

---

# SELF-INTEREST AND SYSTEMIC BENEFITS: EMERGENCE OF COLLECTIVE RATIONALITY IN MIXED AUTONOMY TRAFFIC THROUGH DEEP REINFORCEMENT LEARNING

---

**Di Chen**

Department of Civil and Environmental Engineering  
University of California, Davis  
Davis, CA 95616  
diichen@ucdavis.edu

**Jia Li**

Department of Civil and Environmental Engineering  
Washington State University  
Pullman, WA 99163  
jia.li1@wsu.edu

**Michael Zhang\***

Department of Civil and Environmental Engineering  
University of California, Davis  
Davis, CA 95616  
hmzhang@ucdavis.edu

\* Corresponding author

November 10, 2025

## ABSTRACT

Autonomous vehicles (AVs) are expected to be commercially available in the near future, leading to mixed autonomy traffic consisting of both AVs and human-driven vehicles (HVs). Although numerous studies have shown that AVs can be deployed to benefit the overall traffic system performance by incorporating system-level goals into their decision making, it is not clear whether the benefits still exist when agents act out of self-interest – a trait common to all driving agents, both human and autonomous. This study aims to understand whether self-interested AVs can bring benefits to all driving agents in mixed autonomy traffic systems. The research is centered on the concept of collective rationality (CR). This concept, originating from game theory and behavioral economics, means that driving agents may cooperate collectively even when pursuing individual interests. Our recent research has proven the existence of CR in an analytical game-theoretical model and empirically in mixed human-driven traffic. In this paper, we demonstrate that CR can be attained among driving agents trained using deep reinforcement learning (DRL) with a simple reward design. We examine the extent to which self-interested traffic agents can achieve CR without directly incorporating system-level objectives. Results show that CR consistently emerges in various scenarios, which indicates the robustness of this property. We also postulate a mechanism to explain the emergence of CR in the microscopic and dynamic environment and verify it based on simulation evidence. This research suggests the possibility of leveraging advanced learning methods (such as federated learning) to achieve collective cooperation among self-interested driving agents in mixed-autonomy systems.

**Keywords** mixed autonomy traffic · collective rationality · reinforcement learning · game theory · formation of cooperation · autonomous vehicle behavior design

## 1 Introduction

Autonomous vehicle (AV) is an emerging technology, and it is anticipated that they will co-exist with human-driven vehicles (HVs) in the foreseeable future. AVs are expected to bring various benefits to mixed autonomy traffic systems, e.g. through coordination and cooperation between vehicles and traffic infrastructures (Yu et al. 2019, Yang et al. 2020, Typaldos et al. 2022).

One key assumption often made is that AVs are controlled by one or more shared system-level control objectives, such as system efficiency or energy consumption. However, numerous empirical evidence indicates that individuals tend to be selfish, focusing on maximizing their own utility rather than system-level benefits (Fehr and Gächter 2002, Lange et al. 2014, Komorita 2019). This raises the natural question: when AVs are self-interested and have their own objectives instead of a common one, and there is no enforced external coordination, can cooperation emerge from mixed autonomy traffic intrinsically and can system-level benefits still be achieved? The current literature does not provide a clear answer to this question. The connection between strategic behaviors of AVs and their system level impacts remains elusive (Shang and Stern 2021, Zhong et al. 2020) and there lacks a good understanding of how cooperation emerges in mixed autonomy traffic, at both individual and collective levels.

To fill this gap, our previous work proposed the concept of collective rationality (CR) for mixed autonomy traffic. Originating from game theory and behavioral economics, CR posits that cooperation can emerge at the collective level even when agents pursue individual interests (Weirich 2009). Our earlier works analytically modeled CR as one of the Pareto-efficient equilibria in mixed autonomy traffic (Li et al. 2022) and empirically verified the existence of CR in real-world mixed traffic (Chen et al. 2025). The primary objective of this paper is to explore the attainability of CR in mixed autonomy traffic through deep reinforcement learning (DRL) with a simple reward design in reproducible microscopic simulation environments. In addition, we aim to postulate and verify a mechanism that explains the emergence of CR.

This research is built on two lines of research in literature. Along the first line, in traffic flow theories, a central problem is understanding how microscopic agent interactions shape the collective behaviors of mixed traffic flow at the macroscopic scale, i.e., the micro-macro connections. Early works focused on the longitudinal behaviors of driving agents, without considering lateral interactions or assuming that lateral interactions are exogenous (i.e., regulated by external lane policies). Examples of longitudinal interactions include the dependency of an AV’s car-following mode on the class of the leading vehicles (Qin et al. 2021) and the longitudinal control of AV speeds using mean field games (Huang et al. 2019). When considering the lateral dimension, Chen et al. (2017) and Ghiasi et al. (2017) constructed mixed traffic capacity models that accounted for lane policies, such as AV-dedicated lanes. However, the lateral distribution of traffic agents was exogenously specified, unable to capture the agents’ endogenous choice.

To capture the endogenous lateral interactions of traffic agents, an early work by Daganzo (1997) modeled kinematic waves in a two-lane freeway with a special lane. Daganzo (2002a,b) further explored more complex lateral interactions by modeling macroscopic traffic dynamics in multi-lane freeways, which captures the lane choices between aggressive and timid drivers. The description of discrete lane choices was later generalized to continuous allocation of lateral road spaces. In this context, a division factor, describing the proportion of lateral spaces taken by a class, was introduced to measure lateral interactions between vehicle classes (Logghe and Immers 2008, Qian et al. 2017). The flow-density relations derived from these models implicitly assume a unique equilibrium. However, what has been overlooked is that the complex interactions of mixed traffic agents can lead to multiple equilibria, some of which arise from the rational decisions of self-interested traffic agents. In light of this, Li et al. (2022) formally proposed the concept of collective rationality (also known as collective cooperativeness) to model endogenous interactions between mixed traffic

agents. The empirical analysis by Chen et al. (2025) further verified the existence of CR in real-world mixed traffic of human-driven cars and trucks.

Along another line, research has focused on the pragmatic problem of designing autonomous vehicle behaviors in mixed autonomy environments, with different rationales and purposes. The behavior design problem typically targets either the individual level (micro-scale), aiming to emulate human driving behavior or maximize the vehicle’s own utility, or the system level (macro-scale), aiming to improve overall system performance. At individual level, studies typically train AVs to be “human-like” or “utility-maximizing”, employing a wide range of approaches (Driggs-Campbell and Bajcsy 2016, Driggs-Campbell et al. 2017, Yu et al. 2018, Sama et al. 2020, Xu et al. 2020, Hang et al. 2020). One of the most popular approaches for AV behavior design is DRL, due to its flexibility in incorporating multiple objectives, adaptiveness of controls in dynamic and stochastic environments, and quick online execution. In human-like behavior design, DRL is used to minimize the deviations of AV behaviors from empirical human driving data (Zhu et al. 2018, Zhang et al. 2018, Wei et al. 2018, Shi et al. 2021, Huang et al. 2022).

On the other hand, extensive evidence from simulations and field experiments indicates that AVs, if their behaviors are not properly designed, may cause negative externalities to the system, including reduced capacity (Calvert et al. 2017, James et al. 2019), traffic instability (Milanés and Shladover 2014, Gunter et al. 2019, 2020, Shang and Stern 2021), and negative impacts on HVs (Zhong et al. 2020). Therefore, research explores the idea of using AVs as “regulators” in mixed traffic systems, where AV behaviors are designed to optimize system-level metrics such as travel speed, travel time, fuel consumption, safety, and flow stability (Malikopoulos et al. 2018, Yu et al. 2019, Huang et al. 2019, Goldental and Kanter 2020, Yang et al. 2020, Jiang et al. 2022, Typaldos et al. 2022). However, these system-level improvements are often achieved at the expense of AVs’ own utilities, due to the absence of explicit mechanisms to ensure these individual benefits. Recent literature has attempted to address this issue by adopting locally optimal motion planners and Markov decision processes to minimize fuel and travel time costs for an AV while simultaneously improving the fuel efficiency of surrounding vehicles (Liu et al. 2020, 2022). However, these methods appear heuristic and lack a theoretical explanation of the underlying mechanisms, as well as guarantees for both individual and system-level performance.

The above discussion suggests the potential of integrating the concept of CR in AV behavior design. Towards this goal, this study examines AV behavior design problem and verifies whether CR can emerge with relaxed assumptions in general dynamic settings. Specifically, we demonstrate that CR can be attained among driving agents trained using deep reinforcement learning with a simple reward design. We examine the extent to which self-interested traffic agents can achieve CR without directly incorporating system-level objectives. A mechanism of CR formation in mixed autonomy traffic is also postulated and verified from simulation evidence.

The major finding and contributions are as follows:

- We showed that self-interested traffic agents can attain CR by simple reward design in DRL.
- We verified alignment and difference of mixed traffic behaviors with DRL and game-theoretic models.
- We proposed and analyzed the underlying mechanism to attain CR in the microscopic and dynamic mixed autonomy environment.

The rest of the paper is organized as follows. In Section 2, we introduce prior definitions of CR and provide a practical example of it in real-world settings. In Section 3, we design the self-interested behaviors of AVs using DRL, and present the experiment setup and model training results. In Section 4, we examine the attainability of CR in the mixed autonomy traffic with DRL-trained vehicle behaviors. In Section 5, we explore the mechanisms that lead to CR in the microscopic and dynamic traffic environment. Finally, we conclude the paper with discussions on major findings and outlooks for future research in Section 6.

## 2 Prior definition and physical illustration of collective rationality

In this section, we briefly recapitulate the theoretical definition of CR from our earlier game theoretical model (Li et al. 2022). Following this, we provide an intuitive example of CR in the physical setting.

### 2.1 Collective rationality in equilibrium traffic model

In Li et al. (2022), we proposed a two-player bargaining game model to capture the interplay between two classes of driving agents in mixed autonomy traffic, where CR emerges when these interactions converge to certain Pareto-efficient Nash equilibria (NE). In this game, each class of traffic agents is treated as a player. These players are self-interested and perfectly rational, negotiating the road share (i.e., the proportion of lateral spaces) they occupy to maximize driving speeds. Their interactions are modeled as a collective bargaining process. When settling into NE of the bargaining game, the payoff function is written as,

$$U_i(\rho_1, \rho_2, p_1, p_2) = \begin{cases} u^*(\rho_1, \rho_2) & \text{if } p_1^* + p_2^* > 1, \\ u_i(\rho_i/p_i) & \text{if } p_1^* + p_2^* \leq 1 \end{cases} \quad i = 1, 2 \quad (1)$$

where  $\rho_i$  is the traffic density for class  $i$ ,  $p_i$  is the bid of road share by class  $i$  agent,  $u^*(\cdot, \cdot)$  is the one-pipe equilibrium speed, and  $u_i(\cdot)$  is the nominal speed function for class  $i$ . The term  $p_i^*$  represents the minimum road share taken by class  $i$  at NE, which is computed by  $p_i^* = \frac{\rho_i}{u_i^{-1}(u^*)}$ . The total minimum road share,  $p_1^* + p_2^*$ , determines the type of Pareto-efficient equilibrium reached. If  $p_1^* + p_2^* > 1$ , the 1-pipe equilibrium is Pareto-efficient, where the two classes travel in a fully mixed regime with a synchronized travel speed of  $u^*(\rho_1, \rho_2)$ . If  $p_1^* + p_2^* \leq 1$ , the 2-pipe equilibria are Pareto-efficient, with the two classes traveling separately and being better-off than in a fully mixed regime. CR emerges when the mixed traffic settles into 2-pipe NE and is Pareto-efficient. Below, we present the major conclusions regarding the two types of equilibria in the bargaining game model.

#### 1-pipe equilibrium

At one-pipe equilibrium, all traffic agents move at a synchronized speed, denoted as  $u^*$ . The governing equation for  $u^*$  is,

$$\frac{1}{\rho_{tot}} \left( \frac{\rho_1}{u_1^{-1}(u^*)} \sum_{j=1}^2 \frac{\rho_j}{a_j} + \frac{\rho_2}{u_2^{-1}(u^*)} \sum_{j=1}^2 \frac{\rho_j}{b_j} \right) = 1 \quad (2)$$

where  $\rho_{tot} = \rho_1 + \rho_2$  is the total system density. The terms  $a_j$  and  $b_j$  are the scaling parameters for class 1 and class 2, respectively. These scaling parameters capture the “type-sensitivity” in the mixed traffic, reflecting how a vehicle class’s desired headways or spacings are influenced by the class of the leading vehicle. The 1-pipe equilibrium speed  $u^*$  is computed by solving the implicit function (2), where other parameters and functions can be either assumed or calibrated from data (Chen et al. 2025). This implicit function follows the refinement by Chen et al. (2025), where each class is endowed with a distinct nominal speed function  $u_i(\cdot)$ . The original formulation assumed a single nominal speed function shared by both classes (Li et al. 2022).

#### 2-pipe equilibria

When the 2-pipe equilibria are Pareto-efficient, the condition  $p_1^* + p_2^* \leq 1$  holds. Given that the total road share is 1, this implies that after each class occupies its minimum road share  $p_i^*$ , there is a remaining road share. This remaining road share is referred to as the cooperation surplus (also known as the road share surplus). The formal definition is presented in Definition 1. A positive cooperation surplus ( $s > 0$ ) is a necessary condition for attaining CR.

