



# Investigating the effects of environmental stochasticity on real-time decision-making in a naturalistic driving task using inverse reinforcement learning



Chae-Youn Chung<sup>1</sup> Sang Ho Lee<sup>2</sup> Jinwoo Jeong<sup>3</sup> Min-Hwan Oh<sup>4</sup> Woo-Young Ahn<sup>5,6,7</sup>\*

<sup>1</sup>Interdisciplinary Program in Artificial Intelligence, Seoul Natl. Univ., Seoul, Korea, Republic of; <sup>2</sup>Grad. Sch. of Digital Humanities and Computational Social Sci., KAIST, Daejeon, Korea, Republic of; <sup>3</sup>Dept. of Computer Science, Georgia Inst. of Tech., Atlanta, GA; <sup>4</sup>Grad. Sch. of Data Sci., Seoul Natl. Univ., Seoul, Korea, Republic of; <sup>5</sup>Dept. of Psychology, Seoul Natl. Univ., Seoul, Korea, Republic of; <sup>6</sup>Department of Brain and Cognitive Sciences, Seoul National University, Seoul, Republic of Korea; <sup>7</sup>AI Institute, Seoul National University, Seoul, Republic of Korea

## Introduction

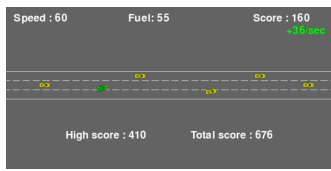
### Research Goal

- Develop a naturalistic behavioral task measures personal traits such as impulsivity.
- Use and validate IRL method to capture and analyze real-time human decision making.

### Research Questions

- How do people with different personality traits behave differently under stochastic environments?
- How are differences in personal traits reflected in the IRL reward?

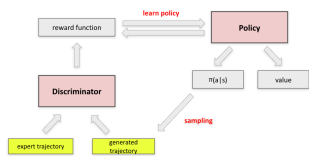
### Highway Task (Lee et al., 2024a)



- Behavioral tasks and self-report questionnaires often showed weak correlation, although they aim to measure the same psychological construct, because of uncommon stimuli and use simple statistics to analysis (Dang et al., 2020).
- Naturalistic tasks provides complex environment with greater ecological validity and large state-action pairs.

- Lee et al (2024a) developed a naturalistic driving task, (static) highway task.
- Lee et al (2024a) used Adversarial Inverse Reinforcement Learning (AIRL) to model underlying individualized reward function from observed behavior. In specific, events such as overtaking or crash, the IRL reward can be a indicator for impulsivity.
- In the fMRI follow-up study, Lee et al. (2024b) showed that BOLD signal in reward circuitry (e.g., prefrontal cortex, striatum, and insula) and IRL reward showed strong correlation.
- In this study, we developed dynamic highway task, where the yellow non-player cars change lane stochastically, which adds ambiguity onto the environment.

### Adversarial Inverse Reinforcement Learning (Fu et al., 2018)



- In inverse reinforcement learning (IRL), agents learn the optimal policy by inferring the underlying reward function from observations of an expert's behavior (i.e., the expert's policy).
- In AIRL, the agent learns the reward function through adversarial learning.

## Method

N = 50 (Male = 25, Female = 25)

- 1) Surveys
  - Impulsivity: Barrat Impulsivity Scale-11 (BIS; Patton et al., 1995)
  - Anxiety: State-Trait Anxiety Inventory-Y (STAI-Y; Spielberger et al., 1971)
- 2) Highway task (the order was counterbalanced)
  - Static environment (17 min \* 2)
  - Dynamic environment (17 min \* 2)

### References

Dang, J., King, K. M., & Inzlicht, M. (2020). Why are self-report and behavioral measures weakly correlated?. *Trends in cognitive sciences*, 24(4), 267-269.

Lee, S. H., Song, M. S., Oh, M. H., & Ahn, W. -Y. (2024a). 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.

Lee, S. H., Oh, M. H., & Ahn, W. -Y. (2024b). Inverse reinforcement learning captures value representations in the reward circuit in a real-time driving task: a preliminary study, 2024 Conference on Cognitive Computational Neuroscience.

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

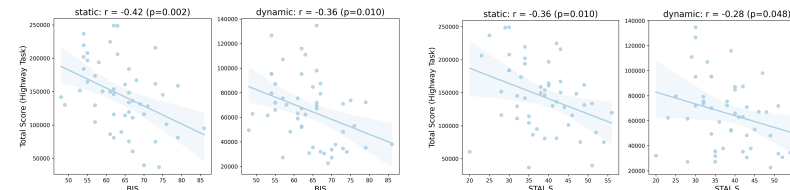
Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of clinical psychology*, 51(6), 768-774.

Spielberger, C. D., Gonzalez-Reigosa, F., Martinez-Uruti, A., Natalicio, L. F., & Natalicio, D. S. (1971). The state-trait anxiety inventory. *Revista Interamericana de Psicologia/Interamerican Journal of psychology*, 5(3 & 4).

