Received October 18, 2019, accepted October 22, 2019, date of publication October 25, 2019, date of current version November 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2949576

Learning Towards Transportation Network Equilibrium: A Model Comparison Study

JIANGUO SONG[1], HANG QI[2], AND GUANGCHAO WANG[3]

[1]School of Information Management, Wuhan University, Wuhan 430072, China
[2]Institute for Advanced Studies in Finance and Economics, Hubei University of Economics, Wuhan 430205, China
[3]School of Information Management, Central China Normal University, Wuhan 430079, China

Corresponding author: Hang Qi (qihang@hbue.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 71801106, and in part by the Science Foundation of Ministry of Education of China under Grant 17YJC630150.

ABSTRACT As an interdisciplinary topic, human travel-choice behavior has attracted the interests of transportation managers, theoretical computer science researchers and economists. Recent studies on tacit coordination in iterated route choice games (i.e., a large number of subjects could achieve the transportation network equilibrium in limited rounds) have been driven by two questions. (1) Will learning behavior promote tacit coordination in route choice games? (2) Which learning model can best account for these choices/behaviors? To answer the first question, we choose a set of learning models and conduct extensive simulations to determine their success in accounting for major behavioral patterns. To answer the second question, we compare these models to one another by competitively testing their predictions on four different datasets. Although all the selected models account reasonably well for the slow convergence of the mean route choice to equilibrium, they account only moderately well for the mean frequencies of the round-to-round switches from one route to another and fail to appropriately account for substantial individual differences. The implications of these findings for model construction and testing are briefly discussed.

INDEX TERMS Route choice game, laboratory experiment, reinforcement learning, tacit coordination, nash equilibrium.

I. INTRODUCTION

In both transportation and communication networks, where the route choices are decentralized, utility-maximizing players facing strategic uncertainty often strive to avoid congestion [1]. Examples include choosing a restaurant on Saturday evening, selecting of a route in a traffic network, and deciding whether to enter a capacitated market. The notion of equilibrium in such scenarios, once they are modeled appropriately as non-cooperative n-person games, leads us naturally to ask how players achieve this "meeting of the minds."

The focus of the present paper is on the choice of routes in directed networks. We focus on computer-controlled experimental studies of a class of network games, called iterative route choice games. These games have multiple equilibria that, depending on the architecture of the network and the number of network users, are counted in thousands or occasionally in millions. The study of such games falls in the intersection of behavioral economics, transportation science [2], computer science [3], and operations management [4].

If tacit coordination in large groups is neither reached by communication nor deduced by introspection, then it is achieved by learning "day by day" [5], [6]. Most previous experimental studies of route choice games largely support this assertion [3]. Variants of Markov adaptive learning models [7], simplified versions of the experience weighted attraction (EWA) learning model [8] and ruled-based learning models [9] been considered separately to account for the dynamics of play in route choice experiments. Our purpose in this paper is to test representative learning models competitively (e.g., [10], [11]). We wish to determine *which learning models best describe the adjustment process over multiple iterations of the stage route choice game.*

In this paper, we compare each competitive model in a particular route-choice context in two stages. In the first stage of our procedure, we select experimental datasets that share the same context and same game type, i.e., non-cooperative n-person games on route choice in directed networks. In the second stage, we select model candidates which exhibit

The associate editor coordinating the review of this manuscript and approving it for publication was Zhengbing He.

alternative approaches to the study of learning in games according to their potential to explain the robust observations commonly emerging in the selected datasets.

The selection of datasets is achieved in the second stage. We selected four datasets that vary from one another in their research purpose, the architecture of the network, sources of uncertainty, and the number of iterations of the stage game, and then compared them to one another in terms of the following two behavioral regularities. These include (1) *gradual convergence* to equilibria, as the mean route choice frequencies approach – but do not necessarily reach – an equilibrium point that is unique up to permutations of the players; and (2) *non-increasing fluctuations over time*, namely, changes in the mean number of round-to-round switches from one route to another.

Previous experimental studies of route choice have demonstrated the moderate success of learning models in accounting for the gradual convergence of choices to equilibrium (1), whereas in this paper we attempt to further explore the adjustment process by studying the additional behavioral regularity (2). Accordingly, we have selected representative learning models that have the potential to account simultaneously for coordination success and the dynamics of switching routes at the aggregate level. These include the Markov adaptive learning model (denoted by MAL) [7], the experience weighted attraction (EWA) learning model [11] and two special cases of the EWA model, namely, reinforcement learning (RL) [12] and belief learning (BL).

After comparing the goodness-of-fit of learning models to account for the subjects' decisions, we further explored our first research question, namely, *whether learning could lead the group to achieve coordination*. To this end, we conducted simulations of the learning models to test how well they replicate observations (1) and (2) regarding the *dynamics* of play [14]. To the best of our knowledge, the present paper is the first to report a systematic comparison of learning models in the context of route choice games in directed transportation networks. We benchmark earlier studies on the comparison of learning models by [12] but depart from them in the following ways.

Not only do we evaluate the goodness-of-fit of models using the methodology of maximum likelihood estimation via several statistical criteria, but we also test their predictive power of replicating the dynamics of play, thereby subjecting the models to additional stress. In model comparison, we conduct in-sample fitting and out-of-sample testing instead of using the entire dataset to estimate the model parameters. In replicating the choices, we further compared the changes in the simulated distribution of choices and simulated switch proportions over time with the experimental results.

