Title: Bridging the gap between self-report and behavioral laboratory measures: A real-time driving task with inverse reinforcement learning

Authors: Sang Ho Lee[1,2], Myeong Seop Song[1], Min-hwan Oh[3], and Woo-Young Ahn[1,2]

[1] Department of Psychology, Seoul National University, Seoul, Korea 08826

[2] Department of Brain and Cognitive Sciences, Seoul National University, Seoul, Korea 08826

[3] Graduate School of Data Science, Seoul National University, Seoul, Korea 08826

Corresponding author: Woo-Young Ahn, Ph.D.

Department of Psychology, Seoul National University, Seoul, Korea 08826

Tel: +82-2-880-2538, Fax: +82-2-877-6428.

E-mail: wahn55@snu.ac.kr

Competing Interest Statement: The authors declare that they have no known financial or non-financial competing interests.

Keywords: Impulsivity, realistic experiment, inverse reinforcement learning, deep learning, driving

Abstract

A major challenge in assessing psychological constructs such as impulsivity is the weak correlation between self-report and behavioral task measures that are supposed to assess the same construct. To address this issue, we developed a real-time driving task called the "highway task", where participants often exhibit impulsive behaviors, such as reckless driving, thereby mirroring real-life impulsive traits captured by self-report surveys. Here, we first show that a self-report measure of impulsivity is highly correlated with performance in the highway task, but not with traditional behavioral task measures of impulsivity. By integrating deep neural networks with an inverse reinforcement learning (IRL) algorithm, we inferred dynamic changes of subjective rewards during the highway task. The IRL results indicated that impulsive participants attribute high subjective rewards to irrational or risky driving behaviors and situations. Overall, our results suggest that using real-time tasks combined with IRL can help reconcile the discrepancy between self-report and behavioral task measures of psychological constructs including impulsivity, with IRL being a practical modeling framework for multidimensional data from real-time tasks.

Introduction

Self-report and behavioral task measures are among the most frequently employed methods for assessing psychological constructs. A prevalent issue across multiple domains such as impulsivity (1), self-control (2), empathy (3), and risk preference (4) is that self-report and behavioral task measures consistently show weak correlations with each other, even when they are assumed to tap the same construct (5). Weak associations between measures of the same construct fosters ambiguity and confusion in assessment, making it challenging to integrate findings across different measures.

We address this problem using impulsivity as a testbed, as it is one of the psychological constructs that is notably affected by the weak association between self-report and behavioral measures. Extensive studies of impulsivity in relation to mental disorders and maladaptive behaviors (6–8) have utilized a range of self-report and behavioral task measures, which are believed to assess the same construct termed "impulsivity". However, large-scale investigations and meta-analyses consistently reported weak correlations between different measures of impulsivity (1, 9, 10).

A widely accepted approach is to view impulsivity as a multidimensional construct with distinct aspects that do not necessarily overlap with each other. For example, MacKillop et al. (9) suggests that measures of impulsivity can be categorized into three distinct domains: impulsive choice, impulsive action, and impulsive personality traits. According to this categorization, behavioral task measures, which reflect impulsive choice (e.g., delay discounting task) (11) and impulsive action (e.g., go/no-go task) (12), do not need to correlate with self-report measures that typically reflect trait impulsivity (e.g., Barratt Impulsiveness Scale) (13). Another explanation for the inconsistency

between self-report and behavioral task measures of impulsivity is that they tap the same construct, but the association between them are obscured by differences in measurement methods (10, 14). Self-reports measure individuals' overall tendencies over a longer duration of time (e.g., for the past week/month), whereas laboratory behavioral tasks usually measure specific behaviors in some discrete states (e.g., go and no-go conditions in the go/no-go task) in highly controlled settings at the time of testing. Thus, behavioral task measures may capture state-specific phenomena that only partly reflect self-reported tendencies of behaviors across situations in real life.

Building on this methodological explanation, we postulate that a laboratory task conducted in real-time, which mimics real-life situations, would yield impulsivity measures that are strongly correlated with self-reported impulsivity. Specifically, we develop and implement a real-time driving task, dubbed the "highway task" (**Figure 1**), in which participants control a car on a highway to drive as fast as possible without crashing into other cars. The performance in the task may reflect the traits that contribute to reckless driving, which is frequently associated with impulsivity (15, 16). Unlike traditional trial-based laboratory tasks with a predefined list of discrete states, the highway task provides trajectories of states that continuously interact with participants' actions (e.g., accelerate, change lanes), with a number of possible states being virtually boundless (see the following sections for details). Behaviors in this task might be a better reflection of trait impulsivity than traditional behavioral task measures, as the task environment resembles complex real-world situations where impulsive behaviors occur (17, 18).

A challenge is how to describe complex data from the highway task beyond simple summary statistics of observed behaviors (e.g., mean speed, number of crashes). Computational modeling is widely recognized as a valuable tool for assessing neurocognitive characteristics underlying behaviors (19). However, traditional computational models may not be readily applicable to data from real-time tasks (e.g., virtual reality, arcade-style games such as Atari), because such models are not typically designed to describe multidimensional behaviors with an immense number of possible states inherent in a real-time task.

We pose this problem as an *inverse* reinforcement learning (IRL) problem, in which the learning algorithm infers the reward function that underlies observed behaviors (20). The objective of IRL is opposite to that of the conventional "forward" approach of reinforcement learning (RL) (21); IRL learns a reward function based on observed behaviors, whereas RL learns a behavioral policy based on observed rewards. Recent advances in algorithmic techniques made IRL well-suited for explaining behaviors in complex environments (22, 23). One of the breakthroughs is the use of deep neural networks (DNN), which can represent complex associations between states and actions in real-time tasks (24), to approximate complex, nonlinear reward functions (22). By integrating DNN with IRL (i.e., deep IRL), we are not restricted to any particular functional form of rewards for observed behaviors. Once learned from the observed data, DNN can calculate rewards for given states and actions. The rewards derived by IRL can be interpreted as a participant's internal reward or preference, making it a valuable tool for modeling human decision making (25).

In the current study, we aim to find indicators of impulsivity in the highway task by comparing IRL-inferred reward functions among participants with varying levels of trait impulsivity. To our knowledge, this is the first study to employ deep IRL to capture individual differences in a psychological construct among human participants. While a few studies modeled human decision making using simpler IRL algorithms using a restricted class of functional forms (25), no other studies have utilized deep IRL to investigate individual differences in reward functions in real-time tasks.

