Understanding Fairness in Congestion Games with Learning Agents

Dimitris Michailidis
University of Amsterdam
Amsterdam, The Netherlands
d.michailidis@uva.nl

Sennay Ghebreab
University of Amsterdam
Amsterdam, The Netherlands
s.ghebreab@uva.nl

Fernando P. Santos
University of Amsterdam
Amsterdam, The Netherlands
f.p.santos@uva.nl

ABSTRACT

To understand the societal impacts of adapting urban transport networks, we must consider their impacts on mobility patterns. Expanding transport networks can worsen congestion, as seen in the Braess Paradox, where adding a route increases travel times. Traditional game-theoretic models often assume rational agents, but real-world behavior is dynamic and influenced by exploration and learning. Moreover, socioeconomic factors such as income can affect exploration rates, leading to disparities in travel times and access. To investigate these issues, we model agents as reinforcement learners and study how disparities in exploration rates impact fairness and efficiency in a toy problem. Our findings reveal that unequal exploration rates can disproportionately harm less explorative groups. Network interventions targeting efficiency can worsen inequities, even when they do not affect the price of anarchy. We highlight the need to account for disparities emerging from individuals' adaptation, when designing transport systems.

KEYWORDS

Braess Paradox, Reinforcement Learning, Multi-Agent Systems, Fairness

1 INTRODUCTION

Transport networks are critical drivers of economic growth and social mobility, yet optimizing their efficiency and inclusivity remains challenging. Expanding infrastructure may lead to unintended consequences, as shown by the Braess's Paradox, where adding new routes increases congestion [8]. When designed for efficiency, transport networks can result in uneven benefits [27]. Understanding how transport infrastructure affects mobility is essential to devise efficient and fair solutions. Fairness in transport is important, as it contributes to more equitable and sustainable cities [33].

Game theory provides a powerful framework for analyzing and understanding the decision-making behavior of citizens within transport systems, and its consequences on the system. By formalizing scenarios where individual choices (e.g. route selection) conflict with collective outcomes, it explains phenomena like the Braess's paradox and suggests tools to address them [2]. Although game theoretical analysis often involves simple toy scenarios — such as the 4-node network considered in the Braess's paradox or the setting in this paper — their insights have been translated to real-world observations. For example, empirical studies confirm the relevance of such simple models in real-world case studies: the insight offered by the Braess's paradox, suggesting that removing transport lines can surprisingly improve travel time, was reported in specific street closures in Boston, NYC, London [43] and Seoul [2].

Mobility patterns in networks are often modelled as congestion games, a type of strategic game where players compete for limited resources, such as road space. In these games, nodes represent locations and edges represent connections with limited capacity that must be shared among commuters. Traditional analyses of congestion games assume agents that maximize their individual utility by adopting optimal strategies under complete information. However, real-world behavior is more dynamic [9, 11, 21, 38]. In this paper, we are investigating environments in which behavior is shaped by learning mechanisms [3, 5, 18, 37]. For example, commuters often rely on trial-and-error learning—exploring different routes, observing travel times, and adjusting their behavior accordingly. This process aligns closely with reinforcement learning (RL), where agents balance exploration (trying new strategies) and exploitation (using known strategies to maximize utility) [40].

Experimental studies demonstrate that simple reinforcement learning (RL) mechanisms can effectively reproduce observed real-world route choices [4, 38]. Furthermore, modeling commuter behavior through learning dynamics has been shown to alleviate some of the inefficiencies traditionally associated with full-information strategic agents [9, 42]. In particular, if we assume that commuters adaptively optimize their travel times using learning mechanisms, many of the classical assumptions underpinning selfish routing models no longer hold. This suggests that incorporating realistic learning behaviors into traffic models can lead to more accurate and potentially more efficient predictions of system-level outcomes.

Such models further reveal that exploration can cause oscillations between Nash equilibria and socially optimal strategies [9]. We explore whether these effects can be amplified in heterogeneous populations. Empirical evidence has shown that wealthier individuals, who can tolerate greater risk, tend to exhibit more exploratory behavior [7, 19]. For instance, the authors of [19] find that residents of low-income neighborhoods are significantly less exploratory in their activities, effectively living within constrained environments. Similarly, research in psychology has shown that individuals from high-resource environments demonstrate greater behavioral variance, with a higher tolerance for uncertainty and risk [7].

These findings suggest that exploration rates can serve as a proxy for resource capacity: individuals with more resources are better positioned to search and consider alternatives. Conversely, resource-limited agents are more likely to be trapped in inefficient outcomes due to tighter constraints on time, mobility, and finances.

Proc. of the Adaptive and Learning Agents Workshop (ALA 2025), Avalos, Aydeniz, Müller, Mohammedalamen (eds.), May 19 – 20, 2025, Detroit, Michigan, USA, ala-workshop.github.io. 2025.

Differences in individuals' behaviors can worsen inequalities introduced by transportation extensions. Considering such sources of heterogeneity in transportation design remains underexplored.

In this paper, we study how disparities in exploration behavior affect fairness and efficiency in urban transport networks. We model mobility as a congestion game in a stylized network model with reinforcement-learning agents. Agents representing individuals belonging to different demographic groups can differ in their exploration rates, simulating differences in adaptability and capacity to test different routing options. Our experiments demonstrate that interventions —such as adding a new route—can widen disparities between more and less exploratory groups. These inequities can emerge even when standard efficiency metrics like the Price of Anarchy remain unchanged. In contrast, more balanced exploration behaviors lead to fairer and more robust outcomes.

These findings highlight the importance of accounting for learning dynamics and behavioral heterogeneity when assessing the societal impacts of transport policy. We propose a new lens for mobility planning—one that considers not only how infrastructure is used, but how people learn to use it. Our model offers a proof of concept for how differences in exploration behavior can drive unequal access to the benefits of new infrastructure. To support reproducibility, we release our code [1].

2 RELATED WORK

Our work relates to two themes in congestion games: learning dynamics within these games and the fairness of outcomes.

2.1 Reinforcement Learning in Congestion Games

Deviations from the standard rational commuter behavior approach have been studied under models such as prospect theory, which considers traveler-centric decision-making under uncertainty [14, 15]. Recent papers have revealed non-Nash convergence and even chaotic behavior. For instance, online learning algorithms such as Multiplicative Weights Update [13] and Follow-the-Regularized-Leader [5] demonstrate how deviations from Nash equilibrium can occur under specific learning dynamics. Learning automata have been used as well [17]

Reinforcement Learning is increasingly used to model travel choice behavior in multiple congestion games [1]. Examples include studies on how commuters adapt over time through learning dynamics [4, 11], the impact of memory [41], as well as the impact of tolls in congestion games [32].