The potential implications of results of this paper include, but not limited to, better understanding the bounded rational behavior of commuters and thus designing routing guaidance strategy given their most possible response patterns to informations and their last experience [13], [14].

The rest of the paper is organized as follows. Section II first gives a brief literature review. Section III introduces the route choice game on directed networks, the datasets and the learning models examined in this study. In Section IV, we discuss the methodology of our model comparison and the results of the goodness-of-fit measures. Section V presents a summary of the results of our model comparison, and Section VI concludes.

II. A BRIEF LITERATURE REVIEW

Considered in the present study are two lines of experiments on route choice in directed networks iterated over time, namely, experiments which focus primarily on the realization of the Braess Paradox [15], and experiments that focus primarily on testing for convergence to equilibrium under different conditions of network topology and information structure. For theoretical studies that introduce route choice games and investigate their properties see [3] and for a review of experimental studies see [6].

The Braess Paradox illustrates quite dramatically that a structural change of the network topology, while the number of network users and the link cost functions are kept fixed, may result in counterintuitive consequences. Experimental, rather than theoretical, research aimed to determine if, and under what conditions, the BP may be realized behaviorally under the fully controlled conditions of the laboratory commenced by [16]. The general findings of this stream of research is that when the stage network game is iterated in time, the players approach to the Pareto deficient equilibrium solution, route switches from round to round diminish with time but do not disappear completely, and individual differences among the group members in the number of switches are substantial. Reinforcement-based learning models and minimization of regret learning models have been tested individually in different experiments but with only moderate success.

Rather than changing the architecture of the network by adding (or deleting) one or more links within a given session, the second stream of experiments on route choice has focused on the social dilemmas in ridesharing [17], the protocol of play (simultaneous vs. sequential choice of routes) [18], the type of uncertainty that the players face [19], and the information structure [19], [20]. A major result of this stream of studies, that complements the results of experiments on the BP, is the robustness of the equilibrium solution as a descriptive model of the aggregate outcomes.

III. DATA AND MODELS

In this section, before introducing the selected experimental datasets, we first introduces the route choice games on directed networks with basic notations.

A. ROUTE CHOICE GAMES ON DIRECTED NETWORKS

We consider in this paper directed networks with a common origin and common destination that are modeled by a graph $G(V, E)$ comprising of a set V of vertices (nodes), a set E

of edges (links, arcs), and a set $K \subseteq (V \times E)$ of origin-destination (OD) routes. The network $G(V, E)$ is *undirected* if for all $p, q \epsilon V$, $(p, q) \epsilon E \leftrightarrow (q, p) \epsilon E$. Otherwise, it is *directed*. The network serves a finite and commonly known number of players (users) n. We denote the finite action space (pure strategies) of each player $i \epsilon N$ by S_i, consisting of m_i discrete choices of routes: $S_i = (s^1, s^2, \ldots, s^{m_i-1}, s^{m_i})$. The joint action space of the game is denoted by $S = S_1 \times \ldots \times S_n$. A strategy combination is denoted by $s = (s_1, \ldots, s_n)$, whereas $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n)$ is a combination of the strategies of all players except player i. We denote by f_{jk} ($f_{jk} = 1, 2, \ldots, n$) the number of players who choose the link jk from vertex j to vertex k in the network G (V, E). In our formulation of the game, the cost of travel incurred by player i in choosing the link jk is denoted by $c_i(f_{jk})$; it is used as a proxy for the delay in traversing link jk. As the n group members are assumed to be homogenous, $c_i(f_{jk})$ has the same value for all the n players. The total individual cost for choosing a route k from O to D, denoted by C_i, is the sum of the costs of links along this route.

In route choice games with negative externalities, also known as *congestion games*, $c_i(f_{jk})$ is increasing in f_{jk}. Mostly for experimental reasons, costs of travel are often modeled by affine functions:

$$c_i(f_{jk}) = a_{jk}(f_{jk}) + b_{jk}, \quad \text{where } a_{jk}, b_{jk} \geq 0. \tag{1}$$

The fixed constant b_{jk} is interpreted as the minimum delay to traverse link jk in the absence of traffic, whereas the variable cost a_{jk} models the effect of congestion [6]. Players are assumed to independently choose an OD route in the set K that minimizes their cost of travel.

B. DATASETS

Our datasets report the results of several experiments conducted by Rapoport, Mak, Gisches, and their collaborators from 2012 to 2015 [7], [8], [20]. Data were collected from a total of 20 sessions, each including 18 or 20 players. Overall, 370 persons participated in four computer-controlled experiments. All were university students, who volunteered to participate in decision-making experiments for payoff contingent on their performance. Payments were cumulative and made in tokens which at the end of the session were converted into US dollars. The mean payment across sessions within the same game ranged between \$19.06 and \$29.50, including a show-up fee of \$5. Each session lasted about two hours. Communication among the subjects was not allowed.

Table 1 displays the design features of the four datasets and the learning models that were tested in the original papers. In all the four experiments, players were symmetric, the games were conducted under complete information of the cost functions and round-to-round travel time distributions, and the individual decisions were submitted under the simultaneous protocol of play. Although each experiment included more than a single condition, we only analyze the data from a single condition in each experiment with complete information, simultaneous play, and choice of OD routes at the network origin.

