# Learning Bidding Strategies for Karma Economies in Realistic Traffic Settings with Multi-Agent Reinforcement Learning

**author names withheld**

**Under Review for NExT-Game 2026**

## Abstract

Karma is a non-monetary resource-allocation mechanism that prioritizes users' needs rather than their financial power. Monetary pricing can effectively reduce congestion by imposing charges on specific road segments, but it may be unfair by favoring higher-income individuals. Prior work has shown that in this context, Karma can achieve similar efficiency while yielding fairer outcomes; however, it has been demonstrated only in a deterministic setting. Demonstrating Karma's applicability under more realistic traffic conditions is therefore important for real-world implementation. Additionally, experimental evidence suggests that humans may struggle to execute optimal bidding strategies in Karma economies. In this paper, we demonstrate the use of Multi-Agent Reinforcement Learning (MARL) to train automated bidding agents for travelers. In a microscopic traffic simulation case study, we show that MARL agents learn effective bidding strategies that yield fairer travel outcomes for drivers than those achieved under monetary pricing schemes.

**Keywords:** Karma, fairness, monetary segment pricing, multi-agent systems, multi-agent reinforcement learning, traffic simulation.

## 1. Introduction

Traffic congestion is a major issue in urban road networks, resulting in substantial travel time losses. To reduce congestion and incentivize travelers to choose alternative routes cities often price specific road segments [6]. While such a mechanism can improve traffic efficiency, it may raise significant equity concerns in societies with uneven income and wealth distributions, as it tends to favor those who can afford to pay while disadvantaging others, thereby exacerbating social inequality. As a result, the implementation of monetary segment pricing is often impeded by public opposition and lower user acceptance [9].

Karma is a resource allocation mechanism that uses an artificial, non-tradable currency and allocates resources based on users' needs rather than their financial means. Originally introduced in the context of file-sharing networks, Karma has since been studied in a wide range of domains [14]. In transport pricing, prior work has demonstrated that Karma can serve as a fairer alternative to monetary pricing [13]. However, this work assumed deterministic travel times and decision-making at the beginning of the day, whereas real-world travel conditions require more complex constraints, such as variations in departure times [2] and stochastic travel-time variability [4].

In this work, we model daily commuting under non-deterministic, i.e., stochastic, traffic dynamics as a repeated congestion game [10], where individuals travel from their homes (origins) to their workplaces (destinations) at different times over many days. Our experiments are conducted leveraging a complex, microscopic traffic simulation tool (SUMO [12]) that captures the real-world

complexity of road transportation. We consider a setting with 300 vehicles and a network in which the fastest route between the origin and destination has limited capacity and becomes congested when about one-third of the vehicles use it. Within this urban framework, our aim is to evaluate the implementation of monetary and Karma pricing mechanisms, assessing their relative efficacy in achieving system-optimal travel times and their respective impacts on fairness in the population travel times.

To find optimal solutions in Karma games, prior works have used optimization methods, including fixed-point computation, the momentum method [3], and evolutionary dynamics-inspired algorithms [13]. The choice of method is important because Karma games can involve large state spaces, and stiff dynamics and this is further amplified in our setting by stochastic traffic dynamics. In addition, learning optimal bidding strategies in Karma economies can be challenging for human users. Experimental evidence indicates that, although a population of human bidders achieves aggregate outcomes superior to random allocation, they still do not reach the theoretical Nash equilibrium solution [8]. This suggests that users may need AI-aided bidding agents to make decisions on their behalf, much as navigation applications help travelers select routes. These considerations motivate the use of MARL for learning bidding strategies in repeated, stochastic environments where multiple agents compete for limited resources. MARL has already been applied to study vehicles' routing decisions [1, 17] and to learn bidding strategies in auction settings [11].