**Definition 1** (Cooperation surplus). *We call the road share left from players’ collective bargaining as the cooperation surplus,  $s := 1 - p_1^* - p_2^*$ .*

Furthermore, unlike the 1-pipe Pareto-efficient NE, which is unique, the 2-pipe Pareto-efficient Nash equilibria are non-unique because there are multiple ways to split the cooperation surplus between the two classes. To further characterize CR, Li et al. (2022) examined how the cooperation surplus is divided between the two classes. This division is quantified by the surplus split factor, formally defined as,

**Definition 2** (Surplus split factor). *The effective split of the cooperation surplus by the two players is called the surplus split factor, denoted as  $\lambda(\rho_1, \rho_2)$ , and  $\lambda \in [0, 1]$ .*

By this definition, the road share allocated to each class becomes,

$$\begin{cases} p_1 = p_1^* + \lambda(\rho_1, \rho_2)s \\ p_2 = p_2^* + (1 - \lambda(\rho_1, \rho_2))s \end{cases} \quad (3)$$

where the total road share satisfies  $p_1 + p_2 = 1$ . The surplus split factor can be empirically estimated when data is available, providing a quantitative measure of the fairness in benefit allocation between two classes of traffic agents (Chen et al. 2025).

## 2.2 Example of collective rationality

We provide an illustrative example of CR in physical world. A cell representation in Figure 1 is used to visually compare the scenarios without and with CR. Type sensitivity is assumed, with AVs having a smaller desired time headway than HVs. In a fully mixed configuration (configuration 1), self-interested traffic agents utilize all available road space and travel at a synchronized speed. In contrast, when CR is present (configuration 2), these traffic agents demonstrate a degree of spatial organization that results in increased spacing. This additional spacing enables both classes to achieve higher traffic speeds compared to the fully mixed scenario (i.e.,  $u'_i > u_i$ ). In other words, configuration 2 leads to a Pareto improvement over configuration 1, demonstrating the existence of CR.

Furthermore, we offer a preliminary postulation regarding the mechanism behind CR attainment using this example. The grid representations of the two configurations intuitively suggest that improved spatial organization is a key factor in achieving CR. This insight is further supported by the proposed metrics, namely,  $p_e$ ,  $p_s$ , and  $H$ , which quantify intra-class platooning and inter-class separation in both the longitudinal and lateral dimensions. Essentially, the emergence of CR in configuration 2 is reflected in higher values of these metrics. Building on this intuition, in later sections, we will propose a mechanism for CR attainment and validate it using simulation evidence. The metrics presented in this figure capture the macroscopic perspective, specifically the spatial organization. To provide a comprehensive understanding of the underlying mechanisms, additional analyses will be conducted at the microscopic and mesoscopic levels.

## 3 Decision-making of self-interested autonomous vehicles

The bargaining game model in Li et al. (2022) demonstrates the existence of CR in idealized traffic scenarios. This model provides a theoretical benchmark for exploring more realistic settings. However, CR's attainability in more complicated and realistic scenarios cannot be directly concluded from this theoretical model. This is because, as a one-shot, macroscopic analytical model, it does not account for the detailed and evolving interactions of traffic agents in a microscopic environment. The model also overlooks the complexities of real-world driving, such as the presence of physical vehicles and longitudinal car-following dynamics. As a result, it lacks insights into how these interactions evolve over time and whether they can lead to CR in realistic driving scenarios.

This study aims to explore the emergence of CR in more complex and realistic scenarios. Our goal is not to align these models with the idealized game theoretical model or design traffic agent behaviors that exactly match the theoretical model's players. Instead, we hypothesize, from an analogous perspective, that CR may also emerge in more realistic mixed traffic environments with self-interested traffic agents. To verify this hypothesis, we will design self-interested deep reinforcement learning (SI-DRL) agent behaviors in a mixed autonomy traffic simulation environment, where

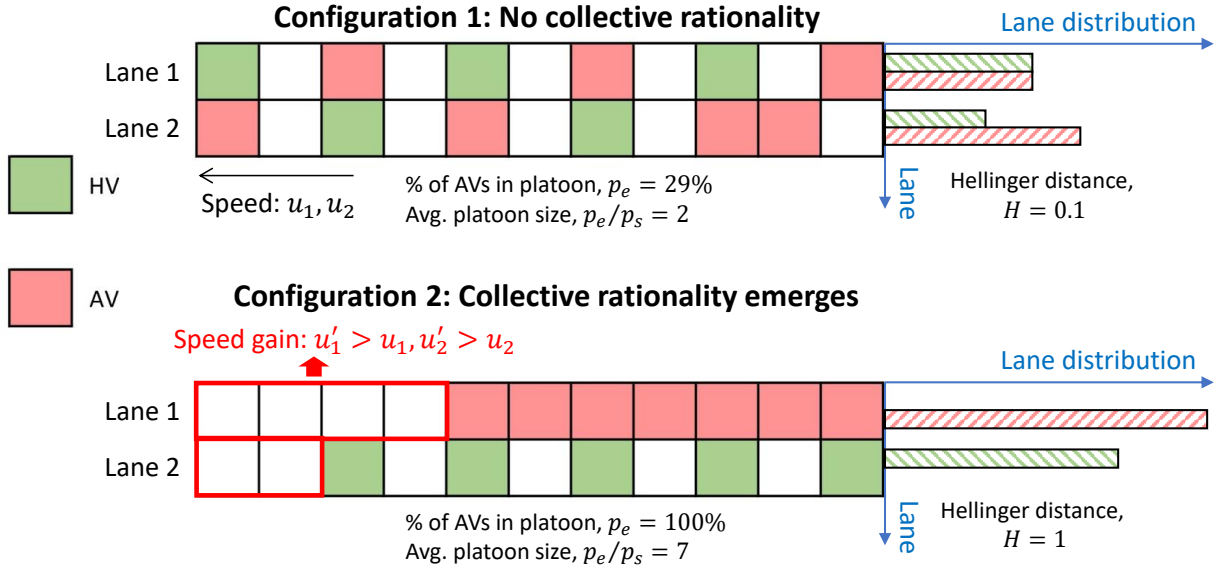


Figure 1: A simple grid illustration of collective rationality in physical world.

AVs are treated as learning agents. These SI-DRL agents' decisions will be tailored using a simple reward design. The DRL framework enables SI-DRL agents to adaptively learn and optimize their driving strategies through continuous interaction with the dynamic traffic environment.

### 3.1 Preliminaries of deep reinforcement learning

The foundation for reinforcement learning (RL) is the Markov decision process (MDP), where the two objects, the RL agent (i.e., the decision making entity) and the environment, continuously interact. The MDP involves four major elements: state, action, policy, and reward. The agent at state  $s$  in the state space  $S$  chooses an action  $a$  from the action space  $A$  based on the policy  $\pi$ , which dictates the probability  $\pi(s, a)$  of taking the action  $a$  at state  $s$ . The action yields an immediate reward  $R(s, a)$ , and the agent transitions to a new state, continuing this cycle until a terminal state is reached. Under a given policy  $\pi$ , the state value function and action value function are respectively expressed as:

$$V^\pi(s) := E_\pi\{R_t | s_t = s\} = E_\pi\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s\right\}, \quad (4)$$

$$Q^\pi(s, a) = E_\pi\{R_t | s_t = s, a_t = a\} = E_\pi\left\{\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s, a_t = a\right\}. \quad (5)$$

The ultimate objective is to discover an optimal policy  $\pi^*$  that maximizes the discounted cumulative reward. To achieve this, we employ the Q-learning algorithm, a model-free method that iteratively updates the action-value function without requiring a model of the environment (Watkins and Dayan 1992). The Q-learning update rule is given by,

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left( r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right) \quad (6)$$

where  $\alpha$  is the learning rate and  $\gamma$  is the discount factor. In practice, large or continuous state spaces often exceed the capabilities of traditional Q-value tables. To address this, we use a deep Q-network (DQN) that approximates  $Q(s, a)$  with an artificial neural network.

### 3.2 Behavior design of self-interested autonomous vehicles

The SI-DRL agents proposed in this study refers to a class of self-interested AVs that make lane change decisions to maximize their individual driving speed, without incorporating with any system-level objectives. These SI-DRL agents have limited perception, being aware only of their local environment without knowledge of the overall system states. Furthermore, we implement a decentralized training scheme, where there is no external coordination among AVs, and each agent independently optimizes its own reward. This study explores whether these self-interested AVs can achieve CR through their continuous interactions with HVs in the mixed autonomy traffic system.

Figure 2 illustrates the decision-making process of SI-DRL agents using DRL in a simulation environment. It incorporates a state representation, a set of actions, and a reward function to guide the learning process.

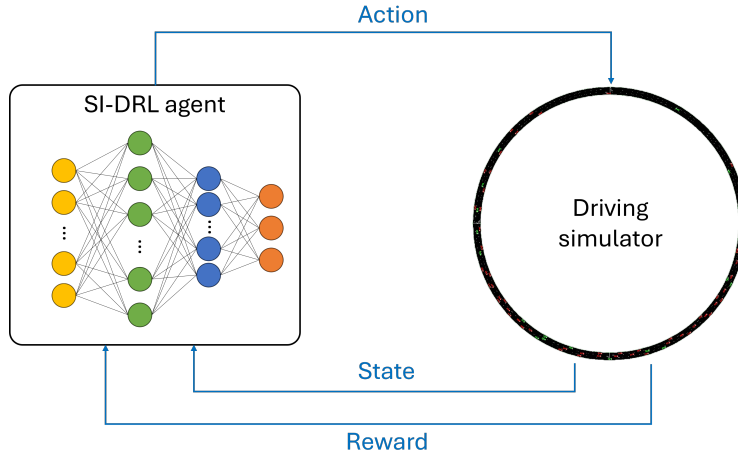


Figure 2: Deep reinforcement learning in the mixed autonomy simulation environment.

#### 3.2.1 State

An SI-DRL AV's lane change decision is influenced by itself and its interactions with surrounding vehicles within its sensing range. For this study, we assume an AV's sensing range to be a circle with a 100-meter radius, as depicted in Figure 3.

The state space is composed of two components. The first component, denoted as  $X_i^{AV}$ , represents the state of the  $i$ th SI-DRL agent itself. The second component, denoted as  $X_i^{surr}$ , captures the state of the surrounding vehicles as perceived by the SI-DRL agent. It is represented as,

$$S_i = \{X_i^{AV}, X_i^{surr}\} \quad (7)$$

where

$$X_i^{AV} = \{x_i, y_i, u_i, l_i\} \quad (8)$$

$$X_i^{surr} = \{x_j^{surr}, y_j^{surr}, u_j^{surr}, l_j^{surr}, c_j^{surr}\}, j = 1, \dots, n \quad (9)$$

Here,  $x, y$  represents the coordinates of a vehicle,  $u$  represents vehicle speed, and  $l$  is an integer representing the lane number a vehicle is traveling at. The variable  $c_j^{surr}$  represents the vehicle type (AV or HV) of the  $j$ th surrounding vehicle, and  $n$  quantifies the total number of surrounding vehicles within an AV's sensing range.

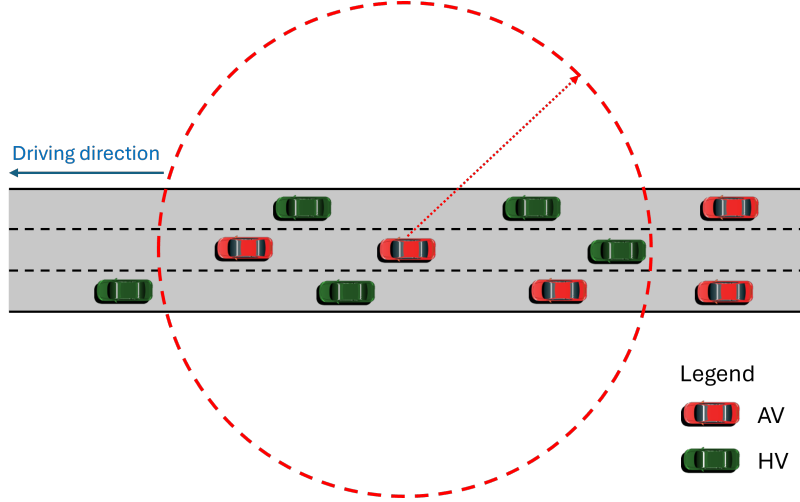


Figure 3: Sensing range of an autonomous vehicle.

### 3.2.2 Action and reward

The action space of the  $i$ th SI-DRL agent, denoted by  $A_i$ , consists of a discrete set of actions representing the agent's lane change decisions, which is written as,

$$A_i = \{\text{change to left, change to right, keep lane}\} \quad (10)$$

The reward function measures the effectiveness of a chosen action in a specific state and is vital in shaping agent behaviors. Since the SI-DRL agent prioritizes increasing individual travel speed, this study formulates a simple reward function to reflect this goal while also considering the external impacts of lane changes on the surrounding traffic. Specifically, the reward function includes two components: the agent's speed, which reflects its self-interest, and a lane change penalty, which discourages frequent lane changes. The reward function is structured as,

$$R_i = w_{\text{speed}} u_i - w_{\text{lane change}} I_i \quad (11)$$

where  $I_i$  is a binary variable indicating whether the agent changes lane or not.