TABLE 1. summary of the data sets examined in this study.

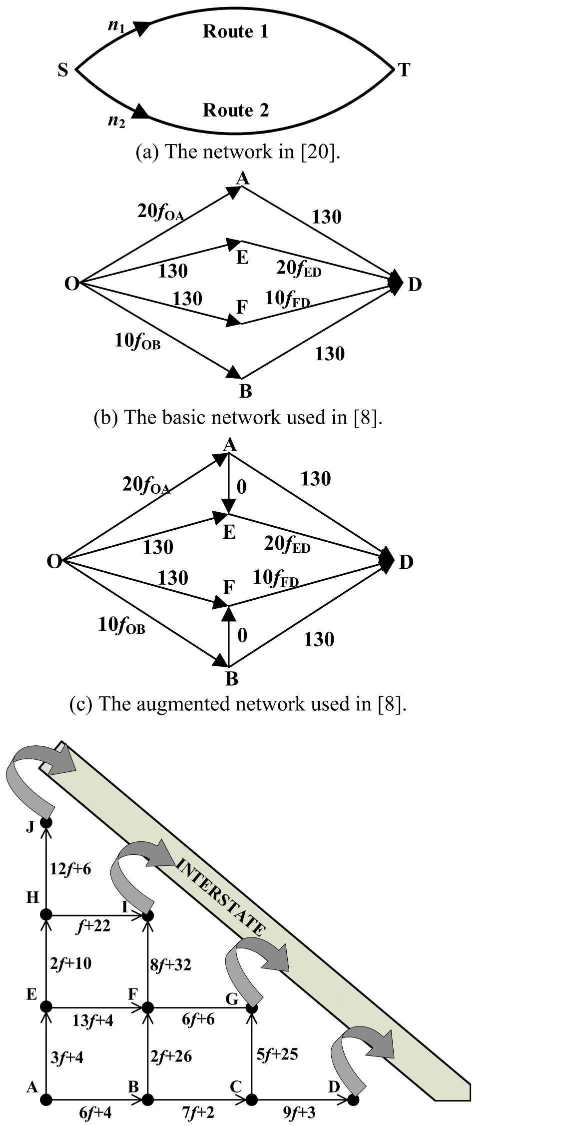
Datasets[a]	Setting of Experiments No. Players	No. Iterations	No Routes	Descriptive Model
Game 2R ([20], fully correlated treatment with full information)	20	80	2	A single-segment two-parameter EWA learning
Game 4R ([8], Condition PUBLIC in the basic network)	18	60	4	A single-segment two-parameter EWA learning with equilibrium payoff as a fixed aspiration level
Game 6R ([8], Condition PUBLIC in the augmented network)	18	60	6	A single-segment two-parameter EWA learning with equilibrium payoff as a fixed aspiration level
Game 8R ([7], Condition RCC)	18	50	8	A Markov adaptive learning model that incorporates regret effect, inertia, and recency effect.

[a] Names of the data sets with a possible abbreviation are in bold, and reference are in parenthesis.

Our choice of datasets is not fortuitous; they have been selected to cover different scenarios. First, the games differ from one another in the topology of their networks, ranging from networks with only two parallel routes that do not intersect each other in Game 2R to networks with eight nonparallel routes in Game 8R (currently the most complex network implemented in route choice experiments). Fig. 1 exhibits the four networks. Except for the network in Game 2R, all the networks induce nontrivial strategy spaces that provide a challenge to learning models. Second, the cost functions in Game 2R are nonlinear, and the players face both strategic and environmental uncertainties. Specifically, travel conditions on the two routes in Game 2R are perfectly and positively correlated, and their cost functions depend on the weather conditions. In contrast, the uncertainty in the other games is only strategic. Third, the four games differ from one another in the number of iterations (rounds) of the stage game. Games with more rounds provide more data for the estimation of model parameters and for detecting individual differences in the frequency of switches, whereas games with fewer rounds decrease boredom and fatigue. Analyzing data jointly from both longer and shorter games places the learning models under yet another source of stress.

C. ALTERNATIVE LEARNING MODELS

It is not plausible that theories of learning alone may explain strategic choices in iterated interactive decisions without the input of behavioral regularities and careful testing of the theories [21], [22]. Moreover, because learning is a complex



(a) The network in [20].

(b) The basic network used in [8].

(c) The augmented network used in [8].

(d) The eight-route network in the experiment conducted by [7].

FIGURE 1. **The networks presented to the players for the selected studies.**

cognitive process that depends on the rules of the game, strategy spaces, number of iterations of the stage game, size of the group of players, degree of sophistication of the players, and their motivation, it is equally unlikely that a single model may account for learning under all circumstances in *all* types of games. Consequently, mostly in psychology and subsequently in experimental economics, alternative approaches to learning in games have been proposed, and the models implementing these approaches have been experimentally tested [12]. Among others, they include evolutionary dynamics, reinforcement learning, belief learning; minimization of regret learning, direction learning, and rule learning.