In this work, MARL agents represent simulated commuters who repeatedly travel in a stochastic traffic environment. Under the Karma mechanism, agents learn a policy that maps their observed state to a bid. In a city employing a Karma scheme, the learned policy can be interpreted as a decision rule that companies could deploy, for example, through a navigation app, to suggest bids to real commuters.

The main research contributions of this work are as follows.

- We demonstrate that multi-agent reinforcement learning (MARL) can effectively be used to learn efficient bidding strategies in Karma economies for road segment pricing.
- We show that MARL agents trained under both Karma and monetary road-pricing schemes equally well mitigate traffic congestion.
- We show that Karma pricing yields fairer outcomes than monetary pricing across four measures that reflect different notions of fairness.

## 2. Methodology

We consider a sequential decision-making process in which individuals commute once per day from home to work in a city operating under either a monetary or a Karma pricing scheme, making decisions in order of their departure times. To create the MARL policy, we specify the RL environment and the agents' actions, rewards, and observations. We use the terms agents, individuals, and commuters interchangeably for the simulated commuter entities. Each agent is assigned an income value, a day-specific urgency, and departure time.

At their own departure time, agents receive an observation from the environment and select an action. They then travel through the SUMO traffic network, and their experienced travel times are recorded, which are part of their reward (cost) that their policy aims to maximize. Our RL environment is based on the RouteRL framework [1], which models vehicle routing decisions using

MARL. We extend it by modifying agents' reward functions and introducing bidding and auction mechanisms.

In the Karma setting, route allocation is determined through a Stackelberg auction, where the city acts as the leader by setting system-optimal (total travel time minimizing) route prices [13], and travelers act as followers by submitting bids in response [18]. This mechanism, unlike first- or second-price auctions, allows individuals to bid just before departure rather than requiring all bids to be submitted at the beginning of the day. Therefore, an agent is assigned to the fastest route if their bid exceeds the centrally defined price for that route. Otherwise, the assignment proceeds to the next-best route by comparing the submitted bid with the corresponding threshold, and this process continues iteratively until a route is allocated. The minimum bids (prices) to get access to the preferred routes are assumed to remain fixed over time.

In both pricing schemes, agents base their decisions on the observed travel times over the previous $\mu$ days, along with their income, departure time, and urgency. Under Karma-based pricing, agents also observe their current Karma balance. They then submit bids for each route $k$ subject to their available Karma, with all agents starting on day one with the same initial endowment of Karma points. Once an agent is assigned to route $k$, the bid submitted for that route is deducted from their Karma balance and, at the end of the day, equally redistributed to all the users in the city.

Before the next round (day), agents evaluate outcomes of their choices by observing their rewards, representing their total experienced costs:

$$r_{i,d}(k) = -c(k) - \beta_i \times u_{i,d} \times \tilde{t}_{i,d}(k), \tag{1}$$

comprising their monetary costs $c(k)$, and the realized (experienced) travel time $\tilde{t}_{i,d}(k)$. Travel time is weighted with its value, being a product of their current urgency $u_{i,d}$ and the income $\beta_i$. In both the monetary and Karma pricing scenarios, agents aim to maximize their reward. Under the monetary pricing scheme, $c(k)$ is fixed, and $\beta_i$ reflects the income distribution within the population. Under Karma pricing scheme, $c(k)$ is always zero, and $\beta_i$ is equal to one for all agents in the reward. Therefore in the monetary pricing wealthy agents may buy priority with their high monetary budgets, whereas in Karma agents have equal power to supply their urgency, independent of monetary budgets.

**MARL.** We train our agents using the Independent Proximal Policy Optimization (IPPO) algorithm, which has shown strong benchmark performance in a variety of tasks [7]. Since our setting involves 300 agents, scalability is important. Parameter sharing, where agents update a single shared neural network, can improve scalability [5] but can hinder convergence in complex environments and limit the ability to capture heterogeneity across agents. We therefore use hypernetworks, which generate agent-specific weights while decoupling observation-driven and agent-conditioned gradients, and train agents with IPPO with hypernetworks [16].