If the behaviors exhibited in the highway task align with the hypothesized trait impulsivity, we would expect the task performance and the rewards inferred by IRL to correlate with measures of trait impulsivity. In our experiment (N = 47; determined by the power analysis described in **SI Methods**), we assessed trait impulsivity using the Barratt Impulsiveness Scale (BIS) (13), which is a widely used self-report measure of impulsivity. The experiment consisted of three behavioral tasks, including the highway task and two traditional behavioral tasks measuring impulsivity: delay discounting and go/no-go tasks. While the latter two tasks measure impulsive choice and impulsive action, respectively, the highway task was chosen to examine its correlation with the BIS score in comparison to other measures.

In the following analysis, we evaluate the credibility of IRL in explaining individual differences in behavior by assessing its accuracy in predicting actions during the highway task. We then investigate the IRL-inferred reward for each state in the task as well as the *real-time trajectory* of the rewards to find indicators of impulsivity. Finally, the behavioral performance measures (e.g., task score) of the highway task and the output of the IRL are used together to predict the BIS score.

Results

We first assessed the validity of the highway task by correlating its task score (see **Materials and Methods** for score calculation) with the total BIS score, which we used as the benchmark measure of trait impulsivity in the current study. The credibility of correlation was assessed using the Bayes factor (BF_{10}) (26) in a Bayesian correlation test (27). Following the classification scheme in Wagenmakers et al. (28), we interpreted BF_{10} of 1-3, 3-10, 10-30, 30-100 and >100 as anecdotal, moderate, strong, very strong, and extreme evidence, respectively. Statistical evidence was strong for the Pearson correlation between the mean score of the highway task and the BIS score ($r = 0.46$, $BF_{10} = 28.41$), suggesting that the task performance improves as the BIS score (i.e., impulsivity) decreases (see **Figure 2**). The task score was also reliable within the task. The split-half reliability between the scores in the first half and the second half of the task assessed by intraclass correlation coefficient (ICC) (29) was 0.72, which is acceptable (30). In contrast, the BIS score did not correlate with widely used behavioral measures of impulsivity from the two traditional laboratory tasks, namely, the delay discounting rate parameter (log k; see **SI Methods** for model description) in the delay discounting task ($r = 0.01$; $BF_{10} = 0.182$) and the no-go error rate in the go/no-go task ($r = 0.07$; $BF_{10} = 0.203$). The results supported our hypothesis that a real-time task in a realistic environment better reflects impulsivity than traditional trial-based tasks. Other statistics derived from the highway task such as the number of crashes ($r = 0.15$; $BF_{10} = 0.3$), mean speed ($r = 0.08$; $BF_{10} = 0.21$), and mean distance from the car ahead ($r = 0.08$; $BF_{10} = 0.21$) did not show any statistical evidence for correlation with the BIS score. These

measures also showed no evidence for correlation with other behavioral task measures, except that the mean speed considerably correlated with the error rate in the go/no-go task (r = -0.38; BF_{10} = 4.87; see **SI Figure S1** for the full correlation matrix).

We used IRL to infer a reward function for each individual based on the individual's observed trajectories of behaviors in the highway task (see **SI Methods** for the details of the algorithm). We investigated the individual differences in the reward functions learned via IRL to identify latent indicators of impulsivity. Prior to interpreting the reward functions, we evaluated the goodness of fit of the model trained by IRL. The model fit was assessed by comparing observed participants' actions with artificial agents' actions generated by the behavioral policies of IRL. If IRL learned the reward functions that accurately explain the data, the actions produced by the agent should closely resemble participants' behaviors. **Figure 3A** shows the mean accuracy of the IRL agents in predicting five possible actions in the highway task; moving up, no action (i.e., no-op), moving down, acceleration, and deceleration. The accuracy was much higher than the chance level (mean accuracy = 0.64; chance level accuracy = 0.2) for all actions except the deceleration.

The IRL agents also showed similar proportions of actions throughout the action trajectory (**Figure 3B**) to observed human actions. A noticeable difference between the IRL agents and the participants was that the participants showed higher mean proportion of no actions. Cross et al. (31) found comparable differences in actions between humans and artificial agents trained by a "forward" RL. In their study, the policy learned via deep Q-learning (DQN) (24) showed a greater proportion of no actions in Atari games, compared to human participants. The authors postulated that humans are more inclined to

abstain from taking action due to metabolic costs and physical constraints (e.g., response speed) for taking action. While the IRL agents learn from human demonstrations that reflect constraints on human behaviors, they might not replicate infrequent inaction due to fatigue or inattention in situations where the participant typically took action.

The similarity between the participants' actions and those of IRL agents suggests that the reward functions derived from IRL reflect subjective rewards underlying observed behaviors. We then assessed whether the IRL reward functions are sensible and interpretable by visually examining the reward functions. The DNN trained by IRL approximated subjective rewards for all possible states in the task. The state in the task was defined as a combination of eleven manually annotated features; the speed of the own car, the lane on which the own car is located, the speed of other cars on each of the three lanes (three features for three lanes), the distance from the closest car ahead on each lane (three features), and the distance from the closest car behind on each lane (three features). **Figure 3C** illustrates the mean reward functions (averaged across participants) in simplified state spaces. To visualize and interpret reward functions in a feasible way, we used mean rewards across the two most seemingly important features, own speed (i.e., y-axis in **Figure 3C**) and the distance from the closest car ahead on the own lane (i.e., x-axis in **Figure 3C**). The high reward states (i.e., dark red area) in the reward function suggests that participants generally favored driving at a low to moderate speed (20 – 60) and a close to moderate distance (10.5 – 63) from the closest car ahead. This reflected a rational strategy to avoid a crash while attempting to overtake a car ahead (i.e., decrease the speed when the distance between the own car and the car ahead is small). By contrast, the state with the smallest distance and the highest speed was associated with extremely

low rewards in that the state would likely result in a crash in the next step. The propensity to avoid a crash, which is the most punishing event in the task, is also reflected in the reward functions marginalized over the speed and distance axes (**Figure 3D**). The mean reward tended to increase with the distance from the car ahead and decrease with the speed, suggesting that participants generally used a safe strategy.