For the competitive testing of learning models in route choice games, we have chosen four models that we describe in detail below. They include a Markov Adaptive Learning (MAL) model that has been proposed and tested [7]. The MAL model is predicated on the assumption that route choices in iterated games are made in an attempt to minimize regret for choosing a route, which is not the best response to the choices of the other group members. The second is the Experience-Weighted Attraction (EWA) model first proposed by [12] and subsequently tested extensively on multiple sets of data mostly collected in matrix games. As noted by [11], reinforcement learning models, which originate in mathematical psychology, assume that players ignore information about foregone payoffs, whereas belief learning models assume that players ignore information about their previous choices. EWA is a hybrid of reinforcement- and belief-based model that incorporates both sources of information. The third (RL) and fourth (BL) models are special cases of the EWA model that either negate the effects of beliefs or negate the effects of reinforcement, respectively.

1) NOTATIONS OF LEARNING MODELS

We begin this section by introducing notation that is shared by the learning models. In the route choice game, players strive to minimize their individual cost of travel by choosing at its origin one of the OD routes of the network. Denote by $C_i(j, t)$ the travel cost that player i incurs if she chooses route j on round t of the game, given the choices of all other players s_{-i} on round t. $C_i(j, t)$ is computed separately for each player i by summing the costs that she incurs in traversing the segments of the route she has chosen. At the end of a round, each player receives a reward E, the same for all the n players, for successfully choosing one of the OD routes. The individual payoff at the end of each round t, which is assumed to be commonly known, is computed from

$$\pi_i(j, t) = E - C_i(j, t), \quad i = 1, 2, \ldots, n. \tag{2}$$

As noted earlier, π_i may assume either positive or negative values.

Next, for each player i and each OD route k, $k \neq j$, we compute the player's *counterfactual travel cost*, namely, the cost that she would have incurred had she chosen route k rather than route j, given the actual route choices of the other n-1 group members on round t. The corresponding *counterfactual payoff* is then denoted by $\pi_i(k, t)$. Clearly, the difference score $\pi_i(j, t)$ - $\pi_i(k, t) > 0$ indicates that route j was a superior choice for player i than route k on round t, whereas $\pi_i(j, t)$ - $\pi_i(k, t) < 0$ indicates the opposite. Route j is said to be player i's *best response* on round t, if $\pi_i(j, t)$ - $\pi_i(k, t) > 0$ for all routes k.

Suppose that the n players are about to start round $t + 1$, $t = 1, 2, 3\ldots, T$-1. Adaptive learning models comprise three major processes that prescribe, respectively, generation of the "attractions" (or "propensities" towards choosing) of all the strategies based on the most recent outcome (round t), weight assignment of all the historical attractions during their experience (on rounds t-1, …, 1), and transformation mapping from the attractions, $\mathbf{A}_i(t)$, into probabilities of the strategies to be chosen on round $t + 1$, $\mathbf{P}_i(t + 1)$. Adaptive learning models differ from each other in the generation of attractions from the player's decision history. Three forms of mapping function are commonly used, i.e., exponential (logit), power, and normal (probit). Because of its advantages in addressing negative attractions and its better fitness compared with the other forms, the logit-form transformation function is applied by all the alternative models. Next, the probability of player i choosing route j in round $t+1$ is computed from the multinomial logit function:

$$P_i^j(t+1) = \frac{\exp(\lambda \cdot A_i^j(t))}{\sum_{k=1}^{m_i} \exp(\lambda \cdot A_i^k(t))}, \quad i \in N \tag{3}$$

where λ is a measure of the sensitivity of the players to payoffs to be estimated and is expected to be positive (when $\lambda = 0$ all the routes are equally likely to be chosen, and when $\lambda = 1$ the strategy with the highest attraction is always chosen). In all the datasets, the players are symmetric with the same set of alternative routes connecting the single OD pair on a network. That is, $m_1 = m_2 = \cdots = m_n$, which means that for all the n players, the total number of strategies is equal to the total number of routes in the experimental networks.

2) MARKOV ADAPTIVE LEARNING MODEL WITH REGRET AND INERTIA

To capture the rapid convergence to the equilibrium solution in Game 8R, [4] proposed a Markov adaptive learning (MAL) model postulating regret minimization and inertia. To evaluate the strategies based on present outcomes, the MAL model introduces reference-dependence effects. Specifically, the player's actual choice and payoff in round t serve as reference points. The evaluation of the actual chosen strategy j is represented by a time-invariant parameter λ_0, indicating inertia with respect to the actual route choice in round t over and above the payoff considerations. The impact of an unchosen strategy k is positively proportional to the difference between the actual and counterfactual costs, which is magnified by a nonnegative regret parameter λ_+. This assumption is based on the players' inclination to minimize regret. As for the weight assignment among all historical evaluations, only the most recent round counts in the MAL model because of its consideration of recency effects. Therefore, the updating rule of attraction of any strategy k for player i after each round t is specified as:

$$A_i^k(t) = \begin{cases} C_i(j,t) - C_i(k,t) \\ + \frac{\lambda_+}{\lambda} \max[0, C_i(j,t) - C_i(k,t)], & if\ k \neq j \\ \frac{\lambda_0}{\lambda}, & if\ k = j \end{cases} \quad for\ \forall k \in S_i \tag{4}$$

where λ, λ_+, λ_0, are three parameters to be estimated for this model and are all expected to be nonnegative. Note that λ is the same parameter in the multinomial logit function in (4).

