Dynamical Driving Interactions between Human and Mentalizing-designed Autonomous Vehicle

1st Yikang Zhang †
Dept. of Biomedical Engineering
SUSTech
Shenzhen, China
yikangzhang@umass.edu

co-1st Shuo Zhang †
Dept. of Biomedical Engineering
SUSTech
Shenzhen, China
12132660@mail.sustech.edu.cn

3rd Zhichao Liang
Dept. of Biomedical Engineering
SUSTech
Shenzhen, China
11930756@mail.sustech.edu.cn

4th Hanran Luna Li
Centre for Cognitive and Brain Sciences and Department of Psychology
University of Macau
Macau, China
hanranli0902@gmail.com

5th Haiyan Wu *
Centre for Cognitive and Brain Sciences and Department of Psychology
University of Macau
Macau, China
haiyanwu@um.edu.mo

6th Quanying Liu *
Dept. of Biomedical Engineering
SUSTech
Shenzhen, China
liuqy@sustech.edu.cn

Abstract—**Autonomous vehicle (AV) is progressing rapidly, but there are still many shortcomings when interacting with humans. To address this problem, it is necessary to study the human behaviors in human-AV interactions, and build a predictive model of human decision-making in the interaction. In turn, modelling human behavior in human-AV interaction can help us better understand human perception of AVs and human driving strategies. In this work, we first train multi-level AV agents using reinforcement learning (RL) models to imitate three mentalizing levels (*i.e.*, level-0, level-1, and level-2), and then design a human-AV driving task that subjects interact with each level of AV agents in a two-lane merging scenario. Both human and AV driving behaviors are recorded. We found that conservative subjects obtain more rewards because of the randomness of the RL agents. Our results indicate that (*i*) human driving strategies are flexible and changeable, which allows to quickly adjust the strategy to maximize the reward when gaming against AV; (*ii*) human driving strategies are related to mentalizing ability, and subjects with higher mentalizing scores drive more conservatively. Our study shed lights on the relationship between human driving policy and mentalizing in human-AV interactions, and it can inspire the next-generation AV.**

Index Terms—**Autonomous vehicle (AV), human-AV interaction, reinforcement learning, human behavior, decision making**

I. INTRODUCTION

In recent years, autonomous driving has made great progress in becoming safer, more efficient, and less time-consuming. Many studies have been focused on designing algorithms for autonomous vehicles (AVs) in terms of trajectory prediction, planning and decision making, such as DIM [1], TNT [2], LaneGCN [3], achieving impressive results. Few studies have focused on AV-human interactions. Many Motion Prediction and Planning methods are proposed for autonomous vehicles Decision Making, for example, [4] [5]. However, traditional rule-based methods usually suffer from unpredictable interaction. Reinforcement learning (RL) is a machine learning method to enable agents learning for taking actions in an environment in order to maximize the notion of cumulative reward. RL, especially the rewards represented by deep neural network (deep RL), has been largely used in AVs [6], [7]. In this study, we train AV agents with deep RL to investigate the interactions between humans and autonomous vehicles.