Analyzing the effects of exploration in congestion games through Reinforcement Learning, and in particular, multi-agent Q-learning hasn't been explored thoroughly. Some efforts have achieved this in a parallel road network, showing that the total social cost varies with different exploration rates [26]. In particular, they show that with sufficiently small exploration, the system will converge to the inefficient equilibrium. Others investigate the effect of exploration on a system's convergence between Nash equilibrium and the social optimum [9]. However, while they assume a homogeneous population with a uniform exploration rate, in reality, commuters have disparate access to and tolerance to exploration. Here, we build upon this work and extend this by modeling two distinct populations with varying exploration rates, allowing for a richer analysis of strategic behavior and its implications for equilibrium outcomes, as well as fairness.

2.2 Fairness in Congestion Games

The challenge of avoiding Braess's Paradox assuming selfish agents has been extensively studied, with efforts focused on network design [24, 34, 36].

Our work contributes to this literature by examining fairness considerations in congestion games, an aspect previously explored primarily in centrally coordinated settings. For example, through the lens of taxation in balancing fairness and efficiency [22], while Pedroso et al. investigated the impact of artificial currency-based pricing mechanisms [31]. Censi et al. proposed a karma-based approach to achieving fairness [12]. Additionally, Oesterle et al. introduced the Double Braess's Paradox in a multi-resource setting with two sources and two destinations, evaluating action-restriction mechanisms as a fairness intervention [30]. Further studies have considered fairness in weighted congestion games, such as [23], who explored mechanisms to reduce disparities in agent outcomes.

Using agent-based frameworks, Belov et al. studied a five-edge Braess paradox network, analyzing the impact of various micro-level parameters on efficiency and fairness [3]. Likewise, Levy et al. explored fairness in travel times between two routes through the lens of exploration but assumed a uniform exploration rate, demonstrating how exploration itself can shape fairness [26].

Our approach examines fairness as an emergent property in a multi-agent system. We define fairness as the disparity in average travel time to the destination across different sources, influenced by varying exploration rates. Our model assumes learning agents who adapt their behavior through reinforcement learning. This framework is designed to capture disparities in exploratory behavior described in the introduction, particularly those arising from differences in income levels.

Unlike previous studies, our work explicitly models two distinct populations with varying exploration capacities. This distinction enables a more nuanced analysis of fairness, shedding light on its interaction with strategic behavior and resource access. Furthermore, our approach allows us to examine the impact of network interventions in different exploration scenarios, providing insights into potential mechanisms for mitigating inequality.

3 PRELIMINARIES

We define congestion games, Nash equilibrium, price of anarchy and fairness.

3.1 Congestion Games

We use the standard definition of a congestion game [36]. Let $G = (V, E)$ be a directed graph, where V is the set of nodes and E is the set of edges. Let $\mathcal{N}$ be the set of players, where each player $i \in \mathcal{N}$ is associated with a source node $s_i \in V$ and a common destination node $d \in V$.

Each edge $e \in E$ is associated with a latency function $f_e : \mathbb{N} \to \mathbb{R}_{\geq 0}$, where $f_e(x)$ represents the latency experienced when x players use edge e. We assume the following properties:

[1] GitHub: https://github.com/dimichai/fairness-braess

Nonnegativity: $f_e(x) \geq 0$ for all $x \in \mathbb{N}$.

Continuity and Monotonicity: Each f_e is continuous and non-decreasing in x [35].

Polynomial-time Evaluation: Each $f_e(x)$ can be evaluated in polynomial time [23].

Each player i chooses a strategy a_i from a strategy set $\mathcal{A}_i$, consisting of all paths from s_i to d. A strategy profile is denoted $a = (a_1, a_2, \ldots, a_{|\mathcal{N}|})$. The edge load $x_e(a)$ denotes the number of players whose paths use edge e under profile a. We denote strategy sets by $\mathcal{A}$ to align with terminology introduced later in the paper, where we incorporate a reinforcement learning (RL) mechanism for strategy selection. In this context, strategies correspond directly to RL actions, with a one-to-one mapping between each strategy and its associated action.

The *social cost* is defined as:

$$C(a) = \sum_{e \in E} f_e(x_e(a)) \cdot x_e(a). \tag{1}$$

3.2 Nash Equilibrium

In congestion games, players are self-interested and aim to minimize their own cost. Let $c_i(a_i, a_{-i})$ denote the latency incurred by player i under profile $a = (a_i, a_{-i})$. A profile a^{NE} is a *Nash equilibrium* if no player can improve their cost by unilaterally deviating:

$$\forall i \in \mathcal{N},\ \forall a_i' \in \mathcal{A}_i : \quad c_i(a_i^{\text{NE}}, a_{-i}^{\text{NE}}) \leq c_i(a_i', a_{-i}^{\text{NE}}). \tag{2}$$

Congestion games are potential games and always admit at least one pure-strategy Nash equilibrium [28, 29].

3.3 Social Optimum and Price of Anarchy

The *social optimum* is the strategy profile $a^\star$ that minimizes the total social cost:

$$a^\star = \arg\min_a C(a). \tag{3}$$

The *Price of Anarchy (PoA)* quantifies the inefficiency of Nash equilibria compared to the social optimum. It is the ratio of the worst-case Nash equilibrium cost to the social optimum:

$$\text{PoA} = \frac{\max_{a \in \text{NE}} C(a)}{C(a^\star)} \geq 1. \tag{4}$$

The PoA measures how much worse the system performs under selfish behavior, compared to optimal coordination.