## 3.3 Experiment setup

### (i) Traffic scenarios and simulation setup

We conduct experiments on a three-lane ring road with a total length of 1000 meters. We use an open-source simulator named Simulation of Urban MObility (SUMO) (Krajzewicz et al. 2012), a micro-traffic simulation software capable of modeling multiple agents in an urban-scale environment, to create the ring road configuration and train the DRL models. The SUMO simulator utilizes a traffic control interface (TraCI), which supports retrieving values from simulated objects and manipulating their behaviors.

To facilitate the study of CR, we examine a wide range of traffic densities and AV penetration rates. We consider four traffic densities: 25 vehicles per mile per lane (vpm/lane), 40 vpm/lane, 55 vpm/lane, and 70 vpm/lane (equivalent to 15 veh/km, 25 veh/km, 34 veh/km and 43 veh/km, respectively). Additionally, we evaluate three AV penetration rates: 25%, 50%, and 75%. This setup results in a total of 12 unique traffic scenario settings. Scenarios corresponding to the

fully congested regime are excluded, as their Pareto-efficient NE are more likely to fall within the 1-pipe regime, where CR does not exist (Li et al. 2022).

Each traffic scenario is input to the simulation experiment respectively, with a 100-second warm-up period to ensure that all vehicles are loaded before the DRL training begins. Departure positions and lane numbers are randomly assigned to each vehicle when loading them into the ring road. Additionally, all vehicles are uniformly 5 m in length and have maximum acceleration and deceleration rates of  $2.6 \text{ m/s}^2$  and  $4.5 \text{ m/s}^2$ , respectively, following the default settings by SUMO. Furthermore, to introduce heterogeneity into the simulation environment, HVs are assigned a random maximum speed within the range of  $[17.5, 25] \text{ m/s}$ , and AVs within  $[21, 30] \text{ m/s}$ .

The lane-changing behaviors of AVs are controlled by the DRL model, while HVs are controlled by SUMO’s built-in dynamics model, LC2013 (Erdmann 2015). Besides, the car-following behaviors of AVs and HVs are both governed by the Intelligent Driver Model (IDM) (Treiber and Kesting 2013).

#### (ii) Baseline scenario

We define the baseline as a no-control scenario, where there is no DRL control over AVs’ lane change decisions. In this case, both lane-changing and car-following maneuvers for AVs and HVs are managed by SUMO’s LC2013 and IDM models, respectively. In this context, CR refers to achieving a Pareto-efficient NE under DRL control, relative to the no-control scenario.

#### (iii) Type-sensitivity parameters

The type-sensitivity between vehicle classes refers to the different desired headways or spacings a class of vehicles maintains based on the class of the leading vehicles. As indicated in Figure 1, it is a crucial property for the attainability of CR in mixed traffic, which has been analytically modeled (Li et al. 2022) and empirically validated in real-world scenarios (Chen et al. 2025). In this study, we assume the presence of type-sensitivity in mixed autonomy traffic, particularly in relation to a vehicle’s minimum desired time headway. Due to the lack of empirical data for AVs with current technology, the desired time headways for AVs are primarily based on assumptions (Jiang et al. 2023, Guo et al. 2023). In this study, by denoting HV as class 1 and AV as class 2, we assume the following minimum headway values:  $h_{11}^* = 1.5 \text{ s}$  for a HV following another HV,  $h_{12}^* = 1.5 \text{ s}$  for a HV following an AV,  $h_{22}^* = 1 \text{ s}$  for an AV following another AV, and  $h_{21}^* = 1.25 \text{ s}$  for an AV following a HV. These values suggest that HVs are type-insensitive to the class of their leading vehicles, whereas AVs exhibit type-sensitivity, tending to follow more closely when their leaders are also AVs, due to their communication capabilities.

#### (iv) Deep-Q network (DQN) and hyperparameters

It is assumed that the SI-DRL AVs belong to a same company, allowing them to utilize a shared DQN for decision-making, thereby reducing computational complexity. The input layer of the DQN processes the state representation, which aligns with the dimensions of the state space. We set the state space size to  $N_s = 200$ . The DQN architecture includes three fully connected hidden layers, containing 256, 128, and 64 neurons, respectively. Each hidden layer uses a rectified linear unit (ReLU) activation function to introduce non-linearity into the model. The output layer has a number of neurons equal to the number of actions, which is 3 in our case.

To improve the training stability and sample efficiency, we use experience replay when updating the online network (Lin 1992). The trajectories  $(s_t, a_t, r_{t+1}, s_{t+1})$  generated by  $Q_{online}$  are stored in memory buffer  $M$  and are randomly sampled to update the online network. This random sampling of trajectories collected across training episodes helps mitigate the temporal correlation of the data. Additionally, to balance the trade-off between action exploration and

exploitation, an  $\epsilon$ -greedy decay function is applied, defined as  $p(a) = \epsilon_{end} + (\epsilon_{start} - \epsilon_{end})e^{-epi/r_{decay}}$ . Here,  $\epsilon_{start}$  and  $\epsilon_{end}$  are the initial and final values of  $\epsilon$  respectively,  $epi$  is the current episode number, and  $r_{decay}$  is the decay rate. We set  $\epsilon_{start} = 1$ ,  $\epsilon_{end} = 0.01$  and  $r_{decay} = 300$ . For the optimization, we use the Adam optimizer (Kingma and Ba 2014), with an exponential learning rate decay strategy starting at  $10^{-3}$  and a decay factor of 0.99.

### 3.4 Model convergence

We perform 2000 training episodes for each traffic scenario, with the first 30 episodes as a warm-up phase during which SI-DRL agents perform random actions for exploration without any model updates. Figure 4 presents the reward curves and their variations, with a smoothing average applied every 20 episodes. The differences in reward scales across traffic scenarios arise from varying traffic conditions and the number of AVs in each scenario. Generally, the DRL model for each scenario converges after approximately 1500 episodes.

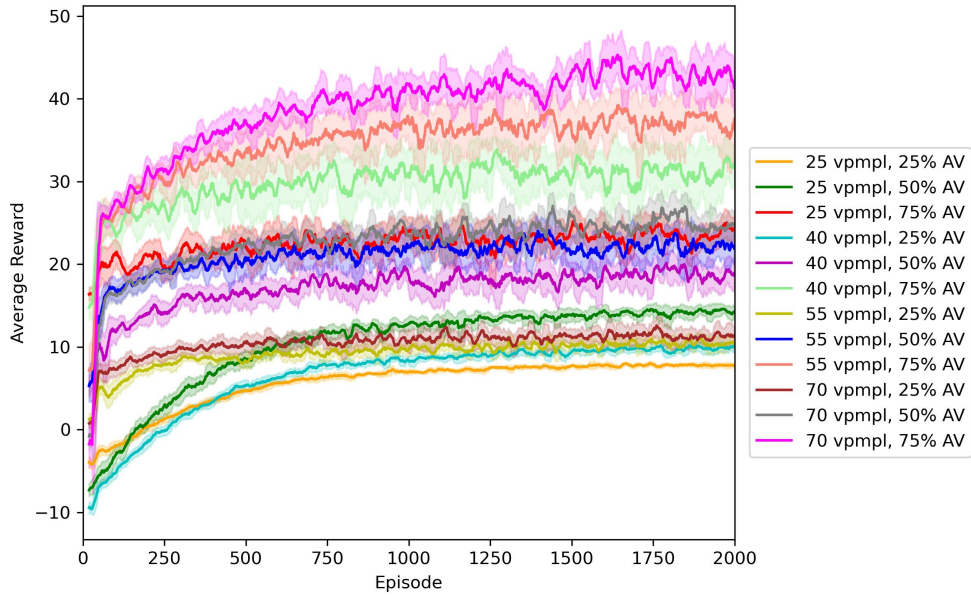


Figure 4: DRL model reward curves for different traffic scenarios.

### 3.5 Model evaluation and traffic measurements

We will evaluate the resulting behaviors in mixed autonomy traffic using these converged models. Specifically, 10 evaluation episodes are run for each traffic density and AV penetration rate setting. Each episode lasts 1200 seconds, with the first 100 seconds designated for vehicle loading time. Besides, each episode uses a unique random seed, different from both each other and those used during training. This approach guarantees that the observed behaviors are not artifacts of seed-specific conditions and ensures the reliability of the proposed DRL models in dynamic and unpredictable environments. The data from the DRL evaluation models will be utilized for the subsequent analysis.

Additionally, to facilitate subsequent analysis, we need to obtain macroscopic traffic measurement data from the simulation. For this purpose, we adopt Edie’s generalized definitions of traffic flow characteristics, including traffic flow, density, and speed (Edie et al. 1963). These definitions are based on a space-time region of interest ( $\Delta x \times \Delta t$ ), where vehicle trajectories are present. The traffic flow characteristics are estimated by averaging vehicle trajectories over these space and time domains. To determine the space-time region of this study, we divide the ring road into 4 subsections, each with a length  $\Delta x = 250$  m. And we choose  $\Delta t = 8$  sec for the space-time region.

## 4 Attainability of collective rationality

In this section, we will examine the attainability of CR in mixed autonomy traffic where SI-DRL AV behaviors have been sufficiently trained. We will assess it from two perspectives: first, the attainability of Pareto-efficient NE (Section 4.1); second, comparisons of various macroscopic traffic characteristics between the DRL simulation output and the game theoretical model predictions (Section 4.2).

### 4.1 Pareto-efficient equilibria

We establish criteria to identify equilibrium conditions in the microscopic and dynamic environment, including variations in vehicle speeds and spacings. We first standardize vehicle speeds and spacings to a mean value of 1 to account for varying scales at different density levels, allowing their standard deviations to reflect variability relative to the mean. Based on this, the criteria to determine the time of reaching NE are defined as follows:

- Speed variation is less than 0.2 m/s.
- Spacing variation is less than 0.35 m.
- The above criteria maintains for at least 20 seconds.

We note that this definition is a relaxed criterion, where achieving NE does not mean that vehicles will stop perform car-following and lane-changing maneuvers. Instead, vehicles will continue these maneuvers more smoothly to meet their speed gain objectives. In the subsequent analysis, we will sample evaluation data within a 600-second time window after NE is reached to assess the attainability of CR.

According to the theoretical definition, CR is achieved when a Pareto-efficient NE is attained (Li et al. 2022). This means that the average speeds for each class of vehicles under DRL control are at least as high as the corresponding vehicle class speeds in the no-control scenario. Unlike the idealized one-shot bargaining game, we acknowledge that perfect Pareto-efficient NE might not be fully attainable in the simulation experiment. This stems from the existence of non-equilibrium behaviors and the stochastic and random nature of traffic dynamics, both of which contribute to the difficulty of accurately measuring traffic regimes. Therefore, we introduce a tolerance window  $\tau \in [0, 1]$ , where a traffic scenario is considered Pareto-efficient if the vehicle speed under DRL control is no less than  $\tau$  times the speed in the no-control scenario. A smaller  $\tau$  value indicates closer proximity to perfect Pareto efficiency, with  $\tau = 0$  representing the ideal Pareto-efficient NE. We expect the tolerance window to be reasonably small to ensure minimal compromise.

Figure 5 illustrates the attainment of Pareto-efficient NE across traffic scenarios under different tolerance windows. The results indicate that HVs are more likely to achieve Pareto efficiency compared to AVs, requiring only a 1% tolerance window. This 1% tolerance window corresponds to a speed reduction of up to 0.15 mph compared to the no-control scenario, which is considered negligible. Furthermore, the right-most figure demonstrates that a Pareto-efficient NE is achievable when class-level average speeds are allowed to be 4% lower than those in the no-control scenario. This equates to a maximum speed reduction of 1.5 mph. This minor speed reduction can be attributed to the slight heterogeneity introduced by the external control (i.e., DRL) in the closed-loop ring road system and is thus considered an acceptable trade-off. Given this, Pareto-efficient NE is considered to be achieved across these scenarios.

