CISTER
Research Centre in
Real-Time & Embedded
Computing Systems

Conference Paper

Reinforcement Learning to Reach Equilibrium Flow on Roads in Transportation System

Hajar Baghcheband

CISTER-TR-190612

2019/03/06

Reinforcement Learning to Reach Equilibrium Flow on Roads in Transportation System

Hajar Baghcheband

CISTER Research Centre

Polytechnic Institute of Porto (ISEP P.Porto)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail: hajar@isep.ipp.pt

https://www.cister-labs.pt

Abstract

Traffic congestion threats the vitality of cities and the welfare of citizens. Transportation systems are using various technologies to allow users to adapt and have a different decision on transportation modes. Modification and improvement of these systems affect commuters 19 perspective and social welfare. In this study, the effect of equilibrium road flow on commuters 19 utilities with a different type of transportation mode will be discussed. A simple network with two modes of transportation will be illustrated to test the efficiency of minority game and reinforcement learning in commuters 19 daily trip decision making based on time and mode. The artificial society of agents is simulated to analyze the results.

Reinforcement Learning to Reach Equilibrium Flow on Roads in Transportation System

Hajar Baghcheband
CISTER Research Centre, ISEP,
Polytechnic Institute of Porto
Porto, Portugal
hajar@isep.ipp.pt

***Abstract*—Traffic congestion threats the vitality of cities and the welfare of citizens. Transportation systems are using various technologies to allow users to adapt and have a different decision on transportation modes. Modification and improvement of these systems affect commuters' perspective and social welfare. In this study, the effect of equilibrium road flow on commuters' utilities with a different type of transportation mode will be discussed. A simple network with two modes of transportation will be illustrated to test the efficiency of minority game and reinforcement learning in commuters' daily trip decision making based on time and mode. The artificial society of agents is simulated to analyze the results.**

Keywords—transportation system, minority game, reinforcement learning, multi-agent system, simulation

I. INTRODUCTION

Transport is an activity where something is moved between the source and destination by one or several modes of transport. There are five basic modes of transportation: road, rail, air, water and pipeline [1]. An excellent transport system is vital for a high quality of life, making places accessible and bringing people and goods together. Information and Communication Technologies (ICT) helps to achieve this high-level objective by enhancing Transportation systems with intelligent systems[2].

Among all type of transport, road transport has an important role in daily trips. People are migrated to cities because of the benefits of services and employment compared to rural areas. However, it is becoming harder to maintain road traffic in smooth working order. Traffic congestion is a sign of a city's vitality and policy measures like punishments or rewards often fail to create a long term remedy. The rise of ICT enables the provision of travel information through advanced traveler information systems (ATIS). Current ATIS based on shortest path routing might expedite traffic to converge towards the suboptimal User Equilibrium (UE) state[3].

Traffic congestion is one of the reasons for negative externalities, such as air pollution, time losses, noise, and decreasing safety. As more people are attracted to cities, future traffic congestion levels are not only decreased but also increased and extending road capacity would not solve congestion problems. While private cars maximize personal mobility and comfort, various strategies have attempted to discourage car travel to use public transportation.

To encourage commuters to shift from private car to public transport or intermodal changes, it is exigent to provide a competitive quality to public transport compared to its private counterpart. This can be measured in different aspects such as safety, comfort, information, and monetary cost, but more importantly, travel times compared to those of private cars[4].

Moreover, Policy measures in transportation planning aim at improving the system as a whole. Changes to the system that result in an unequal distribution of the overall welfare gain are, however, hard to implement in democratically organized societies [5]. Different categories of policies can be considered in urban road transportation: negative incentives [6], positive incentives or rewards[7, 8], sharing economy [3, 9].

In recent researches, positive policies were discussed as discount or money payback to commuters. Kokkinogenis et al. [12], discussed a social-oriented modeling and simulation framework for Artificial Transportation Systems, which accounts for different social dimensions of the system in the assessment and application of policy procedures. They illustrated how a social agent-based model can be a useful tool to test the appropriateness and efficiency of transportation policies[12].