**Fairness.** We compare and evaluate the monetary, Karma pricing, SO, and UE solutions using several notions of quantitative fairness [15]. These fairness measures are computed with respect to a quantity $X$, which may correspond to different formulations of an agent's travel time, or reward. Specifically:

- **Harsanyian fairness**, evaluates fairness by the average welfare across individuals. It is given by:

$$F_{\text{hars}} = \frac{1}{R}\sum_{r=1}^{R}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{K^{(r)}}\sum_{d=1}^{K^{(r)}}\left(X_{i,d}^{(r)}\right)\right)\right], \tag{2}$$

where $R$ is the number of replications, $N$ is the number of agents, $K^{(r)}$ is the number of days in replication $r$, $X_{i,d}^{(r)}$ is the travel time or reward of an agent $i$ on a day (training iteration) $d$.

- **Utilitarian fairness** favors decisions that maximize aggregate welfare at the population level, even if some individuals are disadvantaged. It is given by:

$$F_{\text{util}} = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{1}{K^{(r)}}\sum_{d=1}^{K^{(r)}}\sum_{i=1}^{N}X_{i,d}^{(r)}\right). \tag{3}$$

- **Rawlsian fairness**, evaluates fairness by the welfare of the least advantaged individual. In our setting, we quantify it using the maximum delay experienced in the system, given by:

$$F_{\text{rawls}} = \frac{1}{R}\sum_{r=1}^{R}\max_{i\in\{1,\ldots,N\}}\left[\max_{d\in\{1,\ldots,K^{(r)}\}}\left(X_{i,d}^{(r)}\right)\right]. \tag{4}$$

- **Egalitarian fairness**, emphasizes equality of outcomes across individuals. We evaluate it with the standard deviation of agents' average travel times and the Gini coefficient of travel times below:

$$F_{\text{egal(std)}} = \frac{1}{R}\sum_{r=1}^{R}\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{X}_i^{(r)} - \bar{X}^{(r)}\right)^2}, \tag{5}$$

where $\bar{X}_i^{(r)}$ is the average travel time of an agent $i$ among K days and $\bar{X}^{(r)}$ is the overall average travel time across agents in replication $r$.

$$F_{\text{egal(gini)}} = \frac{1}{R}\sum_{r=1}^{R}\frac{\sum_{i=1}^{N}\sum_{b=1}^{N}\left|\bar{X}_i^{(r)} - \bar{X}_b^{(r)}\right|}{2N^2\left(\frac{1}{N}\sum_{i=1}^{N}\bar{X}_i^{(r)}\right)}. \tag{6}$$

**Experimental setup.** The traffic network used to study this problem is depicted in Appendix 3. It has one origin-destination pair connected by three alternative routes, with only the fastest route (route 0) being priced. In addition, each experiment is repeated 8 times, with different random seed.

We model each MARL episode as lasting 30 days, after which all agents' Karma balances are reinitialized. This prevents excessive concentration of Karma points among a small subset of agents. It is also motivated by an analogy to a monthly budget cycle, in which individuals periodically reassess how they have used their available resources. Historic travel times are computed over the last 10 days ($\mu$ = 10) and defined for each agent as the average travel time on each route observed among individuals departing within a 20-simulation-step window before that agent. The income of

