Accepted to 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021)
Workshop on Autonomous Driving: Perception, Prediction and Planning

Altruistic Maneuver Planning for Cooperative Autonomous Vehicles Using Multi-agent Advantage Actor-Critic

Behrad Toghi[1;2], Rodolfo Valiente[1], Dorsa Sadigh[2], Ramtin Pedarsani[3], Yaser P. Fallah[1]
[1] Connected & Autonomous Vehicle Research Lab (CAVREL), University of Central Florida, Orlando, FL
[2] Intelligent and Interactive Autonomous Systems Group (ILIAD), Stanford University, Stanford, CA
[3] Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA
toghi@knights.ucf.edu

Abstract

With the adoption of autonomous vehicles on our roads, we will witness a mixed-autonomy environment where autonomous and human-driven vehicles must learn to co-exist by sharing the same road infrastructure. To attain socially-desirable behaviors, autonomous vehicles must be instructed to consider the utility of other vehicles around them in their decision-making process. Particularly, we study the maneuver planning problem for autonomous vehicles and investigate how a decentralized reward structure can induce altruism in their behavior and incentivize them to account for the interest of other autonomous and human-driven vehicles. This is a challenging problem due to the ambiguity of a human driver's willingness to cooperate with an autonomous vehicle. Thus, in contrast with the existing works which rely on behavior models of human drivers, we take an end-to-end approach and let the autonomous agents to implicitly learn the decision-making process of human drivers only from experience. We introduce a multi-agent variant of the synchronous Advantage Actor-Critic (A2C) algorithm and train agents that coordinate with each other and can affect the behavior of human drivers to improve traffic flow and safety.

1. Introduction

Autonomous vehicles (AVs) can leverage their superior computation power, precision, and reaction time to avoid errors occurred by human drivers and drive more efficiently. Connecting AVs and human-driven vehicles (HVs) via vehicle-to-vehicle (V2V) communication creates an opportunity for extended situational awareness and enhanced decision making [28, 30]. We are particularly interested in the problem of cooperative decision-making in mixed-autonomy environments where AVs need to share the road infrastructure with human drivers. In such environments, a given AV interacts with other vehicles, whether autonomous or human-driven, and most likely faces conflictive and competitive scenarios where its individual interest does not necessarily align with that of other vehicles.

In his classic, "I, Robot", Isaac Asimov points out these conflicts and lays down the Three Laws of Robotics which establish a non-formal scheme for the co-existence of robots and humans. These rules prescribe principles that an intelligent agent needs to follow in order to protect its own existence and utility while keeping the involving humans safe. The first law emphasizes the importance of protecting humans, "a robot may not injure a human being or, through inaction, allow a human being to come to harm", and the third law defines the local interest of a robot, "A robot must protect its own existence as long as such protection does not conflict with the previous laws" [3]. Despite being rooted in science fiction, this set of rules later inspired the debates on the ethics of artificial intelligence and robotics [12] and is meaningfully relevant in our particular problem. Conflicts and competitions naturally arise in a mixed-autonomy driving environment and it is crucial to assemble cooperative decision-making criteria that ensure safety and efficiency for both human-driven and autonomous vehicles.

The example in Figure 1 helps us to elaborate on our mixed-autonomy driving problem: a vehicle that is attempting to merge into a highway requires the cruising vehicles to open up space for it. Here, the individual interest of the cruising vehicles does not align with that of the merging vehicle and therefore egoistic decision-making will not necessarily lead to an outcome that is optimal for the group of cars. With altruistic decision making, vehicles can optimize for a combination of their individual utility and a social utility. Our key insight is that by working together, a group of vehicles can overcome the physical limitations of a single

Authors B. Toghi and R. Valiente contributed equally.

This material is based upon work supported by the National Science Foundation under Grant No. CNS-1932037.

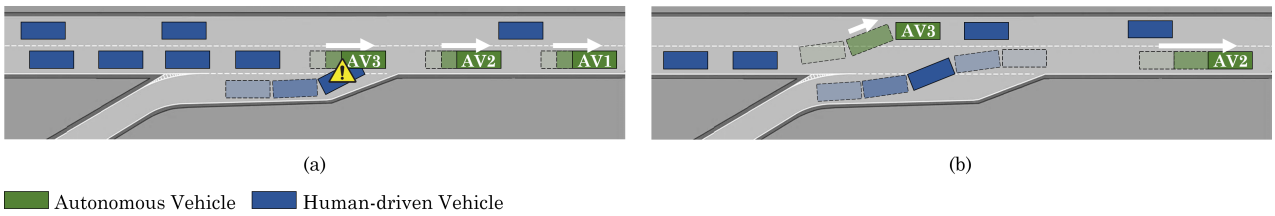

Figure 1: Altruistic AVs can compromise on their individual utility to result socially-desirable behaviors that account for other vehicles. (a) Egoistic AVs solely optimize for their own utility and do not allow the merging vehicle to merge, (b) Altruistic AVs compromise on their individual utility in order to account for the human-driven vehicles and open up space for the merging vehicle.

vehicle and achieve socially-desirable outcomes [20, 22]. However, we believe the willingness of human drivers in demonstrating selflessness and altruism is rather ambiguous and thus cannot be taken for granted. Instead, we rely on the autonomous agents to shape and guide the traffic flow in a fashion that optimizes the social objective.

