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

Anastasia Psarou*[1, 2], Kevin Riehl[3], Matej Jusup[4], Anastasios Kouvelas[3], Michail A. Makridis[3], and Rafał Kucharski[2]

[1]Doctoral School of Exact and Natural Sciences, Jagiellonian University, Krakow, Poland
[2]Jagiellonian University, Faculty of Mathematics and Informatics, Krakow, Poland
[3]Institute for Transportation Planning and Systems, Traffic Engineering Group, ETH Zurich, 8093 Zurich, Switzerland
[4]Independent Researcher, Zurich, Switzerland

Short summary

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, 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 [1].
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 (de Palma & Lindsey, 2011). 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 (Gu et al., 2018).

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 across a wide range of domains (Riehl et al., 2024b). In the context of transport pricing, prior work has demonstrated that Karma can serve as a fairer alternative to monetary pricing (Riehl et al., 2024a). 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 (Bhat & Koppelman, 1999) and stochastic travel-time variability (Chen et al., 2018).

Motivated by these limitations and the need to understand the potential of Karma as a segment-pricing mechanism for real-world implementation, we investigate whether the fairness gains of Karma congestion pricing hold under more realistic traffic conditions and whether users can be AI-assisted in the complex bidding process. We model daily commuting under non-deterministic, i.e., stochastic, traffic dynamics as a repeated congestion game (Holzman & Law-Yone, 1997), where individuals travel from their homes (origins) to their workplaces (destinations) at different times over many days. We experiment with a complex traffic scenario and implement it in SUMO (Lopez et al., 2018), a microscopic traffic simulator. We consider a setting with 300 vehicles and

[1]Code & Resources available on: https://github.com/AnastasiaPsarou/RouteRL-Karma.

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.
To find optimal solutions in Karma games, prior works have used optimization methods, including fixed-point computation, the momentum method (Censi et al., 2019), and evolutionary dynamics-inspired optimization algorithms (Riehl et al., 2024a). The choice of method is important because Karma games can involve large state spaces, 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 (Elokda et al., 2026). 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 (Akman et al., 2025; Thomasini et al., 2023) and to learn bidding strategies in auction settings (Hu et al., 2024).
In this work, MARL agents represent simulated commuters who repeatedly travel in a stochastic traffic environment. Under the Karma mechanism, each agent learns a policy that maps its 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 be used to learn effective bidding strategies in Karma economies for road segment pricing.
- We show that MARL agents trained under the Karma pricing scheme perform comparably to those trained under monetary pricing in reducing 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, travelers, 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 to maximize a reward signal. They then travel through the SUMO traffic network, and their experienced travel times are recorded, which are part of the reward that the policy aims to maximize. Our RL environment is based on the RouteRL framework (Akman et al., 2025), 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 route prices, and travelers act as followers by submitting bids in response (von Stackelberg, 2011). 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. Under the Stackelberg mechanism, 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. For simplicity, we assume that the minimum price to get allocated to a resource is the same across routes and remains fixed over time. When the monetary fee or the minimum bid to get assigned to a route is properly tuned, the system travel times can approach the system-optimal solution (Riehl et al., 2024a).
Before making bids, the individuals in both the monetary and Karma pricing scenarios observe a vector that contains the historical travel times, their income value, their departure time from the origin, their urgency at that day, and, in the case of Karma, their Karma balance. The historical travel times of an agent i represent the average travel times on each route experienced

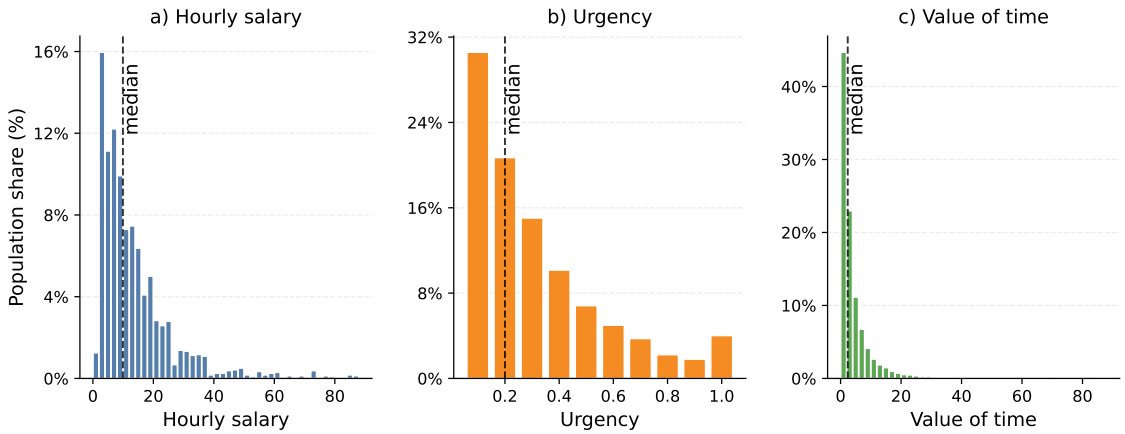


Figure 1: Population-level distributions of hourly salary, urgency, and value of time.