### 4.2 Comparisons with the bargaining game model

In addition to achieving Pareto-efficient NE, observing alignments with the game theoretical model can further support the attainability of CR through DRL. To examine the existence of such alignments, this section compares various macroscopic traffic variables between the DRL evaluation output and the theoretical model’s predictions. We note that exact matches are neither expected nor sought due to the inherent randomness in DRL and the simplified assumptions in the theoretical model. Instead, our primary goal is to identify any existing alignments. We will first establish a

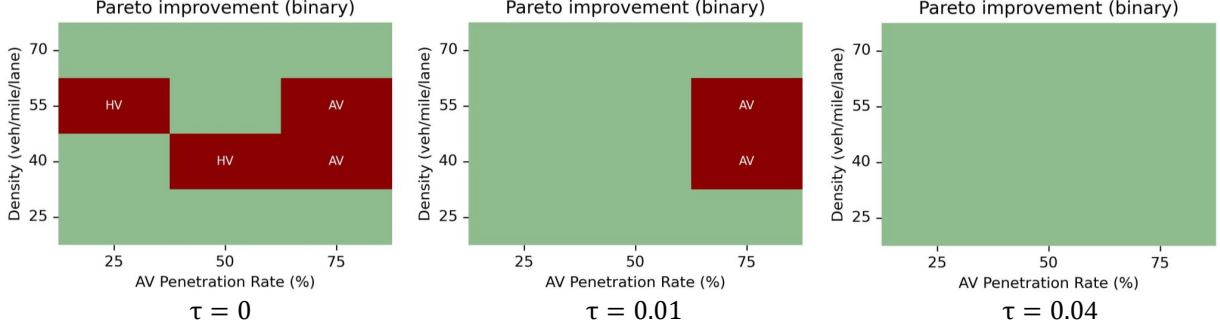


Figure 5: Attainment of Pareto efficiency under DRL control relative to the no-control scenario across different tolerance windows. Green indicates scenarios where Pareto-efficient NE is achieved, while red indicates where it is not. The labels “HV” or “AV” identify the vehicle class that has not reached Pareto-efficient NE in the corresponding scenario.

benchmark game-theoretical model that reaches a Pareto-efficient equilibrium. Then, comparisons will be made from three perspectives: first-order, second-order, and fundamental diagram comparisons.

#### 4.2.1 Establishment of benchmark game theoretical model

According to Li et al. (2022), there are numerous Pareto-efficient equilibria (i.e., CR) in the 2-pipe regime, determined by various ways of splitting the cooperation surplus. It is essential to properly establish a benchmark bargaining game model that aligns with the Pareto-efficient equilibrium achieved in DRL. This involves determining relevant parameters in the theoretical model based on simulation data, which has been shown to be feasible (Chen et al. 2025). Specifically, (i) the nominal behaviors of players will be defined according to those used in the simulation, and (ii) the optimal surplus split factor will be estimated from the simulation.

##### (i) Nominal behaviors of players

In the pay-off functions (1), each player is endowed with a nominal speed-density function,  $u_i(\cdot)$ . In the DRL simulation environment, we use the IDM model for both AVs and HVs, with specific IDM model parameters for each class. To match this setup in the benchmark model, we derive the corresponding speed-density functions of the IDM model under equilibrium conditions and use the same model parameters as in the simulation. The derived functions and the simulation data are plotted in Figure 6, showing a close match for both classes. This validates the feasibility of using the IDM model to represent nominal behaviors in the benchmark game theoretical model.

Additionally, we set the scaling parameters introduced in (2) to be  $a_1 = 1, a_2 = 1, b_1 = 0.8, b_2 = 1$ . This setting suggests that class 1 vehicles (HVs) are not sensitive to the class of their leader, whereas class 2 vehicles (AVs) tend to follow AVs more closely than HVs. This setup is consistent with the minimum time headway parameters used in the DRL simulation settings.

##### (ii) Optimal surplus split factor

According to Li et al. (2022), the way to split the cooperation surplus is denoted as  $\lambda$  (see Definition 2). Building on this, Chen et al. (2025) developed a framework to estimate the surplus split factor from real-world mixed traffic and found that cars and trucks allocate the surplus with cars receiving 81% and trucks 19%. However, it is unclear how this manifests with AVs. We aim to estimate the surplus split in mixed autonomy traffic by solving for the optimal surplus split factor,  $\lambda^*$ .

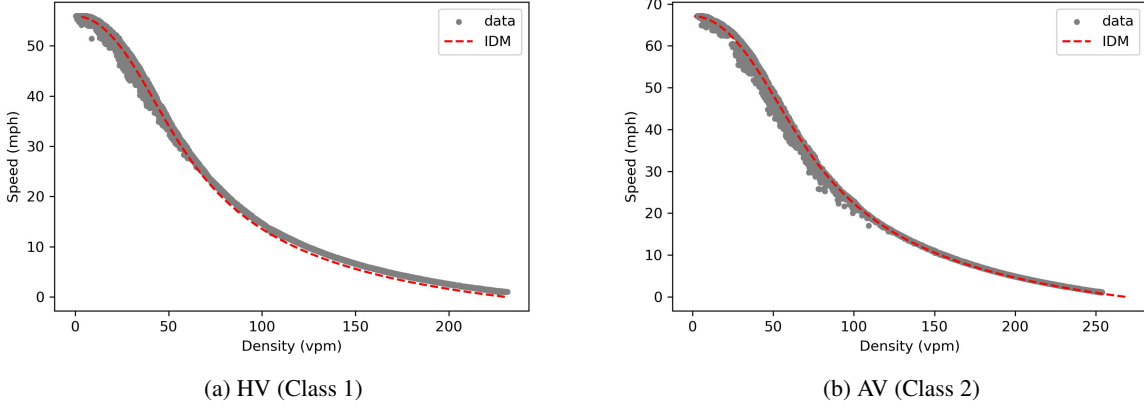


Figure 6: Comparison of speed-density relations between simulation data and the intelligent driver model (IDM).

We define a loss function aiming to best align the game-predicted speeds with the DRL simulated speeds as follows,

$$L(\rho_1, \rho_2, \lambda; \lambda \in [0, 1]) = \sum_{\rho_2=1}^{\rho_2^{max}} \sum_{\rho_1=1}^{\rho_1^{max}} \left[ \sum_{i=1}^2 w_i |\bar{u}_i(\rho_1, \rho_2) - \hat{u}_i(\rho_1, \rho_2; \lambda)| \right]^2 \quad (12)$$

Here, for more convenient calculations, we discretize the class density values into integers ranging from 1 to their respective maximum density values,  $\rho_1^{max}$  and  $\rho_2^{max}$ . The term  $w_i$  is the weighting factor that determines the relative importance between the errors of class 1 and class 2 speeds, and  $\hat{u}_i(\rho_1, \rho_2; \lambda)$  is the speed predicted by the game theoretical model in (1). Additionally,  $\bar{u}_i(\rho_1, \rho_2)$  represents the average speed for class  $i$  for the density pair  $(\rho_1, \rho_2)$ , calculated from the DRL evaluation data. Specifically, let  $S$  be the set of all speed values for the density pair  $(\rho_1, \rho_2)$ , then the average speed is calculated as:

$$\bar{u}_i(\rho_1, \rho_2) = \frac{1}{|S|} \sum_{k=1}^{|S|} u_k^S \quad (13)$$

where  $|S|$  is the size of set  $S$ , i.e., the number of speed values for density pair  $(\rho_1, \rho_2)$ , and  $u_k^S$  is the  $k$ th speed value in  $S$ .

With (12), the goal is to find the optimal  $\lambda$  that minimizes the loss,

$$\lambda^* = \arg \min_{\lambda} L(\rho_1, \rho_2, \lambda; \lambda \in [0, 1]) \quad (14)$$

The estimation yields  $\lambda^* = 0.6484$ , with the mean absolute errors (MAE) for class 1, class 2, and the weighted average being 3.50, 2.70, and 3.10, respectively. This indicates that HVs receive 64.84% of the cooperation surplus, while AVs receive 35.16%. This suggests that in the DRL-attained Pareto-efficient NE, HVs benefit more than AVs. This could be attributed to the other benefits that AV passengers already gain from the autonomous driving technology, such as increased productivity or leisure time, reduced driving stress, and optimized fuel efficiency (Rahman and Thill 2023), making them less inclined to seek further road-based benefits.

Moreover, Figure 7 depicts the loss under different  $\lambda$  values. Generally, the loss function appears symmetric around  $\lambda^*$ . This symmetry indicates that equilibrium can shift to both left and right with equal effort, suggesting that HVs and AVs have similar adaptability in receiving the cooperation surplus.

The settings and solutions of the above parameters result in a Pareto-efficient NE from the bargaining game model that aligns with the DRL-attained NE. This benchmark model will be used for the subsequent analysis.

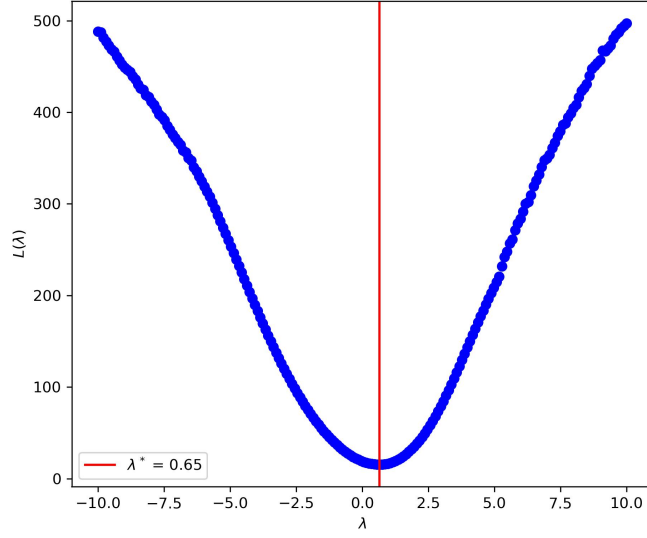


Figure 7: Optimal surplus split factor.

#### 4.2.2 First-order comparison

In the first-order comparisons of the Pareto-efficient NE between the DRL model and the benchmark game theoretical model, we aim to directly compare heatmaps for various traffic variables without performing mathematical calculations.

In Figure 8, panel (a) compares the 1-pipe equilibrium speed from the fully mixed (no control) scenario with the one solved from (2). Panels (b) and (c) compare class-specific speeds, while panels (d) and (e) compare class-specific traffic flow. The total flow of the mixed traffic is compared in panel (f). In the left figure for each panel, the speed value for each grid is calculated using (13), and the flow value is computed by multiplying the speed value by the traffic density. Overall, the macroscopic traffic patterns observed in the simulation closely approximate the predictions from the benchmark model. This alignment indicates that, on one hand, CR is attainable through DRL, and on the other hand, this attainability is not coincidental but is grounded in a theoretical benchmark.

#### 4.2.3 Second-order comparison

The second-order comparison involves simple mathematical subtractions between different traffic variables to reveal their relative relationships.

In Figure 9, we present the speed and flow differences between HVs and AVs, comparing these with the benchmark model's predictions. Overall, both speed and flow exhibit similar patterns to the benchmark model. However, the DRL model demonstrates greater heterogeneity. We seek to explain this phenomenon. Revisiting Figure 8(b)-(c), we denote the speed from DRL simulation as  $\hat{u}_i(\rho_1, \rho_2)$  and the speed from the benchmark model as  $u_i(\rho_1, \rho_2)$ . Then their relationship can be expressed as,

$$\begin{cases} \hat{u}_1(\rho_1, \rho_2) = u_1(\rho_1, \rho_2) + \epsilon_1 \\ \hat{u}_2(\rho_1, \rho_2) = u_2(\rho_1, \rho_2) + \epsilon_2 \end{cases} \quad (15)$$

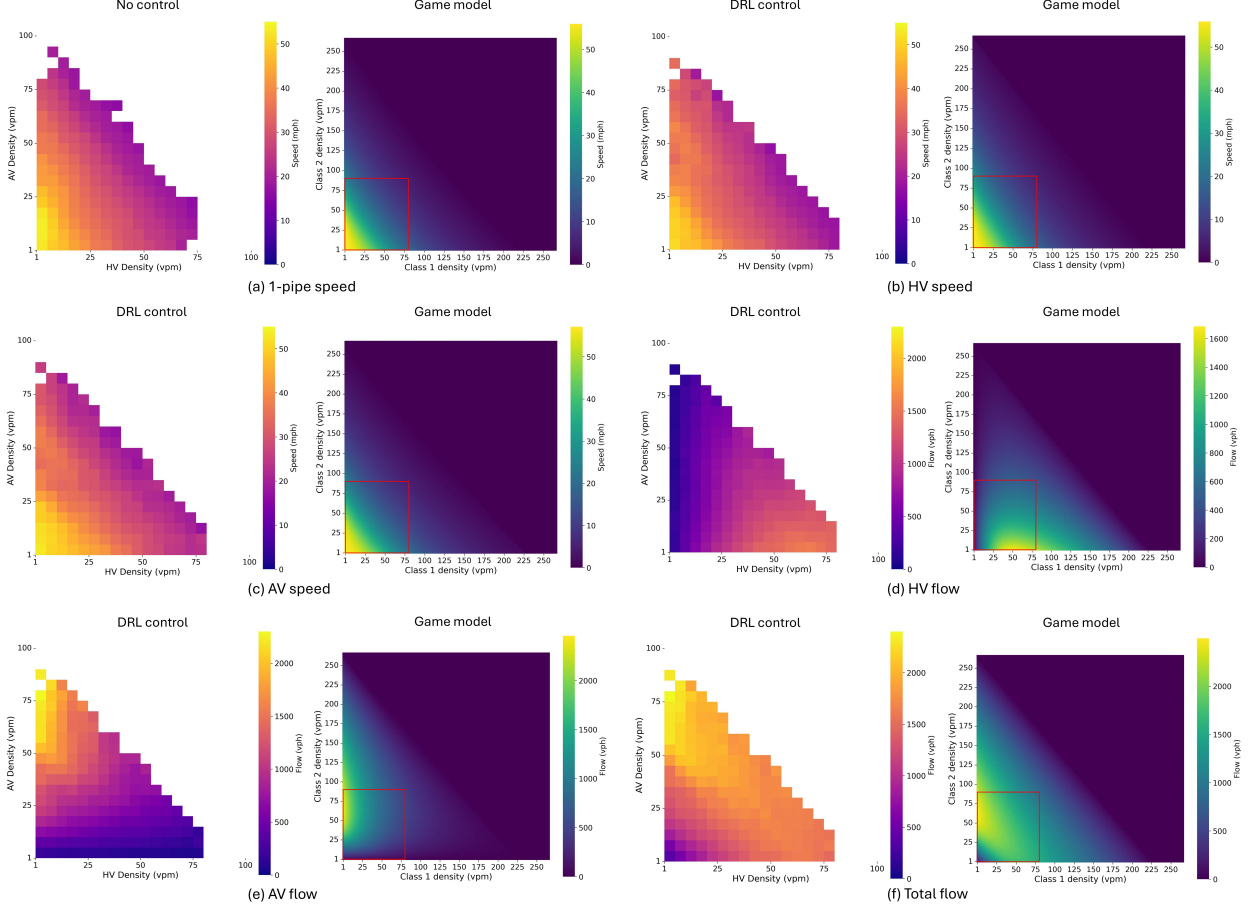


Figure 8: First-order comparisons of macroscopic traffic characteristics achieved by the DRL model and the Pareto-efficient equilibrium (collective rationality) predicted by the game theoretical model.