## Preliminary Results (n=50)

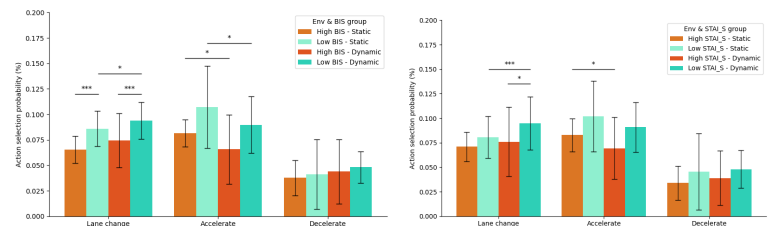
### Descriptive analysis

#### 1. Correlation with highway task scores and self-report measures



- BIS showed significant negative correlation with the highway task score both in static ( $r = -0.42$ ,  $p < 0.01$ ), and dynamic ( $r = -0.36$ ,  $p = 0.010$ ) environments. In particular, BIS-nonplanning sub scale was significantly correlated with task performance in the static ( $r = -0.36$ ,  $p < 0.01$ ) and dynamic ( $r = -0.34$ ,  $p = 0.017$ ) conditions.
- STAI\_S showed a significant negative correlation with highway task score in the static ( $r = -0.36$ ,  $p = 0.01$ ), and dynamic ( $r = -0.28$ ,  $p < 0.05$ ) environment. STAI\_T also showed a significant correlation in the dynamic ( $r = -0.32$ ,  $p = 0.026$ ) condition, while whereas the correlation in the static condition was not significant ( $r = -0.27$ ,  $p = 0.06$ ).

#### 2. Distribution of action selection



#### Environment effect (static vs. dynamic)

- Dynamic condition successfully differentiate High and Low BIS group in choice of acceleration and lane-changing actions, High and Low STAI\_S group in lane changing actions.

#### Trait effect (High vs. Low BIS/STAI group)

- High BIS group chose lane-changing actions fewer than Low BIS group (static and dynamic).
- High STAI\_S group chose lane-changing and acceleration less than Low STAI\_S group (dynamic).

## Discussions

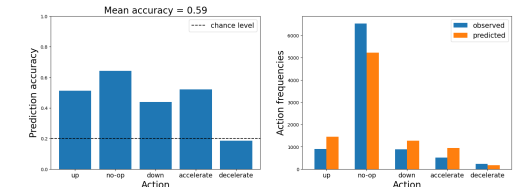
- Adding stochasticity to the environment effectively revealed significant correlation with self-report measures, particularly with BIS-nonplanning, which reflects one's ability for planning and deliberate thinking.
- IRL analysis method provides a framework to examine rewards underlying real-time decision making, which enables to capture time-varying reward of individual, that are not observable through summary statistics.
- Distinct IRL reward trajectories during overtaking and blocked-overtaking across impulsivity groups indicate that IRL reward captures individual differences in how people perceive and evaluate a situation based on personal traits.

### Presenter

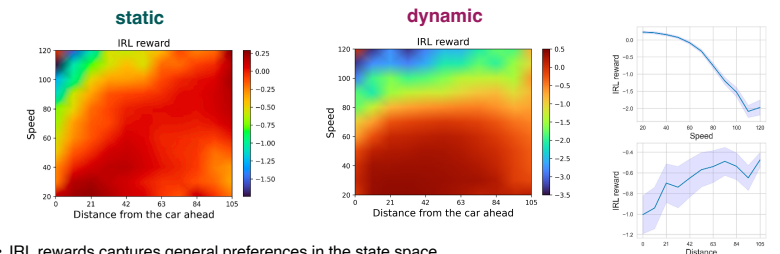
Chaeyoun Chung (chaeyoungchung@snu.ac.kr)  
 LinkedIn: chaeyoun-chung

### IRL analysis (cont, n=43)

#### 1. Model accuracy (dynamic)



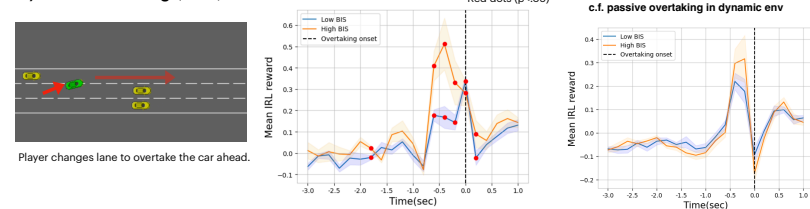
#### 2. Mean IRL reward in speed, distance state space



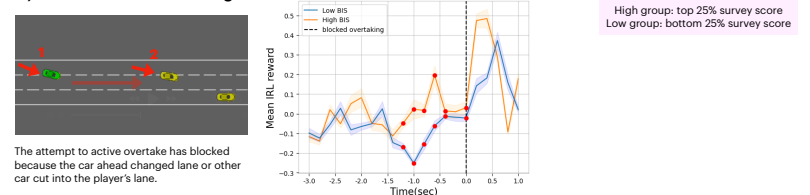
- IRL rewards captures general preferences in the state space.
- Compared to the static environment, high speed is generally less favorable in the dynamic environment. Specifically, high speed combined with close distance from the car ahead leads to substantially lower IRL rewards, suggesting increased penalization in such conditions.

#### 3. (Preliminary) Event-related IRL reward trajectories (dynamic)

##### 1) active overtaking (n=7569)



##### 2) blocked active overtaking (n=1619)



- During active overtaking, which is considered a riskier behavior than passive overtaking, the High BIS group showed an increase in IRL reward before overtaking, indicating that choosing to actively overtake was perceived as rewarding than in Low BIS group.
- When the attempt to actively overtake was blocked due to environmental stochasticity, the Low BIS group showed a decrease in IRL reward, while High BIS group showed an increase in IRL reward, suggesting that they perceived environmental stochasticity as punishing.