Emergence of Collective Rationality in Mixed Autonomous Driving Systems

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

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

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

Abstract

Collective action of agents is essential for steering AI's impacts toward societal benefit. Exhibited in various forms, cooperation is a ubiquitous phenomenon in many socio-technical systems involving interactions of multiple agents. Mixed autonomous driving systems are characterized by complex physical-strategic interactions among agents. It is curious and meaningful to ask whether cooperation can spontaneously emerge in such systems. This paper attempts to answer this question through experimental evidence and lens of collective rationality (CR) – a game-theoretic concept describing emergent cooperation among self-interested agents. We investigate when and how CR arises in mixed autonomous driving systems when Autonomous Vehicles (AVs) are distributedly trained through deep reinforcement learning (DRL) in simulation environments. We show that, intriguingly, simple reward design allows self-interested agents to consistently achieve CR across diverse scenarios, without explicitly including system-level incentives or transfer payments among agents. This finding serves as initial evidence on the emergence and scaling of cooperative behaviors among heterogeneous driving agents in mixed autonomy environments.

1 Introduction

In the era of artificial intelligence (AI), steering development toward social good requires collective action, which encompasses both competitive and cooperative dynamics [de Neufville and Baum, 2021, Hardt et al., 2023, Ben-Dov et al., 2024]. Among them, cooperation is a ubiquitous phenomenon in many socio-technical systems that involve interactions of multiple agents. Autonomous vehicle (AV) is an emerging technology, and it is anticipated that they will co-exist with human-driven

39th Conference on Neural Information Processing Systems (NeurIPS 2025)
Workshop on Algorithmic Collective Action (ACA@NeurIPS 2025).

vehicles (HVs) in the foreseeable future. Harnessing the power of AI for perception and decision-making, 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 governed by one or more shared system-level control objectives, such as system throughput or energy consumption. However, numerous empirical evidence indicates that individuals tend to be self-interested, focusing on maximizing their own utility rather than system-level benefits [Fehr and Gächter, 2002, Lange et al., 2014, Komorita, 2019]. This raises a natural question: when AVs are self-interested and pursue their own objectives without explicit external coordination, can cooperation emerge intrinsically in mixed-autonomy traffic so that system-level benefits are achievable?

The objective of this work is to explore whether cooperation can emerge in dynamic mixed autonomous driving systems. We find that collective rationality (CR) can be attained among heterogeneous self-interested driving agents when a subset of the agents (i.e. AVs) are trained using deep reinforcement learning (DRL) with a simple reward design, even without explicitly incorporating system-level incentives.

2 Experiment

2.1 Collective rationality

Collective rationality (CR) Consider a traffic system with two classes of driving agents. Roughly speaking, CR refers to the set of Pareto-efficient Nash Equilibrium (NE) states when cooperation surplus $S(\rho_1, \rho_2)$ is positive and both classes of agents attain a non-negative share of it according to a split factor $\lambda \in (0, 1)$. We include the formal definitions in Section A.1. The concept of CR in mixed traffic was analytically modeled in [Li et al., 2022] and empirically validated in [Chen et al., 2025].

2.2 Experiment setup

Agent state, action, and reward The state of an AV agent is composed of its own state and perception of surrounding agents in 100-meter radius, in terms of longitudinal and lateral (lane) position, speed, and agent type (AV or HV). The lane-changing behaviors of AV agent i are controlled by DRL, consisting of the following actions,

$$A_i = \{\text{change to left, change to right, keep lane}\} \tag{1}$$

The reward function of AVs is structured as,

$$R_i = w_{speed} u_i - w_{lane\ change} I_i \tag{2}$$

where u_i is speed of agent i, I_i is a binary variable indicating whether lane change occurs, and $w_{\cdot}$ is a weighting coefficient.

Physically, all vehicles are uniformly $5\ m$ in length and have maximum acceleration and deceleration rates of $2.6\ m/s^2$ and $4.5\ m/s^2$, respectively. The car-following of AVs and HVs both obeys Intelligent Driver Model (IDM) [Treiber and Kesting, 2013].