The results suggest that IRL successfully inferred sensible reward functions from participants' behaviors as we proposed. Nonetheless, our primary objective was to identify indicators of impulsive behaviors that may deviate from rational strategies in the highway task. To achieve this goal, we examined the association between the BIS score and the rewards derived by IRL. In the simplified state space shown in **Figure 3C**, we identified states showing statistical evidence for the correlation between the BIS score and the IRL rewards. Higher BIS score (i.e., increased impulsivity) corresponded to higher rewards for apparently irrational states; maximum speed (120) at close distances (0 – 21) and relatively low speed (50) at far distances (74 – 84) ($r = 0.35$-0.39, $BF_{10} > 3$; see **SI Figure S2** for the correlation coefficients across the state space).

We discovered more pronounced indicators of impulsivity from the trajectories of rewards. The reward function generated by IRL provides a simplified representation of how participants' behaviors were interpreted, but it does not depict changes in rewards over time. A real-time task involves trajectories of states and actions. The reward functions inferred by IRL can map these states into reward trajectories, which reveal real-time changes in rewards around significant events. We examined indicators of impulsivity in reward trajectories, with a focus on two salient events in the task: overtaking and crashing. Participants aimed to achieve the highest possible score by

quickly overtaking other cars without crashing into them. Video replays of task performance with real-time display of the IRL reward reveal noticeable changes in rewards at the moments of overtaking and crashing (see **SI Results** for the link to a video replay). This implies that the IRL algorithm identified these events as particularly critical. Subsequent analysis of reward trajectories indicates that IRL rewards during overtaking and crashing moments reflect participants' impulsivity.

The moments of overtaking and crashing were manually specified in the state space. Overtaking was defined as the moment when a car from an adjacent lane goes behind the participant's car. The distances from other cars ahead and behind the participant's own car were the state features used to identify overtaking moments. Further, two types of overtaking were distinguished: "active" overtaking, which occurs within one second of changing lane, and "passive" overtaking without a lane change. Pictures in the bottom of **Figure 4** depict an example of each event. Active overtaking is more hazardous than passive overtaking because changing lanes for overtaking at a close distance can lead to a collision with a car ahead. We hypothesized that success in this risky behavior would be highly rewarding for impulsive individuals. A crash was straightforward to define, as it was the moment when the distance from a car ahead becomes zero.

Figure 4 shows the reward trajectories before and after overtaking (-3 ~ 1 sec from the onset). To compare the reward functions between participants with high and low impulsivity, we grouped participants into two categories based on their BIS scores: the "high BIS" group and the "low BIS" group, consisting of participants in the highest and the lowest quartiles of the BIS score, respectively. The reward trajectories for passive

overtaking (**Figure 4A**) did not differ between the high and low BIS groups, with the IRL rewards showing no correlation with the BIS score across time points. By contrast, active overtaking revealed noticeable differences in the rewards between low and high BIS participants (**Figure 4B**). Compared to the low BIS group, the high BIS group showed a more rapid increase in IRL rewards prior to overtaking. Statistical evidence supported the correlation between the BIS score and the IRL reward at -1.8 – -1.2 seconds (r = 0.46; BF_{10} = 29.5) and -0.2 seconds (r = 0.38; BF_{10} = 5.6) from the moment of active overtaking (red dots in **Figure 4B** show the time points). The positive correlations suggest that impulsive participants favored the states with opportunities for active overtaking within a brief time frame (-1.8 – -1.2 seconds), as well as the moment just prior (-0.2 seconds) to successfully accomplishing it.

Reward trajectories for overtaking is likely to reflect participants' intention, because participants should overtake as many cars as possible to maximize the task score. Another salient event, crashing, is different in that it is an abrupt and unintentional event that should be avoided. The reward function on the distance axis (**Figure 3D**) show that the IRL reward drops at the moment of a crash (i.e., zero distance). Rewards at this moment correlated with the BIS score, but only at specific states where participants decelerated (r = 0.38; BF_{10} = 4). Impulsive participants (i.e., high BIS group) heavily discounted the reward for deceleration immediately before (-0.2 – 0 seconds) a crash, whereas non-impulsive participants (i.e., low BIS group) showed relatively steady rewards until the occurrence of a crash (see **SI Figure S3** for the reward trajectories for crashing). Rewards for other actions around a crash did not show such associations with the BIS score.

Preceding analyses found several indicators of impulsivity, one from a performance measure (i.e., task score) and others from rewards inferred by IRL. The variables that respectively correlate with the BIS score raise a question of whether using them altogether would help explain individual differences in impulsivity. To compare the informativeness of different types of variables, we predicted the BIS scores across individuals with regression analysis (i.e., LASSO) (32) using independent variables of three different categories: Traditional measures, highway task performance measures, and IRL measures. Traditional measures included error rate on no-go trials (variable name in the LASSO models: GNG error) and response time on go trials (GNG RT) in the go/no-go task and the logarithm of the delay discounting rate parameter (log(k)) in the delay discounting task. Highway task performance measures were the task score (Highway score), mean speed (Highway speed), mean distance from the car ahead (Highway distance), and the frequency of crashing (Highway n(crash)). In the models that used IRL rewards, only the rewards that correlated with the BIS score were included. They were the mean of the rewards for the subset of states that correlated with the BIS score in the two-dimensional state space depicted in **Figure 3C** (IRL speed*distance), mean of the rewards marked by red points in the reward trajectory for active overtaking (IRL overtaking), and the reward for deceleration at the moment of a crash (IRL crash).

Figure 5 shows the results from LASSO, which was conducted using glmnet (33) and easyml package (34) in Python. Model prediction was evaluated by the correlation between predicted and observed values of the BIS score in the test set. The histograms in **Figure 5** illustrates the distribution of the correlation coefficients. As expected by the correlation analyses, traditional measures from the delay discounting task and go/no-go

task were unable to predict the BIS score (**Figure 5A**; r = 0.1). The performance measures in the highway task showed the correlation similar to the correlation between the task score and the BIS score (**Figure 5B**; r = 0.48 vs. 0.46), with the task score explaining most of the variance (see the beta coefficients on the bottom row of **Figure 5B**). The selected variables from the IRL reward function were better than the performance measures at predicting the BIS score (**Figure 5C**; r = 0.72). Finally, a model with both performance measures and IRL measures did not show a better correlation score than the "IRL only" model (**Figure 5D**; r = 0.72), suggesting that the behavioral performance measures do not explain additional variance in the BIS score beyond the variance explained by IRL rewards.