where  $\epsilon_i$  is the error between the DRL simulation data and the game model's predictions. Denoting the variance of errors  $\epsilon_1$  and  $\epsilon_2$  as  $\sigma_1$  and  $\sigma_2$ , respectively, the speed difference between HV and AV is computed as,

$$\Delta = \hat{u}_1 - \hat{u}_2 \quad (16)$$

For simplification, assuming  $\hat{u}_1$  and  $\hat{u}_2$  are independent, then the randomness for  $\Delta$  becomes  $\epsilon_1 + \epsilon_2$ , with the variance of randomness being  $\sigma_1 + \sigma_2$ . This accumulation of randomness in  $\Delta$  leads to discrepancies in Figure 9(a) compared to the benchmark model. In contrast, as illustrated in Figure 9(b), the flow difference is less affected by such randomness. The effect of speed variations is smoothed out over the density of vehicles, a deterministic variable that acts as a stabilizing factor, leading to a reduction of the randomness observed in the magnitude of flow.

In Figure 10, we compare the improvements in speed and flow relative to the fully mixed scenario (i.e., no CR). We acknowledge the observation of some differences, due to the inherent heterogeneity in simulation. Nonetheless, notable similarities are also present, highlighted by blue circles indicating regions of similarity and the green curve marking the boundary of positive and non-positive value transitions. For instance, in Figure 10(a), both the DRL-attained NE and the theoretical prediction show positive speed improvement at higher AV densities. Similarly, in Figure 10(c), the boundaries where values transition from positive to non-positive appear similar. Overall, these similarities suggest that the possibility of CR's emergence from the DRL model. Furthermore, the DRL model offers richer outcomes compared to the theoretical model. For example, the DRL model captures negative improvements on speed and flow

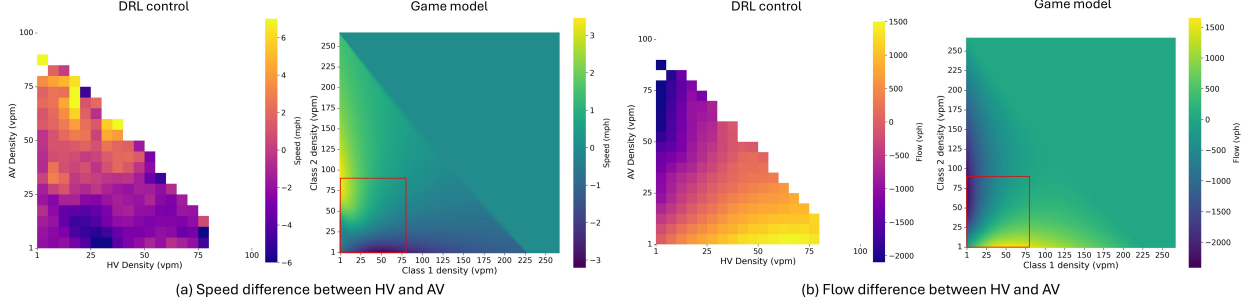


Figure 9: Second-order comparisons of speed and flow differences between HVs and AVs.

due to uncertainties present in both DRL-controlled and no-control simulation environments. While the theoretical model, with its simplified assumptions, fails to account for these variations.

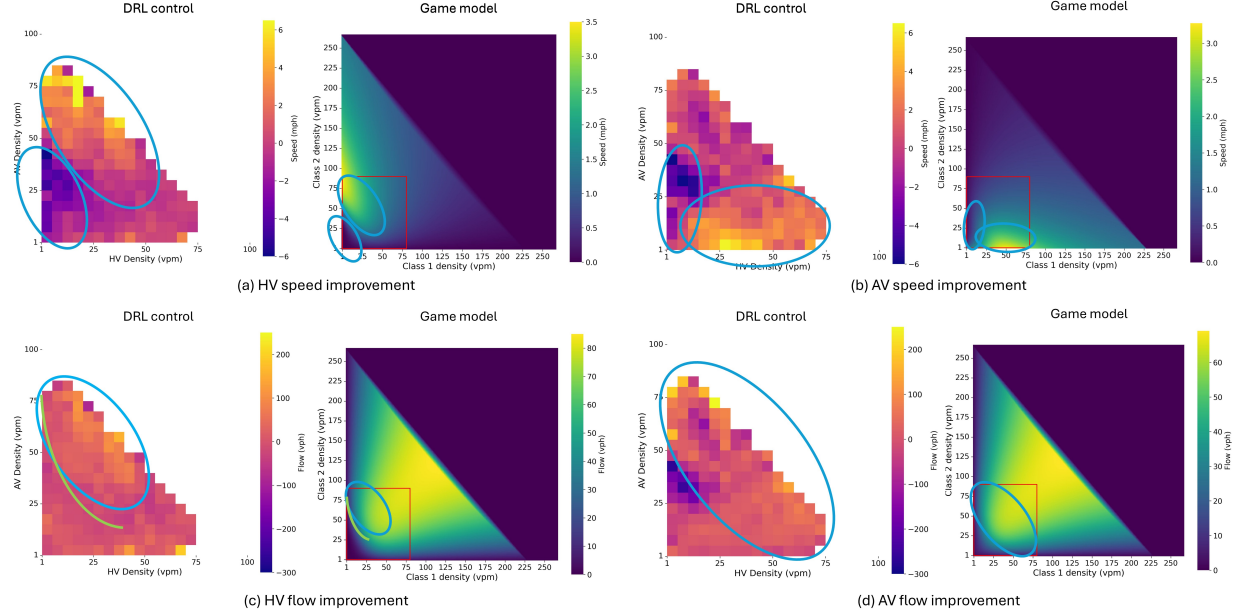


Figure 10: Second-order comparisons of speed and flow improvements between HVs and AVs.

#### 4.2.4 Fundamental diagram comparison

The comparisons of fundamental diagrams are presented in Figure 11, where both flow and density represent the combined totals for the two vehicle classes. As indicated by the black and gray lines, theoretically, the existence of CR improves traffic flow due to the fully utilized cooperation surplus compared with the fully mixed regime (Li et al. 2022). To verify this for the CR attained by DRL, we plot the flow-density relations using the simulation data, as represented by red (DRL control) and blue (no control) scatter points in Figure 11. The maximum flow under DRL control is higher than no DRL control, especially near critical densities, aligning with theoretical predictions. The improvements in maximum flow are 90 vph/lane for 25% AVs, 130 vph/lane for 50% AVs, and 98 vph/lane for 75% AVs. Besides, the maximum flow tends to be higher with increased AV penetration rates for both the theoretical and DRL models. These alignments further support our hypothesis on the attainability of CR through DRL.

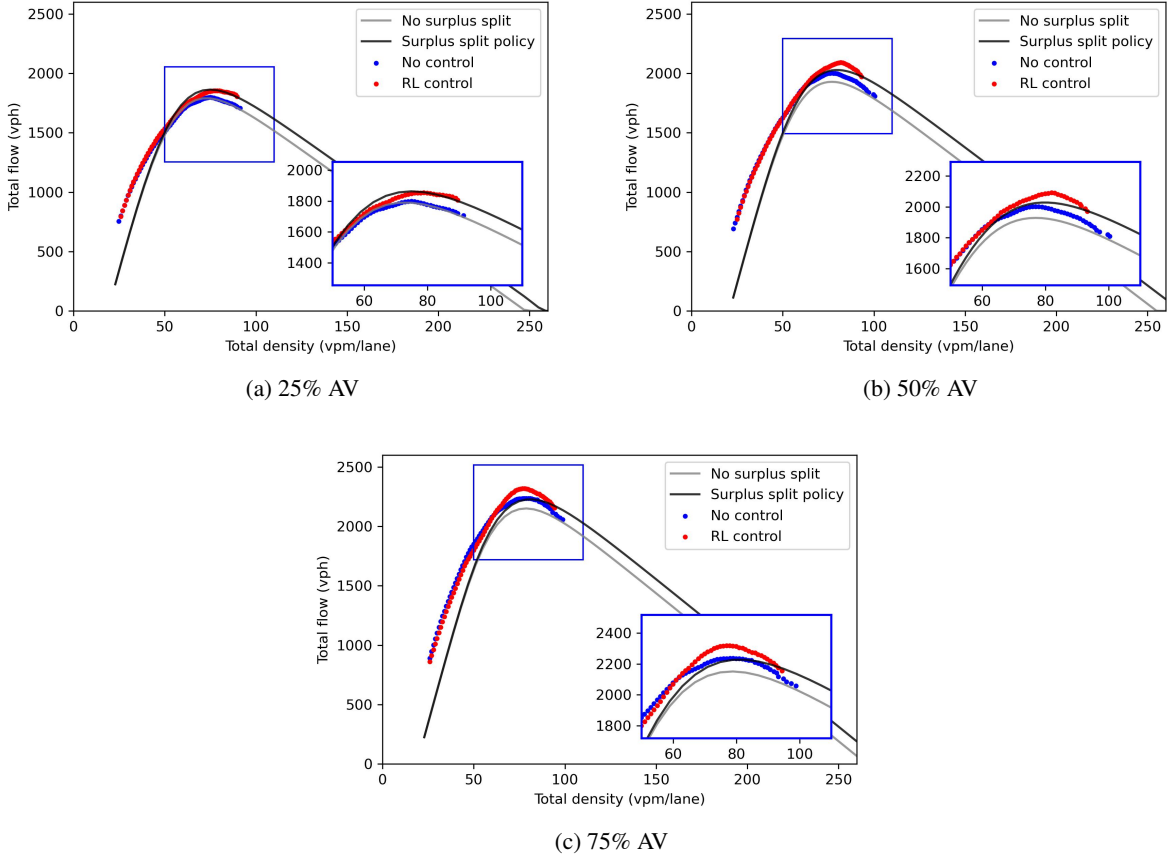


Figure 11: Comparison of fundamental diagrams between the DRL model and game theoretical model under different AV penetration rates.

## 5 Mechanisms to attain collective rationality

### 5.1 Hypothesized mechanism

Drawing on the evidence on the emergence of CR through DRL, it is natural to ask what factors contribute to this attainment in the microscopic and dynamic environment. We hypothesize a mechanism to explain the formation of CR, which is illustrated in Figure 12. In contrast to the game theoretical model, which is a “one-shot” bargaining focused on the macroscopic scale and all possible static equilibria, the DRL method elucidates the dynamic process of reaching CR. The core hypothesis posits that DRL, even with a simple reward design, can achieve better spatial organization through lane-changing maneuvers. This enhanced spatial organization translates to Pareto-efficient NE, indicating the attainment of CR.

As shown in Figure 12, the DRL simulation environment captures the dynamic process of interactions between AVs and HVs. For SI-DRL AVs, the lane change penalty reward leads to fewer lane changes among AVs. Since lane changes impact both the current and target lanes (Jin 2010), this reduction by AVs facilitates easier lane changes for other vehicles. Meanwhile, the speed reward encourages SI-DRL AVs to form platoons and occupy preferred lanes. HVs, which make lane-change decisions based on SUMO’s built-in model, respond to AVs’ lane occupation by making more lane changes to optimize their lane positions. These changes in lane occupation subsequently influence AVs’ decision-making, creating a feedback loop that reflects the dynamic interplay between AVs and HVs. Depending

on whether the feedback loop is self-reinforcing (influenced by decision-making of both HVs and AVs), the system eventually converges to either a CR or non-CR state at the collective level.