3.4 Fairness

We assess fairness using the notion of *source disparity*, defined as the difference in average cost between groups of players originating from different sources. This definition aligns with group-level fairness metrics, which aim to achieve an egalitarian outcome across groups. Specifically, it requires that no group is systematically disadvantaged on average. In the context of artificial intelligence fairness, this corresponds to the group fairness criterion of independence, where the outcome—here, the incurred cost—should be statistically independent of an agent's starting source. [10]. Let $\mathcal{N}_1$ and $\mathcal{N}_2$ denote the subsets of players starting at sources s_1 and s_2, respectively, with cardinalities $|\mathcal{N}_1| = N_1$ and $|\mathcal{N}_2| = N_2$.

$$\text{AvgCost}(s_j) = \frac{1}{N_j} \sum_{i \in \mathcal{N}_j} c_i, \quad j \in \{1, 2\}. \tag{5}$$

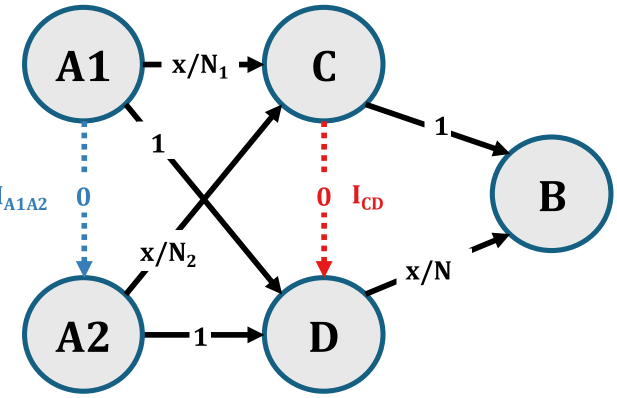

Figure 1: A Braess Paradox game with two sources nodes ($A1$ & $A2$) connecting to a single destination node B. Due to urban segregation, we assume that different demographic groups in a city will start in different source nodes. We examine two interventions: $\mathbf{I}_{A1A2}$ and $\mathbf{I}_{CD}$.

Source disparity is then:

$$\text{SD}(s_1, s_2) = \text{AvgCost}(s_1) - \text{AvgCost}(s_2). \tag{6}$$

A positive SD indicates an advantage for s_2, negative SD favors s_1, and zero SD implies perfect fairness. Note that social optimum is not directly correlated with fairness, as one regards the total cost of the game, and the other the differences between agents starting in different nodes. In fact, optimal routing is not necessarily fair [35].

4 A TWO-SOURCE BRAESS PARADOX GAME

The Braess Paradox is a counterintuitive phenomenon where adding a new edge can increase the overall social cost, contrary to the intuitive expectation that infrastructure expansion reduces congestion. This paradox arises because self-interested agents disrupt the system's equilibrium, converging to a less efficient state [8].

In this paper, we introduce a modified version of the classical Braess paradox game, focusing on a two-source, one-destination network, as illustrated in Figure 1. Such networks are common in real cities, which often develop around a central area, with surrounding districts having multiple routes leading to it. We analyze two distinct instances of this game: (1) the network with the addition of a direct connection between nodes $C \to D$ (Braess Paradox case), and (2) the additional intervention of the direct connection $A1 \to A2$.

In this setting, N_1 number of players start from source $A1$ and N_2 players start from source $A2$, with a common destination B. This network models a typical urban traffic system where most commuters move toward a central location but start from different points and face different costs, depending on their routes.

The latency functions $f_e(x)$ for each edge $e \in E$ depend on the number of players x using the edge and are defined as follows: for edges $A1 \to C$ and $C \to B$, the latency is $f_e(x) = x/N_1$; for edges $A2 \to D$ and $D \to B$, the latency is $f_e(x) = x/N_2$; for edges $A1 \to A2$ and $A2 \to A1$, the latency is a constant $f_e(x) = 1$. In

Scenario	NE	SO	PoA	SD
(0) No Intervention	1.5	1.5	1.0	0.0
(1) I_{CD}	2.0	1.5	1.33	0.0
(2) $\{I_{CD}, I_{A1,A2}\}$	2.0	1.5	1.33	0.0

Table 1: An overview of the equilibria of the two-source game scenarios, assuming no learning. We report total cost (travel time) at the Nash Equilibrium (NE), Social Optimal solution (SO), Price of Anarchy (PoA, Eq. 4) and Source Disparity (SD, Eq. 6).

the scenarios we analyze, additional *fast lanes* are introduced: a blue dashed edge $A1 \rightarrow A2$ and a red dashed edge $C \rightarrow D$, both of which have flow $f_e(x) = 0$.

The social cost of the game is defined as follows:

$$C = \sum_{e \in E} x_e \cdot f_e(x_e) = \\ x_{A1\rightarrow C}\frac{x_{A1\rightarrow C}}{N_1} + x_{A2\rightarrow C}\frac{x_{A2\rightarrow C}}{N_2} + \\ x_{D\rightarrow B}\frac{x_{D\rightarrow B}}{N_1+N_2} + 1\ x_{A2\rightarrow D} + 1\ x_{A1\rightarrow D} + 1\ x_{C\rightarrow B} \quad (7)$$

This game is designed to induce the Braess paradox within a two-source network. In the baseline case without interventions, players from sources $A1$ and $A2$ each have two available strategies, routes $C \rightarrow B$ (Up) and $D \rightarrow B$ (Down), and converge to a socially optimal state. However, the addition of the fast lane $C \rightarrow D$ introduces a third strategy, $C \rightarrow D \rightarrow B$ (Cross), for all players. This intervention changes the Nash equilibrium (NE), causing all players to select the new strategy (Cross), increasing the total cost and the price of anarchy. We analyze the impact of Scenario 2, when an additional intervention is being introduced, assuming exploratory agents in the population. In Scenario 2, agents in source $A2$ still have the same strategies to choose from. However, agents in source $A1$ gain three more strategies: $A1 \rightarrow A2 \rightarrow C \rightarrow B$ (Down-Up), $A1 \rightarrow A2 \rightarrow D \rightarrow B$ (Down-Down), and $A1 \rightarrow A2 \rightarrow C \rightarrow D \rightarrow B$ (Down-Cross). The Nash equilibrium still remains the same as in Scenario 1 (all agents Cross).

It is important to note that, in this context, the interventions are not required to satisfy the fairness constraints as previously defined [23]. In fact, intervention I_{A1A2} deliberately introduces an asymmetry by providing group A1 with more strategic options than group A2. This design is intentional: our goal is to examine the effects of expansions that favors one group in environments with exploratory, learning agents. As shown in Table 1, this intervention does not alter the Nash equilibrium or introduce inherent unfairness, assuming agents are rational, self-interested, and with full information about the network and payoffs. We focus on intervention I_{A1A2} due to the symmetry of the game; equivalent conclusions would follow from introducing I_{A2A1} instead.

5 LEARNING DYNAMICS

In practice, humans often deviate from exact Nash equilibrium strategies, as they typically lack complete information about which strategies yield specific payoffs from the outset [6, 25]. Instead, they explore various options and gradually learn to favor those that maximize their satisfaction [9].

To investigate these effects—particularly the impact of heterogeneous exploration—in the two-source Braess paradox game, we develop a learning framework in which agents repeatedly play the game and adapt their strategies through Reinforcement Learning (RL).

Importantly, we employ RL not as a tool for optimization, but as a framework to simulate the adaptive behaviors of agents interacting within complex environments. This approach enables us to explore emergent dynamics and offer explanatory insights into observed behavioral patterns. Our perspective aligns with the broader *descriptive* agenda in multi-agent learning, which emphasizes modeling learning processes that mirror how real-world agents, such as humans, adapt in social contexts [39].

