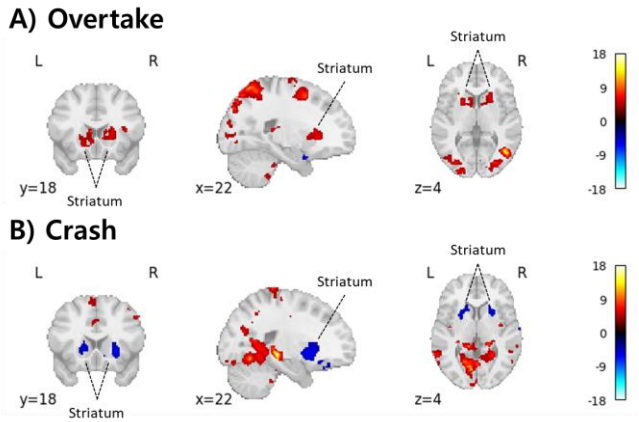

Figure 1: Neural correlates of (A) overtake and (B) crash ($p < 0.001$, uncorrected). Color bars indicate t-statistics.

We then performed a time-series analysis to find an association between the IRL reward and the brain activity, with overtaking and crashing as the events of interest. For the analysis, we generated mean BOLD time-series data for each ROI defined by the whole brain functional parcellation (k = 50) from Neurosynth (neurosynth.org).

In a group-level analysis utilizing the data averaged across participants, the trajectory of IRL rewards for overtaking (-12 to 12 seconds from the onset) showed strong correlation ($p < 0.0001$) with the BOLD time-series in several ROIs including the prefrontal cortex, the insula, and the striatum (O'Doherty, 2004). **Figure 2** illustrates the standardized trajectories of the IRL rewards and the BOLD signals from three ROIs within the striatum, which was of main focus in the GLM analysis. In the nucleus accumbens ($r = 0.49$), the dorsal caudate ($r=0.76$), and the putamen ($r = 0.49$), IRL reward trajectories closely resemble BOLD signals. By contrast, the reward trajectory for crashing did not show considerably strong association with reward-related BOLD signals.

To summarize, this preliminary study showed that the reward inferred by IRL can capture value representations in the brain. This demonstrates the potential utility of IRL as a modeling framework that can account for both behaviors and the brain activity in a real-time task.

Methods

Participants

We recruited ten undergraduate and graduate students at Seoul National University (plan to recruit up to 50 participants in total). The study was approved by the Institutional Review Board at the Seoul National University.

Procedures

Participants performed the highway task (Lee et al., 2024) while undergoing fMRI scanning (Siemens TIM Trio 3T scanner, TR = 1200ms, TE = 30, FOV = 256mm, slices: 64, slice thickness = 2.3mm). Participants were instructed to drive a car on a simulated highway as fast as possible without crashing into other cars. The car was controlled via a four-button response box, with each button corresponding to one of four possible actions: accelerate, decelerate, turn left, and turn right. Participants performed four blocks of the task, each lasting for approximately 9 minutes.

Inverse reinforcement learning

We inferred the reward functions of the participants using adversarial inverse reinforcement learning (AIRL; Fu et al., 2017) algorithm. The algorithm trained deep neural networks for each participant based on their state and action trajectories throughout the task. Once trained, the deep neural networks could calculate the subjective reward of each participant for any possible state.

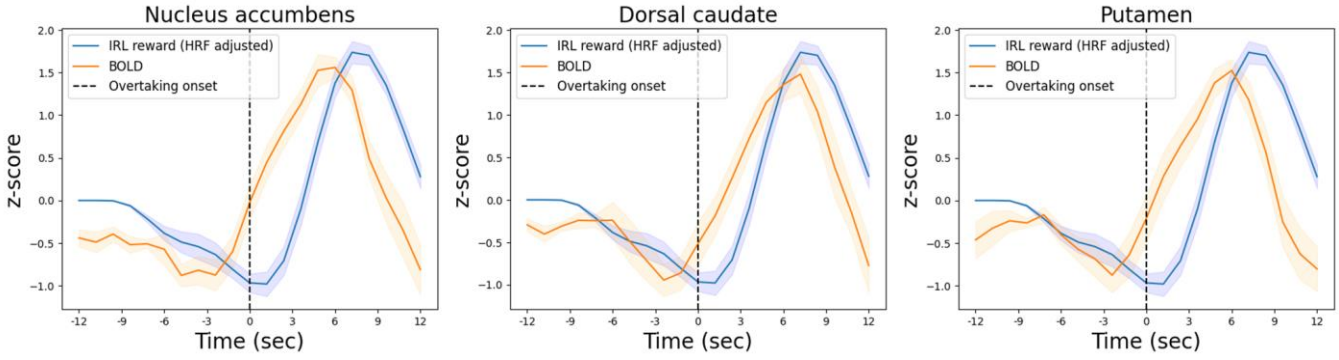

Figure 2: Trajectories of the IRL rewards and the BOLD signals in the striatum around at the onset of overtaking. The shaded areas in the graphs indicate standard errors of the means.

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.); and the Creative Challenge Research Program through National Research Foundation of Korea Grant No. 2022R1I1A1A01066530 funded by the Ministry of Education (to S.L.).

References

Fu, J., Luo, K., & Levine, S. (2017). Learning robust rewards with adversarial inverse reinforcement learning. *arXiv preprint arXiv:1710.11248*.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience, 10*(12), 1625–1633.

Lee, S. H., Song, M. S., Oh, M., & Ahn, W.-Y. (2024). Bridging the gap between self-report and behavioral laboratory measures: A real-time driving task with inverse reinforcement learning. *Psychological Science*, *35*(4), 345-357.

O'Doherty, J. P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current opinion in neurobiology*, *14*(6), 769-776.

O'Doherty, J. P., Cockburn, J., & Pauli, W. M. (2017). Learning, reward, and decision making. *Annual review of psychology, 68*, 73-100.

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience, 9*, Article 660.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Thompson, J. J., McColeman, C. M., Blair, M. R., & Henrey, A. J. (2019). Classic motor chunking theory fails to account for behavioural diversity and speed in a complex naturalistic task. *PloS one*, *14*(6), e0218251.

Wise, T., Emery, K., & Radulescu, A. (2023). Naturalistic reinforcement learning. *Trends in Cognitive Sciences*, *28*(2), 144-158.