Discussion

The lack of correlation between self-report and behavioral task measures of psychological constructs has long been a puzzle. We hypothesized that this discrepancy may be due to the simplicity of traditional behavioral tasks. Our findings regarding impulsivity demonstrate that measures derived from a real-time behavioral task do indeed correlate with a relevant self-report measure. This suggests that behavioral task measures can represent individual traits measured with a self-report questionnaire if the task offers a wide range of states where participants can exhibit diverse behaviors as in real-world situations.

The novelty of the current study stems from employing a deep IRL algorithm to extract participants' reward functions and individual differences in a real-time task. Past studies that associated impulsivity with driving behavior in real-world and simulated

environments typically focused on simple summary statistics of behaviors such as speeding, crashing, and traffic violations (15, 35). However, we found stronger indicators of impulsivity from IRL rewards than from summary statistics (e.g., mean speed, number of crashes). This suggests that IRL offers more than just descriptive analysis, as the reward functions can provide insights into participants' characteristics that may not be apparent in their behaviors. The successful application of IRL in this study highlights the potential of IRL as a modeling framework for real-time tasks that generate multi-dimensional data that are not easily described by conventional computational models.

Black-box machine learning models, which include DNN in the current IRL algorithm (23), have demonstrated high predictive performance, but often lack interpretability in their predictions (36). This absence of explanatory power has restricted the use of black-box models in human behavior research, where explanation is as important as prediction (37). IRL addresses this issue by providing each participant's rewards, which can be interpreted similarly to subjective values in computational models of decision making (e.g., (38, 39)). Participants would choose actions of highest subjective values (or IRL reward) or actions leading to the states with the highest subjective values. This approach enhances the interpretability of the model, making it more suitable for studying human behaviors.

Having behavioral task measures of trait impulsivity might help address some concerns about self-report measures. Concerns regarding the credibility of self-report measures exist due to potential response biases, which include responses influenced by social desirability, a consistent response tendency towards affirmative or negative responses, and a propensity for extreme or mid-point responses (40). In a naturalistic

paradigm including the highway task, participants are less likely to mask their traits or intentionally influence the assessment, because they are not directly questioned about their real-life tendencies and behaviors. A crucial next step for future research is to evaluate the validity of the highway task independently with clinical populations. For instance, we could investigate whether patients with impulsivity-related psychiatric disorders (e.g., substance use disorders; attention deficit hyperactivity disorder) exhibit lower scores in the highway task compared to healthy participants.

A remaining question is whether the rewards inferred by IRL truly represent the internal rewards experienced by participants, as hypothesized. A promising approach to address this question would be to investigate the associations between reward functions learned via IRL and brain activities related to reward (or value) processing. Rewards inferred by IRL might correspond to the representation of subjective values of predicted outcomes in the brain, which has been shown to correlate with functional magnetic resonance imaging (fMRI) activity in the regions such as orbitofrontal cortex (OFC) (41), ventromedial prefrontal cortex (vmPFC) (42), and ventral striatum (39). Predicting real-time changes in brain activities in these areas using the IRL rewards would help interpret IRL reward functions as subjective value functions that underlie human decision making. This approach would also validate the use of IRL in understanding the relationship between rewards and their neural correlates.

The current study focused on impulsivity measures, but our approach can be applied to other real-time tasks assessing different constructs (43). The use of the highway task aligns with recent studies employing realistic and real-time tasks to enhance the ecological validity of neuropsychological assessment (17). The adoption of real-time

tasks and data has increased, as recent technological advances (e.g., virtual reality, mobile devices) have facilitated experiments in realistic settings (44). Our work suggests that deep IRL serves as a practical modeling framework, enabling researchers to fully utilize complex data from real-time tasks without being restricted to simple descriptive analysis. Reward functions inferred by a deep IRL algorithm might reflect participants' subjective rewards or intentions in the task, which are central variables in the theories and models of decision making. In summary, the combination of real-time tasks and deep IRL offers a promising novel approach to improving the assessment of psychological constructs underlying human behaviors and decision making.

Materials and Methods

The experiment was approved by the institutional review board of Seoul National University (IRB No. 2112/004-004).

Participants. Forty-seven undergraduate and graduate students in Seoul National University participated. A Bayesian power analysis determined the number of participants (see **SI Methods** for details of the power analysis).

Procedures. Participants completed one questionnaire (Barratt impulsiveness scale) and three behavioral tasks (highway task, delay discounting task, and go/no-go task) in a dimly lit room. They could take a break between the tasks as long as they desired. An experimenter gave instructions to the participants at the beginning of each task. The questionnaire was controlled by Qualtrics (qualtrics.com) on a web browser. Behavioral experiments were controlled by Python scripts.

Barratt Impulsiveness Scale (BIS-11). We used a Korean version of the Barratt Impulsiveness Scale (45) to measure trait impulsivity. The questionnaire contained 30 questions that corresponds to the question in the English version of BIS-11 (13) Participants answered the questions on a four-point scale (Rarely/Never, Occasionally, Often, Almost Always/Always). Each answer was scored 1-4, with 4 indicating the most impulsive response. The BIS score was calculated by summing the scores across questions. The subscales of BIS (i.e., motor, nonplanning, and attentional impulsivity) were the sum of the scores across subsets of questions (see **SI Results** for the analysis using the BIS subscales).

Highway task. The task was built on a collection of OpenAI gym (46) environments for driving tasks (47). Task display and action input were controlled by Pygame package in Python. The goal of the highway task was to drive the green car on the screen as fast as possible without crashing into yellow cars (see **Figure 1** for the task display). Participants controlled the green car by pressing the arrow keys on the keyboard. The left and right arrow keys decreased and increased the speed by 10 distance/second, respectively. The up and down arrow keys moved the green car to the upper (left) lane and the lower (right) lane, respectively. The score in an episode incremented by $0.2 \times (\frac{speed}{10})^2$ every 0.2 second. An episode continued until the remaining fuel becomes zero or the green car crashes with a yellow car. The remaining fuel, which was displayed on the top of the screen, started as 60 and decreased at the rate of 1/second. Crashing into another car immediately decreased the score by 200. The score was reset to zero at the beginning of each episode. The highest score a participant

achieved in an episode was recorded as the "high score" (text below the road) until the participant score higher in another episode (see **SI Methods** for additional details of the task).