Traditional transport planning tools are not able to provide welfare analysis. In order to bridge this gap, multi-agent microsimulations can be used. Large-scale multi-agent traffic simulations are capable of simulating the complete day-plans of several millions of individuals (agents) [10]. A realistic visualization of agent-based traffic modeling allows creating visually realistic reconstructions of modern or historical road traffic. Furthermore, the development of a complex interactive environment can bring scientists to new horizons in transport modeling by an interactive combination of a traffic simulation (change traffic conditions or create emergencies on the road) and visual analysis[11].

Klein et al. developed a multi-agent simulation model for the daily evolution of traffic on the road that the behavior of agents was reinforced by their previous experiences. They considered various network designs, information recommendations, and incentive mechanisms, and evaluated their models based on efficiency, stability and equity criteria. Their results concluded that punishment or rewards were useful incentives[3].

To improve the behavior of agents, reinforcement learning is one of the key means in multi-agent systems. reinforcement learning techniques recently proposed for transportation applications and they have demonstrated impressive results in game playing. Nallur et al. introduced the mechanism of algorithm diversity for nudging system to reach distributive justice in a decentralized manner. They use minority game as an exemplar of an artificial transportation network and their

result showed how algorithm diversity lead to faired reward distribution[19].

The main goal of this study is to develop a model, based on the concept of minority games and reinforcement learning, to achieve equilibrium flow through public and private transportation. Minority game is applied to consider rewards, positive policy, for winner and learning is a tool to increase the user utility based on rewards. To illustrate, an artificial society of commuters are considered instantiated on a simple network with two modes of transportations, namely public (PT) and private (PR).

The remaining parts are organized as follow. In Section II, the conceptual framework will be discussed and consists of a definition of user utility, minority game, and reinforcement learning algorithms. Illustration scenario of network and commuters and initial setup are explained in Section III. Experiments and results are shown in Section IV and related work are reviewed in Section **Error! Reference source not found.**. Conclusion of the hypothesis and results are drawn in SectionV.

II. Description of the Framework

In this section, the theoretical aspects and methodological ones are described, and also network design and model will be discussed.

A. Traffic Simulation

Traffic simulation models are classified into macroscopic and microscopic models. The hydrodynamic approach to model traffic flow is typical for macroscopic modeling. With this kind of approach, one can only make statements about the global qualities of traffic flow. For observing the behavior of an individual vehicle a microscopic simulation is necessary. Because traffic cannot be seen as a purely mechanical system, a microscopic traffic simulation should also take into consideration the capabilities of human drivers (e.g., perception, intention, driving attitudes, etc.)[2].

B. Network Design

The network is formally represented as graph G (V, L) which V is the set of nodes such as *Origin*, *Destination,* and *middle* nodes and L is the set of roads (edges or links) between nodes[3, 12]. Each line $l_k \in L$ has some properties such as mode, length, and capacity. In addition, the volume-delay function is used to describe the congestion effects macroscopically, that is, how the exceeding capacity of flow in a link affects the time and speed of travel, as below [13]:

$$t_k = t_{0k}\,[1 + \alpha\,(X_k/C_k)^{\beta}] \qquad (1)$$

where t_{0k} is free flow travel time, X_k is the number of vehicle and C_k shows the capacity of the link k, α and β are controlling parameters.

C. Commuters Society

Commuters, agents of the artificial society, have some attributes regarding travel preferences such as time (desired arrival time, desired travel time, mode of transportation, mode flexibility), cost (public transportation fare, waiting time cost, car cost if they have), socioeconomic features (income).

They will learn and make a decision for their dairy plan based on their daily expectation and experience. The iteration module generates the demands of the transportation modes and desired time. Daily trips schedule for a given period of the day and define the set of origins and destinations with the respective desired departure and arrival times to and from each node.