Specifically, we adopt Q-learning, a model-free reinforcement learning algorithm, to capture some simple properties of human learning such as trial-and-error adaptation, habit formation and bounded rationality. This choice is grounded in the observation that commuters often lack comprehensive or accurate models of how their actions influence congestion dynamics. Instead, they rely on repeated experience to guide future choices. Model-free learning allows us to study emergent behaviors under simple, reactive update rules, without assuming agents possess detailed planning capabilities or perfect knowledge of system dynamics. Finally, we also use Q-Learning to better align with recent literature aiming to model learning behavior in congestion games [9].

5.1 Q-Learning Framework

Let the game be played over T discrete time steps, indexed by $t = 0, 1, 2, \ldots, T-1$. Each agent $i \in N$ seeks to minimize their experienced travel time (latency) by iteratively updating their strategy based on observed outcomes.

Action Space Each agent i selects a strategy (route) $a_i \in \mathcal{A}(s)_i$, where $\mathcal{A}(s)_i$ is the set of available strategies for agent i, starting in source s. In the two-source Braess base network:

$$\mathcal{A}(A1)_i = \mathcal{A}(A2)_i = \{\text{Up, Down}\}.$$

In Scenario 1 — intervention I_{CD}:

$$\mathcal{A}(A1)_i = \mathcal{A}(A2)_i = \{\text{Up, Down, Cross}\}.$$

In Scenario 2 — interventions $\{I_{CD}, I_{A1A2}\}$:

$$\mathcal{A}(A1)_i = \{\text{Up, Down, Cross, Down-Up, Down-Down, Down-Cross}\},$$
$$\mathcal{A}(A2)_i = \{\text{Up, Down, Cross}\}.$$

Note that in Scenario 2, agents starting in source $A1$ have three more strategies than agents starting in source $A2$.

Q-Values The Q-value $Q(i, a_i)$ represents agent i's estimate of the cumulative latency for selecting strategy a_i. These values are updated iteratively to reflect the observed latency for each strategy. We create two Q-tables, one for each source, with rows representing agents and columns representing strategies. Therefore, for Scenario 1, $Q_{A1}, Q_{A2} \in \mathbb{R}^{|N_1| \times 3}$ and for Scenario 2, $Q_{A1} \in \mathbb{R}^{|N_1| \times 6}, Q_{A2} \in \mathbb{R}^{|N_1| \times 3}$.

Learning At each time step t, the Q-value for an agent i is updated as:

$$Q_{s_i}(i, a_i) = Q_{s_i}(i, a_i) + \alpha\left[r_i(t) + \gamma \max_{a'} Q_{s'_i}(i, a') - Q_{s_i}(i, a_i)\right], \quad (8)$$

where $\alpha \in (0, 1]$ is the learning rate, and $r_i(t)$ is the observed latency for the chosen strategy a_i at time t. Specifically, the observed latency is calculated as follows:

$$r_i(t) = \sum_{e \in P_i(t)} f_e\big(x_e(t)\big), \quad (9)$$

where $P_i(t)$: the path (sequence of edges) that correspond to the selected strategy (action) of agent i at time t.

Note that here we omit the state, as the game is stateless and rewards are observed immediately. The observed latency $r_i(t)$ is determined by the congestion on the edges, $f_e(x)$. During the learning process, observed latencies are represented as negative values. As a result, the maximization function incentivizes actions that minimize the expected latency.

Exploration vs. Exploitation To balance exploration (trying new strategies) and exploitation (using the best-known strategy), each agent follows an ϵ-greedy policy. At time t, agent i selects a strategy $a_i(t)$ as:

$$a_i(t) = \begin{cases} \text{random action from } \mathcal{A}_i, & \text{with probability } \epsilon, \\ \arg\max_{a \in \mathcal{A}_i} Q_{s_i}(i, a), & \text{with probability } 1 - \epsilon. \end{cases}$$

Here, ϵ is the exploration rate, which determines how often an agent will explore. To investigate the effects of disparate exploration rates, we define two distinct rates, ϵ_{A1} and ϵ_{A2} for agents starting in $A1$ and $A2$ respectively.

Convergence At each step, agents collectively contribute to network congestion. The latency experienced by any individual agent depends not only on their own strategy choices but also on the aggregate behavior of the population. Each agent estimates the expected value of selecting a given strategy based on the cumulative experience gathered over time. With sufficient sampling, ongoing exploration, and continued learning, agents' value estimates for each strategy converge to some approximated values. However, these do not necessarily correspond to a Nash Equilibrium (NE); convergence is shaped by factors such as the learning rate, exploration level, and the environment's structural properties [9]. In this paper, we assume a constant exploration rate throughout the duration of the game to reflect realistic human behavior — such as the inherent stochasticity in commuter route choices, driven by exploration of new modes. Unlike traditional reinforcement learning aimed at optimization, our focus here is on describing and understanding emergent phenomena in a population of learning agents.

5.2 Experimental Setup

We ran the experiments with $|\mathcal{N}| = 200$ agents, evenly divided between two sources ($N_1 = N_2 = 100$). Each agent's Q-values were initialized uniformly at random within the interval $[0, 1]$. Our goal is to assess the impact of differing exploration rates between sources on the game's dynamics. To isolate this effect, we fixed the learning rate at $\alpha = 0.01$ for all experiments, following [9]. While the learning rate influences both convergence and convergence speed, we refer readers to [9] for a detailed analysis. Here, we focus specifically on the effects of exploration rate heterogeneity within the population. Each experiment was run for 10,000 steps and repeated 40 times with different random seeds. Performance was measured by averaging metrics over the final 2,000 steps to reduce the influence of transient exploration.

6 RESULTS

In this section we report the results obtained when two different interventions (I_{CD} and I_{A1A2}) are introduced in the network. From the onset we note that, as in the original Braess's paradox, our augmented network – with two source nodes instead of one – induces selfish agents to split equally, with a total social cost equal to the social optimum and with a price of anarchy of 1. From this baseline result, we focus on scenarios (1) and (2), where I_{CD} and I_{A1A2} are introduced, respectively. We assume that agents learn, over time, to select different routes, according to Q-Learning. Agents can differ in their starting node and exploration rates.