Traditional behavioral tasks. See **SI Methods** for details of the delay discounting task and the go/no-go task.

Inverse reinforcement learning (IRL) algorithm. We inferred the reward functions underlying the trajectories of states and actions in the highway task using adversarial inverse reinforcement learning (AIRL) (23), which is an IRL algorithm that achieves state-of-the-art performance. IRL is a challenging problem because multiple policy and reward functions can explain a given set of observed behaviors, leading to ambiguity in the learned reward function (20). AIRL is built on maximum entropy IRL (22, 48), which mitigates the ambiguity in solution by identifying a single reward function that maximizes the entropy of the policy derived from the rewards (49). AIRL also addresses the complexity of behaviors in a real-time task by using DNN to approximate nonlinear reward functions, whereas a majority of IRL methods (e.g., (48, 50)) assume linear reward functions that might be simplistic in complex tasks (see **SI Methods** for the details of the algorithm).

Acknowledgments

This work was supported by the National Research Foundation grant funded by the Ministry of Science, Information and Communication Technologies and Future Planning, the Korean Government (Grant No. 2021M3E5D2A0102249311 [to W-YA]); the BK21 FOUR Program (Grant No. 5199990314123 [to W-YA]); the Creative-Pioneering

Researchers Program through Seoul National University [to W-YA]; and the Creative Challenge Research Program through the National Research Foundation of Korea funded by the Ministry of Education (Grant No. 2022R1I1A1A01066530 [to SHL]). We thank Adam Gazzaley, Jay Myung, Mark Pitt, and Robert Whelan for their constructive feedback on the earlier version of the manuscript.

References

1. L. Sharma, K. E. Markon, L. A. Clark, Toward a Theory of Distinct Types of "Impulsive" Behaviors: A Meta-Analysis of Self-Report and Behavioral Measures. *Psychol Bull* 140, 374–408 (2014).

2. B. Saunders, *et al.*, Reported Self-control is not Meaningfully Associated with Inhibition-related Executive Function: A Bayesian Analysis. *Collabra: Psychol.* 4 (2018).

3. B. A. Murphy, S. O. Lilienfeld, Are Self-Report Cognitive Empathy Ratings Valid Proxies for Cognitive Empathy Ability? Negligible Meta-Analytic Relations With Behavioral Task Performance. *Psychol. Assess.* 31, 1062–1072 (2019).

4. R. Frey, A. Pedroni, R. Mata, J. Rieskamp, R. Hertwig, Risk preference shares the psychometric structure of major psychological traits. *Sci. Adv.* 3, e1701381 (2017).

5. J. Dang, K. M. King, M. Inzlicht, Why Are Self-Report and Behavioral Measures Weakly Correlated? *Trends Cogn. Sci.* 24, 267–269 (2020).

6. J. L. Perry, M. E. Carroll, The role of impulsive behavior in drug abuse. *Psychopharmacology* 200, 1–26 (2008).

7. S. P. Whiteside, D. R. Lynam, The Five Factor Model and impulsivity: using a structural model of personality to understand impulsivity. *Pers Indiv Differ* 30, 669–689 (2001).

8. J. MacKillop, *et al.*, Multidimensional Examination of Impulsivity in Relation to Disordered Gambling. *Exp Clin Psychopharm* 22, 176–185 (2014).

9. J. MacKillop, *et al.*, The latent structure of impulsivity: impulsive choice, impulsive action, and impulsive personality traits. *Psychopharmacology* 233, 3361–3370 (2016).

10. M. A. Cyders, A. Coskunpinar, The relationship between self-report and lab task conceptualizations of impulsivity. *J Res Pers* 46, 121–124 (2012).

11. L. Green, J. Myerson, A Discounting Framework for Choice With Delayed and Probabilistic Rewards. *Psychol Bull* 130, 769–792 (2004).

12. C. M. Hartung, R. Milich, D. R. Lynam, C. A. Martin, Understanding the Relations Among Gender, Disinhibition, and Disruptive Behavior in Adolescents. *J Abnorm Psychol* 111, 659–664 (2002).

13. J. H. Patton, M. S. Stanford, E. S. Barratt, Factor structure of the barratt impulsiveness scale. *J. Clin. Psychol.* 51, 768–774 (1995).

14. M. A. Cyders, A. Coskunpinar, Measurement of constructs using self-report and behavioral lab tasks: Is there overlap in nomothetic span and construct representation for impulsivity? *Clin Psychol Rev* 31, 965–982 (2011).

15. J. Hatfield, A. Williamson, E. J. Kehoe, P. Prabhakharan, An examination of the relationship between measures of impulsivity and risky simulated driving amongst young drivers. *Accid Analysis Prev* 103, 37–43 (2017).

16. F. O'Brien, M. Gormley, The contribution of inhibitory deficits to dangerous driving among young people. *Accid Analysis Prev* 51, 238–242 (2013).

17. K. Robertson, M. Schmitter-Edgecombe, Naturalistic tasks performed in realistic environments: a review with implications for neuropsychological assessment. *Clin Neuropsychologist* 31, 16–42 (2017).

18. A. Verdejo-Garcia, *et al.*, A unified online test battery for cognitive impulsivity reveals relationships with real-world impulsive behaviours. *Nat Hum Behav* 5, 1562–1577 (2021).

19. S. Palminteri, V. Wyart, E. Koechlin, The Importance of Falsification in Computational Cognitive Modeling. *Trends Cogn Sci* 21, 425–433 (2017).

20. S. Arora, P. Doshi, A survey of inverse reinforcement learning: Challenges, methods and progress. *Artif Intell* 297, 103500 (2021).

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

22. M. Wulfmeier, P. Ondruska, I. Posner, Maximum Entropy Deep Inverse Reinforcement Learning. *Arxiv* (2015) https:/doi.org/10.48550/arxiv.1507.04888.

23. J. Fu, K. Luo, S. Levine, Learning Robust Rewards with Adversarial Inverse Reinforcement Learning. *Arxiv* (2017) https:/doi.org/10.48550/arxiv.1710.11248.

