

# Learning Generalizable Multi-Lane Mixed-Autonomy Behaviors in Single Lane Representations of Traffic

Extended Abstract

Abdul Rahman Kreidieh  
University of California, Berkeley  
aboudy@berkeley.edu

Samyak Parajuli  
University of California, Berkeley  
samyak.parajuli@berkeley.edu

Yibo Zhao  
University of California, Berkeley  
brentzhao@berkeley.edu

Alexandre M. Bayen  
University of California, Berkeley  
bayen@berkeley.edu

## ABSTRACT

This paper tackles the problem of learning generalizable congestion-mitigation strategies in simple representations of traffic. In particular, we look to mixed-autonomy ring roads as depictions of instabilities common to many generic settings, and ask the question: What features are needed to ensure that policies here can be adapted to typical multi-lane highways? To answer this, we study the implications of the scale of the source task and the modeling of pseudo-lane change events within it on the transferability of policies learned to complex networks. Our findings suggest that negating the effects of boundary conditions and introducing lane changes that approximately match trends in more complex systems can significantly improve the generalizability of learned behaviors.

## KEYWORDS

Reinforcement Learning; Social Simulation; Traffic Control

### ACM Reference Format:

Abdul Rahman Kreidieh, Yibo Zhao, Samyak Parajuli, and Alexandre M. Bayen. 2022. Learning Generalizable Multi-Lane Mixed-Autonomy Behaviors in Single Lane Representations of Traffic: Extended Abstract. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022)*, Online, May 9–13, 2022, IFAAMAS, 3 pages.

## 1 INTRODUCTION

Reinforcement learning techniques can provide substantial insights into the desired behaviors of future autonomous driving systems. By solving for societal metrics of traffic such as increased throughput and reduced fuel consumption, such methods can derive maneuvers that may significantly improve the quality of traffic. These methods, however, are hindered in practice by the difficulty of designing accurate large-scale models of traffic, as well as the challenges associated with learning behaviors for dozens of interacting agents.

In this paper, we demonstrate that policies learned in simple and computationally efficient ring road settings, if properly defined, can effectively generalize to more complex problems. To enable proper generalization, we identify two limiting factors that exacerbate the dynamical mismatch between the two problems and devise methods for addressing each of them. For one, to address mismatches that arise from variations in the boundary conditions, we construct

a curriculum learning paradigm that scales the performance of policies learned to larger rings, where boundaries pose less of a concern. Next, to address perturbations that arise from lane changes and cut-ins, we introduce a simple approach for simulating such disturbances in single-lane settings. We validate our method on a calibrated model of the I-210. Our findings suggest that learning in larger rings and introducing pseudo-lane changes can greatly improve the generalizability of resultant behaviors.

## 2 EXPERIMENTAL SETUP

We define the mixed-autonomy ring road problem as an extension of the standard model depicted in [2] in which the actions of decentralized agents dictate the desired accelerations of individual AVs. Let  $\mathcal{S}_{AV} \subset \{1, \dots, n\}$  be the set of AVs whose actions are dictated by a learning agent,  $\pi_\theta$  with shared parameters  $\theta$ . The system of differential equations depicting the dynamics of the network is:

$$\begin{cases} \dot{x}_i(t) = f(\dot{x}_i(t), \dot{x}_{i+1}(t), \Delta x_i(t)) & i = 1, \dots, n-1 \mid i \notin \mathcal{S}_{AV} \\ \dot{x}_i(t) = \pi_\theta(s_i(t)) & i \in \mathcal{S}_{AV} \\ \dot{x}_n(t) = f(\dot{x}_n(t), \dot{x}_1(t), \Delta x_n(t)) \end{cases} \quad (1)$$