6.1 Scenario 1 — Intervention I_{CD}

Equal exploration leads to efficient and fair outcomes: We first focus on Scenario 1, investigating learning dynamics and equilibria in the game under varying exploration rates. We assume that agents in both sources have an equal exploration rate (ϵ). In Figure 2 (top) we observe that, with sufficient exploration, the learning agents do not approach the Nash Equilibrium (NE). Instead, agents oscillate between the Social Optimum (SO) and the NE, depending on the value of ϵ. As ϵ increases, agents converge toward a state close to the SO where the Price of Anarchy (PoA) approaches 1. These findings align with the recently identified influence of exploration rates on game equilibria (notably in the classic Braess's paradox) [9].

As illustrated in Figure 2 (bottom), under equal exploration rates starting in different source nodes does not result into systematic advantages. Fairness fluctuates due to the dynamics of the learning process and stochasticity introduced by exploration and multi-agent routing, but overall, fairness is maintained across the two sources. This observation is crucial before addressing Scenario 2, which involves additional interventions and disparities in resources and exploration rates.

Disparity in exploration leads to unfair outcomes: Rather than being an inherent network property, unfairness arises when agents have disparity in exploration. In Figure 3, we present heat maps that illustrate the price of anarchy (PoA) and source disparity for various combinations of exploration rates between the two sources. First, we observe that increasing the exploration rate consistently reduces PoA. This trend can be seen along the diagonal of Figure 3 (top). Even when there is a disparity in exploration rates, higher average exploration across the system improves efficiency and lowers the price of anarchy. However, as shown in Figure 3 (bottom), disparities in exploration rates result in unfairness in the average travel times experienced by agents from different sources. Specifically, when $\epsilon_{A1} > \epsilon_{A2}$, agents in $A1$ explore more often and the source disparity is negative, indicating that agents from $A2$

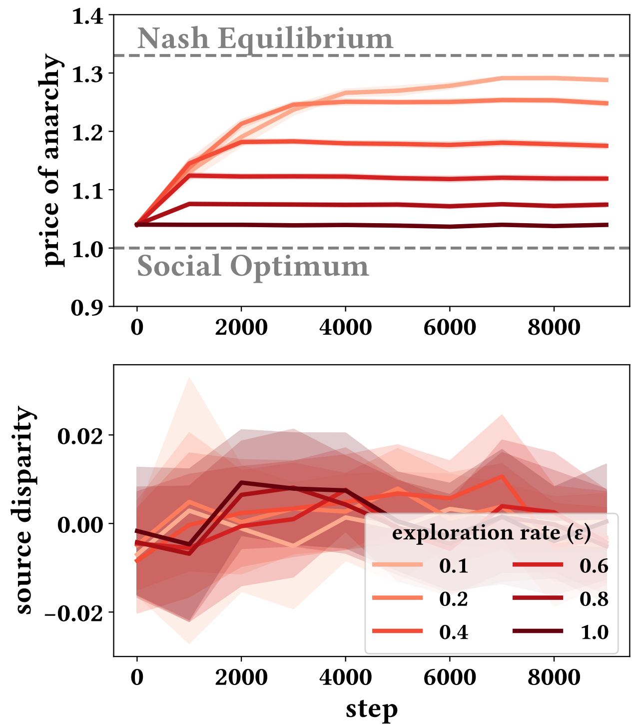

Figure 2: Price of Anarchy and source disparity for Intervention I_{CD}, for different exploration rates. Higher ϵ leads to lower price of anarchy, without significantly increasing the source disparity. Source disparity (Eq. 6) is the difference between average costs of individuals starting in two different source nodes.

experience longer average travel times. This trend is symmetrically reversed when $\epsilon_{A2} > \epsilon_{A1}$.

6.2 Scenario 2 — Adding Intervention I_{A1A2}

We now examine the effects of adding intervention I_{A1A2}. This intervention grants agents from source $A1$ three additional strategies. However, under selfish routing without exploration, neither the Nash Equilibrium nor the Socially Optimal configurations change (Table 1), as $A1$ agents lack an incentive to adopt the new strategies.

Figure 4 illustrates the impacts on price of anarchy (top) and source disparity (bottom), as functions of exploration ϵ. We compare Scenario 1 (intervention I_{CD}, red) with Scenario 2 (both I_{CD} and I_{A1A2}, blue). As we can observe, adding I_{A1A2} leads to similar maximum PoA, however the resulting source disparity is remarkably different. We use this intervention as a proof-of-concept to highlight that even if two interventions result in similar PoA, they might result in drastically different source-based costs and inequality.

Introducing I_{A1A2} increases inefficiency and unfairness: Introducing I_{A1A2} raises the price of anarchy across all exploration rates, indicating a systemic increase in inefficiency. It also amplifies disparity, with $A2$ agents experiencing longer travel times than $A1$. However, at high exploration ($\epsilon > 0.5$), this unfairness diminishes

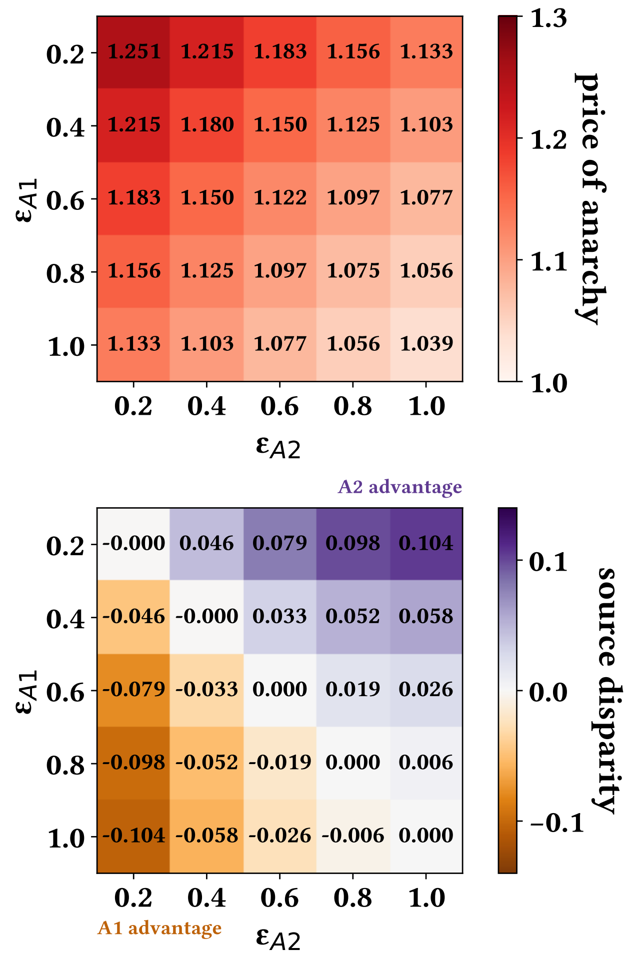

Figure 3: Heatmap showing the relationship between PoA and source disparity for intervention I_{CD} under varying exploration rates for agents starting in source node $A1$ (ϵ_1) or $A2$ (ϵ_2). Higher exploration rates reduce PoA, while an exploration advantage for one source increases its travel time advantaged and total unfairness.

as $A1$ agents take suboptimal routes, inadvertently reducing congestion and balancing travel times. The price of anarchy peaks as ϵ increases from 0 to 0.1, consistent with prior findings [9]. Beyond this range, higher exploration rates reduce inefficiency.