24. V. Mnih, *et al.*, Human-level control through deep reinforcement learning. *Nature* 518, 529–533 (2015).

25. R. Zhang, *et al.*, Modeling sensory-motor decisions in natural behavior. *Plos Comput Biol* 14, e1006518 (2018).

26. R. E. Kass, A. E. Raftery, Bayes Factors. *Journal of the American Statistical Association* 90, 773–795 (1995).

27. R. Wetzels, E.-J. Wagenmakers, A default Bayesian hypothesis test for correlations and partial correlations. *Psychon B Rev* 19, 1057–1064 (2012).

28. E.-J. Wagenmakers, *et al.*, Bayesian inference for psychology. Part II: Example applications with JASP. *Psychon B Rev* 25, 58–76 (2018).

29. T. K. Koo, M. Y. Li, A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Medicine* 15, 155–163 (2016).

30. D. V. Cicchetti, Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. *Psychol Assessment* 6, 284–290 (1994).

31. L. Cross, J. Cockburn, Y. Yue, J. P. O'Doherty, Using deep reinforcement learning to reveal how the brain encodes abstract state-space representations in high-dimensional environments. *Neuron* 109, 724-738.e7 (2021).

32. R. Tibshirani, Regression Shrinkage and Selection Via the Lasso. *J Royal Statistical Soc Ser B Methodol* 58, 267–288 (1996).

33. J. Friedman, T. Hastie, R. Tibshirani, Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 33, 1–22 (2010).

34. W.-Y. Ahn, P. Hendricks, N. Haines, Easyml: Easily Build and Evaluate Machine Learning Models. *Biorxiv*, 137240 (2017).

35. P. Bıçaksız, T. Özkan, Impulsivity and driver behaviors, offences and accident involvement: A systematic review. *Transp Res Part F Traffic Psychology Behav* 38, 194–223 (2016).

36. C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 1, 206–215 (2019).

37. T. Yarkoni, J. Westfall, Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspect Psychol Sci* 12, 1100–1122 (2017).

38. C. S. Sripada, R. Gonzalez, K. L. Phan, I. Liberzon, The neural correlates of intertemporal decision-making: Contributions of subjective value, stimulus type, and trait impulsivity. *Hum. Brain Mapp.* 32, 1637–1648 (2011).

39. J. W. Kable, P. W. Glimcher, The neural correlates of subjective value during intertemporal choice. *Nat Neurosci* 10, 1625–1633 (2007).

40. A. Furnham, M. Henderson, The good, the bad and the mad: Response bias in self-report measures. *Pers Indiv Differ* 3, 311–320 (1982).

41. J. A. Gottfried, J. O'Doherty, R. J. Dolan, Encoding Predictive Reward Value in Human Amygdala and Orbitofrontal Cortex. *Science* 301, 1104–1107 (2003).

42. M. P. Paulus, L. R. Frank, Ventromedial prefrontal cortex activation is critical for preference judgments. *Neuroreport* 14, 1311–1315 (2003).

43. J. A. Anguera, *et al.*, Video game training enhances cognitive control in older adults. *Nature* 501, 97–101 (2013).

44. T. D. Parsons, Virtual Reality for Enhanced Ecological Validity and Experimental Control in the Clinical, Affective and Social Neurosciences. *Front Hum Neurosci* 9, 660 (2015).

45. S.-R. Lee, *et al.*, The Study on Reliability and Validity of Korean Version of the Barratt Impulsiveness Scale-11-Revised in Nonclinical Adult Subjects. *Journal of the Korean Neuropsychiatric Association* 51, 378–386 (2012).

46. G. Brockman, *et al.*, OpenAI Gym. *Arxiv* (2016) https:/doi.org/10.48550/arxiv.1606.01540.

47. E. Leurent, An Environment for Autonomous Driving Decision-Making. *GitHub repository* (2018).

48. B. D. Ziebart, A. Maas, J. A. Bagnell, A. K. Dey, Maximum Entropy Inverse Reinforcement Learning in *AAAI Conference on Artificial Intelligence*, (2008), pp. 1433–1438.

49. A. J. Snoswell, S. P. N. Singh, N. Ye, Revisiting Maximum Entropy Inverse Reinforcement Learning: New Perspectives and Algorithms. *Arxiv* (2020) https:/doi.org/10.48550/arxiv.2012.00889.

50. P. Abbeel, A. Y. Ng, Apprenticeship learning via inverse reinforcement learning. *Twenty-first Int Conf Mach Learn - Icml '04*, 1 (2004).

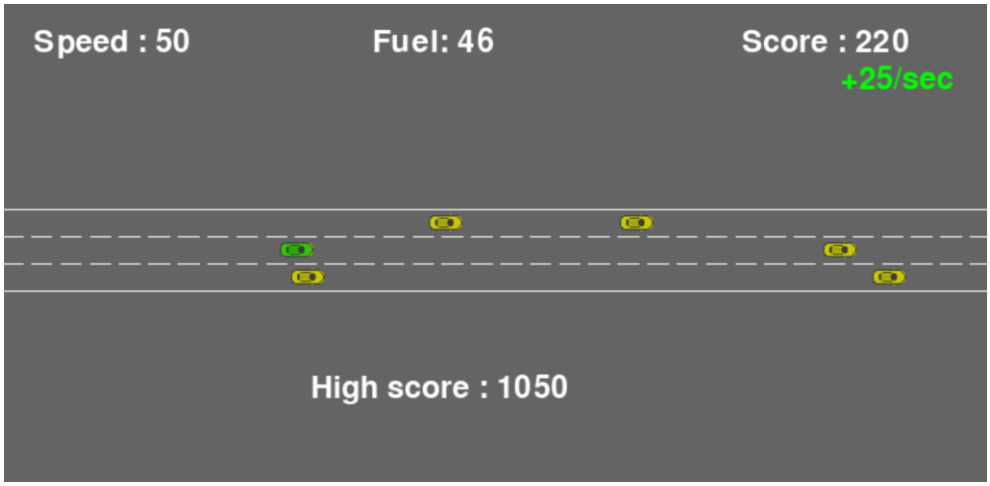

Figure 1. A screenshot of the highway task. Participants control the green car to drive as fast as possible without crashing into yellow cars. Score per second increases with speed. An episode (or trial) continues until the car crashes or runs out of fuel. High score indicates the highest score achieved in an episode during the highway task.