In order to achieve the above goal, an intelligent agent needs to 1) be able to understand and predict the behav-ior of human drivers and 2) coordinate with its allies, i.e., other AVs, to construct formations that eventually bene-fit the group. Understanding and anticipating the actions of human drivers is a challenging task as it is often dif-ficult to model human behaviors. The models that keep track of beliefs are mostly not scalable and partner model-ing techniques predominantly are intractable in capturing a belief over partners' behaviors that often vary due to factors such as fatigue, distraction, etc. [37]. The core differentiating idea that we emphasize on is that the human driver models which are extracted in the absence of autonomous agents, are not necessarily valid when humans confront autonomous agents. To address these challenges, we take an end-to-end approach and induce altruism into the decision-making process of autonomous agents using a decentralized reinforcement structure. Despite not having access to an explicit model of the human drivers, the trained autonomous agents learn to implicitly model the environment dynamics, including the behavior of human drivers, which enables them to interact with HVs and guide their behavior.

Through our experiments in a highway merging scenario, we demonstrate that our agents are able to learn from scratch not only to drive, but also understand the behavior of HVs and coordinate with them. Our main contributions are as follows:

- Using a multi-agent variant of Advantage Actor-Critic (A2C), we implement a decentralized reinforcement learning algorithm that induces altruism into the decision-making process of autonomous vehicles and incentivizes them to account for the utility of other vehicles in order to result socially-desirable behaviors.

- We choose a complex and conflictive driving scenario and show that our altruistic autonomous agents are able to plan long sequences of actions that reduce the number of crashes and improve the overall traffic flow as compared to egoistic autonomous agents.

2. Literature Review

A fundamental challenge in multi-agent learning and training agents that evolve concurrently is the intrinsic non-stationarity of the environment. Some approaches to address this issue, such as the work by Arel et al., assume that all agents have access to the global state of the world or that they could share their states among the neighbors [2]. However, these assumptions are not always practical in real-world problems. Egorov et al. attempt to overcome this challenge using a centralized Critic function that can mitigate the effect of non-stationarity in the learning process [5]. Foerster et al. propose the counterfactual multi-agent (COMA) algorithm which utilizes a set of joint actions of all agents as well as the full state of the world during the training [6]. A global centralized reward function is then used to calculate the agent-specific advantage function. In contrast, we assume partial observability and a decentralized reward function during both training and execution that is expected to promote cooperative and sympathetic behavior among autonomous vehicles. Lowe et al. present a general-purpose multi-agent learning algorithm that enables agents to conquer simple cooperative-competitive games with access to local observations of the agents [13]. An adaptation of Actor-Critic methods with a centralized action-value function is employed that uses the set of actions of all agents and local observations as its input. In our work, however, agents do not have access to the actions of their allies and/or opponents.

Within the social navigation literature, Omidshafiei et al. and Lauer et al. study the multi-agent navigation and focus on solving cooperative and competitive problems by making assumptions on the nature of interactions between agents (or agents and humans) [9,19]. However, in our case, we are interested in the emerging sympathetic cooperative behavior that enables the agents to cooperate among themselves as well as with their competitors, i.e., humans. Re-

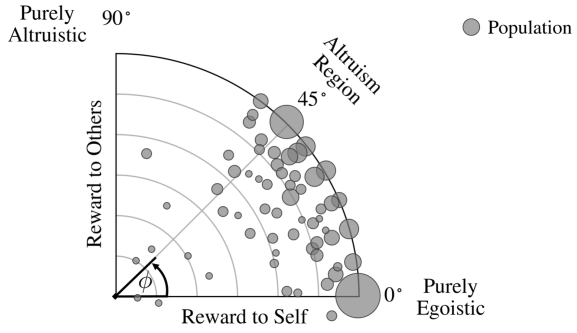

Figure 2: The Social Value Orientation ring demonstrates different behaviors based on a human/robot's preference to account for others. The diameter of the circles show the likelihood of a specific behavior. Figure is based on the data from [7]

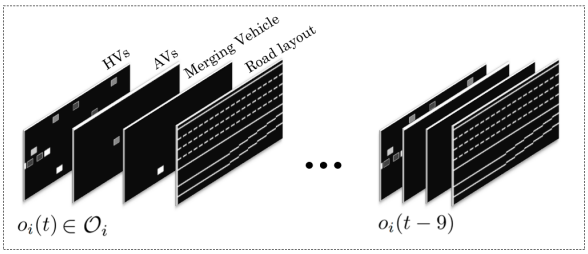

Figure 3: Stacked multi-channel VelocityMap state representation embeds the speed and position of vehicles. Each observation o_i is a tensor of size 10 (4 64 512).