The subsequent sections will validate the hypothesized mechanism based on simulation evidence. This includes analyzing lane-changing behaviors at the behavioral layer (Section 5.2), examining platooning and lane occupations at the mesoscopic layer (Section 5.3), and introducing a spatial organization metric to assess macroscopic spatial organization in the collective layer (Section 5.4). Additionally, we will further elucidate the relationship between the proposed spatial organization metric and the game theoretical model in the collective layer (Section 5.5).

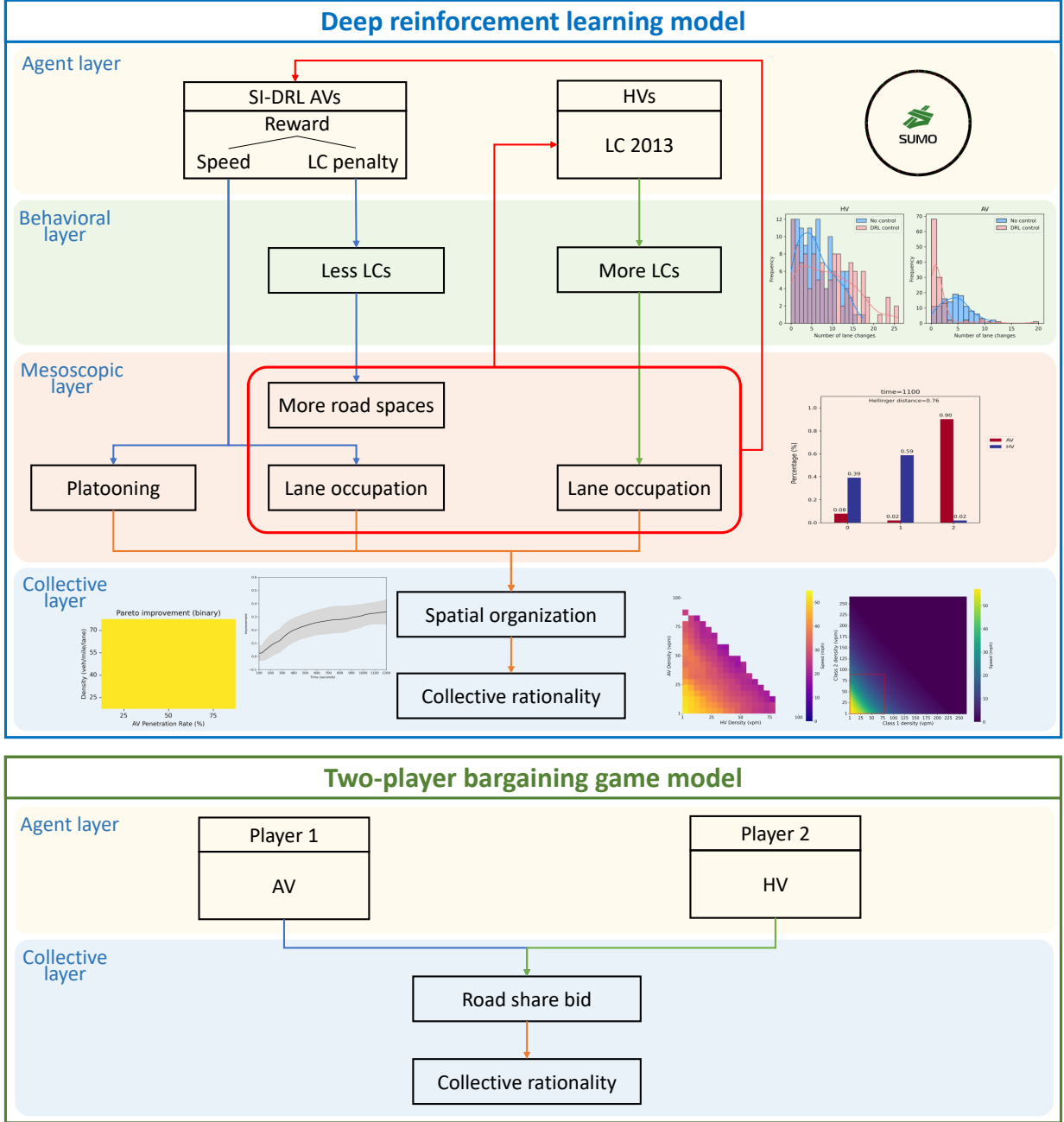


Figure 12: Hypothesized mechanisms for attaining collective rationality through deep reinforcement learning and comparisons with the game theoretical model. The term “lane change” is referred to LC in the figure.

## 5.2 Behavioral analysis

As postulated in Figure 12, the SI-DRL AVs engage in fewer lane change maneuvers, leading to HVs making more lane changes during their interactions. To assess these lane change behaviors, we measure the lane change frequency for each class and compare the distributions with the no-control scenario. The lane change frequency is calculated as the average number of lane changes per vehicle over each evaluation period, excluding the vehicle loading time. Figure 13 illustrates lane change frequency distributions for the system, HVs, and AVs. The results indicate significant differences between DRL control and no-control scenarios for both AVs and HVs. Specifically, in the DRL-controlled environment, SI-DRL agents perform fewer lane changes due to their optimized decision models, while HVs show more lane changes to adapt to AV impacts and find optimal lanes. These interactions suggest the potential for different spatial distributions compared to the no-control scenario.

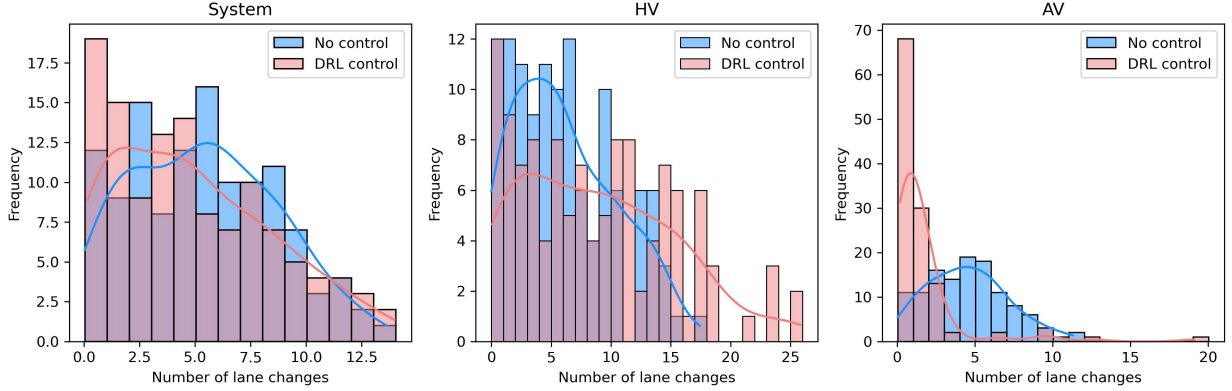


Figure 13: Comparisons of lane change frequency distributions between DRL control and no control.

## 5.3 Mesoscopic analysis

To analyze the AV-HV interactions at the mesoscopic level, we capture screenshots from the DRL model evaluation process in SUMO and examine the lane occupation and platooning behaviors of vehicles. Figure 14 illustrates a typical scenario with a traffic density of 40 vpm/lane and 75% AVs. At timestamp  $t = 100$ , vehicles have just been loaded onto the road, and the system is in a fully mixed state, with both vehicle classes approximately evenly distributed across lanes and AVs not yet forming large platoons. At timestamp  $t = 451$ , equilibrium is reached and maintained. Different lane occupations between the two classes start to emerge. AVs begin to occupy their preferred lanes, specifically lane 1 (middle lane) and lane 2 (leftmost lane). In response to the lane occupation by AVs, more HVs shift to the rightmost lane. Additionally, AVs start forming platoons, but the platoon sizes are still small and some AVs remain outside of platoons. At timestamp  $t = 600$ , AVs maintain their lane occupation, while HVs continue to explore different lane positions. Larger platoons are also forming among AVs. Finally, at  $t = 1200$ , AVs and HVs predominantly occupy separate lanes while sharing the middle lane. In the middle lane, AVs are platooned together, traveling separately from HVs.

## 5.4 Spatial organization of mixed traffic

Building on our previous work (Li et al. 2022), which showed that CR emerges in traffic regimes where the two classes of vehicles are spatially separated, we postulate that when CR is present in this study, AVs and HVs may exhibit different spatial organization at the collective level, as illustrated in Figure 12. This section seeks to validate this hypothesis by examining the spatial organization of mixed traffic through simulation evidence. To this end, we propose a spatial organization metric that evaluates the evolution of spatial arrangements, quantifying both intra-class platooning and inter-class segregation behaviors across longitudinal and lateral dimensions.

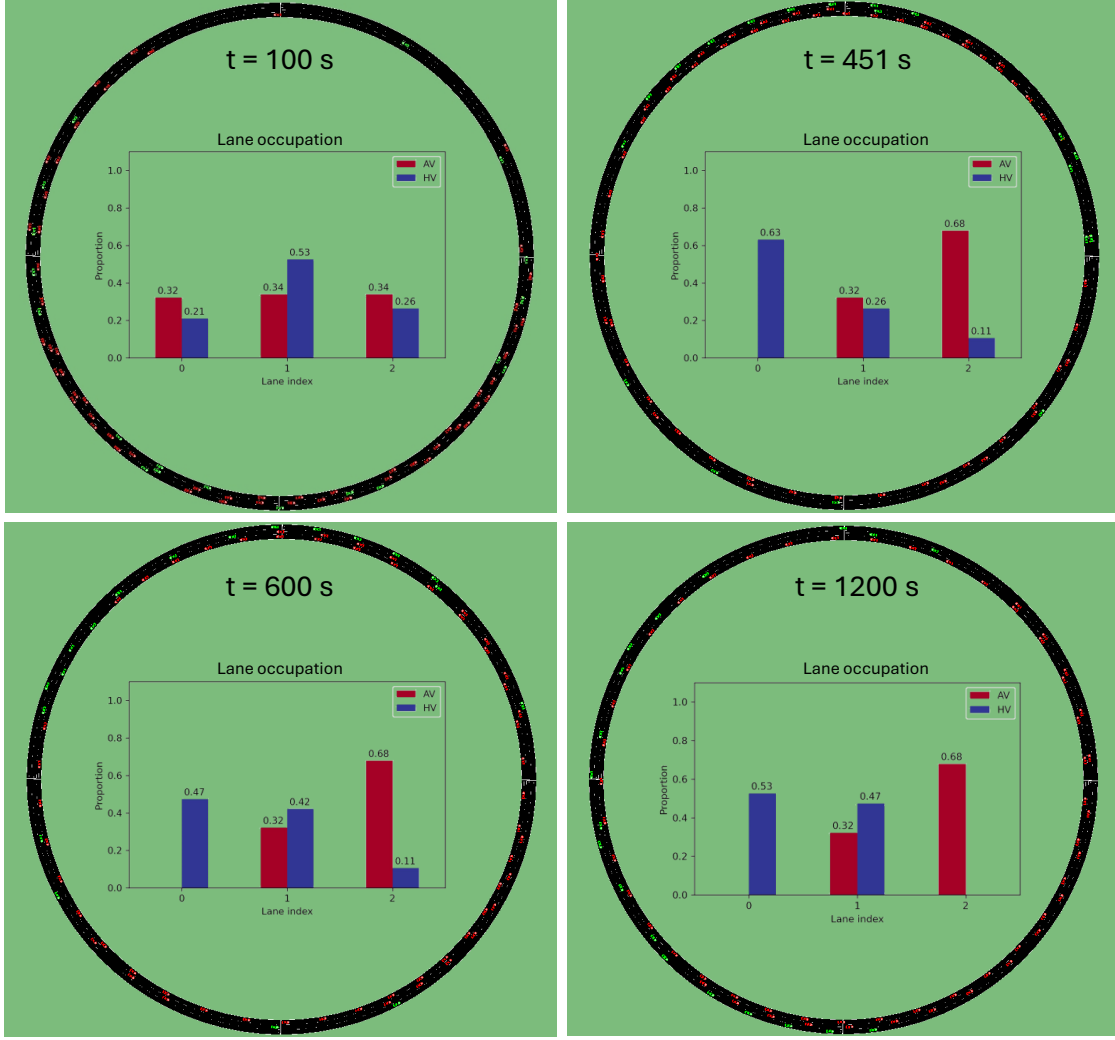


Figure 14: Screenshot of evaluation scenario for 40 vpm/lane and 75% AVs in SUMO. Vehicles in red and green represent AVs and HVs, respectively.

#### 5.4.1 The proposed spatial organization metric

In mixed traffic, the spatial distribution between vehicle classes impact the overall system performance (Ma et al. 2021, Yao et al. 2022). Our previous work (Li et al. 2022) also analyzed different configurations of mixed traffic in the analytical modeling of CR, finding that CR is present in a regime where the two vehicle classes are spatially separated from each other. Therefore, to understand the underlying mechanisms by which DRL achieves CR, we propose a spatial organization metric to evaluate the evolution of spatial arrangements in mixed traffic as it progresses towards CR. This metric quantifies both intra-class platooning and inter-class segregation behaviors, encompassing both longitudinal and lateral dimensions.

##### (i) Intra-class metric

As shown in Figure 1, the benefits of CR arise from more AVs platooning together, where AVs' smaller desired headways result in road space savings. Consequently, the intra-class metric, which focuses specifically on AVs, aims to quantify their platooning behaviors.