In traffic scenarios where humans and AVs coexist, merging is a frequent but challenging task that is critical for safe and efficient driving. Completing a merge requires understanding the surrounding environment, modeling the driving strategy of the opponent, and learning to make actions corresponding to the driving behavior of opponents. Therefore, merging is a commonly used task for training and testing AVs. [8] proposed a probabilistic and interactive prediction approach on ramp merging scenario. [9] analyses the significance of different
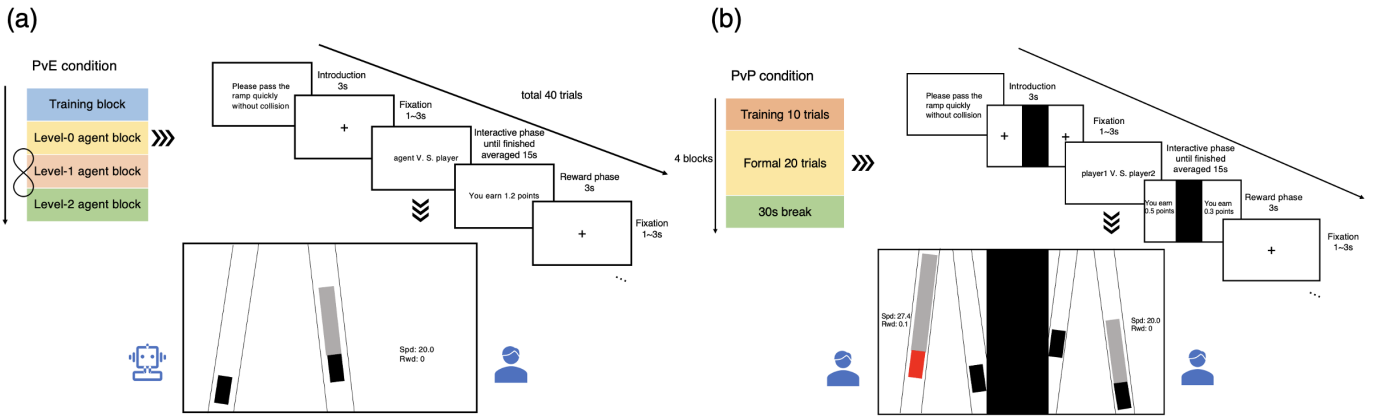
Fig. 1: The experimental design of human-AV interactive task and human-human interactive task. (a) PvE (*i.e.*, Player versus Environment) condition. The human subject is asked to interact with the AV agent to arrive the destination as quickly as possible without collision. The human subjects can receive a training block before the formal experiment. We use the block design, including three blocks corresponding to 3 levels of AV agents. A block begins with an introduction and then 40 trials with a 1∼3-second jitter in inter-trial interval (ITI). Human subjects press the number buttons to imitate driving scenario and interact with an AV in each block. The reward is informed in the end of each trial. (b) PvP (*i.e.*, Player versus Player) condition. Two human subjects are asked to interact with each other with the goal to arrive the destination as quickly as possible without collision. In PvP condition, The monitor is divided into two windows separated with a black screen. We instruct the human subjects to only focus on the window on their own side. The bottom panel of (a) and (b) are examples of real driving moments. The color of car is black when maintaining a constant speed, red when accelerating, and blue when slowing down. The gray shadowed area is the position to arrive in the next 1 second, indicating the speed of the car. Note that both human and AV can only see the color and the gray area of the car controlled by themselves and cannot see the the accelerating and speed information of the opponent.

variables for driving style classification when merging. [10] defined merging behavior as a multi-agent dynamic game, and proposed an iterative algorithm to learn the interactive costs.

It is unclear whether AV can deal with humans with different personality, and in return, whether humans can learn to safely and efficiently interact with AV agents with different driving styles in a short period of time.

To operate safely, drivers need to observe the other vehicles and infer the mental state or driving style of the opponents in the traffic flow, which inevitably needs mentalizing. Mentalizing is a fundamental process in which we infer the inner thoughts and intentions of others. Wu et al proposed interactive mentalizing theory (IMT) [11] and its related interactive mentalizing questionnaire (IMQ) [12] to measure different mentalizing abilities in social interaction. Specifically, it measures self-self mentalization (SS), self-other(SO) and other-self mentalization (OS). The SO score quantifies the ability to assess the mental state of others from the perspective of the self. The SS score quantifies the ability to assess self-generated mental states from the perspective of the self. The OS score quantifies the ability to evaluate mentalization of self-generated mental states from the perspective of others. Meanwhile, a k-ToM model (e.g., level-0, level-1, level-2) is proposed, which predicts that the performance of agents engaged in competitive repeated interactions [13]. In the social interaction like merging, mentalizing other drivers requires both prediction of other's operations (SO), and also inference

of how the others think about oneself's intention (OS). Unpredictable reactions during driving (*e.g.*, speeding, slowing down and right-/left-turn in the merging) may cause catastrophic accidents. Although large amounts of efforts have been made on human's mentalizing process in the in-lab experiments in the psychology community [14]–[16], and the decision-making algorithms in the merging task in AV community [17]–[19], few studies have combined the two to investigate mentalizing on AVs and human subjects, and the trade-off in time-saving and safety considerations.