cent works by Pokle et al. have revealed the potential for collaborative planning and interaction with humans [21]. Alahi et al. have proposed the social-LSTM framework to learn general human movement and predict their future trajectories [1]. Toghi et al. present a maneuver-based dataset and model on human driving that can be used to classify driving maneuvers [27]. Nikolaidis et al. optimize a common reward function in order to enable joint trajectory planning for humans and robots [18]. In contrast, we seek altruistic behaviors without having an explicit model of the human driver's behavior or relying on their aid. Sadigh et al. also demonstrate an approach based on imitation learning [22] to learn a reward function for the human drivers and then employ that to affect the behavior of human-driven vehicles. Mahjoub et al. took a statistical machine learning approach to the problem of driver behavior prediction and were able to predict sub-maneuvers within a short time-horizon [14–16]. Works by Wuet al. and Lazaret al. study the human-autonomy problem from a macro-traffic perspective and have demonstrated emerging human behaviors within mixed-autonomy scenarios and the possibility to leverage these patterns to control and stabilize the traffic flow [10, 36].

3. Problem Statement

Humans and intelligent agents interacting in a mixed-autonomy environment can hold different levels of altruism. The preference of a person to account for the inter-est of others in spending its individual resources, e.g. time, money, energy, is formally studied as that person's Social Value Orientation (SVO) in social psychology [7, 24]. This concept can also be employed to quantify an autonomous agent's willingness to act egoistically or altruistically [4]. As demonstrated in Figure 2, the behavior of a human, or similarly an autonomous agent, can lay on anywhere from absolutely egoistic to absolutely altruistic depending on the weight they assign to the utility of others. In order to achieve a set of socially-desirable maneuvers, autonomous agents must act based on the decisions made by their allies, i.e., other AVs, and human drivers.

It is well established in the behavioral decision theory that the SVO of humans is rather heterogeneous. Murphy and Ackermann [17] formalize the human decision-making process as maximizing an expected utility with weights governed by the human's individual preferences. However, these preferences are typically unknown which makes the SVO of a human ambiguous and unclear. The existing works on social navigation for AVs and robots in general, often make assumptions on the humans' willingness to cooperate with autonomous agents, whereas Figure 2 points out the large likelihood of having an egoistic human with no will to account for others. Therefore, we define our problem in a worst-case scenario configuration and assume that all human drivers are egoistic and hence cannot be relied on in terms of cooperating with autonomous agents. Autonomous agents must coordinate with each other and when it is necessary, shape the behavior of human drivers around them to realize a social goal that benefits the group of the vehicle. The desired outcome for our particular driving problem would be enabling seamless and safe highway merging while maximizing the distance traveled by all vehicles and obviously, avoiding collisions.

4. Proposed Solution

Our key insight is bypassing the need for prior information on human drivers' behavior and relying on them to cooperate with AVs. Instead, we adjust the AVs' SVO through a decentralized reinforcement learning algorithm which promotes altruistic behavior in their decision-making process. As opposed to the existing works which rely on derived behavior models of human drivers, we follow our methodology in [31, 32] which allows the autonomous agents to implicitly model human drivers SVO in real-time and through experience.

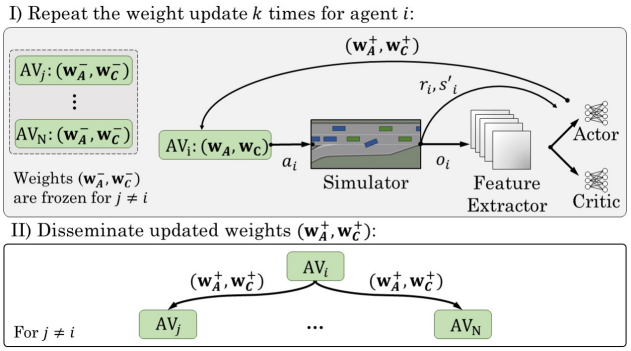

Figure 4: The multi-agent Advantage Actor-Critic framework and policy dissemination process.

4.1. Formalism

Consider a highway merging ramp and a straight highway with a group of human-driven and autonomous vehicles cruising on it, similar to what is depicted in Figure 1. All vehicles are connected together using V2V communication are able to share their observations [23, 25, 29]. The highway merging scenario is particularly interesting to us due to its competitive nature as the interest of the merging vehicle does not align with the interest of the cruis-ing vehicles. We formally annotate a set of autonomous vehicles I, a set of human-driven vehicles V, and the human-driven merging vehicle, $M \in V$ which is attempting to merge into the highway. AVs share their situational awareness through cooperative perception to overcome the limitations of occluded and non-line-of-sight vision [34, 35]. Thus, we assume each AV has a local partial observation of the world constructed using the shared situational awareness with its allies and is able to observe a subset of AVs $\in I$, in addition to a subset of HVs $\in V$.