In the MAL model, low-cost routes attract more switches in their direction than in a model without postulated reference-dependence effects. Thus, in the early periods of the game, when cost differences among routes are considerable, the regret parameter λ_+ speeds up the elimination of cost differences thereby induce faster convergence. With inertia, once players move toward the equilibrium choice distribution, they have an additional tendency to stick with it compared to the case of no inertia. Thus, during later periods,the inertia parameter λ_0 serves to effectively stabilize convergence.

3) THE EWA LEARNING MODEL

The multiparametric EWA learning model proposed by [10] is selected because of its generality, simplicity and most importantly its relative success in accounting for learning in three out of the four route choice games that we consider. In their original papers [10], the authors applied simplified variants of the EWA model with $\phi = 1$ and $\kappa = 1$, and successfully reproduced some of the main findings in terms of mean route-choice distribution over time and observed differences between the conditions.

In the EWA model, with an indicator function $I(k,j)$ that equals to 1 if $k = j$, and 0 otherwise, the attraction of route k at the end of round t is updated as:

$$A_i^k(t) = \frac{\phi \cdot N(t-1) \cdot A_i^k(t-1) + [\delta + (1-\delta) \cdot I(k,j)] \cdot \pi_i(k,t)}{N(t)} \quad for\ \forall k \in S_i \tag{5}$$

where the imagination parameter δ measures the relative weight between foregone payoffs and actual payoffs, the parameter ϕ is an index of (perceived) changes that regulates the decay of past attractions, and the parameter $N(t)$ is the weighted experience. It is updated by

$$N(t) = \phi \cdot (1-\kappa) \cdot N(t-1) + 1, \quad for\ \forall t \geq 1 \tag{6}$$

where the parameter κ determines the growth rate of attractions. When $\kappa = 1$, attractions cumulate past payoffs quickly, while if $\kappa = 0$, attractions are weighted averages of lagged attractions and past payoffs.

In total, four parameters (δ, ϕ, κ, λ) of the EWA model are to be estimated. In addition, we are also interested in the explanatory power of two special cases of the EWA model, namely, reinforcement learning (RL) and belief-based learning (BL) with parameters $\delta = 0$ and $\delta = 1$, respectively.

We estimate the RL and BL models using the same initialization procedure and estimation methodology as with the EWA model.

IV. MODEL COMPARISON

A. METHODOLOG

We start this section by describing the methodologies associated with parameter estimation and model evaluation. The method of maximum likelihood estimation (MLE) was used to estimate the parameters of the various models. Several general features merit mention.

First, only the first 70% of the data is applied to calibrate the models and the last 30% to predict, which is a tougher test than using the whole dataset to estimate parameters [23]. That is, for each dataset (pooled five sessions of each game), we searched for a single parameter combination that maximized the total likelihood value over the first 70% rounds of the data across all players. Thus, the log-likelihood (LL) function for the single-representative agent EWA model for a dataset is computed as follows:

$$LL\left(\delta, \varphi, \kappa, \lambda\right) = \sum_{i=1}^{N} \sum_{t=1}^{0.7 \cdot T} \ln\left(P_i^j\left(t\right)\right) \tag{7}$$

where $P_i^j(t)$ is the predicted probability for player i to choose strategy j as the same with the actual choice in round t in (2). Similarly, the LL function for the MAL model is specified as:

$$LL\left(\lambda_0, \lambda_+, \lambda\right) = \sum_{i=1}^{N} \sum_{t=1}^{0.7 \cdot T} \ln\left(P_i^j\left(t\right)\right) \tag{8}$$

Second, to keep it simple and parsimonious, we estimated initial attractions $\mathbf{A}_i(0)$ from the first period of actual data, rather than allowing them to be free parameters in likelihood functions. In addition, for EWA models, we set $N\ (0) = 1$ (common to all players) due to the parsimony of the model and the psychological interpretability of its parameters [24].

To evaluate model accuracy in the in-sample calibration, two criteria we report are: Log likelihoods and the Bayesian information criteria (BIC) that penalize theories with more free parameters. Specifically, BIC is calculated by $LL - k/2 \cdot LogM$, where k is the number of freedom degrees and M is the size of the calibration sample. In this study, we also reported two criteria for out-of-sample prediction, a tougher test than in-sample fitting, to evaluate model fitness on the validation sample. That is, Log likelihoods, a mean squared deviation (MSD), which is defined as:

$$MSD = \sum_{i=1}^{N} \sum_{0.7T+1}^{T} \sum_{j=1}^{m_i} \frac{[P_i^j(t) - I(s_i^j, s_i(t))]^2}{0.3 \cdot T \cdot N \cdot m_i} \tag{9}$$

Note that the MSD does not average observations across individuals. In addition, we report a criterion for overall goodness of fit for models based on the whole sample, that is, pseudo-R^2, denoted ρ^2, based on the Bayesian measure, indicating how much better the models do than random choice. The measure of ρ^2 is the difference between the BIC and the log-likelihood of a random-choice model in which all strategies are chosen equally often in each round, normalized by the random-model log-likelihood. As a rule of thumb, values of ρ^2 from 0.2 to 0.4 basically indicate excellent model fit [25].

B. GOODNESS OF FIT

Table 2 previews and summarizes results in all criteria of model fitness. Within each game and measure, the best fit statistics are printed in italics and the best fit model(s) is marked with an asterisk. Close inspection of these results yields three observations.