Recent work has integrated the social value orientation (SVO) of drivers into AV algorithms to quantify an agent's degree of selfishness or altruism, resulting in reduced errors in predictions for algorithms incorporated with SVO [20]. It brings a potential direction to AV algorithms, that is, taking personality traits and mentalized predictions into account in the model. In the current work, we designed a task under the framework of mentalizing level, with the incorporation of mentalizing in the human-human and human-AV interactions (see Fig. 1). That is, with consideration of mentalizing in AV and humans, we trained AV agents with different mentalizing levels and measure the first (SO) and higher-order mentalizing (OS) abilities in human drivers. Here, we aimed to test the following two hypotheses: (i) humans have more diverse driving policy compared with the RL-based AV, and can flexibly adjust driving policy to maximize rewards (**Hypo 1**); (ii) in human-AV game, human subjects with higher IMQ

scores can earn more rewards than the lower ones since subjects with higher IMQ scores can better predict the policy of the opponent AV resulting in more rewards (**Hypo 2**). Testing these two hypotheses would help us understand the relationship between human driving policy and mentalizing in human-AV interactions.

II. METHOD

A. Simulator Design

The environment is designed as shown in Figure 1. Two cars start in two lanes and end up merging in one lane. The opponent car in the left is controlled by the trained RL agents while the ego car in the right is controlled by subjects. Different colors on the cars indicate acceleration (red), constant speed (black) and deceleration (blue) respectively. Shadow indicates the future trajectory of the ego car. The **State Space** for each agent is defined as $S = \{dx, dy, dv, x, v\}$, where dx, dy are the relative longitude and latitude distances of opponent, dv means the velocity difference, and x, v are the global position and velocity of ego car.

There are five discrete actions in the **Action Space** A to keep the speed of the vehicle at 0, 10, 20, 30 and 40. To avoid discontinuous state change, instead of setting constant acceleration values directly, we use Model Predictive Controller to Control the speed change of the cars in the next 3 seconds. The one-step control input is optimized with a minimum jerk. In our study, four reward terms are considered: $R_1 = 2/R_2 = 1$, the reward for the first/second car that reaches the merge point; $R_{col} = -10$, the punishment for both cars if a collision happens; and $R_{spd} = \alpha|v - v_{standard}|$, the punishment for speed deviation that We set a standard speed 20 and the car receives punishments proportional to the deviation of the current speed from the standard speed. The weights are carefully designed so that the strategy to keep full speed for R_1 would gain lower reward than those who keep standard speed and get R_2. As a result, the best strategy is to stay at standard speed for most time, while overtake a little bit if the opponent takes no aggressive actions.

B. Three levels of Agents trained with Deep Q-Learning

In our study, we design the deep Q-learning network for training the multi-level intelligent autonomous vehicles with mentality. To overcome the requirement of the huge memory storage in the traditional Q-function table of model-free reinforcement learning, the introduced deep Q-learning network includes the neural network module (three fully connected layers followed with the Rectified Linear Unit activation function) for learning a function that characterizes the relationship between the state of the vehicle and its Q-value with different actions. The main objective of the deep Q-Learning Network tries to minimize the mean square error of the predicted Q-value and the target Q-value Q^* (the optimal action-value function) [21], [22]. In other words, the autonomous agent tries to maximize the cumulative future reward with the action selection. In our study, we iteratively train three mentalizing levels (level-0, level-1, level-2) of autonomous vehicles by
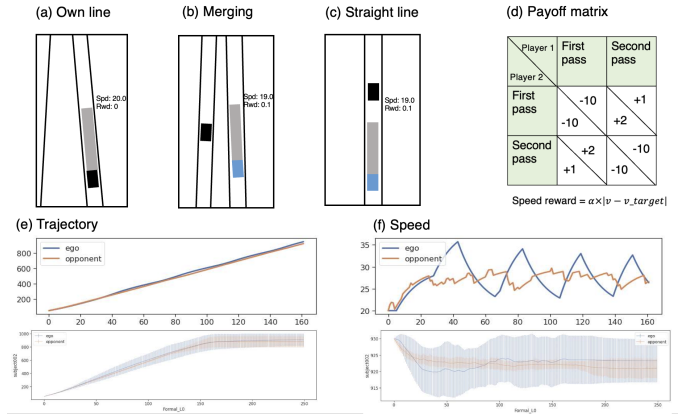