Deep-Q learning for AV agents We assume that AVs share a common Deep-Q network (DQN) for decision-making. The DQN architecture includes three fully connected hidden layers, respectively containing 256, 128, and 64 neurons. Each hidden layer uses a rectified linear unit (ReLU) activation function. The output layer has 3 neurons (equal to A_i). We set the state space size to $N_s = 200$. Additionally, to balance action exploration and exploitation, an ϵ-greedy decay function is applied, defined as $p(a) = \epsilon_{end} + (\epsilon_{start} - \epsilon_{end})e^{-epi/r_{decay}}$. Here, ϵ_{start} and ϵ_{end} are the initial and final values of ϵ 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. To improve the training stability and sample efficiency, we adopt experience replay: 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.

Mixed autonomous driving system The system is a 1000-meter ring road consisting of three lanes. We consider four traffic densities: 25 vehicles per mile per lane (vpm/lane), 40 vpm/lane, 55 vpm/lane, and 70 vpm/lane. We evaluate three AV penetration rates: 25%, 50%, and 75%. There are 12 traffic scenarios in total. The baseline scenario is defined as the no-control/learning scenario, where all agents follow prescribed driving rules. We use open-source simulator Simulation of Urban MObility (SUMO) [Krajzewicz et al., 2012] to implement the experiments.

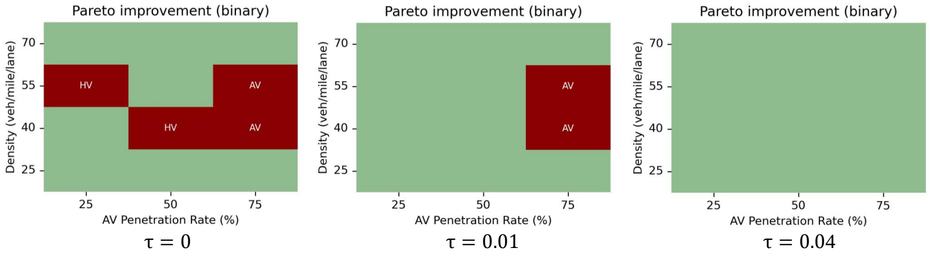
Compute resources All experiments were executed on a local Windows machine equipped with an Intel Core i7-12700 (12th Gen) CPU and an NVIDIA GeForce RTX 3080 GPU. Each individual training run required approximately 24 hours of wall-clock time. Across the 12 traffic scenarios (with replications per scenario), the estimated total compute amounts to 288 GPU hours.

3 Findings

We examine the attainability of CR in mixed autonomy traffic with the trained AVs from two perspectives: (1) the attainability of Pareto-efficient NE; and (2) comparisons of simulated collective characteristics of agents with game-theoretic predictions.

3.1 Pareto-efficient equilibria

Dynamic simulation rarely reaches exact NE. Therefore, we adopt a set of criteria to relax the conditions and identify ϵ-NE from simulation data, where ϵ is a tolerance window. Figure 1 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. Pareto-efficient NE is considered to be achieved across these scenarios.



$\tau = 0$ $\tau = 0.01$ $\tau = 0.04$

Figure 1: 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.

3.2 Comparisons with game-theoretic baseline

This section compares various macroscopic traffic variables between the DRL evaluation output and the theoretical model's predictions based on Li et al. [2022].

3.2.1 Split factor of cooperation surplus

We define a loss function to compare the game-theoretic predicted speeds with the DRL simulated speeds as follows,

$$L(\rho_1, \rho_2, \lambda; \lambda \in [0, 1]) = \sum_{\rho_2=1}^{\rho_2^{max}} \sum_{\rho_1=1}^{\rho_1^{max}} \left[\sum_{i=1}^{2} w_i |\bar{u}_i(\rho_1, \rho_2) - \hat{u}_i(\rho_1, \rho_2; \lambda)| \right]^2 \tag{3}$$

where ρ_i is the density of class i agents, w_i is the weighting factor, and $\hat{u}_i(\rho_1, \rho_2; \lambda)$ is the speed predicted by the game theoretic model. Additionally, $\bar{u}_i(\rho_1, \rho_2)$ represents the average speed for class i for the density pair (ρ_1, ρ_2), calculated from the DRL evaluation data. Numerical value of split factor λ is calculated as minimizer of the loss (3),