First, the explanatory power of the four models cluster into two subgroups, that is, MAL very similar to BL, and RL similar to EWA learning, but the former group is worse than the latter. The MAL model with minimum *cognitive sophistication* assumes that players only respond to the most recent environmental stimulus (i.e., payoff distribution over routes) such as automatons, while others with higher cognitive sophistication assume that players, to some extent, utilize their memory or past experience, including payoff history and decision history. The relatively poor performance of MAL seems to imply that human subjects' decisions are more likely to be *history-dependent*. Moreover, switching from MAL to BL models did not improve the performance pretty much. Recall that the most critical difference between the former group (MAL and BL models) and latter group is whether the dependence on what *other* players have done took on an important role. This might be one of the reasons differenting the performance between them.

Second, in terms of the overall performance measured by ρ^2, EWA substantially outperforms others in all games, although RL parallels it in part of the games. If EWA was overfitting, it would do relatively better in calibration than in validation, but this is not the case, judging from the similar values of ρ^2 based on overall data to that on in-sample data (comparing columns 6 and 9 in Table 2). Importantly, the overall ρ^2 of estimated EWA models fit the real data perfectly (higher than 0.2), regardless of the changes in the network topology. Third, EWA does better relative to RL in the calibration phase but the latter performs better in the validation phase in terms of LL value. Reinforcement learning assumes the utilization of two kinds of information, that is, the decision maker's own strategy and payoff. However, EWA learners are assumed to utilize the choices made by other players, aside from the above two pieces

TABLE 2. Model calibration and valuation of four learning models in all datasets.

Dataset	Model	No. of Parameters	In -sample Calibration[a]			Out-of-sample Validation[b]		Overall Validation[c]
			LL	BIC	ρ^2	LL	MSD	ρ^2
Game 2R	EWA[d]	4	*-3045*	*-3062*	*0.22*	-1302	0.18	*0.22*
	RL	3	-3162	-3175	0.19	*-1238*	*0.17*	0.21
	BL	3	-3873	-3886	0.00	-1662	0.25	0.00
	MAL	3	-3394	-3407	0.13	-1370	0.19	0.14
	Random	0	-3882	-3882	0.00	-1664	0.25	0.00
Game 4R	EWA[d]	4	*-3722*	*-3738*	*0.29*	-1545	0.15	*0.30*
	RL	3	-3818	-3831	0.27	*-1502*	*0.15*	0.29
	BL	3	-4502	-4514	0.14	-1926	0.19	0.14
	MAL	3	-4556	-4569	0.13	-1887	0.19	0.14
	Random	0	-5240	-5240	0.00	-2246	0.19	0.00
Game 6R	EWA[d]	4	*-4886*	*-4902*	*0.28*	-1672	0.14	*0.32*
	RL	3	-4916	-4929	0.27	*-1639*	*0.14*	*0.32*
	BL	3	-5240	-5252	0.23	-1902	0.17	0.26
	MAL	3	-5846	-5859	0.14	-2164	0.16	0.17
	Random	0	-6773	-6773	0.00	-2903	0.14	0.00
Game 8R	EWA[ds]	4	*-4903*	*-4919*	*0.25*	*-1611*	*0.12*	*0.30*
	RL	3	-4920	-4932	0.25	-1629	0.12	*0.30*
	BL	3	-5518	-5530	0.16	-2113	0.14	0.18
	MAL	3	-5723	-5735	0.13	-2173	0.14	0.16
	Random	0	-6550	-6550	0.00	-2807	0.11	0.00

[a] Based on the first 70% data; [b] based on the last 30% data; [c] based on the whole data; [d] represents the model with best performance among the alternatives in each dataset.

of information. It would thus seem that in the early stage of the game, the additional consideration of the ''forgone payoff'' of unchosen options, could improve the descriptive power to some extent, while the edge vanishes over time.

Previous literature indicates the RL appears to fit better in games with relatively low dimensional strategy spaces that have a mixed strategy equilibrium [26], whereas BL models appear to make relatively better predictions in coordination games [27]. However, this may not apply to the route choice games we investigated here. Because of the large size of groups comprising of 18 to 20 individuals, keeping an eye on all others' choices throughout the repeated games is overly complicated and far beyond the cognitive abilities of the participants. Moreover, route-choice decisions in recurrent games could be regarded as ''small decision'' problems, in which the expected consequence of each decision is relatively small. As such, participants would quickly realize that such efforts to track choices made by other players did not pay off and represent high inertia rates.

Next, we looked further into the parameter estimates of EWA models, which may provide some clues to the above results. Among the four datasets, the estimated δ (and ϕ) range intensively between 0 and 0.30 (and 0.88 and 1), respectively; whereas the value of k varies relatively much in different games. Recall that the parameter δ measures the relative weight given to foregone payoffs, compared to actual payoffs, in updating attractions. Estimated values of delta leaning towards 0 means that subjects' choices are more likely to be governed by the ''law of *actual* effect'', compared to the ''law of *simulated* effects''. Moreover, the parameter ϕ depreciates past attractions, $A(t)$. Its value, much close to 1 among all datasets, indicates that subjects did not realize other players were adapting as well, so the old observations of what others did become less useful but with a very slow decay rate. Experimental evidence of route choice games, in fact, favors the dependence on the overall decision history over *recency effect*, in which players were highly influenced by what happened in the most recent round. Note that MAL model is set up based on the assumption of this recency effect, which may be the main reason for its relative worse performance compared to EWA models.