Fig. 2: Examples of 3 types of interactions in merging task and record results. Each car starts driving in its own line (a), then merges in the merging points (b) and eventually drive in the straight line (c). (d) Payoff matrix. The recorded behavioral data include the trajectory (e) and the driving speed (f) of individual human subject and AV.

setting the opponent vehicle with the lower mentalizing level. The trained autonomous agents have iterative driving policies.

C. Experiment

We recruited 40 human participants to take part in the experiment, 30 (8 females) for the PvE condition and 10 (2 females) for the PvP condition(aged 19-33 years (22.75±2.99). There are no significant difference between age and genders of the participants in two conditions (Age: F=6.280, p=0.017; Gender: F=0.1696, p=0.683). The experiment was divided into two parts: in the first part, the participants needed to complete the Interactive Mentalizing Questionnaire (IMQ) [12] which evaluates the mentalizing ability of the participants; the second part was the interactive experiment. In our study, we randomly assigned participants to two conditions (PvE or PvP). All participants signed the informed consent before the formal experiment, and the experimental protocol was approved by the Institutional Review Board of the Southern University of Science and Technology.

In PvE condition, the human participants are informed to play with self-play AV agents. They can have a training block with 20 trials before the formal interactive experiment to ensure understanding the whole experiment. The formal experiment consists of three blocks, corresponding to the AV agents with 3 different mentalizing levels (*i.e.*, level-0, level-1, level-2). The blocks are randomized across subjects, and each block consists of 40 trials. The participant is instructed to maximize the reward by controlling his/her own car. Participants can press buttons 0, 1, 2, 3, and 4 to reach the desired speed. The larger the number of keys, the faster of the speed. The speed of car is indicated by the gray area in front of the car in the monitor (Fig. 1(a)). The ego car controlled by the human driver is in the right lane, while the op car controlled by AV is in the left lane. Each trial started with a 1-3s jittered ITI, and then the participant can control the car by pressing buttons. The

speed of the ego car and the current reward are displayed on the right-side, as described in II-A. The trial is ended when the ego car reached the end, and the reward of the current trial is shown for 3 seconds (Fig. 1 (a)). In the end of each block, the total rewards of the participant is presented. The participant's final payment is proportional to the total rewards in three blocks.

In PvP condition, we designed 4 blocks with 30 trials in each block (the first 10 trials for training, and the rest 20 trials for the formal experiment) and a 30s break after each block (Fig. 1 (b)). Two human participants are involved in the task. In the training trials, participant 1 is asked to free driving for 5 trials and participant 2 has to drive his/her car at a constant speed. The free driving car alternates after 5 trials. The formal experiment in PvP condition is similar to that in PvE condition. The only difference is the monitor is split into two windows with a black screen in the centre. Subjects are instructed to only look at his/her own side of the window.

III. RESULTS

In the PvE condition, the three agents which control the opponent car had very different driving policy that level-0 agent and level-2 agent drove aggressively and level-1 agent drove conservatively. The participants needed to learn the driving policy of the agent and change their own driving policy correspondingly to maximize rewards. The best driving policy for level-0 agent and level-2 agent is to press 2 the whole time and the best driving policy for level-1 agent is to press 3 or 4 to stay a safe distance in front of the opponent car and then press 2 till the end. In the PvP condition, two participants drove against each other in their own policy to maximize their rewards. Different from the PvE condition, subjects' driving policy could change much across trials and one participant could consider what would the other participant do to maximize their rewards. The PvP condition was similar to the Chicken Game and the best policy for the two participants was taking turns showing weakness maximizing group rewards.

