Impact of RL-enabled autonomous vehicle route choice behaviors on urban traffic dynamics

Ahmet Onur Akman*, Anastasia Psarou, Zoltán György Varga, Grzegorz Jamróz, Rafał Kucharski

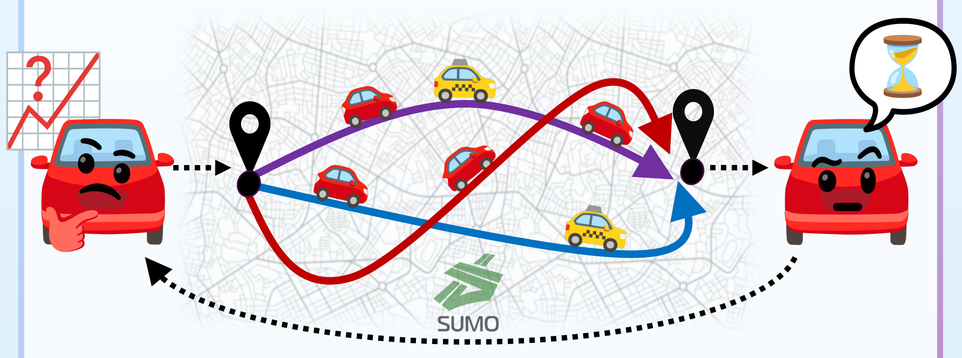
Faculty of Mathematics and Computer Science, Jagiellonian University, Kraków, Poland

RL-enabled autonomous vehicles can optimize behavioral objectives by learning to choose better routes!

...and this may have a noticeable impact on human drivers and traffic efficiency.

Introduction

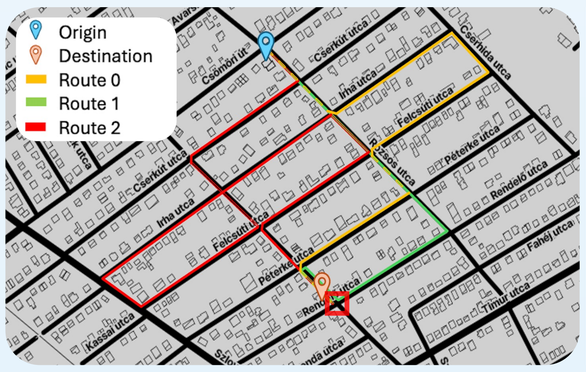
- We investigate the coexistence of **human drivers** and **RL-enabled autonomous vehicles** (AVs) in a **day-to-day route choice** scenario.
- Every day, drivers **pick one of the routes** connecting their origin-destination (OD) pairs. Day by day, they refine their route preferences according to **experienced travel times**.

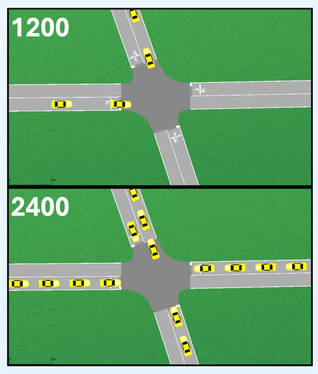


- Each AV agent uses individual DQNs and learns from individual experiences. Human agents are modeled with a state-of-the-art behavioral model.
- We explore the scenarios in which AVs adopt different **behavioral objectives**.
- Our experiments are conducted using our MARL framework: **PARCOUR**.

Problem

- Each episode simulates the commute of **1000 drivers** at a **rush hour** in traffic.
- There are two origins, two destinations, and **three predefined routes** for each of the **four ODs**.
- An episode consists of a single-step decision of each agent, which is their route choice for the day.
- **Humans seek to minimize individual travel times.** AVs aim to optimize their returns.





Left: Csömör traffic network used in our experiments, with routes connecting OD (0,0).

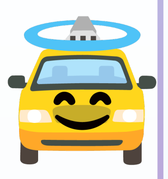
Right: The congestion at the highlighted intersection on indicated timesteps within an episode.

AV Behaviors

- We define six behaviors for the AVs with different reward formulations:

SELFISH
Minimize self travel time.

ALTRUISTIC
Minimize everyone's travel time.



SOCIAL
Minimize self and everyone's travel time.

COLLABORATIVE
Minimize the AV fleet's travel time.

COMPETITIVE
Minimize own, maximize human travel time.

MALICIOUS
Maximize human travel time.

- The reward function is a linear combination of the mean travel times of different subsets of agents, **weighted by the behavioral coefficients**.
- In each experiment, all AVs **uniformly** adopt a selected behavior.

Experimental Setting

Phase I: Settle

There are **only human drivers** in the traffic. They tune their cost expectations to make better decisions, until the system stabilizes.

Phase II: Shock

AVs replace 1/3 of the vehicles. Now AVs optimize their policies, while the remaining **humans stick to their learned preferences**.