by individuals who departed from their origin within a specific time window of simulation steps prior to agent i's departure, over the last μ days.
Agents submit bids based on their available karma balance (on the first day, all agents receive an equal number of Karma points). When an agent is assigned to a route k, the submitted bid for that route is deducted from their Karma balance and agents travel along the assigned route k.
Before the next round (day), agents evaluate outcomes of their choices by observing their rewards:

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

composed of the 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 β_i. In both the monetary and Karma pricing scenarios, agents aim to maximize their reward. Under monetary pricing scheme, $c(k)$ is fixed, and β_i reflects the income distirubtion within the population. Under Karma pricing scheme, $c(k)$ is always zero, and β_i is equal to one for all agents. Therefore in the monetary pricing wealthy agents may buy priority with their high monetary budgets, whereas in Karma each agent has the same power to supply urgency.

MARL. We train our agents using the Independent Proximal Policy Optimization (IPPO) algorithm, which has shown strong benchmark performance in a variety of tasks (de Witt et al., 2020). Since our setting involves 300 agents, scalability is important. Parameter sharing, where agents update a single shared neural network, can improve scalability (Christianos et al., 2021) 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 (Tessera et al., 2025), and train agents with IPPO with hypernetworks.

Fairness. We compare and evaluate the monetary, karma pricing, SO, and UE solutions using several notions of quantitative fairness (Riehl et al., 2025). 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:

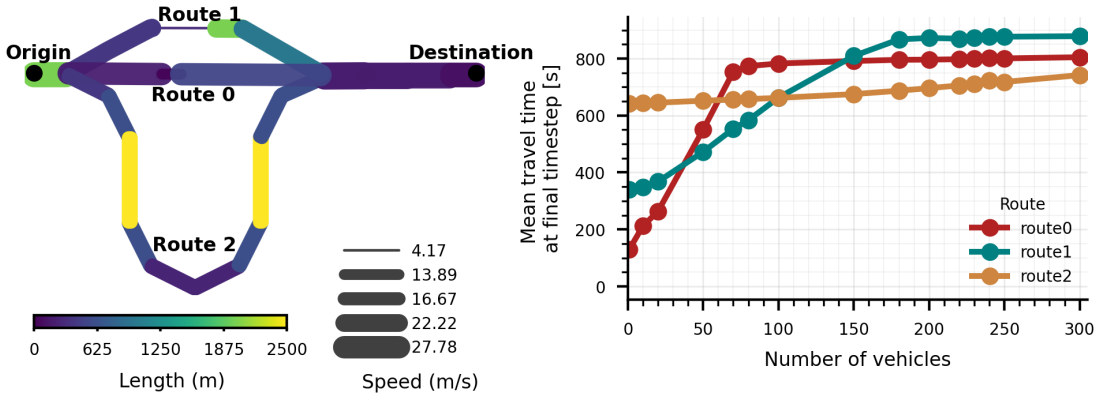


Figure 2: **Case study network.** Left: Traffic network. Right: Mean travel time at the final simulation timestep for different numbers of vehicles.

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

- **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:

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

- **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 Figure 2. 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 a different random seed in each. This means that, in each replication, the agents' income values may vary.
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 for the last 10 days ($\mu = 10$) and, for each agent, are defined as the average travel times on each route observed among individuals departing within a 20-simulation-step window before that agent. The income of the agents is drawn from the European income distribution reported in the World Inequality Report[2], and the urgency is drawn from a geometric distribution with parameter $p = 0.3$ following (Riehl et al., 2024a). The population distributions of hourly salary, urgency, and value of time are depicted in Figure 1a, b, c.

[2]Derived from https://wir2022.wid.world/methodology.

In addition, we consider two baseline scenarios, the approximated System-Optimal (SO) solution and the approximated user-equilibrium (UE) solution, in which drivers choose the fastest route based on historical travel times from previous days.

3 Results and discussion

Both monetary and Karma pricing reduce congestion, achieving average system travel times close to the SO solution, as shown in Figure 3a. In Figure 3b, the average system rewards increase during training and then stabilize, suggesting that the MARL algorithms converged to stable policies.