4.2. Observations and Actions

As mentioned before, our interest is in maneuver planning for AVs and studying their interaction with human drivers. This fact motivates us to view the problem from an abstracted view that determines the meta-actions for an autonomous agent. These meta-actions then further get translated into optimized trajectories and then low-level control signals, e.g., throttle level and steering angle. Specifically, we define an AV's action-space as a set of discrete meta-actions a_i:

$$a_i \in A_i = \begin{bmatrix} \text{Lane Left} \\ \text{Idle} \\ \text{Lane Right} \\ \text{Accelerate} \\ \text{Decelerate} \end{bmatrix} \quad (1)$$

We design a multi-channel VelocityMap embedding, as shown in Figure 3, to represent the state of an autonomous agent. VelocityMaps embed the relative position of AVs and HVs into two channels with pixel values proportional to their relative speed. Additionally, we embed the road layout, position of the observer, and the relative position of the merging vehicle (only if it fits within the observer's perception range) in separate layers. Driving data carries important temporal patterns which can help in modeling human behavior and predictive maneuver planning. Hence, we stack a group of VelocityMaps recorded at consecutive time-steps to capture this temporal information, as Figure 3 depicts.

4.3. Decentralized Reinforcement Structure

Inspired by the work done on social autonomous driving by Schwarting et al. [24] we define the decentralized reward function of an autonomous agent, $I_i \in I$, based on its SVO. For the scenario defined in Section 4.1,

$$\begin{aligned} R_i(s_i; a_i; \varphi_i) &= \cos(\varphi_i) R_i^{E} + \sin(\varphi_i) R_i^{A} \\ &= \cos(\varphi_i) r_i(s_i; a_i) + \\ &\quad \sin(\varphi_i) \Big[\sum_j r_{i;j}(s_i; a_i) + \sum_k r_{i;k}(s_i; a_i) \Big] \end{aligned} \quad (2)$$

where φ_i is the SVO angular preference of the agent I_i and determines its level of altruism. Summations are over $j \in \{I \setminus I_i\}$, $k \in \{V \setminus M\}$. The individual performance of a vehicle, r, can be quantified using metrics such as distance traveled, average speed, and a negative cost for lane-change or acceleration/deceleration to minimize the jerk and unnecessary maneuvers. The altruistic part of the reward, R^A, sums over the individual performance of the other vehicles. Using such a decentralized reward function in our multi-agent reinforcement learning setup, we can train agents that optimize for a social utility governed by their SVO, i.e., we induce altruism in their decision-making process without knowledge on the SVO of the humans.

4.4. Multi-agent Advantage Actor-Critic

As we discussed in Section 3, a major challenge in multi-agent reinforcement learning is dealing with the set of learning agents that are evolving concurrently and therefore making the environment non-stationary. Particularly, this creates an issue for the methods based on the experience replay mechanism, such as Deep Q-networks. In order to avoid the need for the replay buffer, we employ a synchronous Advantage Actor-Critic (A2C) algorithm in which the Critic approximates the state-value function, $V(s)$, and the Actor update the stochastic policy based on the gradient from the Critic. The difference between the action-value function, $Q(s; a)$, and $V(s)$ can be used as a baseline, commonly re-

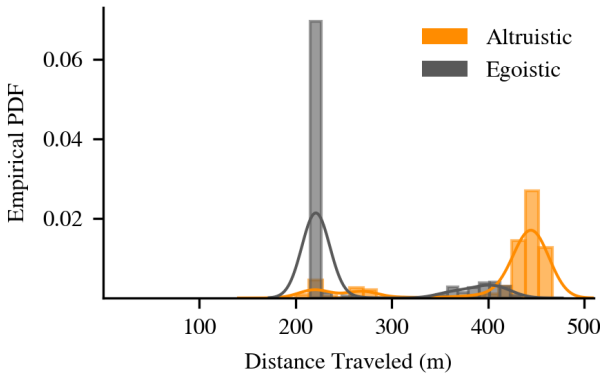

Figure 5: Distribution of distance traveled by the merging vehicle when the AVs act egoistically (Gray) as compared to the case with altruistic AVs (Orange).

ferred to as the Advantage Function [26], i.e.,

$$A(s_t; a_t) = r_{t+1} + V_{w_c}(s_{t+1}) \quad V_{w_c}(s_t) \qquad (3)$$

where w_C denotes the weights of the neural network for the Critic. Thus, the update rule can be simplified as,

$$r J(\,) \qquad \sum_{t=0}^{1} r \log(a_t | s_t) A(s_t; a_t) \qquad (4)$$

where is the policy parameterized by parameters .

As demonstrated in Figure 4, we use a 3D convolutional network (Conv3D) as the feature extractor with the stacked VelocityMaps as its input and two separate multi-layer perceptron networks for the Actor and the Critic. A 3D convolutional architecture enables us to extract the temporal information embedded in the VelocityMaps, details of the feature extractor network are listed in Table 1. The Actor network outputs a probability distribution corresponding to

Table 1:

Architecture of A2C and Feature Extractor Networks.