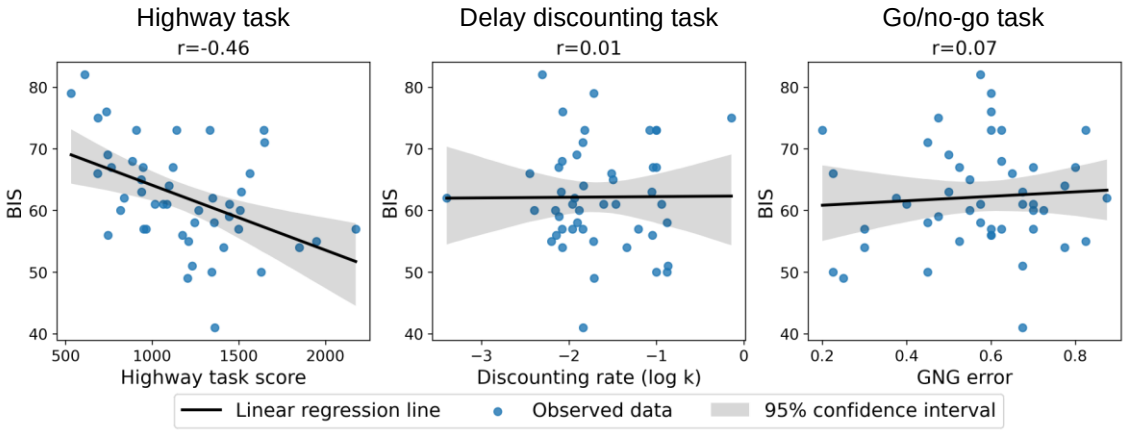

Figure 2. The correlation between the BIS score and behavioral task measures of impulsivity.

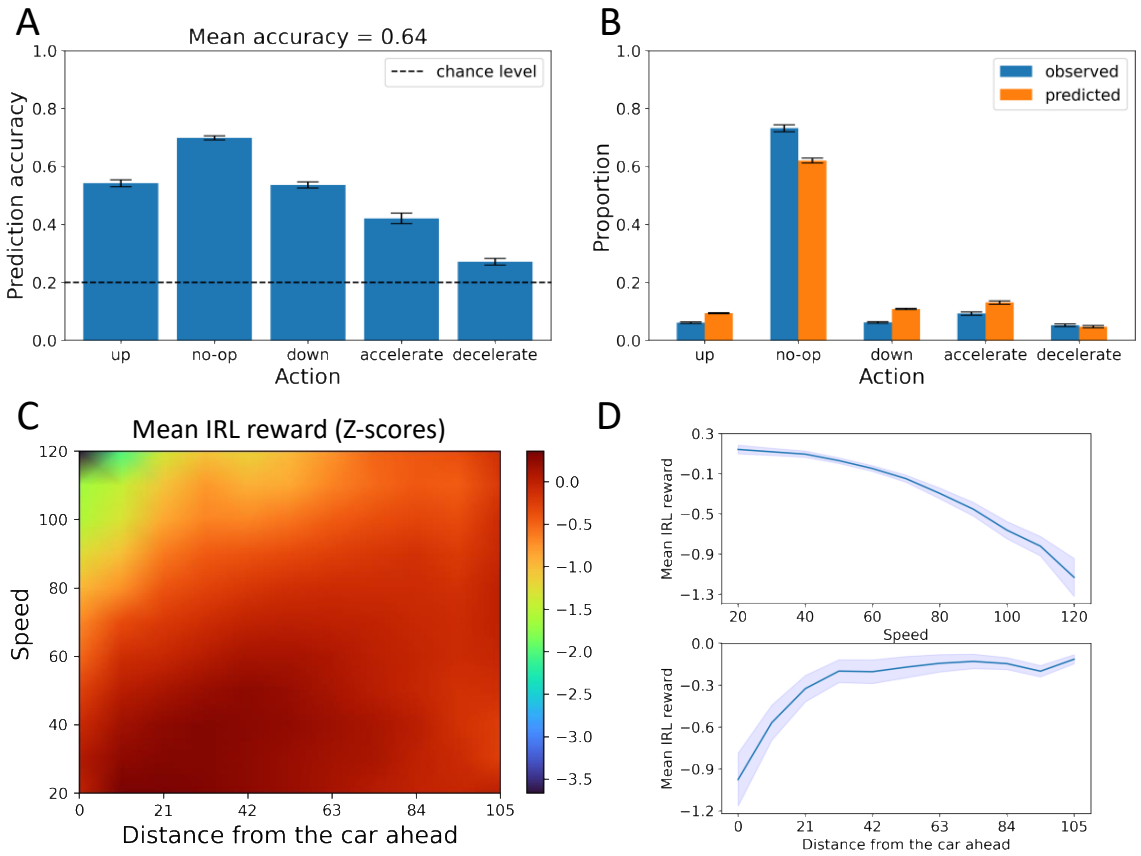

Figure 3. Performance of the IRL algorithm. A) Mean accuracy of the five possible actions produced by the IRL agents. B) Mean proportions of actions in the observed data (blue) and the action trajectories generated by the IRL agents (orange). C) Mean IRL rewards in a state space defined by the combination of the speed of the green car (y-axis) and the distance from the closest car on the same lane (x-axis). D) Mean IRL rewards marginalized over the speed and distance axes. Error bars in A-B) and the blue areas in D) indicate standard errors of the means.