The utility-based approach is considered to evaluate travel experience and help agents make decisions. Total utility is computed as the sum of individual contribution as follow and is the combination based on previous researches [5, 12]:

$$U_{total} = \sum_{i=1}^{n} U_{perf,\,i} + \sum_{i=1}^{n} U_{time,\,i} + \sum_{i=1}^{n} U_{cost,\,i} \qquad (2)$$

where U_{total} is the total utility for a given plan; n is the number of activities, which equals the number of trips (the first and the last activity are counted as one); $U_{Pref,\,i}$ is the utility earned for performing activity i; $U_{time,\,i}$ is the (negative) utility earned by the time such as travel time and waiting time for activity i; and $U_{cost,\,i}$ is the (usually negative) utility earned for traveling during trip i.

1) Performance Utility

To measure the utility of selecting activity *i,* each mode of transportation has different variables. For public mode, comfort level and bus capacity, and for private, pollution and comfort level are considered.

2) Time Utility

The measurement of the travel time quantifies the commuter's perception of time based on various components like waiting and in-vehicle traveling. Waiting time indicates the service frequency of public transportation. In-vehicle traveling time is an effective time to travel from origin to destination.

3) Monetary Cost Utility

Monetary cost can be defined as fare cost of public transportation, cost of fuel, tolls (if exists), car insurance, tax, and car maintenance. This kind of cost will be measured based on the income of commuters.

Regard to three different utilities, the total utility of public and private can be measured as follow:

$$U^{total}_{private} = \sum_{i=1}^{N} U^{i}_{private} \qquad (3)$$

$$U^{i}_{private} = \left(\alpha_{time} * (t^{i}_{tt,exp}/t^{i}_{tt})\right) + \left(\beta_{PR} * (cost\_PR/income_i)\right) + \left(\alpha_{pollution} * t^{i}_{tt} * pollution\right) + \alpha_{com\_PR} * \left(t^{i}_{tt,exp}/t^{i}_{tt}\right) \qquad (4)$$

$$U^{total}_{public} = \sum_{i=1}^{N} U^{i}_{public} \qquad (5)$$

$$U^{i}_{public} = \left(\alpha_{time} * (t^{i}_{tt,exp}/t^{i}_{tt})\right) + \left(\beta_{PT} * (cost\_PT/income_i)\right) + \alpha_{com\_PT} * \left(t^{i}_{wt,exp}/t^{i}_{wt}\right) + \left(\alpha_{cap} * t^{i}_{tt} * capacity^{\,i}_{exp}/bus\_capacity\right) \qquad (6)$$

where t^{i}_{tt} and $t^{i}_{tt,exp}$ are total travel time and expected total travel time of agent i, *cost_PR* is the monetary cost of private transportation (fuel, car maintenance and etc.), *cost_PT*, the fare of public transportation, $income_i$, the agent's income per day, *pollution* is the amount of pollution is produced by private vehicles, $capacity^{\,i}_{exp}$, and *bus_capacity* are expected capacity of the bus and the total capacity of each bus respectively, $t^{i}_{wt,exp}$, expected waiting time and t^{i}_{wt} is the waiting time by agent i. α_{time}, β_{PT}, β_{PR}, $\alpha_{pollution}$, α_{com_PR}, α_{com_PT} and α_{cap} are considered as marginal utilities or preferences for different components.

D. Minority Game

The minority game, introduced by Challet and Zhang (1997)[14], consisting of N agents (N is an odd number). They have to choose one of two sides independently and those on

the minority side win. Winner agents get reward points, nothing for others. Each agent draws randomly one out of his *S* strategies and uses it to predict the next step. To choose what strategy to use each round, each is assigned a score based on how well it has performed so far, the one with the leading score is used at a time step.

It was originally developed as a model for financial markets, although it has been applied in different fields, like genetics and transportation problems[15]. While simple in its conception and implementation, it has been applied in various fields of transportation such as public transportation[16], route choosing[17], road user charging scheme[18]. It can be useful in traffic management which travelers try to find less crowded and congestion roads.