Platooning refers to a fleet of vehicles traveling with a small desired inter-vehicle spacing with the help of advanced automated driving systems. Typical definitions for such inter-vehicle spacing include the constant spacing (CS) policy and the constant time headway (CTH) policy. The CS policy sets a fixed inter-vehicle distance, whereas the CTH policy sets a fixed time headway and adjusts the inter-vehicle distance based on vehicle speed. We adopt the CTH policy as it better addresses safety concerns by varying the spacing with travel speed. Under the CTH policy, the maximum spacing considered as part of a platoon is defined as,

$$s_p^* = uh_{AA}^* + d_0 \quad (17)$$

where  $s_p^*$  is the equilibrium inter-vehicle spacing,  $u$  is the vehicle travel speed,  $h_{AA}^*$  is the desired equilibrium time headway between AVs, and  $d_0$  is the standstill gap. In this study, we adopt  $h_{AA}^* = 1 \text{ sec}$  and  $d_0 = 1 \text{ m}$ .

Vehicles in different platoon positions experience varying benefits, with platoon members generally receiving more advantages, while the platoon leader benefits less (Tsugawa et al. 2016, Shladover et al. 2015). To focus primarily on the mechanisms of CR, we simplify the platooning problem by assuming that only platoon members gain benefits. Accordingly, the intra-class metric quantifies the percentage of AVs benefiting from platooning and consists of two components: one measures the percentage of AVs traveling within platoons, denoted as  $P_e \in [0, 1]$ , and the other measures the percentage of AVs acting as platoon leaders, denoted as  $P_s \in [0, 1/2]$ . The objective is to encourage the formation of platoons while penalizing the creation of numerous small platoons. The benefit of platooning, denoted as  $B$ , is given by,

$$B = \frac{N_{AV}^p - N_p}{N_{AV}} = P_e - P_s \quad (18)$$

where  $N_{AV}$  denotes the total number of AVs,  $N_{AV}^p$  denotes the number of AVs in platoons, and  $N_p$  represents the number of platoons, which is equivalent to the number of platoon leaders.

## (ii) Inter-class metric

The inter-class metric assesses the spatial distribution between HVs and AVs. In measuring the spatial distributions of mixed traffic, studies often employ statistical methods such as probability theory to quantify the likelihood of platooning (Chang et al. 2020, Ma et al. 2021). However, these metrics typically focus solely on the longitudinal dimension and tend to overlook the lateral distribution (i.e., lane distribution) of mixed traffic agents.

To quantify the spatial distribution between the two classes both longitudinally and laterally, we utilize a statistical metric known as the Hellinger distance (Hellinger 1909). This metric measures the similarity between two probability distributions, either continuous or discrete, without requiring specific assumptions about the distributions. Additionally, it can handle multivariate distributions (Tamura and Boos 1986, Cutler and Cordero-Brana 1996), making it suitable for comparing two sets of bivariate distributions.

To begin with, we define a one-dimensional Hellinger distance to measure the differences in lane distributions between vehicle classes. Let  $p_i$  and  $q_i$  represent the percentage of AVs and HVs in the  $i$ th lane relative to the total number of AVs and HVs across all lanes, respectively. And denote  $P = \{p_i \mid i = 1, \dots, N\}$  and  $Q = \{q_i \mid i = 1, \dots, N\}$  as two discrete probability distributions for AVs and HVs respectively, and it follows that  $\sum_{i=1}^N p_i = 1$  and  $\sum_{i=1}^N q_i = 1$ , where  $N$  is the total number of lanes. Then, the Hellinger distance is given by:

$$H(P, Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^N (\sqrt{p_i} - \sqrt{q_i})^2} \quad (19)$$

This distance measures the extent to which AVs and HVs are laterally separated from each other. It satisfies  $H(P, Q) \in [0, 1]$ , with 0 represents the minimum distance (fully mixed) and 1 represents the maximum distance (fully separate). Examples of these extreme cases are illustrated in Figure 1.

Such lateral organization of mixed traffic was also discussed in Li et al. (2022) in terms of fully mixed and fully separate regimes, which conceptually aligns with what the one-dimensional Hellinger distance captures. However, Li et al. (2022) overlooks longitudinal spatial organization due to its simplified assumptions, which should not be neglected in the more realistic settings of this study. To address this, we extend the Hellinger distance to a two-dimensional case.

We partition the multi-lane ring road into  $M \times N$  cells, where  $M$  represents the number of cells in each lane longitudinally. Each cell contains a certain number of AVs and HVs. The proportion of AVs in cell  $(i, j)$  over the total number of AVs is denoted as  $p_{ij}$ ; similarly of  $q_{ij}$  for HVs. Then the probability distributions for AVs and HVs are respectively represented as  $P = \{p_{ij} \mid i = 1, \dots, N; j = 1, \dots, M\}$  and  $Q = \{q_{ij} \mid i = 1, \dots, N; j = 1, \dots, M\}$ . The two-dimensional Hellinger distance is therefore written as,

$$H(P, Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^N \sum_{j=1}^M (\sqrt{p_{ij}} - \sqrt{q_{ij}})^2} \quad (20)$$

Here, the relation follows  $\sum_i \sum_j p_{ij} = 1$  and  $\sum_i \sum_j q_{ij} = 1$ . As a special case of this definition, by setting  $M = 1$ , meaning each lane is considered a cell, (20) reduces to the one-dimensional Hellinger distance as defined in (19). For the studied ring road scenario, we set  $N = 3$  and  $M = 10$ .

### (iii) Integrated spatial organization metric

By integrating the proposed metrics in (18) and (20), we can assess the degree of spatial organization at any timestamp  $t$ . The spatial organization metric  $M_t$  is expressed as,

$$M_t = a_1 B_t + a_2 H_t \quad (21)$$

where the coefficients  $a_1$  and  $a_2$  determine the emphasis on micro-scale (intra-class) versus system-level (inter-class) performance. In this study, we select  $a_1 = 0.5$  and  $a_2 = 1$ .

## 5.4.2 Spatial organization evaluation

The evolution of spatial organization between HVs and AVs is depicted in Figure 15. The y-axis represents the difference in metric (21) between the DRL-controlled scenario and the no-control baseline (i.e., fully mixed). The metric generally remains positive and increases over time, indicating a tendency for mixed traffic to become more spatially organized in both longitudinal and lateral dimensions. An exception is observed at a density of 25 vpm/lane with 25% AVs. This occurs because, under free-flow conditions with ample road space available, self-interested agents can benefit without the need for spatial organization. Despite this exception, the overall trend across all traffic conditions shows a tendency towards spatial organization, as indicated by Figure 15b.

## 5.5 Relations with the game theoretical model

The improvement of the proposed spatial organization metric (21) indicates a more organized use of road space in the DRL-controlled scenario compared to fully mixed traffic. This concept is similar to the cooperation surplus (see Definition 1), where players achieve a more organized traffic regime to be better off than in a fully mixed regime. Therefore, it is natural to explore if these two variables are related and how they correlate.

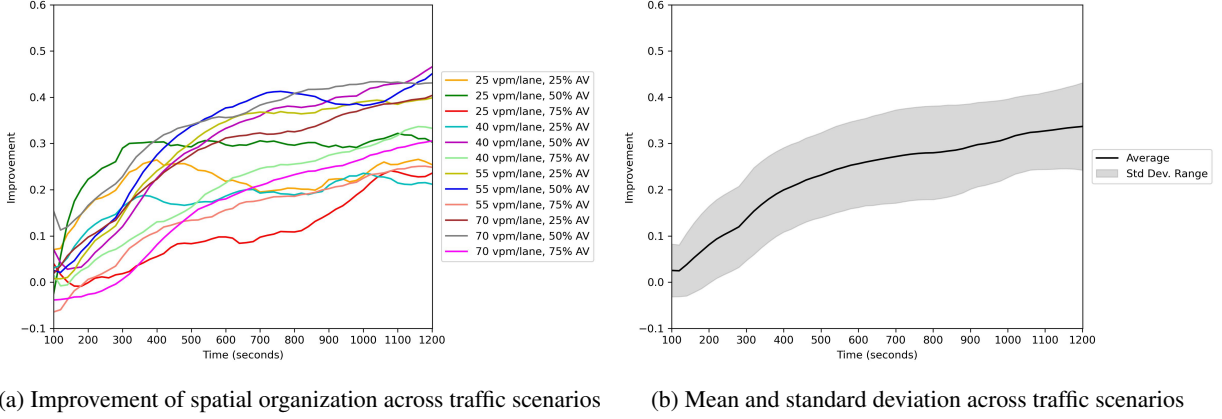


Figure 15: Improvement of the spatial organization metric (21) under DRL control compared to no control. The first 100 seconds are excluded as they represent the vehicle loading time.

To connect these two variables, we align the traffic densities of each class from the DRL simulation output with the corresponding traffic densities from the game theoretical model. The relationship between the predicted cooperation surplus and the improvement in spatial organization is illustrated in Figure 16. A positive linear relationship is observed, where the cooperation surplus (in %) increases by 2.34 units for every unit increase in spatial organization improvement, with a y-intercept of 3.18. The p-value of 0 indicates that this relationship is statistically significant, indicating a extremely low probability of observing this data by chance. Besides, the Pearson’s correlation coefficient of 0.53 indicates a moderate positive correlation between these two variables.

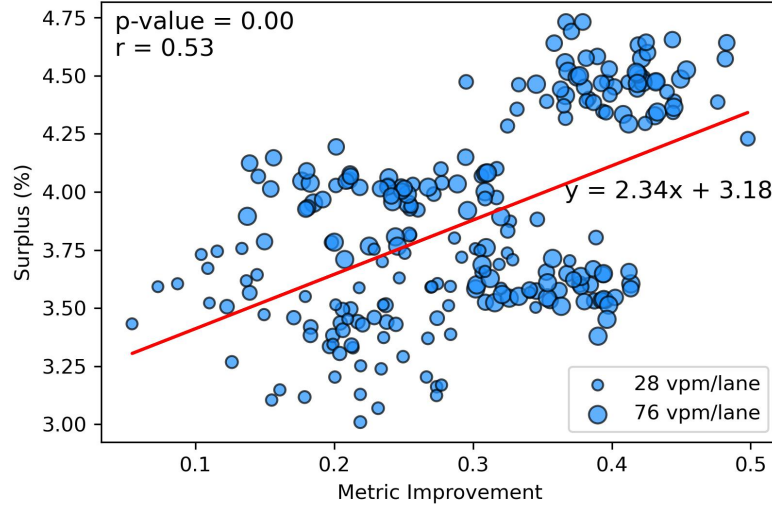


Figure 16: Relationship between cooperation surplus and improvement of the spatial organization metric.

Furthermore, the size of the scatter points in Figure 16 represents total traffic densities. Smaller points are mainly clustered in the lower-left corner, while larger points are in the upper-right corner. This suggests that within the range of the experimented traffic densities, CR tends to be more evident at higher traffic densities.

Overall, this positive relationship implies the robustness of CR’s emergence from both analytical models and simulation-based models. Moreover, it suggests that enhancing spatial organization is a viable approach to achieving CR in real-world mixed traffic scenarios.

## 6 Conclusion

This study examines how self-interested AV behaviors can positively impact mixed traffic systems, through the lens of collective rationality (CR). Using DRL with a simple reward design, we demonstrate that CR can emerge in microscopic and dynamic mixed autonomy environment. Comparisons between the DRL-based results and our earlier game-theoretical predictions (Li et al. 2022) show the first-order alignment between the two, and richer second-order outcomes from DRL-driven learning process. The analysis of fundamental diagrams shows a maximum capacity improvement of 130 vph/lane under 50% AVs when collective rationality is attained. Moreover, the surplus split factor estimation reveals HVs gain 64.84% of benefits from collective rationality, with AVs taking 35.16%.

We further hypothesized and validated an underlying mechanism for the emergence of CR in the physical driving environment. Microscopically, significant differences in lane-changing behaviors are observed for both AVs and HVs when collective rationality is attained. Macroscopically, the proposed spatial organization metric shows a stable increasing trend in both inter-class segregation and intra-class platooning behaviors. These results reveal that collective rationality emerges as mixed traffic agents perform lane-changing and car-following maneuvers in a way that results in a more organized traffic system. Overall, these findings suggest that enhancing spatial organization in real-world mixed traffic has the potential to benefit all road users.

We envision several extensions for future work. Firstly, advanced learning methods, such as federated learning, can be adopted to achieve collective rationality among self-interested driving agents while ensuring data privacy for AVs. This is achieved by allowing each AV to train its local model using local data while only sharing the model parameters to the central aggregator. Secondly, the lane-changing decisions for HVs in this study are controlled by SUMO’s LC2013 model, a rule-based and heuristic decision policy. It is possible to learn human-like lane-changing decisions from empirical data and implement them to explore collective rationality in more realistic settings. Lastly, this study assumes that all AVs belong to the same company. However, different decision-making policies from various manufacturers can result in diverse AV behaviors. Therefore, it is worthwhile to explore collective rationality when multiple AV classes are considered using both analytical and learning-based methods.