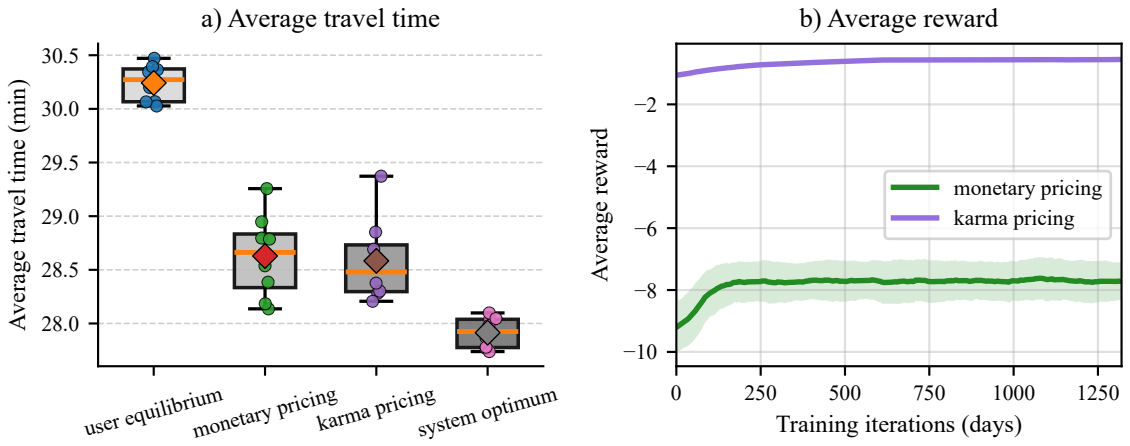


Figure 1: **Learning outcomes of congestion pricing.** a) Monetary and Karma pricing achieve similar average system travel times that are substantially lower than in the User Equilibrium (UE) case and close to the theoretical System Optimum (SO) solution. Circles denote individual replications, rhombus the mean travel time across replications, and orange lines the median travel time. b) During training, the average agent reward increases and then stabilizes, suggesting that the MARL agents converge to a stable solution.

the agents is drawn from the European income distribution reported in the World Inequality Report[1], and the urgency is drawn from a geometric distribution with parameter $p = 0.3$ following [13]. The population distributions of hourly salary, urgency, and value of time are depicted in Appendix B.

We consider two baseline scenarios: the approximated System-Optimal and User-Equilibrium solutions. In the latter, the agents choose the fastest route based on historical travel times.

## 3. Results and discussion

Both monetary and Karma pricing reduce congestion, achieving average system travel times close to the SO solution, shown in Figure 1a. In Figure 1b, the average system rewards increase during training and then stabilize, suggesting that the MARL algorithms converged to stable policies. Figure 2a shows that monetary pricing can be unfair, as it favors individuals with higher Value of Time (VoT), whereas under Karma, travel times remain nearly constant across VoT bins. In this figure, we fix departure times across days to reduce noise from daily demand variation. Figure 2b shows that agents with higher urgency experience lower average travel times under both pricing mechanisms.

Table 1 compares Karma and monetary pricing schemes across the fairness ideologies, discussed in Section 2. We evaluate fairness in different quantities, including agent travel time, travel time weighted by urgency, and travel time weighted by VoT. The Harsanyian and Rawlsian fairness measures are quantified using the agent's experienced average delay, defined as the difference between the agent's experienced travel time and the shortest historical travel time it has observed. The results show that, under the travel-time and urgency-weighted quantities, Karma yields lower fair-