E. Reinforcement Learning Method

Reinforcement learning (RL) is a class of machine learning concerned with how agents ought to take actions in an environment so as to maximize cumulative reward [19]. Alvin Roth and Ido Erev developed a new algorithm, which is called "Roth-Erve"[20], to model how humans perform in competitive games against multiple strategic players. The algorithm specifies initial propensities (q_0) for each of N actions and based on reward (r_k) for action (a_k) the propensities at the time ($t+1$) are defined as[20]:

$$q_j(t+1) = (1-\varphi)\, q_j(t) + E_j(\varepsilon, N, k, t) \tag{7}$$

$$E_j(\varepsilon, N, k, t) = \begin{cases} r_k(t)[1-\varepsilon] & \text{if } j=k \\ r_k(t) * (\varepsilon/N-1) & \text{otherwise} \end{cases} \tag{8}$$

Where φ is recency as forgetting parameter and ε is exploration parameter. The probability of choosing action j at time t is:

$$P_j(t) = q_j(t) / \sum_{n=1}^{N} [q_n(t)] \tag{9}$$

III. Illustrative Scenario

In the simulation step, the perspective of the conceptual framework was considered a simple scenario where commuters make a decision over transportation mode and time during morning high-demand peak hour. Simulation model implemented through NetLogo [21] agent-based simulation environment.

A. Network and Commuters

In this study, two different links of two modes (PT or PR) consist of two middle nodes on each link. As it is shown in Fig. 1, to simplify the upper link is for private and the other for public transportation where each road is composed of one-way links.

Commuters, as type of agents, is defined by a number of state variables which are: (1) desired departure and arrival times, (2) experienced travel time, (3) the uncertainty they experienced during the trip with a given transportation mode, (4) a set of preferences about the transportation mode, (5) the perceived comfort as personal satisfaction for the mode choice, and (6) a daily income variable. While the agent experiences its travel activities, the costs associated with the different transportation mode, the perceived satisfaction of traveling (expressed in terms of travel times and comfort) and rewards earned by winners will have a certain impact on its mode and time choices.

Commuters can choose between traveling by PT or PR modes based on own-car value. The decision-making process of each agent is assumed to maximize the utility and flow equilibrium on roads. They perceive current traffic condition as well as previous experience and use this information in making a decision.

At the end of the travel each commuter stores the experienced travel time, costs, and crowding level (for PT mode users only) and emissions. These variables will be used to calculate the following day's utility. After that each agent evaluates its own experience, comparing the expected utility to the effective utility.

Based on minority game, we considered the number of commuters on each road and type of transportation and regard to Roth-Erve learning, the reward assigned to the winner who is in minority number and has the following criteria:

1) Their obtained utility ($U_{effective}$) is greater than the utility prediction ($U_{expected}$) as below:

$$U_{effective} > \alpha * U_{expected} \tag{10}$$

α is marginal preference.

2) The obtained utility of agent is higher than mean utility in the whole network :

$$U_{effective} \geq U_N \quad \text{where } U_N = \frac{1}{N} \sum_{i=1}^{N} U^i_{effective} \tag{11}$$

Based on reward, effective utility they earned in their daily trip, car-ownership and mode-flexibility, each commuter decides about their new mode and time.

B. Initial Setup

As it is written in TABLE I, The capacity for all links was considered 150 and max capacity for each bus was 70 people. Population consist of 201 commuters was created, the odd number to coordinate with minority game, and they iterated their daily trips in 60 days. They were characterized by a number of attributes such as departure and arrival times, mode, daily income, car-ownership, and flexibility. Car-ownership is a Boolean variable and indicates if the agent is a private or public transportation user. Flexibility reflects the willingness of a private mode user to change its mode.