Layer	Type	Size	Activation
Feature Extractor Network			
0	Input	10 (4 64 512)	-
1	Conv3D	32 @ (1 8 8)	ReLU
2	Conv3D	64 @ (3 4 4)	ReLU
3	Conv3D	64 @ (3 3 3)	ReLU
4	FC	512	ReLU
Actor and Critic Networks			
1	FC	256	ReLU
2	FC	5	Softmax

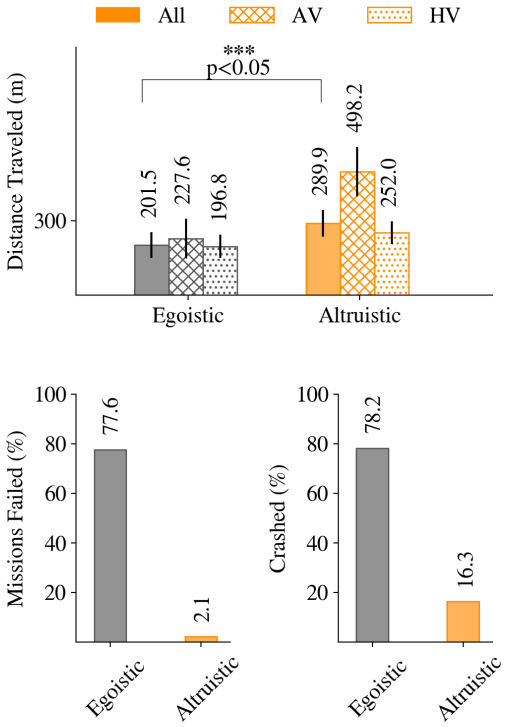

Figure 6: Our experiments demonstrate that our altruistic autonomous agents are able to take sequences of actions that reduce the number of crashes and improve the overall traffic flow as compared to egoistic autonomous agents.

each action. The Critic network maps each state to its corresponding state-value. To stabilize the multi-agent training process, we train each AV in a semi-sequential fashion by freezing the networks of all other AVs (w_A; w_C), and then disseminating the updated weights, ($w_A^{\pm}$; $w_C^{\pm}$), periodically.

5. Experiments

We use an abstract traffic simulator for our experiments. An OpenAI Gym environment provided by Leurent et al. [11] is customized to suit our particular driving scenario and multi-agent problem. A Kinematic Bicycle Model governs the kinematics of the vehicles and a proportional–integral–derivative (PID) controller renders the meta-actions into low-level steering and throttle control signals. The dynamics of the HVs and their decision-making process is implemented using the lateral and longitudinal driver models proposed by Treiber et al. and Kesting et al. [8, 33].

Manipulated Factors. Equation 2 characterizes the decentralized reward that an AV, I_i 2 I receives by taking an action a_i at the state s_i. The agent's SVO angular preference, $_i$, determines the importance that AV is willing to put

on other vehicles' utility. Thus, we experiment with the values of $_i$ = 0, i.e., purely egoistic behavior, and $_i$ = =4, i.e., altruistic behavior, to examine the benefits of altruistic maneuver planning in comparison with AVs that act egoistically.

Performance Metrics. As discussed in Section 1, we aim to leverage the altruistic behavior of AVs to 1) enhance the overall traffic flow, 2) improve driving safety, and last but not least 3) generate socially-desirable outcomes. Hence, we gauge our results using an intuitive traffic flow metric that shows the average distance traveled by all vehicles in the experiment. The percentage of episodes with a collision in them is also measured to assess the safety aspect of the resulting behaviors. Episodes with a successful and seamless merging are also counted to provide an understanding of the AVs' ability to consider social utility.

Hypothesis. We examine a key hypothesis that we initially stated in Section 1. By varying the SVO angular preference towards the purely altruistic behavior, i.e., = =2, AVs learn to work together and affect the behavior of human drivers to open up a safe corridor for the merging vehicle and allow it to merge safely and seamlessly. We hypoth-esize despite some AVs compromising on their individual utility and acting altruistically, the overall traffic-level met-rics improve as a result of their altruistic behavior. More importantly, we believe that our multi-agent A2C algorithm enables AVs to take socially desirable sequences of actions.

Findings. Using the proposed multi-agent A2C algorithm structure explained in Section 4.4, we train the autonomous agents for 30,000 episodes and evaluate their performance over 3,000 test episodes. The learning performance and the evolution of our metrics during the training are shown in Figure 7. We ran multiple rounds of training to ensure that agents converge to similar points in terms of training stability.

The overall performance of both egoistically and altruistically trained autonomous agents is demonstrated in Figure 6. Confirming our hypothesis, we can observe that altruistic AVs manage to significantly reduce the number of collisions while improving the average distance traveled by all vehicles, which expresses the improved traffic flow.