A. Changes in the ego car speed and the op car speed

We computed the standard deviation of the car speed to capture the change in speed within a single trial. The results are illustrated in Figure 3. We found that the speed of ego car varied in a larger range than that of the opponent car, suggesting human flexibly utilizes a diversity of driving policies. The comparisons between PvP and PvE conditions show great similarity (Pearson Correlation: $R = 0.7563, p = 2.45 \times 10^{-75}$) between the behavior in person-to-person interactions and the game behavior when the policy of AV agent maintains while the policy of human flexibly varies. The cross shows the change of the driving policy that location of the cross converges to a region within 20 trials in the PvE condition and do not converge after 40 trials in the PvP condition.

B. Relationship between rewards and mentalizing ability

We investigate the relationship between the policy in the game (e.g., total rewards) and the mentalizing ability of
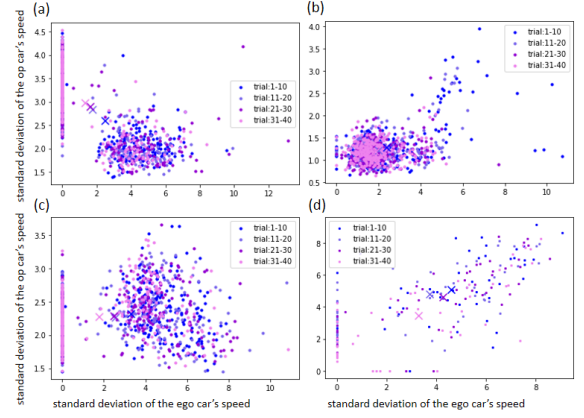

Fig. 3: The behavioral results of the game. (a-c) In PvE condition, games between the human subject (i.e., ego car) and the 3 levels of AV opponent (op car), i.e., level-0 (a), level-1 (b) and level-2 (c), respectively. (d) In PvP condition, the game between two human subjects. The x-axis and y-axis are the variance of the speed of the ego car and op car in a single trial, with each dot representing a trial. The four gradient colors represent the 1-10, 11-20, 21-30, 31- 40 trials respectively, and the symbol × represents the average of speed variance across 10 trials.

human subjects (e.g., IMQ scores) in PvE condition. We first applied the regression analysis with all IMQ scores and further breakdown with correlation analysis on the total rewards and the IMQ scores. For multiple linear regression, the result shows different predictive effect to average final results from IMQ scores. The fitted equations are as below :

$$R(\text{level-0}) = -0.0003\text{SS} - 0.0533\text{SO} + 0.0269\text{OS} + 0.7844 \tag{1}$$

$$R(\text{level-1}) = 0.0011\text{SS} - 0.0213\text{SO} + 0.0069\text{OS} + 1.7818 \tag{2}$$

$$R(\text{level-2}) = -0.0373\text{SS} - 0.0852\text{SO} + 0.0344\text{OS} + 2.1183 \tag{3}$$

The R^2 are 0.079, 0.068, 0.451, respectively. For the OS score, it is with regression coefficient 0.0269/0.0069/0.0344 for three different levels of AV, and a significant negative correlation between the averaged final rewards and SO score (with regression coefficient $-0.0533/-0.0213/-0.0852$).

C. Relationship between policy switching and mentalizing ability

We perform a linear regression on the state variable of each trial in the Level-2 agent block and the subject's action:

$$ACTION = \sum a_i * STATE_i + b \tag{4}$$

State: δx, the longitudinal distance between the two cars; δy, the lateral distance between the two cars; δv, the speed difference between the two cars; s, the distance traveled by the ego car; v_1, the speed of the ego car; v_2, the speed of the opponent car.
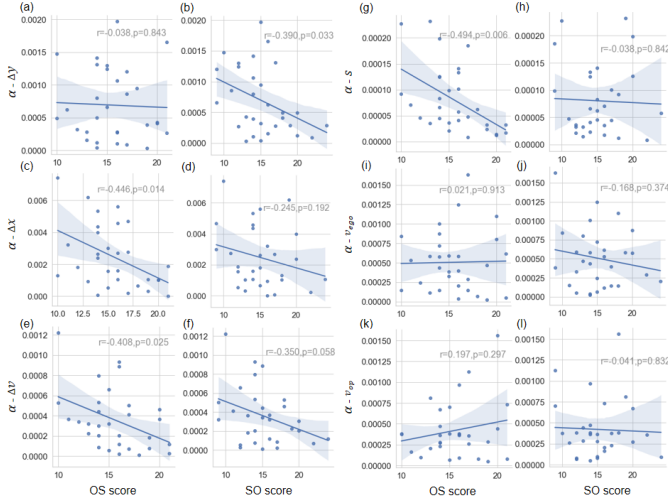
Fig. 4: Double regression parameters α and IMQ scores. $ACTION = \sum a_i * STATE_i + b$, $a_{in} = \alpha_i * Trial_n + b'$

The fitted parameter a can describe the importance of different state variables to the subject's decision-making in each trial. In the experiment, the subjects will learn the driving policy of the opponent car in order to maximize the reward, and then change their own driving strategy according to the driving strategy of the opponent car. We perform linear regression on the parameter a and the number of trials to further characterize the strategy switching between subjects:

$$a_{in} = \alpha_i * Trial_n + b' \qquad (5)$$

The absolute value of the fitted parameter $|\alpha|$ can depict how fast the participant's policy changes between trials. In Figure 4, we can see that the SO and OS scores are negatively correlated with the $|\alpha|$ corresponding to almost every state. Under the assumption that subjects adopt near-optimal driving strategies at the end of the Level-2 agent block, the subjects with higher OS and SO scores can better predict the driving policy of the opponent car, and the driving policies of the subjects in the early trials are closer to the optimal driving policy.

IV. Discussion and conclusion

Although early AV models of decision-making largely ignored mentalizing, the last few decades have detailed the central role mentalizing play in guiding social interaction choices. Take the humans' and AV's mentalizing level into account, the current work investigate the driving patterns in both the human-av interaction with different mentalizing level AVs, and human-human interaction with mentalizing ability measures. Our findings first confirm the diverse strategy of humans in human-human interactions, compared to human-AV interactions (**Hypo 1**). Meanwhile, we observe the interesting prediction effect of human mentalizing ability in policy of the opponent AV or human (**Hypo 2**).

A. Different driving policies in human-human interactions and human-AV interactions

Through experimental study [23], it indicated a difference between human-human and human-AV interactions on the road. For example, human drivers felt more comfortable following the AV than human driver [24].AV that can predict risk level (Driver's Risk Field (DRF) model [25]) can be more reliable and human-like.Under an aggressive driving scenario that using a professional racing simulator [26],they found different features are more important between human drivers and AV. Echos to previous study indicate the human-human interaction is more complex than human-AV interaction [27], our findings show similar pattern in the policy or strategy of merge. Figure 3 showed the speed of human-driving car varied within a larger range in the PvE condition, indicating that human subjects' driving strategies were more diverse and flexible. By comparing PvE and PvP conditions, we uncover the differences in human-human interactions and human-AV interactions. We found that the variance of the speed of the two subjects in PvP condition has a linear correlation, indicating that human subjects will adjust their strategies according to each other's strategies, while in the PvE condition the strategy of AV agent was relatively monotonous and did not rely on the change of the subject's strategy. More specifically, in the PvE condition, when the human subject adopt the no-press strategy, the velocity variance of the AV agent continued varying, indicating that the AV agent has greater randomness. We found that the averaged variance converged within the first 20 trials in PvE condition, indicating that human subjects can quickly learn the AV agent's strategy and adjust their own strategies to maximize rewards. In the PvP condition, the averaged variance did not converge even after 40 trials, implying the high complexity of human-human interactions. The constant adjustment of the driving strategy during human-human game leads to the instability of the environment, which requires more learning to achieve the optimal strategy.