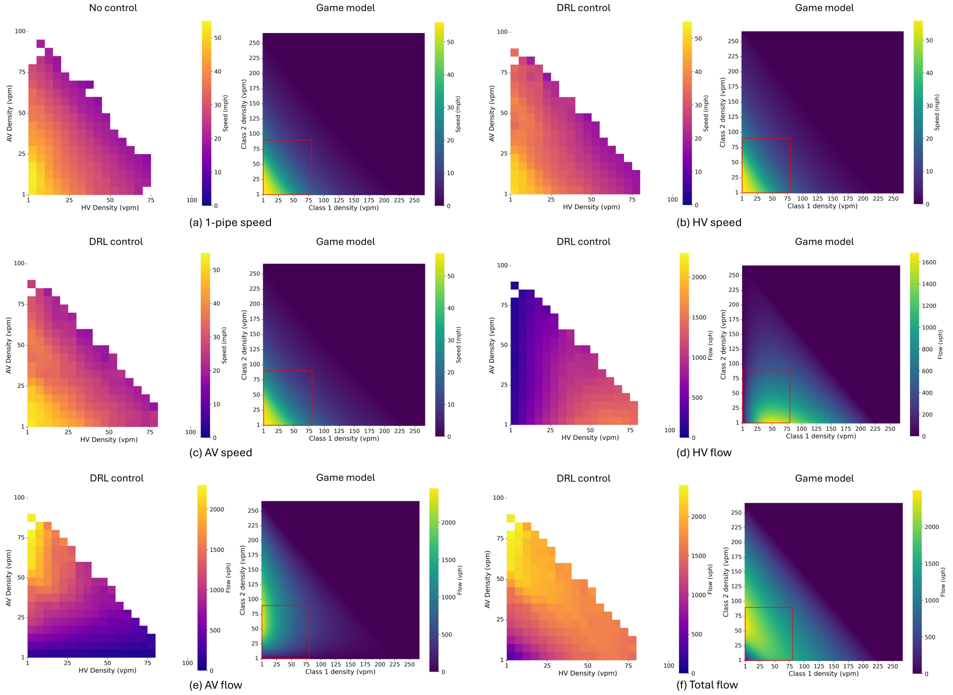
$$\lambda^* = \arg\min_{\lambda} L(\rho_1, \rho_2, \lambda; \lambda \in [0, 1]) \tag{4}$$

The estimation yields $\lambda^* = 0.6484$, with the mean absolute errors (MAE) for class 1, class 2, and the weighted average being 3.50 mph, 2.70 mph, and 3.10 mph, 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.

3.2.2 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 2, panel (a) compares the 1-pipe equilibrium speed (defined in Appendix A.1) from the fully mixed (no control) scenario with the one solved from (8). 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). 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.



(a) 1-pipe speed

(b) HV speed

(c) AV speed

(d) HV flow

(e) AV flow

(f) Total flow

Figure 2: 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.

3.2.3 Second-order comparison

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

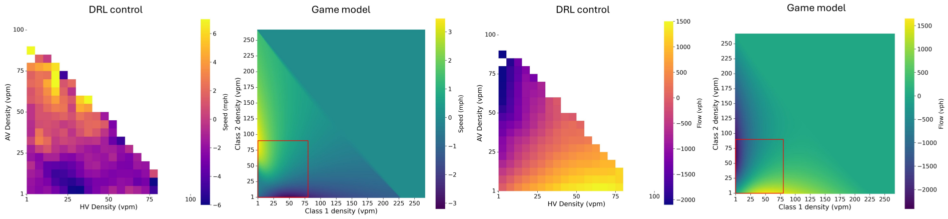
In Figure 3, we present the speed and flow differences between HVs and AVs, comparing these with the benchmark model's predictions. Overall, both speed and flow exhibit similar patterns to the benchmark model. However, the DRL model demonstrates greater heterogeneity. We seek to explain this phenomenon. Revisiting Figure 2(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,