## Acknowledgment

This research is supported by NSF grant “EAGER: Collaborative Research: Fostering Collective Rationality Among Self-Interested Agents to Improve Design and Efficiency of Mixed Autonomy Networks and Infrastructure Systems” (Award numbers: 2437983, 2437982).

## References

- SC Calvert, WJ Schakel, and JWC Van Lint. Will automated vehicles negatively impact traffic flow? *Journal of advanced transportation*, 2017, 2017.
- Xin Chang, Haijian Li, Jian Rong, Xiaohua Zhao, et al. Analysis on traffic stability and capacity for mixed traffic flow with platoons of intelligent connected vehicles. *Physica A: Statistical Mechanics and Its Applications*, 557:124829, 2020.
- Danjue Chen, Soyoung Ahn, Madhav Chitturi, and David A Noyce. Towards vehicle automation: Roadway capacity formulation for traffic mixed with regular and automated vehicles. *Transportation research part B: methodological*, 100:196–221, 2017.
- Di Chen, Jia Li, and Michael Zhang. Evidence and quantification of cooperation of driving agents in mixed traffic flow. *Transportation Research Part B: Methodological*, 200:103285, 2025.
- Adele Cutler and Olga I Cordero-Brana. Minimum hellinger distance estimation for finite mixture models. *Journal of the American Statistical association*, 91(436):1716–1723, 1996.
- Carlos F Daganzo. A continuum theory of traffic dynamics for freeways with special lanes. *Transportation Research Part B: Methodological*, 31(2):83–102, 1997.

- Carlos F Daganzo. A behavioral theory of multi-lane traffic flow. part i: Long homogeneous freeway sections. *Transportation Research Part B: Methodological*, 36(2):131–158, 2002a.
- Carlos F Daganzo. A behavioral theory of multi-lane traffic flow. part ii: Merges and the onset of congestion. *Transportation Research Part B: Methodological*, 36(2):159–169, 2002b.
- Katherine Driggs-Campbell and Ruzena Bajcsy. Communicating intent on the road through human-inspired control schemes. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3042–3047. IEEE, 2016.
- Katherine Driggs-Campbell, Vijay Govindarajan, and Ruzena Bajcsy. Integrating intuitive driver models in autonomous planning for interactive maneuvers. *IEEE Transactions on Intelligent Transportation Systems*, 18(12):3461–3472, 2017.
- Leslie C Edie et al. *Discussion of traffic stream measurements and definitions*. Port of New York Authority New York, 1963.
- Jakob Erdmann. Sumo’s lane-changing model. In *Modeling Mobility with Open Data: 2nd SUMO Conference 2014 Berlin, Germany, May 15-16, 2014*, pages 105–123. Springer, 2015.
- Ernst Fehr and Simon Gächter. Altruistic punishment in humans. *Nature*, 415(6868):137–140, 2002.
- Amir Ghiasi, Omar Hussain, Zhen Sean Qian, and Xiaopeng Li. A mixed traffic capacity analysis and lane management model for connected automated vehicles: A markov chain method. *Transportation Research Part B: Methodological*, 106:266–292, 2017.
- Amir Goldental and Ido Kanter. A minority of self-organizing autonomous vehicles significantly increase freeway traffic flow. *Journal of Physics A: Mathematical and Theoretical*, 53(41):414001, 2020.
- George Gunter, Caroline Janssen, William Barbour, Raphael E Stern, and Daniel B Work. Model-based string stability of adaptive cruise control systems using field data. *IEEE Transactions on Intelligent Vehicles*, 5(1):90–99, 2019.
- George Gunter, Derek Gloudemans, Raphael E Stern, Sean McQuade, Rahul Bhadani, Matt Bunting, Maria Laura Delle Monache, Roman Lysecky, Benjamin Seibold, Jonathan Sprinkle, et al. Are commercially implemented adaptive cruise control systems string stable? *IEEE Transactions on Intelligent Transportation Systems*, 22(11):6992–7003, 2020.
- Mengting Guo, Yang Bai, Xia Li, Wei Zhou, Chunyang Wang, Xinwei Ma, Huixin Gao, and Yuewen Xiao. Freeway capacity modeling and analysis for traffic mixed with human-driven and connected automated vehicles considering driver behavior characteristics. *Physica A: Statistical Mechanics and its Applications*, 623:128894, 2023.
- Peng Hang, Chen Lv, Yang Xing, Chao Huang, and Zhongxu Hu. Human-like decision making for autonomous driving: A noncooperative game theoretic approach. *IEEE Transactions on Intelligent Transportation Systems*, 22(4):2076–2087, 2020.
- Ernst Hellinger. Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *Journal für die reine und angewandte Mathematik*, 1909(136):210–271, 1909.
- Kuang Huang, Xuan Di, Qiang Du, and Xi Chen. A game-theoretic framework for autonomous vehicles velocity control: Bridging microscopic differential games and macroscopic mean field games. *arXiv preprint arXiv:1903.06053*, 2019.
- Zhiyu Huang, Haochen Liu, Jingda Wu, and Chen Lv. Conditional predictive behavior planning with inverse reinforcement learning for human-like autonomous driving. *arXiv preprint arXiv:2212.08787*, 2022.
- Rachel M James, Christopher Melson, Jia Hu, and Joe Bared. Characterizing the impact of production adaptive cruise control on traffic flow: An investigation. *Transportmetrica B: Transport Dynamics*, 7(1):992–1012, 2019.
- Liming Jiang, Yuanchang Xie, Nicholas G Evans, Xiao Wen, Tienan Li, and Danjue Chen. Reinforcement learning based cooperative longitudinal control for reducing traffic oscillations and improving platoon stability. *Transportation Research Part C: Emerging Technologies*, 141:103744, 2022.
- Yangsheng Jiang, Tingting Ren, Yuqin Ma, Yunxia Wu, and Zhihong Yao. Traffic safety evaluation of mixed traffic flow considering the maximum platoon size of connected automated vehicles. *Physica A: Statistical Mechanics and its Applications*, page 128452, 2023.
- Wen-Long Jin. A kinematic wave theory of lane-changing traffic flow. *Transportation research part B: methodological*, 44(8-9): 1001–1021, 2010.
- Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- Samuel S Komorita. *Social dilemmas*. Routledge, 2019.
- Daniel Krajzewicz, Jakob Erdmann, Michael Behrisch, and Laura Bieker. Recent development and applications of sumo-simulation of urban mobility. *International journal on advances in systems and measurements*, 5(3&4), 2012.

- Paul Alphons Maria Lange, Daniel Balliet, Craig D Parks, and Mark Van Vugt. *Social dilemmas: the psychology of human cooperation*. Oxford University Press, 2014.
- Jia Li, Di Chen, and Michael Zhang. Equilibrium modeling of mixed autonomy traffic flow based on game theory. *Transportation research part B: methodological*, 166:110–127, 2022.
- Long-Ji Lin. Self-improving reactive agents based on reinforcement learning, planning and teaching. *Machine learning*, 8(3): 293–321, 1992.
- Xiangguo Liu, Guangchen Zhao, Neda Masoud, and Qi Zhu. Trajectory planning for connected and automated vehicles: Cruising, lane changing, and platooning. *arXiv preprint arXiv:2001.08620*, 2020.
- Xiangguo Liu, Neda Masoud, Qi Zhu, and Anahita Khojandi. A markov decision process framework to incorporate network-level data in motion planning for connected and automated vehicles. *Transportation Research Part C: Emerging Technologies*, 136: 103550, 2022.
- S Logghe and Lambertus H Immers. Multi-class kinematic wave theory of traffic flow. *Transportation Research Part B: Methodological*, 42(6):523–541, 2008.
- Ke Ma, Hao Wang, and Tiancheng Ruan. Analysis of road capacity and pollutant emissions: Impacts of connected and automated vehicle platoons on traffic flow. *Physica A: Statistical Mechanics and its Applications*, 583:126301, 2021.
- Andreas A Malikopoulos, Seongah Hong, B Brian Park, Joyoung Lee, and Seunghan Ryu. Optimal control for speed harmonization of automated vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 20(7):2405–2417, 2018.
- Vicente Milanés and Steven E Shladover. Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data. *Transportation Research Part C: Emerging Technologies*, 48:285–300, 2014.
- Zhen Sean Qian, Jia Li, Xiaopeng Li, Michael Zhang, and Haizhong Wang. Modeling heterogeneous traffic flow: A pragmatic approach. *Transportation Research Part B: Methodological*, 99:183–204, 2017.
- Yanyan Qin, Hao Wang, and Daiheng Ni. Lighthill-whitham-richards model for traffic flow mixed with cooperative adaptive cruise control vehicles. *Transportation Science*, 55(4):883–907, 2021.
- Md Mokhlesur Rahman and Jean-Claude Thill. Impacts of connected and autonomous vehicles on urban transportation and environment: A comprehensive review. *Sustainable Cities and Society*, 96:104649, 2023.
- Kyle Sama, Yoichi Morales, Hailong Liu, Naoki Akai, Alexander Carballo, Eijiro Takeuchi, and Kazuya Takeda. Extracting human-like driving behaviors from expert driver data using deep learning. *IEEE transactions on vehicular technology*, 69(9): 9315–9329, 2020.
- Mingfeng Shang and Raphael E Stern. Impacts of commercially available adaptive cruise control vehicles on highway stability and throughput. *Transportation research part C: emerging technologies*, 122:102897, 2021.
- Haotian Shi, Yang Zhou, Keshu Wu, Xin Wang, Yangxin Lin, and Bin Ran. Connected automated vehicle cooperative control with a deep reinforcement learning approach in a mixed traffic environment. *Transportation Research Part C: Emerging Technologies*, 133:103421, 2021.
- Steven E Shladover, Christopher Nowakowski, Xiao-Yun Lu, and Robert Ferlis. Cooperative adaptive cruise control: Definitions and operating concepts. *Transportation Research Record*, 2489(1):145–152, 2015.
- Roy N Tamura and Dennis D Boos. Minimum hellinger distance estimation for multivariate location and covariance. *Journal of the American Statistical Association*, 81(393):223–229, 1986.
- Martin Treiber and Arne Kesting. Traffic flow dynamics. *Traffic Flow Dynamics: Data, Models and Simulation*, Springer-Verlag Berlin Heidelberg, pages 983–1000, 2013.
- Sadayuki Tsugawa, Sabina Jeschke, and Steven E Shladover. A review of truck platooning projects for energy savings. *IEEE Transactions on Intelligent Vehicles*, 1(1):68–77, 2016.
- Panagiotis Typaldos, Markos Papageorgiou, and Ioannis Papamichail. Optimization-based path-planning for connected and non-connected automated vehicles. *Transportation Research Part C: Emerging Technologies*, 134:103487, 2022.
- Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3):279–292, 1992.
- Shouyang Wei, Yuan Zou, Tao Zhang, Xudong Zhang, and Wenwei Wang. Design and experimental validation of a cooperative adaptive cruise control system based on supervised reinforcement learning. *Applied sciences*, 8(7):1014, 2018.
- Paul Weirich. *Collective rationality: Equilibrium in cooperative games*. Oxford University Press, 2009.

- Donghao Xu, Zhezhang Ding, Xu He, Huijing Zhao, Mathieu Moze, François Aioun, and Franck Guillemard. Learning from naturalistic driving data for human-like autonomous highway driving. *IEEE Transactions on Intelligent Transportation Systems*, 22(12):7341–7354, 2020.
- Xianfeng Terry Yang, Ke Huang, Zhehao Zhang, Zhao Alan Zhang, and Fang Lin. Eco-driving system for connected automated vehicles: Multi-objective trajectory optimization. *IEEE Transactions on Intelligent Transportation Systems*, 22(12):7837–7849, 2020.
- Zhihong Yao, Qiufan Gu, Yangsheng Jiang, and Bin Ran. Fundamental diagram and stability of mixed traffic flow considering platoon size and intensity of connected automated vehicles. *Physica A: Statistical Mechanics and its Applications*, 604: 127857, 2022.
- Chunhui Yu, Yiheng Feng, Henry X Liu, Wanjing Ma, and Xiaoguang Yang. Corridor level cooperative trajectory optimization with connected and automated vehicles. *Transportation Research Part C: Emerging Technologies*, 105:405–421, 2019.
- Hongtao Yu, H Eric Tseng, and Reza Langari. A human-like game theory-based controller for automatic lane changing. *Transportation Research Part C: Emerging Technologies*, 88:140–158, 2018.
- Yi Zhang, Ping Sun, Yuhang Yin, Lin Lin, and Xuesong Wang. Human-like autonomous vehicle speed control by deep reinforcement learning with double q-learning. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 1251–1256. IEEE, 2018.
- Zijia Zhong, Earl E Lee, Mark Nejad, and Joyoung Lee. Influence of cav clustering strategies on mixed traffic flow characteristics: An analysis of vehicle trajectory data. *Transportation Research Part C: Emerging Technologies*, 115:102611, 2020.
- Meixin Zhu, Xuesong Wang, and Yinhai Wang. Human-like autonomous car-following model with deep reinforcement learning. *Transportation research part C: emerging technologies*, 97:348–368, 2018.