Two hints about the human subjects' learning suggested by the model comparison analysis are that, (1) the past of experience is of importance to make present decision, (2) subjects' choices seem to be governed more by the ''law of *actual* effect'' than by the ''law of *simulated* effects''.

V. RESULTS

The model comparison analysis above indicates that EWA learning models could in general account for the subjects' decisions rather well in terms of statistic criteria at the population level. To further explore whether learning could lead people to coordination in route choice games, we next move on to the *dynamics* of population decisions predicted by learning models via simulations. Specifically, we generated data for artificial subjects using the best-fitting parameter values and compared simulated to actual behavior in terms of two statistics at the aggregate level: (1) the mean choice distribution over time and (2) the proportion of subjects who switched from their last choice over time.

A minor reason to proceed with such simulations of the dynamic process is to complement or strengthen the results based on the MLE method, in which decisions in two successive rounds are assumed to be mutually independent while learning processes are always dependent on time and history. That is also the reason that although we could reduce the difference between real and simulated players by re-estimating parameters to meet the new criteria in this section, we have opted not to do so. Rather, we applied the parameter estimates estimated using the in-sample fitting method to maximize the log-likelihood scores in the last section. If the models could stand still in the new criteria regarding the adjustment process, our results of model comparison would be more convincing.

A. SIMULATION SETTING

As all learning models include stochastic elements, we simulated 1,000 groups of artificial agents (18 or 20 per group) for each game to avoid the influence of statistical outliers. In each simulation run, all agents act in accordance with the same learning model using the same parameters estimated by the *in-sample* fitting method to avoid potential over-fitting problems [25]. The 1,000 groups of artificial players started the game with randomization of equal probability on each choice in the first round. They differed from one another only in the value of the seed number that was used to generate the random numbers, which, in turn, determined the probability of route choices and the path of route choice distribution over time. In each round, the route choice distribution is generated *endogenously* by the artificial agents' decisions; after each round, the agents receive feedback information regarding the choice distribution and payoffs distribution over all alternative routes, in parallel with the complete information setting in experiments.

B. AGGREGATE CHOICE DISTRIBUTION

In a simulation of each game, we calculated the mean frequency of each route j in the first half and second half of rounds, respectively, termed by f_j^s. Similarly, f_j^{ob} and f_j^{eq} are calculated for each game and either of the two blocks, which represent the average frequency of route j observed in data and predicted by equilibrium solutions, respectively. By means of the quadratic distances, defined as (10), we will examine whether the learning models could reproduce the observed route choice distribution and capture the convergence tendency to equilibria. The mean quadratic distances Q_{ob} and Q_{eq} are the average quadratic distance from equilibrium prediction and from observed data over $R(=1,000)$ simulations and over the four games, respectively,

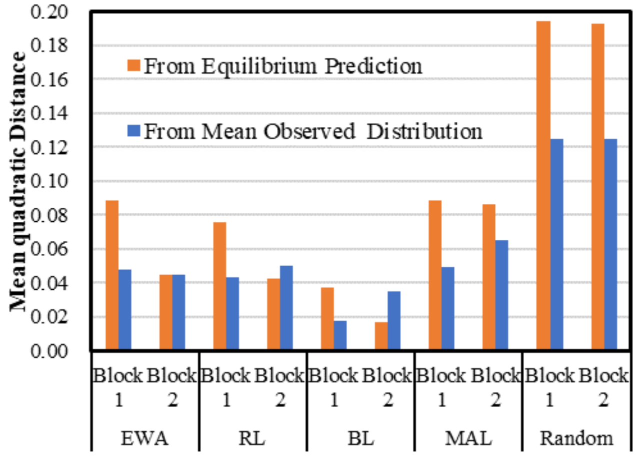


FIGURE 2. The mean quadratic distances Q_{eq} (left bars) and Q_{ob} (right bars) across four experimental games, by learning model and by block, with Random play as a benchmark.

defined as:

$$Q_{ob} = \frac{1}{4R}\sum_{g=1}^{4}\sum_{r=1}^{R}\sum_{j=1}^{m_i}(\bar{f}_j^s - \bar{f}_j^{ob})^2$$
$$Q_{eq} = \frac{1}{4R}\sum_{g=1}^{4}\sum_{r=1}^{R}\sum_{j=1}^{m_i}(\bar{f}_j^s - \bar{f}_j^{eq})^2 \tag{10}$$

where m_j indicates the total number of routes in networks that vary in different games. The predictive success of a learning model increases with a decreasing mean quadratic distance, i.e., the smaller the mean quadratic distance is, the better the learning theory fits the experimental data at the aggregate level.

Fig. 2 displays the mean quadratic distances Q_{eq} (left bars) and Q_{ob} (right bars) across four experimental games, by learning model and by block, with Random play as a benchmark. Inspection of Figure 2 yields two observations:

(1) For all learning models, the values of Q_{eq} seem to decrease over time, while the values of Q_{ob} seems to be stable across blocks.
(2) Both the distances of all learning models in either block are much lower than the Random play benchmark, that is, at most roughly half of the latter prediction, implying that all the alternative models outperform the Random play benchmark;
(3) In terms of capturing the observed distributions over all games Q_{ob}, the performance of the EWA and RL models are very similar to each other and both better than that of MAL, and the performance of all three models are, in turn, worse than that of the BL models.