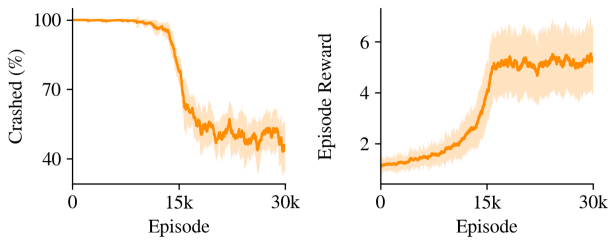

Figure 7: Evolution of our altruistic agents during training using a decentralized reinforcement learning algorithm.

More importantly, it is clear that egoistic AVs did not allow the merging vehicle to safely merge into the highway and caused sub-optimal and unsafe situations, whereas altruistic AVs worked together in order to open up a safe space for the merging HV and allow it to merge.

To grasp a better understanding of how altruism changes the decision-making process of AVs, we dive deeper into the results and particularly measure the average distance traveled by the merging vehicle in both altruistic and egoistic settings. Figure 5 illustrates the empirical distribution of the distance traveled by the merging vehicle. It is evident that when AVs are egoistic, in the majority of episodes the merging vehicle either crashes after traveling 220m or gets stuck in the merging ramp and hence not able to move. In contrast, in the episodes that contain altruistic AVs, the merging vehicle is able to safely merge into the highway and consequently travel approximately as twice as the previous case, 450, and reach farther distances on the highway.

We narrow down our investigation particularly to the behavior of the merging vehicle and the AV that directly meets the merging vehicle at the merging point, i.e., the Guide AV. Although the merging vehicle mostly has a direct interaction only with the vehicles that are cruising on the right-most lane of the highway but the existence of other vehicles and their interactions, indirectly affects the sequence of actions that both the merging vehicle and the Guide AV take. We deliberately perform our study in a multi-lane highway with a fairly large mixed population of AVs and HVs to investigate these indirect relations. It is important to note that although the size of the discrete action-space defined in Equation 1 is relatively small but a vehicle in such a complex multi-lane highway faces a large combination of sequences of actions that it can take.

The eventual outcome of an episode is determined by the sequence of maneuvers taken by vehicles and not their momentarily actions. Thus, an AV must be able to make decisions that lead to a desirable outcome in time-steps ahead. In Figure 8, we illustrate three sample behaviors recorded during the training of autonomous agents. In the first example (the left-most plots), the Guide AV (orange) sees a group of HVs (blue) blocking its left-hand-side lane and thus has no choice but to stay in its lane and slow down to block the traffic behind and open up space for the merging vehicle (purple). This behavior is observable in the speed profile of the Guide AV as well as the provided snapshots. In the second example (the middle plots), the Guide AV's path is blocked by another HV but it finds the left lane open and thus makes a lane-change to open up space for the merging vehicle, resulting in a safe merging. In the third example (the right-most plots), the Guide AV has mastered a way to both enable the merging vehicle to merge, i.e., account for the utility of others, and also perform another lane-change

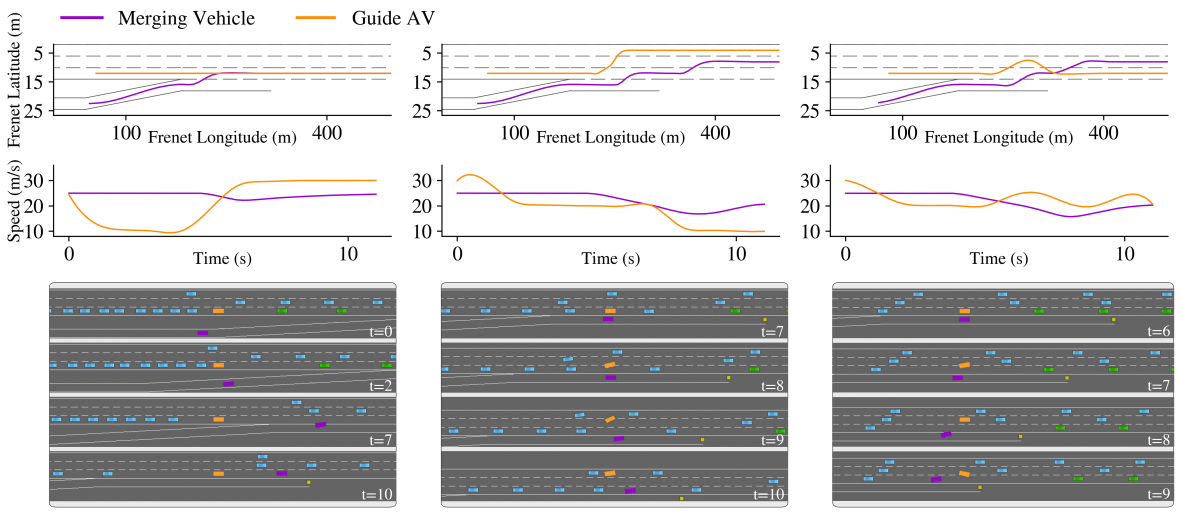

Figure 8: Three example training episodes demonstrate how altruistic AVs learn to take a sequence of decisions that eventually lead to safe, efficient, and socially-desirable outcomes. AVs and HVs are annotated with green and blue color, respectively.

to speed up and optimize for its local utility.