All agent's plans were done in rush hours of the day from 6:30 a.m. to 10:30 a.m., with a normal distribution to simulate peak times. It was observed a high demand in peak duration between 8- 9:30 a.m., on both roads. The range of income was 20 to 70 Euro per day. The routes between nodes *Origin* and *Destination* had both a length of 19 km.

The free-flow travel time from node *Origin* to *Destination* was approximately 25 minutes in the PR mode and for the public transportation, around 35 minutes plus the waiting time at the bus stop and walking time. The bus frequency service was 10 minutes before the rush hour and 5 minutes during the rush hour.

IV. Results and Experiments

We performed sixty iterations of the model, Roth-Erve learning was used to establish the equilibrium commuters between both roads along the departure time interval. During simulation steps, we monitored agents' expected and effective utilities, average travel times of public and private transportation, average total travel times, number of

commuters of each mode and differences between the average of total travel time in public and private transportation.

TABLE I. DEFAULT VALUE OF NETWORK AND LEARNING PARAMETERS

Variable	Value
Number of commuters	N=201
Capacity of links	L_i= 150
Capacity of bus	B=70
Time	6:30 a.m. to 10:30 a.m.
Range of income	20 to 70 € per day
Simulation period	60 days
Recency (φ)	0.3
Exploration (ε)	0.6

The propensity of commuters to select public and private were set by normal distribution random and updated based on recency and exploration learning parameters. Earned scores and two propensities were observed during all days.

In Fig. 2, the distribution of all commuters is depicted among all days, where green line and the red line show the number of agents on roads with public and private transportation respectively. The number of commuters on different modes of transportation converged by use of learning tools and rewards. In the first day, most of the people had a tendency to use public transportation, which was decreased in the last day of the simulation period.

Total time of daily trips for both mode of transportation which was selected by agents, measured and the differences between these two times for all day long was calculated. In Fig. 3, the result is shown for the simulation period. This fluctuation was related to different factors such as traffic on road, departure time and waiting time for public transportation each day. However, in final days, the difference time between public and private transportation was less than 10 minutes by reaching equilibrium flow on transportation mode and, it seemed to be stable.

Based on rewards and decision making of departure time and transportation mode, the commuters' utilities were changed daily. Fig. 4 represents daily changes in effective utilities earned by each commuter among the whole period. In this chart, it is shown that both public and private utilities were increased with day-to-day variation.

The number of commuters on different modes, effective utility, average time of public and private and average of total time which were observed at last day are described in TABLE II , and it shows the average of total travel time of each mode were roughly similar to the average of total travel time of both modes and effective utilities of commuters had a bit difference with ones they expected.

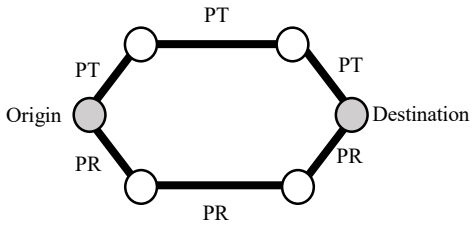

Fig. 1. Network

TABLE II. THE RESULT OF THE LAST DAY

Variable	Value
No. of commuters on public transportation	102
No. of commuters on public transportation	99
Average total time on public transportation	34.26 min
Average total time on private transportation	25.14 min
Average total time of both mode	29.771 min
Average effective utility on public transportation	15.502
Average effective utility on private transportation	14.920
Average expected utility on public transportation	15.710
Average expected utility on private transportation	15.012