B. From mentalizing to driving policy: impacts of mentalizing abilities

As we predicted, beliefs about our own mentalizing and other's intention is critical in social interaction, including human-human and human-AV scenarios [28]. We investigate the relationship between the human driving policy and link it to one's mentalizing ability. Our study provide evidence to confirm that the mentalizing ability affect the driving policy both when one interacts with AV or interacts with another human subject. Our results show that subjects with higher mentalizing scores drive more conservatively and thereby resulting in more total rewards (Fig. 5). According to these findings, an initial summary is that all three sub-components play role in predicting human interaction with other agents from level-0 to level-2. However, it shows that the higher-order ability (OS score) [12] may be with stronger influence in this prediction effect to both the optimal strategy and the gained reward.
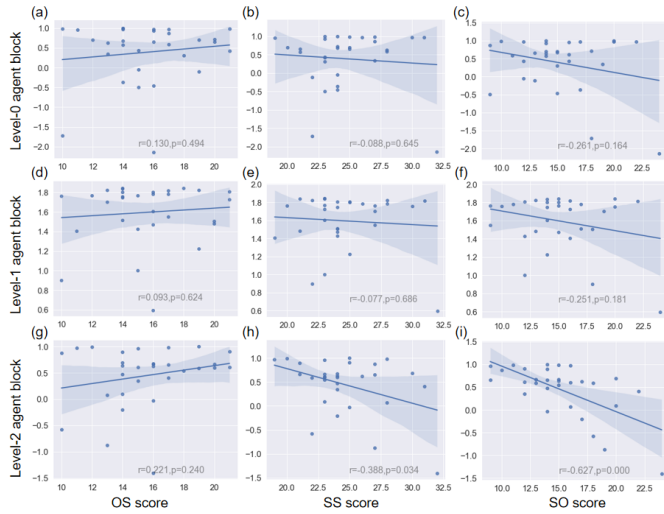
Fig. 5: Reward as a function of mentalizing score. From left to right, the relationship between the accumulated rewards in the block and the SS score (left), SO score (middle) and OS score (right). From top to bottom, it present the results with level-0 agent (top), level-1 agent (middle), and level-2 agent (bottom).

C. Future work

The three agents in PvE condition were not very 'smart'. We may expect an uncertainty module to regular the agents' actions and train more high-level agents with different initial conditions and construct a recurrent neural network to capture temporal information of each trial. In addition to behavioral experiments, we may do some ERP experiments in the same driving environment. Future work with brain signal recording can further provide evidence of which brain regions will be engaged in different interactions during driving scenarios.

REFERENCES

[1] Nicholas Rhinehart, Rowan McAllister, and Sergey Levine. Deep imitative models for flexible inference, planning, and control. *arXiv preprint arXiv:1810.06544*, 2018.

[2] Hang Zhao, Jiyang Gao, Tian Lan, Chen Sun, Benjamin Sapp, Balakrishnan Varadarajan, Yue Shen, Yi Shen, Yuning Chai, Cordelia Schmid, et al. Tnt: Target-driven trajectory prediction. *arXiv preprint arXiv:2008.08294*, 2020.

[3] Ming Liang, Bin Yang, Rui Hu, Yun Chen, Renjie Liao, Song Feng, and Raquel Urtasun. Learning lane graph representations for motion forecasting. In *European Conference on Computer Vision*, pages 541–556. Springer, 2020.

[4] Stéphanie Lefèvre, Dizan Vasquez, and Christian Laugier. A survey on motion prediction and risk assessment for intelligent vehicles. *ROBOMECH journal*, 1(1):1–14, 2014.

[5] David González, Joshué Pérez, Vicente Milanés, and Fawzi Nashashibi. A review of motion planning techniques for automated vehicles. *IEEE Transactions on intelligent transportation systems*, 17(4):1135–1145, 2015.

[6] Ahmad EL Sallab, Mohammed Abdou, Etienne Perot, and Senthil Yogamani. Deep reinforcement learning framework for autonomous driving. *Electronic Imaging*, 2017(19):70–76, 2017.

[7] Aidin Ferdowsi, Ursula Challita, Walid Saad, and Narayan B Mandayam. Robust deep reinforcement learning for security and safety in autonomous vehicle systems. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 307–312. IEEE, 2018.