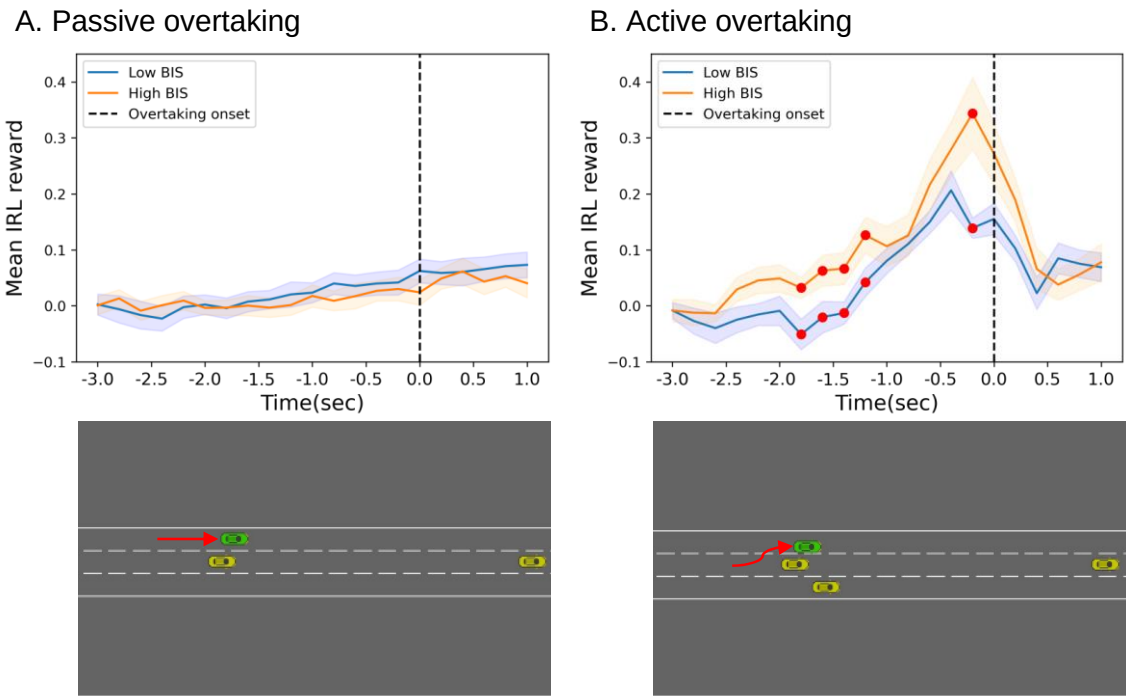

Figure 4. Reward trajectories before and after overtaking moments. The red points on the lines in B indicate the time points at which the rewards credibly correlated with the BIS score ($BF_{10} > 3$). The pictures on the bottom are examples of passive and active overtaking in the highway task. Red arrowed lines in the pictures illustrate the movement trajectories of the own (green) car.

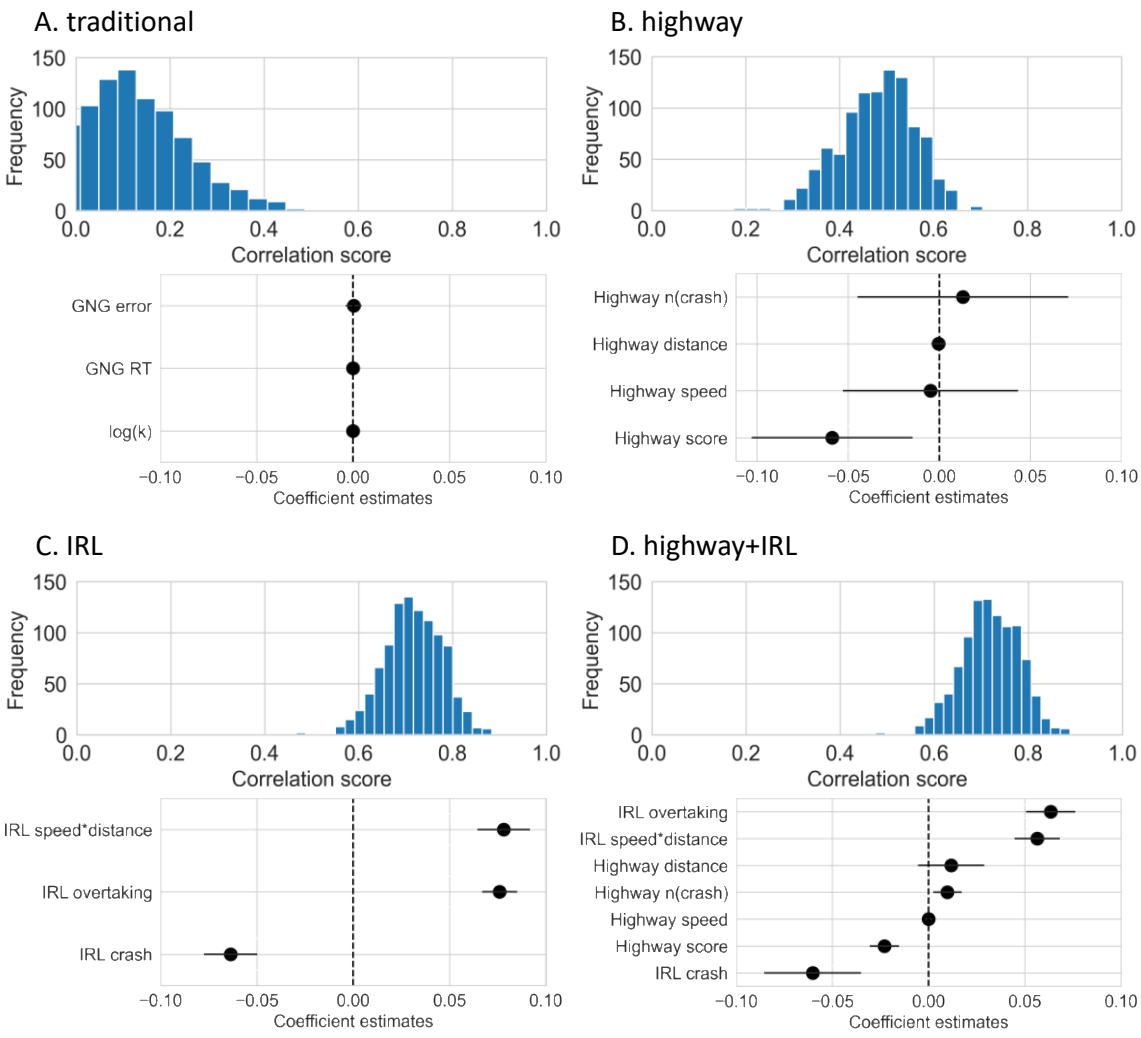

Figure 5. Model fit and beta coefficients of the Lasso models that predicted the BIS score using different independent variables. The histograms show the distributions of the correlation score, which is the correlation coefficients between observed and predicted values of the BIS score. The graphs displayed below the histograms depict the beta coefficient for each variable. The error bars represent the 95% confidence intervals.