6. Concluding Remarks

Overall, we observe that using our proposed decentralized multi-agent learning scheme, we are able to induce altruism into the decision-making process of autonomous vehicles and adjust their social value orientation. Our altruistic agents not only learn to drive on the highway environment from scratch but also are able to coordinate with each other and affect the behavior of humans around them to realize socially-desirable outcomes that eventually improve traffic safety and efficiency. Our work on social navigation is limited in the sense that we have not used actual human drivers in our simulations nor realistic human driving data and instead have implemented a simple driving model from the literature. However, we believe that our solution can be extended using more realistic human behaviors and our trained agents should be able to adapt accordingly. Additionally, we used a simple remedy to grasp the temporal dependencies in driving data using our stacked VelocityMaps, however, more sophisticated solutions such as recurrent architectures can further improve the capabilities of our methods in maneuver planning.

References

[1] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social lstm: Human trajectory prediction in crowded spaces. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 961–971, 2016. 3

[2] Itamar Arel, Cong Liu, Tom Urbanik, and Airton G Kohls. Reinforcement learning-based multi-agent system for network traffic signal control. IET Intelligent Transport Systems, 4(2):128–135, 2010. 2

[3] Isaac Asimov. I, robot. Spectra, 2004. 1

[4] Noam Buckman, Alyssa Pierson, Wilko Schwarting, Sertac Karaman, and Daniela Rus. Sharing is caring: Socially-compliant autonomous intersection negotiation. In 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 6136–6143. IEEE, 2019. 3

[5] Maxim Egorov. Multi-agent deep reinforcement learning. CS231n: convolutional neural networks for visual recognition, pages 1–8, 2016. 2

[6] Jakob Foerster, Gregory Farquhar, Triantafyllos Afouras, Nantas Nardelli, and Shimon Whiteson. Counterfactual multi-agent policy gradients. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 32, 2018. 2

[7] Alexis Garapin, Laurent Muller, and Bilel Rahali. Does trust mean giving and not risking? experimental evidence from the trust game. Revue d'économie politique, 125(5):701–716, 2015. 3

[8] Arne Kesting, Martin Treiber, and Dirk Helbing. General lane-changing model mobil for car-following models. Transportation Research Record, 1999(1):86–94, 2007. 5

[9] Martin Lauer and Martin Riedmiller. An algorithm for distributed reinforcement learning in cooperative multi-agent systems. In In Proceedings of the Seventeenth International Conference on Machine Learning. Citeseer, 2000. 2

[10] Daniel A Lazar, Erdem Bıyık, Dorsa Sadigh, and Ramtin Pedarsani. Learning how to dynamically route autonomous vehicles on shared roads. arXiv preprint arXiv:1909.03664, 2019. 3

[11] Edouard Leurent, Yann Blanco, Denis Efimov, and Odalric-Ambrym Maillard. Approximate robust control of uncertain dynamical systems. arXiv preprint arXiv:1903.00220, 2019. 5

[12] Patrick Lin, Keith Abney, and George Bekey. Robot ethics: Mapping the issues for a mechanized world. Artificial Intelligence, 175(5-6):942–949, 2011. 1

[13] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. arXiv preprint arXiv:1706.02275, 2017. 2

[14] H. N. Mahjoub, B. Toghi, and Y. P. Fallah. A driver behavior modeling structure based on non-parametric bayesian stochastic hybrid architecture. In 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall), pages 1–5, 2018. 3

[15] H. N. Mahjoub, B. Toghi, and Y. P. Fallah. A stochastic hybrid framework for driver behavior modeling based on hierarchical dirichlet process. In 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall), pages 1–5, 2018. 3

[16] Hossein Nourkhiz Mahjoub, Behrad Toghi, SM Osman Gani, and Yaser P Fallah. V2x system architecture utilizing hybrid gaussian process-based model structures. In 2019 IEEE International Systems Conference (SysCon), pages 1–7. IEEE, 2019. 3

[17] Ryan O Murphy and Kurt A Ackermann. Social preferences, positive expectations, and trust based cooperation. Journal of Mathematical Psychology, 67:45–50, 2015. 3

[18] Stefanos Nikolaidis, Ramya Ramakrishnan, Keren Gu, and Julie Shah. Efficient model learning from joint-action demonstrations for human-robot collaborative tasks. In 2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pages 189–196. IEEE, 2015. 3

[19] Shayegan Omidshafiei, Jason Pazis, Christopher Amato, Jonathan P How, and John Vian. Deep decentralized multi-task multi-agent reinforcement learning under partial observability. In International Conference on Machine Learning, pages 2681–2690. PMLR, 2017. 2