[8] Liting Sun, Wei Zhan, and Masayoshi Tomizuka. Probabilistic prediction of interactive driving behavior via hierarchical inverse reinforcement learning. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 2111–2117. IEEE, 2018.

[9] Huanjie Wang, Wenshuo Wang, Shihua Yuan, Xueyuan Li, and Lijun Sun. On social interactions of merging behaviors at highway on-ramps in congested traffic. *IEEE Transactions on Intelligent Transportation Systems*, 2021.

[10] Negar Mehr, Mingyu Wang, and Mac Schwager. Maximum-entropy multi-agent dynamic games: Forward and inverse solutions. *arXiv preprint arXiv:2110.01027*, 2021.

[11] Haiyan Wu, Xun Liu, Cindy C. Hagan, and Dean Mobbs. Mentalizing during social InterAction: A four component model. *Cortex*, 126:242–252, 2020.

[12] Haiyan Wu, Bowen J Fung, and Dean Mobbs. Mentalizing During Social Interaction: The Development and Validation of the Interactive Mentalizing Questionnaire. *Frontiers in Psychology*, 12, 2022.

[13] Marie Devaine, Guillaume Hollard, and Jean Daunizeau. The social bayesian brain: does mentalizing make a difference when we learn? *PLoS computational biology*, 10(12):e1003992, 2014.

[14] Alan N Hampton, Peter Bossaerts, and John P O'Doherty. Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences*, 105(18):6741–6746, 2008.

[15] Frank Van Overwalle. A dissociation between social mentalizing and general reasoning. *Neuroimage*, 54(2):1589–1599, 2011.

[16] Sahba Besharati, Stephanie J Forkel, Michael Kopelman, Mark Solms, Paul M Jenkinson, and Aikaterini Fotopoulou. Mentalizing the body: spatial and social cognition in anosognosia for hemiplegia. *Brain*, 139(3):971–985, 2016.

[17] Jian Sun, Jixiang Ouyang, and Jianhao Yang. Modeling and analysis of merging behavior at expressway on-ramp bottlenecks. *Transportation Research Record*, 2421(1):74–81, 2014.

[18] Jinxian Weng and Qiang Meng. Rear-end crash potential estimation in the work zone merging areas. *Journal of Advanced Transportation*, 48(3):238–249, 2014.

[19] Jinxian Weng, Shan Xue, and Xuedong Yan. Modeling vehicle merging behavior in work zone merging areas during the merging implementation period. *IEEE Transactions on Intelligent Transportation Systems*, 17(4):917–925, 2015.

[20] Wilko Schwarting, Alyssa Pierson, Javier Alonso-Mora, Sertac Karaman, and Daniela Rus. Social behavior for autonomous vehicles. *Proceedings of the National Academy of Sciences*, 116(50):24972–24978, 2019.

[21] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

[22] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

[23] Yalda Rahmati and Alireza Talebpour. Towards a collaborative connected, automated driving environment: A game theory based decision framework for unprotected left turn maneuvers. In *2017 IEEE Intelligent Vehicles Symposium (IV)*, pages 1316–1321. IEEE, 2017.

[24] Yalda Rahmati, Mohammadreza Khajeh Hosseini, Alireza Talebpour, Benjamin Swain, and Christopher Nelson. Influence of autonomous vehicles on car-following behavior of human drivers. *Transportation research record*, 2673(12):367–379, 2019.

[25] Sarvesh Kolekar, Joost de Winter, and David Abbink. Human-like driving behaviour emerges from a risk-based driver model. *Nature communications*, 11(1):1–13, 2020.

[26] Adrian Remonda, Eduardo Veas, and Granit Luzhnica. Comparing driving behavior of humans and autonomous driving in a professional racing simulator. *PLoS one*, 16(2):e0245320, 2021.

[27] Yi-Ching Lee, Ali Momen, and Jennifer LaFreniere. Attributions of social interactions: Driving among self-driving vs. conventional vehicles. *Technology in Society*, 66:101631, 2021.

[28] Giorgio M Grasso, Chiara Lucifora, Pietro Perconti, and Alessio Plebe. Evaluating mentalization during driving. In *VEHITS*, pages 536–541, 2019.