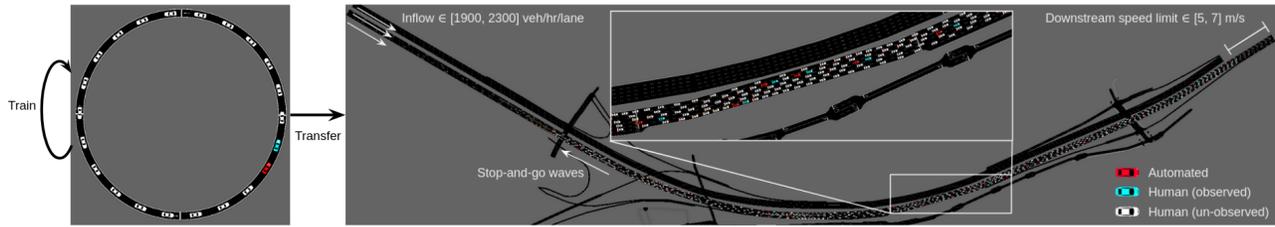
where  $x_i(t)$  is the position of vehicle  $i$  at time  $t$ ,  $f$  is a car-following model<sup>1</sup>, and  $s_i(t)$  is the state of AV  $i$  at time  $t$ , defined as the ego speed, leader speed, and space headway across multiple timesteps. Within this task, agents are incentivized to solve for socially optimal driving behaviors by rewarding speeds near the network's uniform-flow equilibrium [7]. We expand on this in our main paper.

### 2.1 Limitations to Generalizability

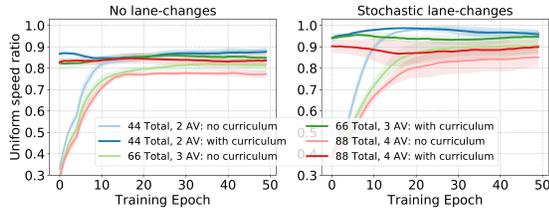
Learning traffic regulation policies within simulated ring roads presents several notable benefits in the context of RL. For one, the simplicity of the dynamics renders the problem easy to reconstruct and computationally efficient to simulate. In addition, notions of stability and social optimality render the definition of the reward function relatively straightforward. The question, nevertheless, remains: Are policies learned within these tasks usable in more complex settings? To study this, we identify and address two features deemed key in improving the performance of learned policies.

**2.1.1 Boundary conditions.** The first of these challenges considers the effect of periodic boundary conditions unique to closed (circular) networks. These boundaries produce strong couplings between

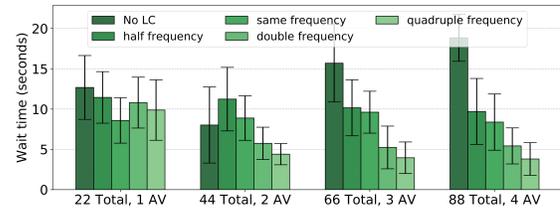
<sup>1</sup>We consider the Intelligent Driver Model [5], a popular model for simulating string-instabilities in human driving, and set the parameters to match the work of [1].



**Figure 1: Explored problem and transfer learning procedure. This paper aims to learn policies in ring road settings that are transferable to more complex tasks. The target network is a section of I-210 in which outflow restrictions result in the formation of stop-and-go congestion.**



**Figure 2: Learning curves for different methods/environments.**



**Figure 3: Delays incurred by different policies on the I-210.**

the actions of an individual vehicle and the response of *all* other vehicles. This is contrary to the unidirectional nature of traffic in open networks [6], and likely results in the formation of undesirable behaviors. Larger diameter rings with additional vehicles, as noted in [3], can help dilute the effects of these boundaries. Learning in these settings, however, poses a challenge to multiagent RL. To assist policies learn in larger tasks, we design a curriculum learning paradigm that exploits the extendability of mixed-autonomy rings. Within this paradigm, policies are originally trained in rings of length  $L$  with 1 AV. This solution fails in much larger rings, but serves as a proper initialization in rings of length  $2L$  with 2 AVs. As such, we retrain a policy in the 2 AV setting using the prior warm start, and continue to repeat this process until a proper initialization may be provided to an  $n$ -AV problem.

**2.1.2 Perturbations by adjacent lanes.** In addition to the effects of boundaries, perturbations induced by lane changes also present a significant challenge. These perturbations, which include sudden reductions in gaps from overtaking actions and other fluctuations in network densities, expose the learned policy to unexpected disturbances which may destabilize the system if not addressed. To resolve this concern, we attempt to model the effects of lane changes from the perspective of individual lanes. To do so, we take inspiration from the work of [8] and model lane changes as stochastic insertions and deletions of vehicles within the ring. These behaviors are modeled using a uniform distribution for the entry and exit probabilities,  $p_{\text{enter}}$  and  $p_{\text{exit}}$  respectively, with entry vehicles being initially placed at the midpoint between two vehicles and driving at the average speed of the network. This represents the simplest approach for modeling such effects in single-lane traffic. In the context of transfer learning, this closest to domain randomization [4].