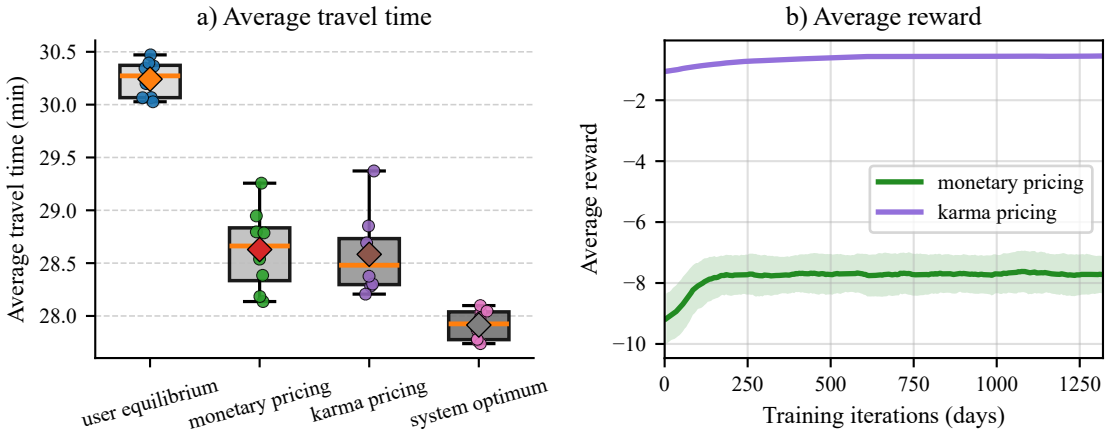


Figure 3: **Learning outcomes of congestion pricing.** a) Monetary and Karma pricing achieve similar average system travel times that are substantially lower than in the UE case and close to the SO solution. Circles denote individual replications, diamonds the mean travel time across replications, and 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 error bars represent variability among replications.

Figure 4a shows that monetary pricing can be unfair, as it favors individuals with higher VoT, whereas under Karma, travel times remain nearly constant across VoT bins. For this analysis, and only in this figure, we fix departure times across days to isolate the effect of the pricing mechanism and reduce noise from daily demand variation. Figure 4b shows that agents with higher urgency experience lower average travel times under both monetary and Karma pricing, consistent with (Riehl et al., 2024a). Table 1 compares Karma and monetary pricing schemes across the four fairness ideologies, discussed in Section 2. We evaluate fairness under different quantities, including agents' travel time, travel time weighted by urgency, and travel time weighted by value of time (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 fairness 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.

4 Conclusions

This work demonstrates that Karma-based segment pricing is an equally efficient and fairer alternative to monetary segment pricing in realistic traffic environments with stochastic traffic dynamics, where individuals commute at different times each day. We further show that Multi-Agent Reinforcement Learning (MARL) can successfully learn optimal bidding strategies in Karma economies and can be used by individuals as automated bidding agents, overcoming the difficulty human users may face in bidding optimally in such settings. Overall, these findings highlight the potential of Karma as both an efficient and fair mechanism for real-world congestion management.

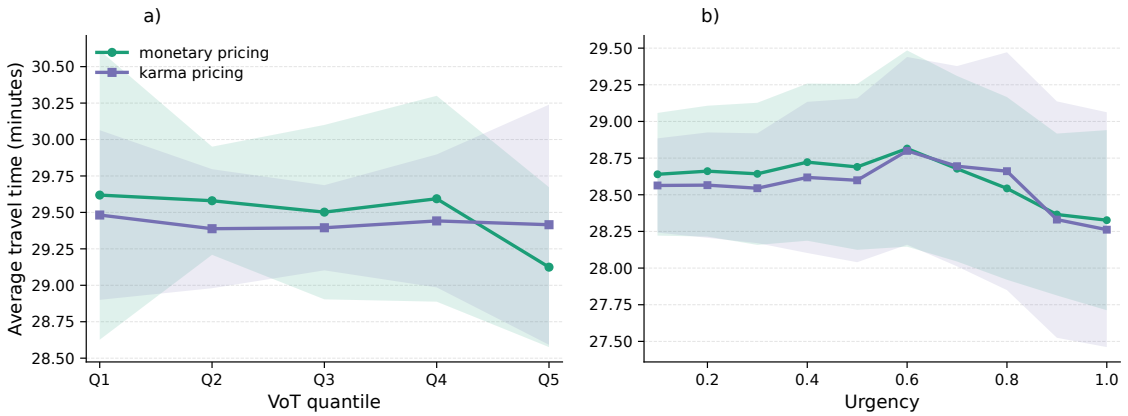


Figure 4: a) Under monetary pricing, average travel times drop in the highest quantile, whereas under Karma, they remain nearly constant across quantiles. b)Under both Karma and monetary pricing, average travel times decrease with urgency.

Table 1: Fairness measures across scenarios.