1. Derived from https://wir2022.wid.world/methodology.

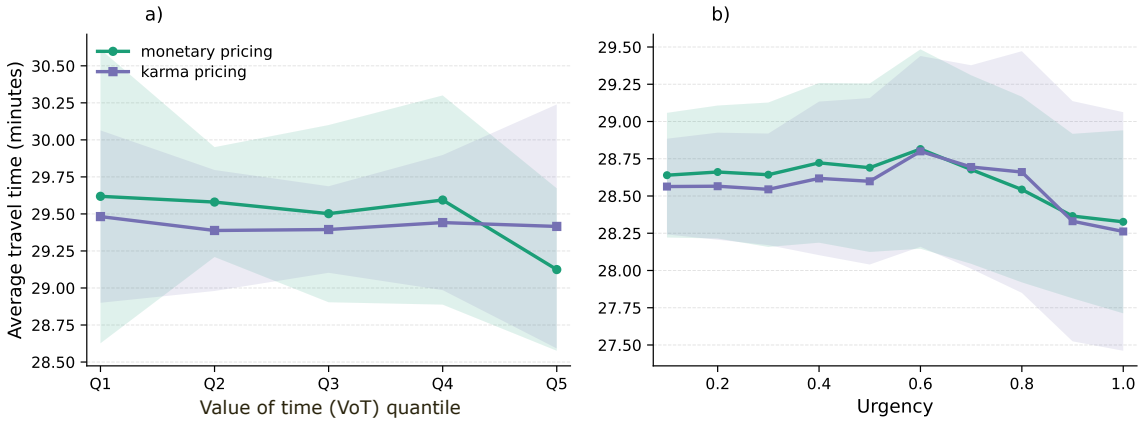


Figure 2: **Average travel times by urgency and Value-of-Time quintile.** (a) Under monetary pricing, average travel times decrease for travelers in the highest Value-of-Time quintile, whereas under Karma they remain nearly constant across quintiles. (b) Under both Karma and monetary pricing, average travel times decrease with urgency.

Table 1: Fairness measures across scenarios.

| | Fairness Measures | System optimum | Monetary Pricing | Karma Pricing | User equilibrium |
|---|---|---|---|---|---|
| Travel time | Egalitarian (std) | - | 1.23 (0.08) | **1.21 (0.04)** | 1.41 (0.06) |
| | Egalitarian (gini) | - | 0.024 (0.00) | **0.023 (0.00)** | 0.026 (0.00) |
| | Utilitarian | 8374 (36.02) | 8588.60 (116.54) | **8575.04 (116.80)** | 9072.49 (52.09) |
| Delay | Rawlsian | - | 39.44 (4.73) | **36.58 (2.63)** | 45.73 (3.40) |
| | Harsanyian | - | 25.28 (1.25) | **25.24 (1.24)** | 27.33 (1.41) |
| Travel time × Urgency | Egalitarian (std) | - | 0.98 (0.04) | **0.96 (0.06)** | 1.05 (0.95) |
| | Egalitarian (gini) | - | 0.06 (0.00) | **0.05 (0.00)** | 0.06 (0.00) |
| | Utilitarian | - | 2754.74 (139.62) | **2752.73 (163.51)** | 2922.44 (165.07) |
| | Rawlsian | - | 43.41 (1.55) | **43.07 (1.09)** | 47.73 (1.22) |
| | Harsanyian | - | 9.18 (1.08) | **9.17 (1.08)** | 9.74 (1.17) |
| Travel time × VoT | Egalitarian (std) | - | **116.63 (11.37)** | 117.73 (14.43) | 124.77 (14.16) |
| | Egalitarian (gini) | - | **0.44 (0.01)** | 0.44 (0.02) | 0.44 (0.02) |
| | Utilitarian | - | **37678.10 (2288.61)** | 37711.20 (2693.02) | 40022.7 (2620.85) |
| | Rawlsian | - | **3058.81 (158.03)** | 3232.51 (315.70) | 3679.62 (696.75) |
| | Harsanyian | - | **125.59 (117.16)** | 125.70 (118.63) | 133.41 (125.56) |

ness measures than monetary pricing, indicating fairer outcomes. Under the VoT-weighted quantity, monetary pricing yields higher fairness values than Karma, since VoT depends on agents' income. Overall, these findings indicate the potential of Karma as an efficient and fair congestion management mechanism.

## References

[1] Ahmet Onur Akman, Anastasia Psarou, Łukasz Gorczyca, Zoltán György Varga, Grzegorz Jamróz, and Rafał Kucharski. Routerl: Multi-agent reinforcement learning framework for urban route choice with autonomous vehicles. *SoftwareX*, 31:102279, 2025. ISSN 2352-7110. doi: 10.1016/j.softx.2025.102279.

[2] Chandra R. Bhat and Frank S. Koppelman. *Activity-Based Modeling of Travel Demand*, pages 35–61. Springer US, Boston, MA, 1999. ISBN 978-1-4615-5203-1. doi: 10.1007/978-1-4615-5203-1_3.