[20] Praveen Palanisamy. Multi-agent connected autonomous driving using deep reinforcement learning. In 2020 International Joint Conference on Neural Networks (IJCNN), pages 1–7. IEEE, 2020. 2

[21] Ashwini Pokle, Roberto Martín-Martín, Patrick Goebel, Vincent Chow, Hans M Ewald, Junwei Yang, Zhenkai Wang, Amir Sadeghian, Dorsa Sadigh, Silvio Savarese, et al. Deep local trajectory replanning and control for robot navigation. In 2019 International Conference on Robotics and Automation (ICRA), pages 5815–5822. IEEE, 2019. 3

[22] Dorsa Sadigh, Shankar Sastry, Sanjit A Seshia, and Anca D Dragan. Planning for autonomous cars that leverage effects on human actions. In Robotics: Science and Systems, volume 2. Ann Arbor, MI, USA, 2016. 2, 3

[23] Md Saifuddin, Mahdi Zaman, Behrad Toghi, Yaser P Fallah, and Jayanthi Rao. Performance analysis of cellular-v2x with adaptive & selective power control. In 2020 IEEE 3rd Connected and Automated Vehicles Symposium (CAVS), pages 1–7. IEEE, 2020. 4

[24] 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. 3, 4

[25] Ghayoor Shah, Rodolfo Valiente, Nitish Gupta, SM Osman Gani, Behrad Toghi, Yaser P Fallah, and Somak Datta Gupta. Real-time hardware-in-the-loop emulation framework for dsrc-based connected vehicle applications. In 2019 IEEE 2nd Connected and Automated Vehicles Symposium (CAVS), pages 1–6. IEEE, 2019. 4

[26] Richard S Sutton and Andrew G Barto. Reinforcement learning: An introduction. MIT press, 2018. 5

[27] Behrad Toghi, Divas Grover, Mahdi Razzaghpour, Rajat Jain, Rodolfo Valiente, Mahdi Zaman, Ghayoor Shah, and Yaser P. Fallah. A maneuver-based urban driving dataset and model for cooperative vehicle applications, 2020. 3

[28] Behrad Toghi, Md Saifuddin, Yaser P Fallah, and MO Mughal. Analysis of distributed congestion control in cellular vehicle-to-everything networks. In 2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall), pages 1–7. IEEE, 2019. 1

[29] Behrad Toghi, Md Saifuddin, Hossein Nourkhiz Mahjoub, MO Mughal, Yaser P Fallah, Jayanthi Rao, and Sushanta Das. Multiple access in cellular v2x: Performance analysis in highly congested vehicular networks. In 2018 IEEE Vehicular Networking Conference (VNC), pages 1–8. IEEE, 2018. 4

[30] Behrad Toghi, Md Saifuddin, MO Mughal, and Yaser P Fallah. Spatio-temporal dynamics of cellular v2x communication in dense vehicular networks. In 2019 IEEE 2nd Connected and Automated Vehicles Symposium (CAVS), pages 1–5. IEEE, 2019. 1

[31] Behrad Toghi, Rodolfo Valiente, Dorsa Sadigh, Ramtin Pedarsani, and Yaser P Fallah. Cooperative autonomous vehicles that sympathize with human drivers. arXiv preprint arXiv:2107.00898, 2021. 3

[32] Behrad Toghi, Rodolfo Valiente, Dorsa Sadigh, Ramtin Pedarsani, and Yaser P Fallah. Social coordination and altruism in autonomous driving. arXiv preprint arXiv:2107.00200, 2021. 3

[33] Martin Treiber, Ansgar Hennecke, and Dirk Helbing. Congested traffic states in empirical observations and microscopic simulations. Physical review E, 62(2):1805, 2000. 5

[34] Rodolfo Valiente, Mahdi Zaman, Yaser P. Fallah, and Sedat Ozer. Connected and Autonomous Vehicles in the Deep Learning Era: A Case Study on Computer-guided Steering, chapter 2.10, pages 365–384. 4

[35] Rodolfo Valiente, Mahdi Zaman, Sedat Ozer, and Yaser P Fallah. Controlling steering angle for cooperative self-driving vehicles utilizing cnn and lstm-based deep networks. In 2019 IEEE Intelligent Vehicles Symposium (IV), pages 2423–2428. IEEE, 2019. 4

[36] Cathy Wu, Aboudy Kreidieh, Eugene Vinitsky, and Alexandre M Bayen. Emergent behaviors in mixed-autonomy traffic. In Conference on Robot Learning, pages 398–407. PMLR, 2017. 3

[37] Annie Xie, Dylan Losey, Ryan Tolsma, Chelsea Finn, and Dorsa Sadigh. Learning latent representations to influence multi-agent interaction. In Proceedings of the 4th Conference on Robot Learning (CoRL), November 2020. 2