	Fairness Measures	SO	Monetary Pricing	Karma Pricing	UE
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)

Acknowledgements

Matej Jusup acknowledges that the initial phases of this research were conducted while affiliated with ETH Zurich. Anastasia Psarou and Rafał Kucharski acknowledge funding by the Horizon Europe Framework Programme (ERC Starting Grant COeXISTENCE no. 101075838). Kevin Riehl, Anastasios Kouvelas, and Michail A. Makridis acknowledge funding from the European Union's Horizon Europe research and innovation program under grant agreement No. 101203465 (project "FEDORA"), and the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract No. REF-1131-52301.

References

Akman, A. O., Psarou, A., Łukasz Gorczyca, Varga, Z. G., Jamróz, G., & Kucharski, R. (2025). Routerl: Multi-agent reinforcement learning framework for urban route choice with autonomous vehicles. *SoftwareX*, *31*, 102279. doi: 10.1016/j.softx.2025.102279

Bhat, C. R., & Koppelman, F. S. (1999). Activity-based modeling of travel demand. In R. W. Hall (Ed.), *Handbook of transportation science* (pp. 35–61). Boston, MA: Springer US. doi: 10.1007/978-1-4615-5203-1_3

Censi, A., Bolognani, S., Zilly, J. G., Sadat Mousavi, S., & Frazzoli, E. (2019). Today me, tomorrow thee: Efficient resource allocation in competitive settings using karma games. In *2019 ieee intelligent transportation systems conference (itsc)* (p. 686-693). doi: 10.1109/ITSC.2019.8916911

Chen, P., Tong, R., Lu, G., & Wang, Y. (2018). Exploring travel time distribution and variability patterns using probe vehicle data: case study in beijing. *Journal of Advanced Transportation*, *2018*(1). doi: 10.1155/2018/3747632

Christianos, F., Papoudakis, G., Rahman, A., & Albrecht, S. V. (2021). Scaling multi-agent reinforcement learning with selective parameter sharing. In *International conference on machine learning (icml).*

de Palma, A., & Lindsey, R. (2011). Traffic congestion pricing methodologies and technologies. *Transportation Research Part C: Emerging Technologies*, *19*(6), 1377-1399. doi: 10.1016/j.trc.2011.02.010

de Witt, C. S., Gupta, T., Makoviichuk, D., Makoviychuk, V., Torr, P. H. S., Sun, M., & Whiteson, S. (2020). *Is independent learning all you need in the starcraft multi-agent challenge?* doi: 10.48550/arXiv.2011.09533

Elokda, E., Bolognani, S., Dörfler, F., & Nax, H. H. (2026). *Dynamic resource allocation with karma: An experimental study.* doi: 10.48550/arXiv.2404.02687

Gu, Z., Liu, Z., Cheng, Q., & Saberi, M. (2018). Congestion pricing practices and public acceptance: A review of evidence. *Case Studies on Transport Policy*, *6*(1), 94-101. doi: 10.1016/j.cstp.2018.01.004

Holzman, R., & Law-Yone, N. (1997). Strong equilibrium in congestion games. *Games and Economic Behavior*, *21*(1), 85-101. doi: 10.1006/game.1997.0592

Hu, Y., Han, C., Guo, T., & Xiao, H. (2024). Applying opponent modeling for automatic bidding in online repeated auctions. In *Proceedings of the 23rd international conference on autonomous agents and multiagent systems* (p. 843–851). doi: 10.5555/3635637.3662938

Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., ... Wießner, E. (2018). Microscopic traffic simulation using SUMO. In *The 21st ieee international conference on intelligent transportation systems.* IEEE.

Riehl, K., Kouvelas, A., & Makridis, M. A. (2024a). Karma economies for sustainable urban mobility – a fair approach to public good value pricing. *npj Sustainable Mobility and Transport*, *1*, 14. doi: 10.1038/s44333-024-00014-4

Riehl, K., Kouvelas, A., & Makridis, M. A. (2024b). Resource allocation with karma mechanisms—a review. *Economies*, *12*(8). doi: 10.3390/economies12080211

Riehl, K., Kouvelas, A., & Makridis, M. A. (2025). Quantitative fairness—a framework for the design of equitable cybernetic societies. *Computers in Human Behavior: Artificial Humans*, *6*, 100236. doi: 10.1016/j.chbah.2025.100236

Tessera, K.-a. A., Rahman, A., Storkey, A., & Albrecht, S. V. (2025). Hypermarl: Adaptive hypernetworks for multi-agent rl. In *The thirty-ninth annual conference on neural information processing systems.* doi: 10.48550/arXiv.2412.04233

Thomasini, L. A., Alegre, L. N., Ramos, G. O., & Bazzan, A. L. C. (2023). Routechoiceenv: a route choice library for multiagent reinforcement learning. In *Adaptive and learning agents workshop at aamas.*

von Stackelberg, H. (2011). *Market structure and equilibrium* (None ed.) (No. 978-3-642-12586-7). Springer. doi: 10.1007/978-3-642-12586-7