Comparing the two interventions, I_{CD} causes a steeper rise in the price of anarchy within $0 \leq \epsilon \leq 0.1$ but outperforms I_{A1A2} at higher exploration levels. For $0.6 \leq \epsilon \leq 1.0$, both interventions approach the social optimum, though I_{CD} remains the more efficient option, while I_{A1A2} plateaus.

Disparity in exploration rates can amplify or alleviate unfairness: To better understand how exploration disparity affects the price of anarchy and fairness, we conducted experiments where the total exploration rate remained constant while varying the ratio between ϵ_{A1} and ϵ_{A2}. Figure 5 presents the results for two scenarios: a total exploration rate of 0.2 (top) and 0.6 (bottom).

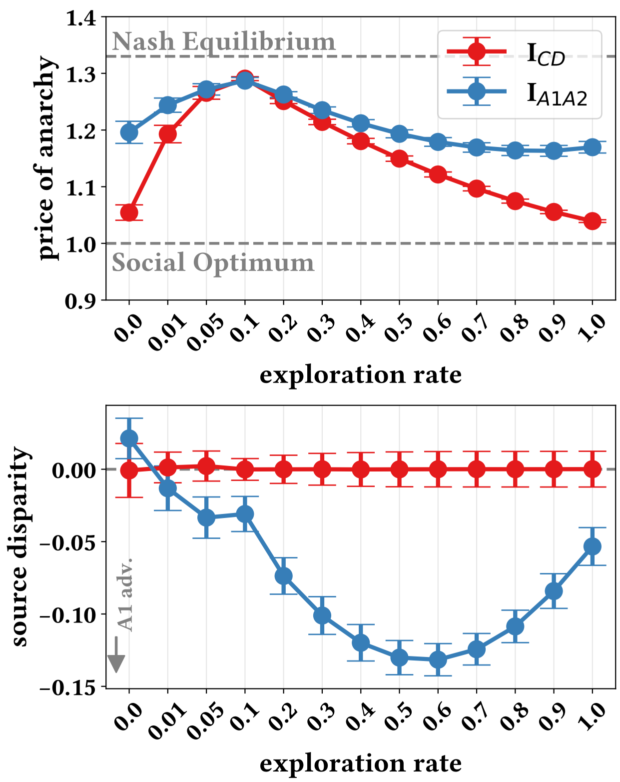

Figure 4: The impact of interventions on PoA and SD under constant, equal exploration rates. Introducing intervention I_{A1A2}, gives $A1$ a strategic advantage by increasing its available routes, altering network efficiency and fairness dynamics. On the other hand, introducing connection I_{CD} has a limited effect on disparity (as already concluded in Fig. 2). The observed levels of efficiency and fairness are contingent on agents' exploration rate. Importantly, we observe that while I_{CD} and I_{A1A2} lead to similar values of (max) PoA, they result in drastically different levels of disparity.

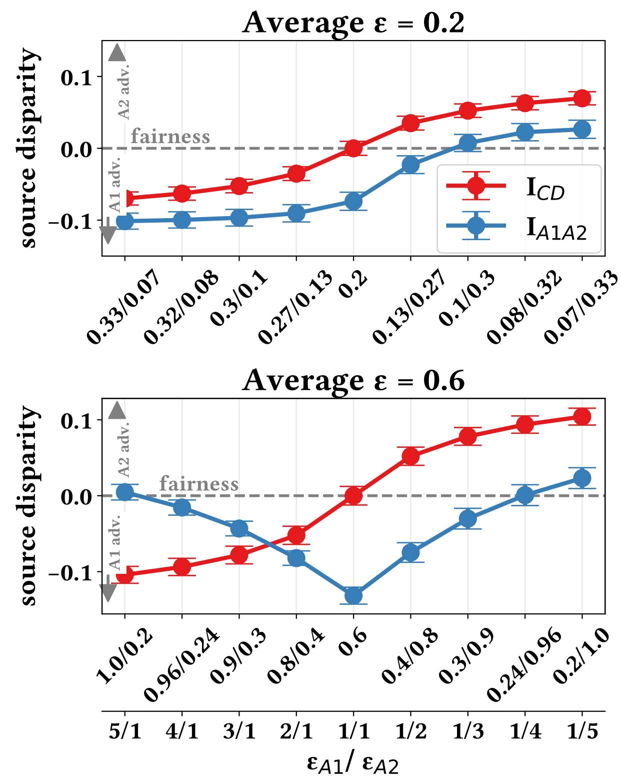

Figure 5: The effect of exploration disparity on fairness, for a population with an average exploration rate of 0.2 (top) and 0.6 (bottom). The ratio $\epsilon_{A1}/\epsilon_{A2}$ represents the relative exploration rates of $A1$ and $A2$. At low base exploration, higher exploration for $A1$ leads to a travel-time advantage for $A1$, even without I_{A1A2}. However, at high base exploration, I_{A1A2} can counteract this disparity.

In the low-exploration scenario (top), greater exploration by $A1$ leads to unfairness, both before and after intervention I_{A1A2}. I_{A1A2} disproportionately benefits $A1$ when its exploration rate is up to five times that of $A2$ or even when it is halved relative to $A2$. Fairness is restored when $A2$ explores at least three times more than $A1$. This contrasts with Scenario 1, where equal exploration rates yield equal outcomes.

In the high-exploration scenario (bottom), counterintuitively, unfairness peaks when both sources explore equally. As $A1$ explores more relative to $A2$, fairness improves. This occurs because increased exploration by $A1$ disperses traffic, reducing congestion on key routes for $A2$ (e.g., CD and DB). Meanwhile, $A2$'s lower exploration allows it to capitalize on the less-congested Crossing path. When both explore equally, however, $A1$ consistently clogs routes that disproportionately penalize $A2$ (e.g., A2C). Notably, when $A1$ explores five times more than $A2$, the proportion of $A2$ agents using DB and A2C decreases by 16%.