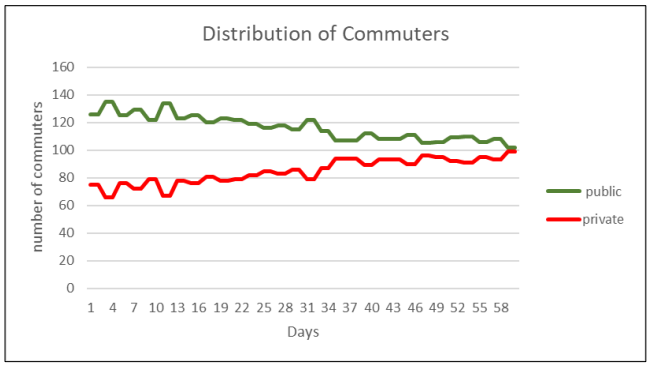

Fig. 2. Distribution of commuters on roads

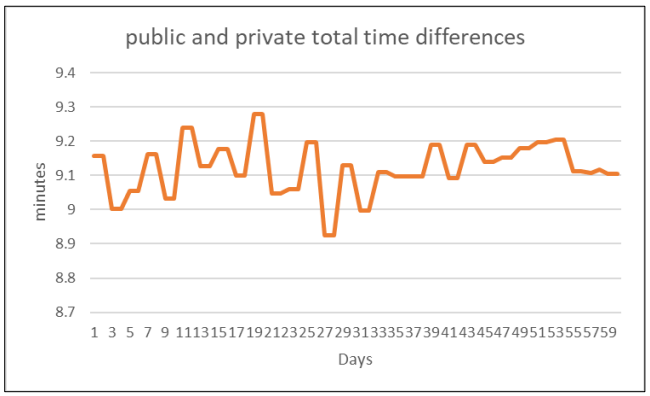

Fig. 3. Public and private total time differences

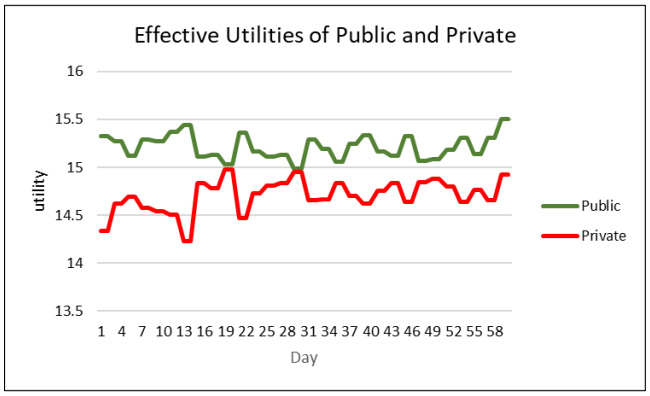

Fig. 4. Effective Utilities of Public and Private

V. Discussion and Conclusion

In this paper, we have proposed the framework for evaluating the reinforcement learning and effect of minority games on equilibrium flow on roads. We suggested applying agent-based modeling and simulation as a platform to implement our framework.

To illustrate, a simple network consists of two different types of mode (PT and PR) were considered and a population of commuters with the memory of travel experiences was generated. They performed their daily plan in morning high-demand hours and their activities iterated for sixty days. Their experience, expected and effective utilities, expected and effective travel time and rewards were observed and analyzed.

In regard to results, the commuters learned to predict total travel time in both modes which their exception was similar to obtained total travel time in each mode. By balancing number of commuters on each type of transportation, they gained higher utilities rather than first days.

From the illustrative example, the hypothesis of the study, which was to use RL and minority game to reach equilibrium flow, was reached and it is concluded that equilibrium flow can follow higher utilities and more precise time prediction in daily trips.

For future work, we will consider a realistic large-scale network and demands, different types of incentives and roads with combination type of transportations so as to better study and analysis of commuters' behavior and performance of the transportation system. With such improvements, we are confident that our framework can be proper and accurate to increase the commuters' pleasant and also the performance of the road transportation system.

Acknowledgment

This work was partially supported by National Funds through FCT/MCTES (Portuguese Foundation for Science and Technology), within the CISTER Research Unit (UID/CEC/04234).

References

[1] M. Ebrahimi, "License Plate Location Based on Multi Agent Systems," *Intell. Eng. Syst.*, 2007.

[2] K. Modelewski and M. Siergiejczyk, "Application of multi-agent systems in transportation," *Appl. Multiagent Syst. Transp.*, 2013.