[3] Andrea Censi, Saverio Bolognani, Julian G. Zilly, Shima Sadat Mousavi, and Emilio Frazzoli. Today me, tomorrow thee: Efficient resource allocation in competitive settings using karma games. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 686–693, 2019. doi: 10.1109/ITSC.2019.8916911.

[4] Peng Chen, Rui Tong, Guangquan Lu, and Yunpeng Wang. Exploring travel time distribution and variability patterns using probe vehicle data: case study in beijing. *Journal of Advanced Transportation*, 2018(1), 2018. doi: 10.1155/2018/3747632.

[5] Filippos Christianos, Georgios Papoudakis, Arrasy Rahman, and Stefano V. Albrecht. Scaling multi-agent reinforcement learning with selective parameter sharing. In *International Conference on Machine Learning (ICML)*, 2021.

[6] André de Palma and Robin Lindsey. Traffic congestion pricing methodologies and technologies. *Transportation Research Part C: Emerging Technologies*, 19(6):1377–1399, 2011. ISSN 0968-090X. doi: 10.1016/j.trc.2011.02.010.

[7] Christian Schröder de Witt, Tarun Gupta, Denys Makoviichuk, Viktor Makoviychuk, Philip H. S. Torr, Mingfei Sun, and Shimon Whiteson. Is independent learning all you need in the starcraft multi-agent challenge? *CoRR*, abs/2011.09533, 2020.

[8] Ezzat Elokda, Saverio Bolognani, Florian Dörfler, and Heinrich H. Nax. Dynamic resource allocation with karma: An experimental study. *arXiv preprint arXiv:2404.02687*, 2026. doi: 10.48550/arXiv.2404.02687.

[9] Ziyuan Gu, Zhiyuan Liu, Qixiu Cheng, and Meead Saberi. Congestion pricing practices and public acceptance: A review of evidence. *Case Studies on Transport Policy*, 6(1):94–101, 2018. ISSN 2213-624X. doi: 10.1016/j.cstp.2018.01.004.

[10] Ron Holzman and Nissan Law-Yone. Strong equilibrium in congestion games. *Games and Economic Behavior*, 21(1):85–101, 1997. ISSN 0899-8256. doi: 10.1006/game.1997.0592.

[11] Yudong Hu, Congying Han, Tiande Guo, and Hao Xiao. Applying opponent modeling for automatic bidding in online repeated auctions. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '24, page 843–851, 2024. ISBN 9798400704864. doi: 10.5555/3635637.3662938.

[12] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wießner. Microscopic traffic simulation using SUMO. In *The 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE, 2018. doi: 10.1109/ITSC.2018.8569938.

[13] Kevin Riehl, Anastasios Kouvelas, and Michail A. Makridis. Karma economies for sustainable urban mobility – a fair approach to public good value pricing. *npj Sustainable Mobility and Transport*, 1:14, 2024. doi: 10.1038/s44333-024-00014-4.

[14] Kevin Riehl, Anastasios Kouvelas, and Michail A. Makridis. Resource allocation with karma mechanisms—a review. *Economies*, 12(8), 2024. ISSN 2227-7099. doi: 10.3390/economies12080211.

[15] Kevin Riehl, Anastasios Kouvelas, and Michail A. Makridis. Quantitative fairness—a framework for the design of equitable cybernetic societies. *Computers in Human Behavior: Artificial Humans*, 6:100236, 2025. ISSN 2949-8821. doi: 10.1016/j.chbah.2025.100236.

[16] Kale-ab Abebe Tessera, Arrasy Rahman, Amos Storkey, and Stefano V Albrecht. Hypermarl: Adaptive hypernetworks for multi-agent rl. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. doi: 10.48550/arXiv.2412.04233.