Figure 6 shows how the price of anarchy and source disparity evolve as exploration shifts between sources. A greater exploration advantage for $A2$ reduces the price of anarchy, as $A2$ avoids congested routes while $A1$ exploits them, leading to more efficient network utilization.

These findings suggest that even when $A1$ has a strategic advantage, increasing exploration resources for $A2$ reduces overall inefficiency and unfairness by expanding its routing options. Ultimately, exploration allocation plays a crucial role in infrastructure-induced inequality.

7 CONCLUSION

We investigated the impact of exploration on network congestion and fairness in a Braess's Paradox game, where agents from two sources share a common destination. Our findings demonstrate that exploration influences both the price of anarchy and fairness. Specifically, an exploration rate advantage leads to unfairness towards those who explore less. We also examine the consequences of interventions, showing that they can yield either benefits or

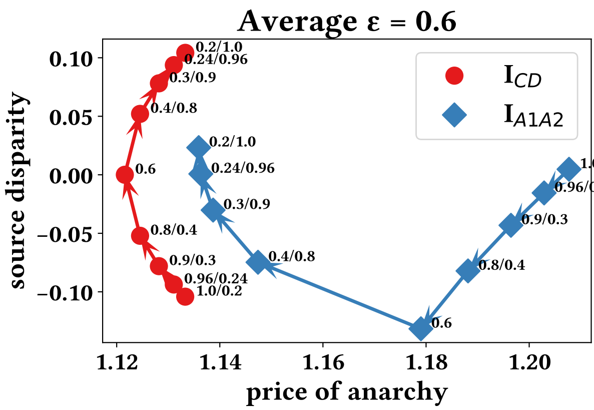

Figure 6: A detailed examination of how intervention I_{A1A2} affects PoA and fairness in a population of highly-explorative agents. Arrows indicate shifts from $A1$'s advantage to $A2$'s advantage in exploration rate. Before the intervention introducing connection I_{A1A2}, exploration disparities did not affect PoA. After intervention, an exploration advantage for $A2$ improves efficiency and fairness, while equal exploration rates produce the worst fairness outcomes.

unintended harms. Finally, we showed that if exploration disparity exists, accessibility interventions might not be enough to alleviate unfairness.

This paper contributes to the ongoing research on fair transport network design, by emphasizing the importance of considering new sources of heterogeneity that dictate mobility behavior. Our analysis highlights the interplay between individual exploration and systemic fairness. Our results suggest that interventions aimed at mitigating congestion should also account for the heterogeneity in agents' learning dynamics, as overlooking disparities in exploration may reinforce inequities rather than resolve them.

To isolate the role of exploration, we focused on a symmetrical network in which, under rational conditions, optimal and fair routing coincide. In this setting, unfairness arises solely through exploratory behavior. Future work could extend this analysis to asymmetric networks where, even with fully rational agents, a trade-off exists between optimal and fair routing—as demonstrated in [35]. Furthermore, it would be much relevant to consider alternative learning algorithms, inspired in psychology and economics literature, such as Roth-Erev learning [20] or model-based algorithms [16]. Investigating how such inequities might be mitigated in the presence of learning agents remains an important direction for research.

ACKNOWLEDGMENTS

This research was supported by the Innovation Center for AI (ICAI, The Netherlands) and the City of Amsterdam.

REFERENCES

[1] Furkan Ahmad and Luluwah Al-Fagih. 2024. Travel behaviour and game theory: A review of route choice modeling behaviour. *Journal of choice modelling* 50 (2024), 100472.

[2] Linda Baker. 2009. Removing roads and traffic lights speeds urban travel. *Scientific American* 1 (2009).

[3] Aleksandr Belov, Konstantinos Mattas, Michail Makridis, Monica Menendez, and Biagio Ciuffo. 2022. A microsimulation based analysis of the price of anarchy in traffic routing: The enhanced Braess network case. *Journal of Intelligent Transportation Systems* 26, 4 (2022), 448–460. https://doi.org/10.1080/15472450.2021.1904920 arXiv:https://doi.org/10.1080/15472450.2021.1904920

[4] Eran Ben-Elia and Yoram Shiftan. 2010. Which road do I take? A learning-based model of route-choice behavior with real-time information. *Transportation Research Part A: Policy and Practice* 44, 4 (2010), 249–264. https://doi.org/10.1016/j.tra.2010.01.007

[5] Jakub Bielawski, Thiparat Chotibut, Fryderyk Falniowski, Grzegorz Kosiorowski, Michał Misiurewicz, and Georgios Piliouras. 2021. Follow-the-regularized-leader routes to chaos in routing games. In *International Conference on Machine Learning*. PMLR, 925–935.

[6] Sarah Bonau. 2017. A Case for Behavioural Game Theory. 6 (02 2017), 7. https://doi.org/10.5923/j.jgt.20170601.02

[7] Melusine Boon-Falleur, Nicolas Baumard, and Jean-Baptiste André. 2024. The Effect of Income and Wealth on Behavioral Strategies, Personality Traits, and Preferences. *Perspectives on psychological science : a journal of the Association for Psychological Science* (01 2024), 17456916231201512. https://doi.org/10.1177/17456916231201512

[8] Dietrich Braess, Anna Nagurney, and Tina Wakolbinger. 2005. On a Paradox of Traffic Planning. *Transportation Science* 39, 4 (2005), 446–450. http://www.jstor.org/stable/25769266

[9] Cesare Carissimo. 2024. Counter-Intuitive Effects of Q-Learning Exploration in a Congestion Dilemma. *IEEE Access* 12 (2024), 15984–15996. https://doi.org/10.1109/ACCESS.2024.3358608

[10] Alessandro Castelnovo, Riccardo Crupi, Greta Greco, Daniele Regoli, Ilaria Giuseppina Penco, and Andrea Claudio Cosentini. 2022. A clarification of the nuances in the fairness metrics landscape. *Scientific Reports* 12, 1 (March 2022), 4209. https://doi.org/10.1038/s41598-022-07939-1

[11] Oded Cats and Jens West. 2020. Learning and adaptation in dynamic transit assignment models for congested networks. *Transportation Research Record* 2674, 1 (2020), 113–124.

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

[13] Thiparat Chotibut, Fryderyk Falniowski, Michał Misiurewicz, and Georgios Piliouras. 2020. The route to chaos in routing games: When is price of anarchy too optimistic? *Advances in Neural Information Processing Systems* 33 (2020), 766–777.

[14] Ioannis Vasileios Chremos and Andreas A Malikopoulos. 2022. The design and analysis of a mobility game. *arXiv preprint arXiv:2202.07691* (2022).