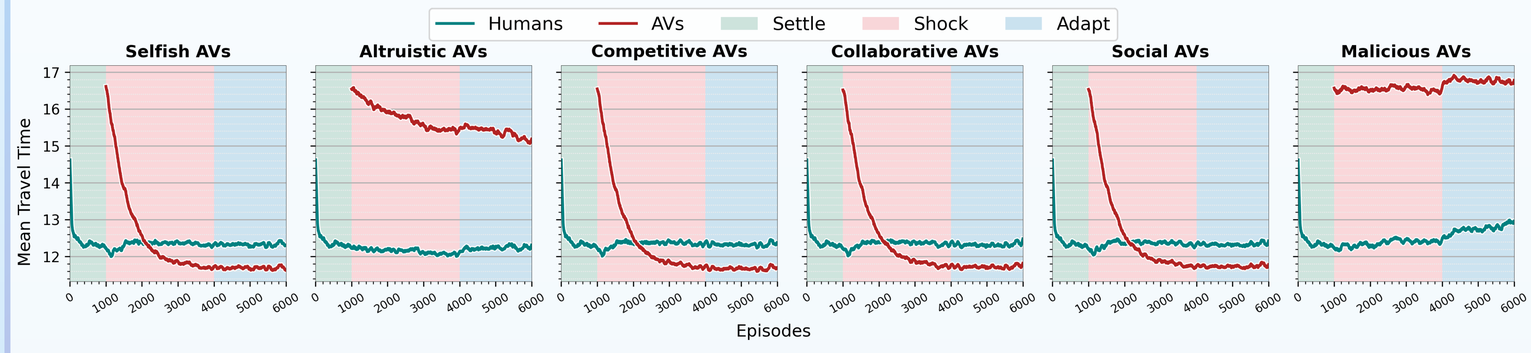
Phase III: Adapt

Humans are now reacting to the changes, by updating their preferences. This creates a shift for the AVs, and they **adjust their policies**.

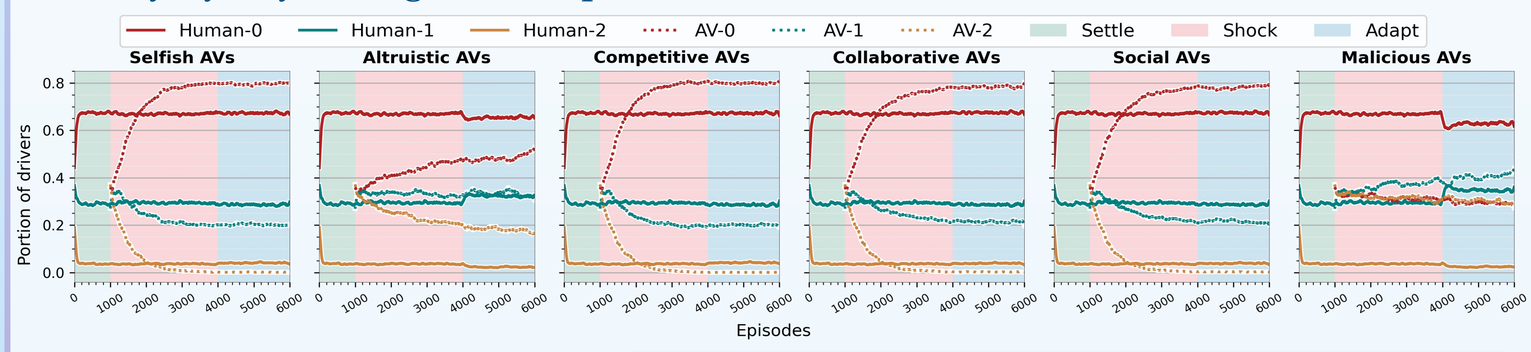
Days (Episodes)

Results

1) Changes in human and AV travel times over the episodes for each AV behavior.



2) Day by day, changes in the preferences of human drivers and AVs. (OD 0, 0)



3) Consequences of AV deployment in each scenario. We compare the **mean travel times** of **humans at the end of Phase 1** and **the indicated group at the end of Phase 3**. A positive effect indicates reduced travel time.

Behavior \ Effect on	AV Travel Time	Human Travel Time	Traffic Efficiency
Altruistic	-23.1%	+0.3%	-7.1%
Collaborative	+4.3%	-0.7%	+0.9%
Competitive	+4.6%	-0.8%	+0.9%
Malicious	-36.3%	-5.4%	-15.1%
Selfish	+4.6%	-0.7%	+1.0%
Social	+4.2%	-0.7%	+0.9%

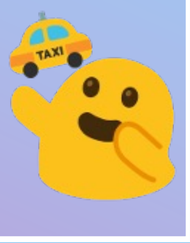
Takeaways

AV users enjoyed shorter commutes with each self-travel-time optimizing AV behavior.

In most cases, the AV deployment improved the traffic efficiency.

Malicious and altruistic behaviors caused great delays for AV users.

Whenever AVs aimed to reduce their travel times, switching was preferable to manual driving.



This research is financed by the European Union within the Horizon Europe Framework Programme, ERC Starting Grant COeXISTENCE no. 101075838.

EXPLORE COeXISTENCE