[17] Luiz A. Thomasini, Lucas N. Alegre, Gabriel O. Ramos, and Ana L. C. Bazzan. Route-choiceenv: a route choice library for multiagent reinforcement learning. In *Adaptive and Learning Agents Workshop at AAMAS*, 2023.

[18] Heinrich von Stackelberg. *Market Structure and Equilibrium*. Number 978-3-642-12586-7 in Springer Books. Springer, none edition, January 2011. doi: 10.1007/978-3-642-12586-7.

## Appendix A. Case study network

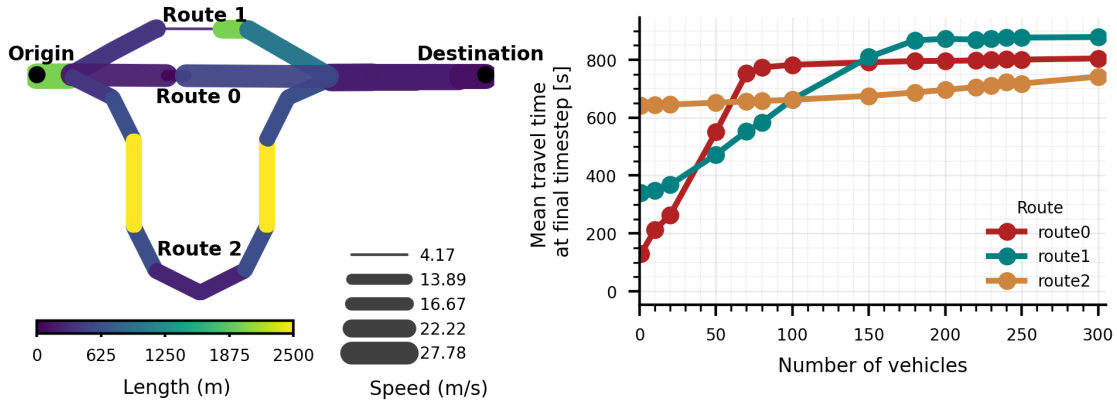


Figure 3: **Case study network.** Left: Traffic network. Right: Mean travel time at the final simulation timestep for different numbers of vehicles.

## Appendix B. Population-level distributions of hourly salary, urgency, and value of time

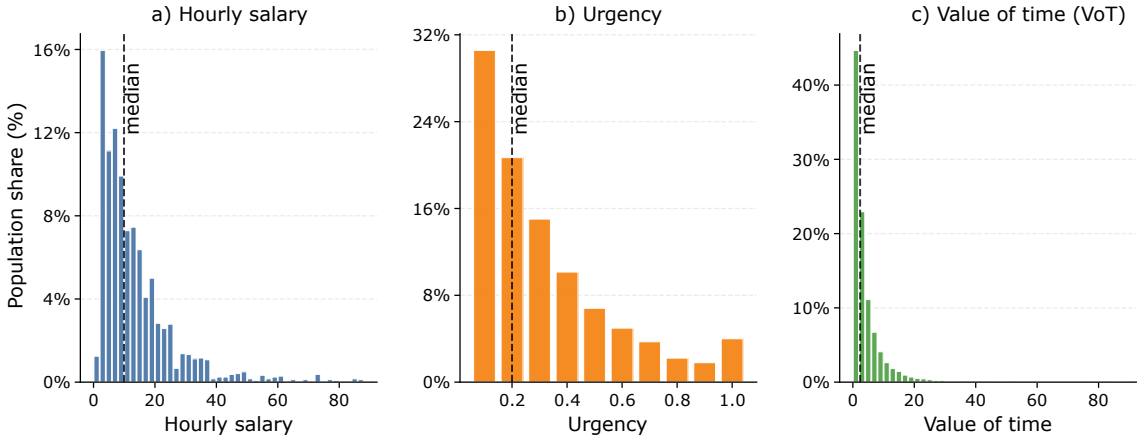


Figure 4: **Population-level distributions of hourly salary, urgency, and value of time.**