[15] Ioannis Vasileios Chremos and Andreas A. Malikopoulos. 2023. A traveler-centric mobility game: Efficiency and stability under rationality and prospect theory. *PLOS ONE* 18, 5 (May 2023), 1–32. https://doi.org/10.1371/journal.pone.0285322 Publisher: Public Library of Science.

[16] Nathaniel D Daw, Samuel J Gershman, Ben Seymour, Peter Dayan, and Raymond J Dolan. 2011. Model-based influences on humans' choices and striatal prediction errors. *Neuron* 69, 6 (2011), 1204–1215.

[17] Gabriel de O. Ramos and Ricardo Grunitzki. 2015. An Improved Learning Automata Approach for the Route Choice Problem. In *Agent Technology for Intelligent Mobile Services and Smart Societies*, Fernando Koch, Felipe Meneguzzi, and Kiran Lakkaraju (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 56–67.

[18] Arthur Dolgopolov. 2024. Reinforcement learning in a prisoner's dilemma. *Games and Economic Behavior* 144 (2024), 84–103. https://doi.org/10.1016/j.geb.2024.01.004

[19] Xiaowen Dong, Alfredo J Morales, Eaman Jahani, Esteban Moro, Bruno Lepri, Burcin Bozkaya, Carlos Sarraute, Yaneer Bar-Yam, and Alex Pentland. 2020. Segregated interactions in urban and online space. *EPJ Data Science* 9, 1 (2020), 20.

[20] Ido Erev and Alvin E Roth. 1998. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review* (1998), 848–881.

[21] Dick Ettema, Guus Tamminga, Harry Timmermans, and Theo Arentze. 2005. A micro-simulation model system of departure time using a perception updating model under travel time uncertainty. *Transportation Research Part A: Policy and Practice* 39, 4 (2005), 325–344.

[22] Bryce L Ferguson and Jason R Marden. 2021. The impact of fairness on performance in congestion networks. In *2021 American control conference (ACC)*. IEEE, 4521–4526.

[23] Miriam Fischer, Martin Gairing, and Dario Paccagnan. 2023. Fair Interventions in Weighted Congestion Games. *arXiv preprint arXiv:2311.16760* (2023).

[24] Martin Gairing, Tobias Harks, and Max Klimm. 2017. Complexity and Approximation of the Continuous Network Design Problem. *SIAM Journal on Optimization* 27, 3 (2017), 1554–1582. https://doi.org/10.1137/15M1016461 arXiv:https://doi.org/10.1137/15M1016461

[25] Daeyeol Lee. 2008. Game theory and neural basis of social decision making. *Nature neuroscience* 11, 4 (April 2008), 404–409. https://doi.org/10.1038/nn2065

[26] Nadav Levy, Ido Klein, and Eran Ben-Elia. 2018. Emergence of cooperation and a fair system optimum in road networks: A game-theoretic and agent-based modelling approach. *Research in Transportation Economics* 68 (2018), 46–55.

[27] Karel Martens. 2016. *Transport justice: Designing fair transportation systems*. Routledge.

[28] Igal Milchtaich. 1996. Congestion games with player-specific payoff functions. *Games and economic behavior* 13, 1 (1996), 111–124.

[29] Dov Monderer and Lloyd S. Shapley. 1996. Potential Games. *Games and Economic Behavior* 14, 1 (1996), 124–143. https://doi.org/10.1006/game.1996.0044

[30] Michael Oesterle, Tim Grams, Christian Bartelt, and Heiner Stuckenschmidt. 2024. RAISE the Bar: Restriction of Action Spaces for Improved Social Welfare and Equity in Traffic Management. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*. 1492–1500.

[31] Leonardo Pedroso, Andrea Agazzi, WPMH Heemels, and Mauro Salazar. 2024. Fair artificial currency incentives in repeated weighted congestion games: Equity vs. equality. *arXiv preprint arXiv:2403.03999* (2024).

[32] Gabriel de O Ramos, Roxana Rădulescu, Ann Nowé, and Anderson R Tavares. 2020. Toll-based learning for minimising congestion under heterogeneous preferences. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*. 1098–1106.

[33] Kevin Riehl, Anastasios Kouvelas, and Michail Makridis. 2024. Towards fair roads – Why we should & how to improve the fairness in traffic engineering. arXiv:2408.01309 [econ.GN] https://arxiv.org/abs/2408.01309

[34] Charlotte Roman and Paolo Turrini. 2019. Multi-population congestion games with incomplete information. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence, 565–571.

[35] Tim Roughgarden. 2002. How unfair is optimal routing?. In *Symposium on Discrete Algorithms: Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms*, Vol. 6. 203–204.

[36] Tim Roughgarden. 2006. On the severity of Braess's Paradox: Designing networks for selfish users is hard. *J. Comput. System Sci.* 72, 5 (2006), 922–953. https://doi.org/10.1016/j.jcss.2005.05.009

[37] Yuzuru Sato, Eizo Akiyama, and J. Doyne Farmer. 2002. Chaos in learning a simple two-person game. *Proceedings of the National Academy of Sciences* 99, 7 (2002), 4748–4751. https://doi.org/10.1073/pnas.032086299 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.032086299

[38] R. Selten, T. Chmura, T. Pitz, S. Kube, and M. Schreckenberg. 2007. Commuters route choice behaviour. *Games and Economic Behavior* 58, 2 (2007), 394–406. https://doi.org/10.1016/j.geb.2006.03.012

[39] Yoav Shoham, Rob Powers, and Trond Grenager. 2007. If multi-agent learning is the answer, what is the question? *Artificial Intelligence* 171, 7 (2007), 365–377. https://doi.org/10.1016/j.artint.2006.02.006 Foundations of Multi-Agent Learning.

[40] Richard S. Sutton and Andrew G. Barto. 2018. *Reinforcement learning: an introduction* (second edition ed.). The MIT Press, Cambridge, Massachusetts.

[41] Fangfang Wei, Shoufeng Ma, and Ning Jia. 2014. A Day-to-Day Route Choice Model Based on Reinforcement Learning. *Mathematical Problems in Engineering* 2014, 1 (2014), 646548. https://doi.org/10.1155/2014/646548 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1155/2014/646548

[42] Michael Wunder, Michael L Littman, and Monica Babes. 2010. Classes of multiagent q-learning dynamics with epsilon-greedy exploration. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 1167–1174.

[43] Hyejin Youn, Michael T Gastner, and Hawoong Jeong. 2008. Price of anarchy in transportation networks: efficiency and optimality control. *Physical review letters* 101, 12 (2008), 128701.