### 3 NUMERICAL RESULTS

In this section, we present numerical results for the proposed training procedures. Through these results, we aim to answer the following: 1) Does the curriculum learning procedure improve the

scalability of learning algorithms? 2) What effect do larger rings and simulated lane changes have on the transferability of policies?

**Ring Experiments** Figure 2 depicts the learning performance on the ring road task with and without the use of lane changes and curricula, evaluated the policy’s ability to achieve uniform-flow speeds. As we can see, while early stage problems perform well without curricula, policies struggle to achieve a similar degrees of traffic stabilization as the size and complexity of the problem increases. Conversely, when trained via a curriculum of gradually growing rings, policies continue to achieve scores near to the ideal value of 1.

**Realistic Network Experiments** Next, we study the generalizability of the learned policy by evaluating its zero-shot transfer to an calibrated model of human-driving. The network considered (Figure 1) is a 1-mile section of the I-210 in Los Angeles, CA. We here exploit the work of [1], which explores the role of mixed-autonomy systems in improving the energy-efficiency of the I-210.

Figure 3 depicts the performance of the different learned policies once transferred onto the I-210, evaluated on the policy’s ability to minimize stopped traffic. To evaluate the effect of chosen lane change distributions, we assume values for  $p_{\text{enter}}$  and  $p_{\text{exit}}$  that are multiples of the true values experienced within the I-210. As we can see, introducing lane changes reduces in the duration of stopped traffic in the target network. This is also the case as the size of the problem grows, until the boundary stops plays a significant role. It must be noted however, that lane changes, while reducing stopped traffic, also reduce the speed of traffic once severely off-distribution. You can learn more about this in our full paper.

### ACKNOWLEDGMENT

This material is based upon work supported by the U.S. Department of Energy’s Office of Energy Efficiency and Renewable Energy (EERE) award number CID DE-EE0008872. The views expressed herein do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

**REFERENCES**

- [1] Jonathan W Lee, George Gunter, Rabie Ramadan, Sulaiman Almatrudi, Paige Arnold, John Aquino, William Barbour, Rahul Bhadani, Joy Carpio, Fang-Chieh Chou, et al. 2021. Integrated Framework of Vehicle Dynamics, Instabilities, Energy Models, and Sparse Flow Smoothing Controllers. In *Proceedings of the Workshop on Data-Driven and Intelligent Cyber-Physical Systems*. 41–47.
- [2] Gábor Orosz and Gábor Stépán. 2006. Subcritical Hopf bifurcations in a car-following model with reaction-time delay. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 462, 2073 (2006), 2643–2670.
- [3] Gábor Orosz, R Eddie Wilson, Róbert Szalai, and Gábor Stépán. 2009. Exciting traffic jams: Nonlinear phenomena behind traffic jam formation on highways. *Physical review E* 80, 4 (2009), 046205.
- [4] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. 2017. Domain randomization for transferring deep neural networks from simulation to the real world. In *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 23–30.
- [5] Martin Treiber, Ansgar Hennecke, and Dirk Helbing. 2000. Congested traffic states in empirical observations and microscopic simulations. *Physical review E* 62, 2 (2000), 1805.
- [6] Martin Treiber and Arne Kesting. 2011. Evidence of convective instability in congested traffic flow: A systematic empirical and theoretical investigation. *Procedia-Social and Behavioral Sciences* 17 (2011), 683–701.
- [7] Cathy Wu, Aboudy Kreidieh, Eugene Vinitsky, and Alexandre M Bayen. 2017. Emergent behaviors in mixed-autonomy traffic. In *Conference on Robot Learning*. PMLR, 398–407.
- [8] Cathy Wu, Eugene Vinitsky, Aboudy Kreidieh, and Alexandre Bayen. 2017. Multi-lane reduction: A stochastic single-lane model for lane changing. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 1–8.