$$\begin{cases} \hat{u}_1(\rho_1, \rho_2) = u_1(\rho_1, \rho_2) + \epsilon_1 \\ \hat{u}_2(\rho_1, \rho_2) = u_2(\rho_1, \rho_2) + \epsilon_2 \end{cases} \tag{5}$$

where ϵ_i is the error between the DRL simulation data and the game model's predictions. Denoting the variance of errors ϵ_1 and ϵ_2 as σ_1 and σ_2, respectively, the speed difference between HV and AV is computed as,

$$\Delta = \hat{u}_1 - \hat{u}_2 \tag{6}$$

For simplification, assuming $\hat{u}_1$ and $\hat{u}_2$ are independent, then the randomness for Δ becomes $\epsilon_1 + \epsilon_2$, with the variance of randomness being $\sigma_1 + \sigma_2$. This accumulation of randomness in Δ leads to discrepancies in Figure 3(a) compared to the benchmark model. In contrast, as illustrated in Figure 3(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.



(a) Speed difference between HV and AV

(b) Flow difference between HV and AV

Figure 3: Second-order comparisons of speed and flow differences between HVs and AVs.

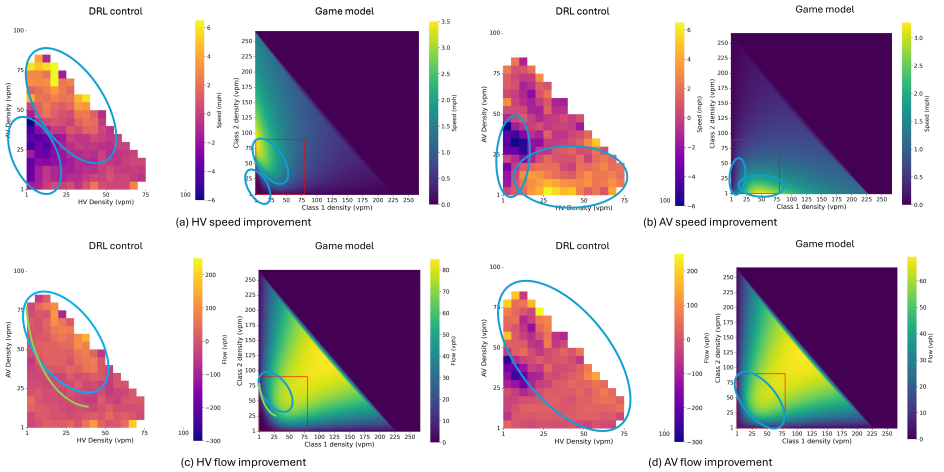
In Figure 4, we compare the improvements in speed and flow relative to the fully mixed scenario (i.e., no CR). We acknowledge the observation of some differences, due to the inherent heterogeneity in simulation. Nonetheless, notable similarities are also present, highlighted by blue circles indicating regions of similarity and the green curve marking the boundary of positive and non-positive value transitions. For instance, in Figure 4(a), both the DRL-attained NE and the theoretical prediction show positive speed improvement at higher AV densities. Similarly, in Figure 4(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 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.

3.2.4 Fundamental diagram comparison

The comparisons of fundamental diagrams are presented in Figure 5, 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 5. 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.

4 Conclusion and Future Work

This study examines the emergence of collective rationality in mixed autonomous driving systems, through the lens of collective rationality (CR). Using DRL with a simple reward design, we demon-