[3] I. Klein, N. Levy, and E. Ben-Elia, "An agent-based model of the emergence of cooperation and a fair and stable system optimum using ATIS on a simple road network," *Transp. Res. Part C Emerg. Technol.*, vol. 86, no. November 2017, pp. 183–201, 2018.

[4] M. Tlig and N. Bhouri, "A multi-agent system for urban traffic and buses regularity control," *Procedia - Soc. Behav. Sci.*, vol. 20, pp. 896–905, 2011.

[5] D. Grether, B. Kickhöfer, and K. Nagel, "Policy Evaluation in Multiagent Transport Simulations," *Transp. Res. Rec. J. Transp. Res. Board*, vol. 2175, no. 1, pp. 10–18, 2010.

[6] J. Rouwendal and E. T. Verhoef, "Basic economic principles of road pricing: from theory to applications," *Transp. Policy*, vol. 13, pp. 106–114, 2006.

[7] E. Ben-Elia and D. Ettema, "Rewarding rush-hour avoidance: a study of commuters' travel behavior.," *Transp. Res. Part A Policy Pract.*, vol. 45, p. 567–582., 2011.

[8] M. C. J. Bliemer and D. H. van Amelsfort, "Rewarding instead of charging road users : a model case study investigating effects on traffic conditions," *Eur. Transp.*, vol. 44, pp. 23–40, 2010.

[9] A. M. Kaplan and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," *Bus. Horiz.*, vol. 53, no. 1, pp. 59–68, 2010.

[10] M. N. K. Grether, D; Chen, Y.; Rieser, "Effects of a simple mode choice model in a large-scale agent-based transport simulation.," in *Complexity and Spatial Networks. In Search of Simplicity, Advances in Spatial Science,* P. Reggiani, A.; Nijkamp, Ed. Springer, 2009, pp. 167–186.

[11] K. Golubev, A. Zagarskikh, and A. Karsakov, "A framework for a multi-agent traffic simulation using combined behavioural models," *Procedia Comput. Sci.*, vol. 136, pp. 443–452, 2018.

[12] Z. Kokkinogenis, N. Monteiro, R. J. F. Rossetti, A. L. C. Bazzan, and P. Campos, "Policy and incentive designs evaluation: A social-oriented framework for Artificial Transportation Systems," *2014 17th IEEE Int. Conf. Intell. Transp. Syst. ITSC 2014*, pp. 151–156, 2014.

[13] J. de D. Ortúzar and L. G. Willumsen, *Modelling Transport*. Chichester: John Wiley & Sons, Ltd, 2011.

[14] D. Challet and Y. C. Zhang, "Emergence of cooperation and organization in an evolutionary game," *Phys. A Stat. Mech. its Appl.*, vol. 246, no. 3–4, pp. 407–418, 1997.

[15] A. Physics, "The Minority Game : evolution of strategy scores," 2015.

[16] P. C. Bouman, L. Kroon, P. Vervest, and G. Maróti, "Capacity, information and minority games in public transport," *Transp. Res. Part C Emerg. Technol.*, vol. 70, pp. 157–170, Sep. 2016.

[17] T. Chmura, T. Pitz, and M. Schreckenberg, "Minority Game - Experiments and Simulations," in *Traffic and Granular Flow '03*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 305–315.

[18] T. Takama and J. Preston, "Forecasting the effects of road user charge by stochastic agent-based modelling," *Transp. Res. Part A Policy Pract.*, vol. 42, no. 4, pp. 738–749, May 2008.

[19] V. Nallur, E. O. Toole, N. Cardozo, and S. Clarke, "Algorithm Diversity - A Mechanism for Distributive Justice in a Socio-Technical MAS," in *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, 2016, pp. 420–428.

[20] Roth AE and Erev I, "Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term," *Games Econ. Behav.*, vol. 8, no. 1, pp. 164–212, 1995.

[21] Uri Wilensky, “Netlogo: Center for connected learning and computerbased modeling.” Northwestern University, 1999.