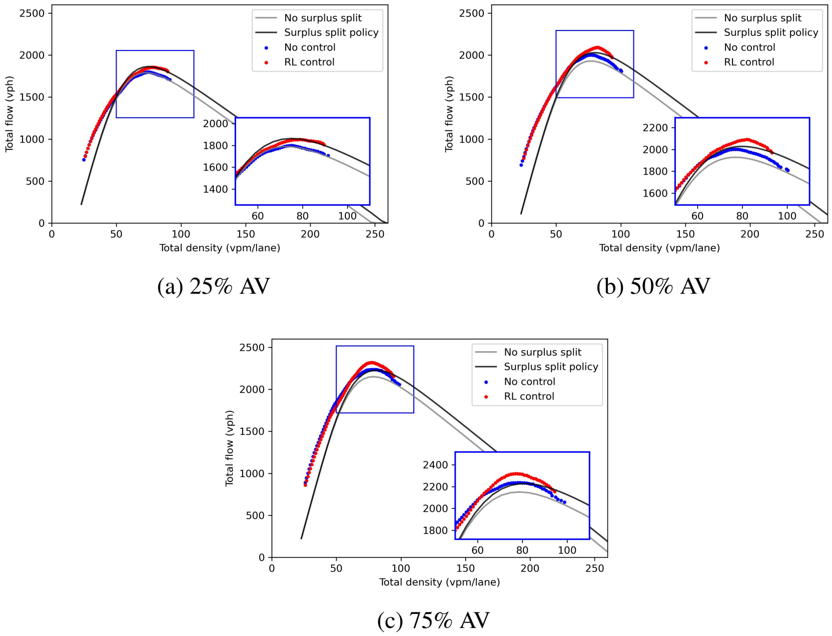
(a) HV speed improvement

(b) AV speed improvement

(c) HV flow improvement

(d) AV flow improvement

Figure 4: Second-order comparisons of speed and flow improvements between HVs and AVs.



(a) 25% AV

(b) 50% AV

(c) 75% AV

Figure 5: Comparison of fundamental diagrams between the DRL model and game theoretical model under different AV penetration rates.

strate that CR can emerge in microscopic and dynamic mixed autonomy environment. Comparisons between the DRL-based results and game-theoretic predictions show the first-order alignment between the two, and richer second-order outcomes from DRL-driven learning processes. The analysis of fundamental diagrams shows a maximum capacity improvement of 130 vph/lane under 50% AVs when collective rationality is attained. The surplus split factor estimation reveals HVs gain 64.84% of benefits from collective rationality, with AVs taking 35.16%.

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 promising 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 relax this assumption in the future.

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

Omri Ben-Dov, Jake Fawkes, Samira Samadi, and Amartya Sanyal. The role of learning algorithms in collective action. *arXiv preprint arXiv:2405.06582*, 2024.

Di Chen, Jia Li, and Michael Zhang. Estimate collective cooperativeness of driving agents in mixed traffic flow. Manuscript submitted to Transportation Research Part B: Methodological (in press), https://arxiv.org/abs/2408.07297, 2025.

Robert de Neufville and Seth D Baum. Collective action on artificial intelligence: A primer and review. *Technology in Society*, 66:101649, 2021.

Ernst Fehr and Simon Gächter. Altruistic punishment in humans. *Nature*, 415(6868):137–140, 2002.

Moritz Hardt, Eric Mazumdar, Celestine Mendler-Dünner, and Tijana Zrnic. Algorithmic collective action in machine learning. In *International Conference on Machine Learning*, pages 12570–12586. PMLR, 2023.

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.

Martin Treiber and Arne Kesting. Traffic flow dynamics. *Traffic Flow Dynamics: Data, Models and Simulation, Springer-Verlag Berlin Heidelberg*, pages 983–1000, 2013.

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.

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.

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.

A Concept of Collective Rationality (CR)

In this section, we briefly introduce 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.

A.1 Collective rationality in equilibrium traffic model

Li et al. [2022] 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,

$$U_i(\rho_1, \rho_2, p_1, p_2) = \begin{cases} u^*(\rho_1, \rho_2) & \text{if } p_1^* + p_2^* > 1, \\ u_i(\rho_i / p_i) & \text{if } p_1^* + p_2^* \leq 1 \end{cases} \quad i = 1, 2 \tag{7}$$

where ρ_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,

$$\frac{1}{\rho_{tot}} \left(\frac{\rho_1}{u_1^{-1}(u^*)} \sum_{j=1}^{2} \frac{\rho_j}{a_j} + \frac{\rho_2}{u_2^{-1}(u^*)} \sum_{j=1}^{2} \frac{\rho_j}{b_j} \right) = 1 \tag{8}$$

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 (8), where other parameters and functions can be assumed or calibrated from data [Chen et al., 2025].

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,

$$\begin{cases} p_1 = p_1^* + \lambda(\rho_1, \rho_2)s \\ p_2 = p_2^* + (1 - \lambda(\rho_1, \rho_2))s \end{cases} \tag{9}$$

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].