Next, we proceeded with statistic tests to examine these observations. To this end, we broke down the distance scores Q_{eq} and Q_{ob} separately for each game, instead of averaging over all games. Our first test of observation (1) concerns how the route choice distribution changed over time as predicted by learning models.

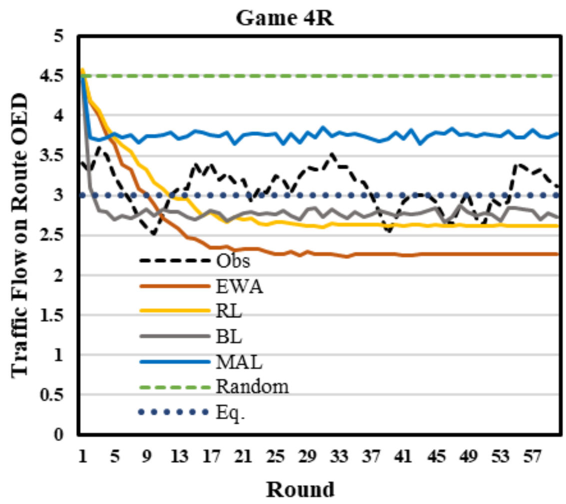


FIGURE 3. The typical development of the mean frequency of choosing route OED in Game 4R (the network referring to Fig. 1(b)) over all 1,000 simulation runs, compared with the moving average of the choice frequency on it (step = 5 rounds) over all groups in the experiments (dashed line), with two benchmarks as Nash equilibrium (dotted line) and Random play (dash-dot line).

To check the capability to capture the convergence towards equilibrium, for each learning model we compared their distance score from equilibrium, Q_{eq}, in the first block with that in the second block of periods, using a game as the unit of analysis. We cannot reject the null hypothesis that the distance from equilibrium does not decrease over blocks according to the Wilcoxon Signed Ranks tests (one-sided) for all learning models ($p > 0.19$). One possible reason is that the models' prediction converges towards equilibrium rapidly within the first block (see Fig. 3 for an example). To check the stability of the explanatory power of learning models over time, we implemented similar tests but in terms of the distance scores from the observed choice distribution Q_{ob}. Similarly, we cannot reject the null hypothesis that no difference exists between the two blocks according to the Wilcoxon Signed Ranks tests (two-sided) for all learning models ($p > 0.46$). This result implies that the explanatory power of these models is rather stable over time, especially considering that the predictions are based on parameter estimates obtained using the in-sample fitting method.

Further analysis provides support for observations (2) and (3). Using a simulation for each game as the unit of analysis, we conducted Fisher-Pitman permutation tests for every paired distance score Q_{ob} generated by a learning model and by Random play benchmark in either of the two blocks respectively. All comparison results were significant ($p < 0.001$), which supports that all learning models explain the observed route choice distribution over time significantly better than the Random play benchmark. We proceed with the same tests that compared the distance scores between any pair of learning models. Consistent with the inspection of Fig. 2, the difference between the EWA and RL models was very small and irrelevant, the corresponding test yielded no significant result with no surprise ($p > 0.5$). Except for that pair, all the results were significant ($p < 0.01$). Because of the large number of observations (4,000 per learning type), the order of explanatory power in the dynamics (in Fig. 2) was statistically robust. The order from the best to the worst is, BL, followed by EWA and RL with rare difference, and then MAL, Random play.

Take Game 4R for example, Fig. 3 demonstrates the difference in predictions made by the alternative learning models in more detail. Fig. 3 exhibits the typical adjustment process of the mean frequency for choosing route OED in Game 4R over all simulation runs, compared with the moving average of the choice frequency on it (step = 5 rounds) over all groups in the experiments (dashed line), with two benchmarks as Nash equilibrium (dotted line) and Random play (dash-dot line). The inspection of Fig. 3 suggests that all the learning models except for the MAL model simulates the observed mean choices on Route OED rather closely. However, the BL model converges more quickly than RL and EWA, all of which have significantly lower distance scores than the MAL model. That implies once again that the learning process shared by human subjects is more likely to be history-dependent, instead of a direct response to the latest information or situation.

C. AGGREGATE ROUTE SWITCHES

However, a more detailed analysis, which also focuses directly on the adjustment process postulated by the learning models, considers total switches in decisions among subjects over blocks. Paired t-tests (with a group as the unit of observation) that compared the mean number of switches in the two halves in Games 6R and 8R yielded significant results ($t(3) > 6.36$, $p < 0.01$), while there were no significant results for Games 2R and 4R ($p > 0.1$). As such, we observed a significant decreasing tendency of the number of switches over blocks in part of the datasets, but not in others. These diverse patterns are of particular importance for assessing the performance of any learning model that presumes to capture the dynamics.

In a simulation of each game, we calculated for each round the proportion of subjects who switched from their last choice, and then took the average for each simulated game over the rounds in the first and second half, respectively. Table 3 presents the means and standard deviations of the simulated proportion of switches by game and block, compared with the observed ones. Using each simulation as a unit of analysis, pair t-tests that compared mean proportion of switches predicted by the EWA, RL and BL models in the two halves in all games yielded a significant decreasing tendency ($t > 96.09, p < 0.01$), while the same tests indicate that there is no significant difference between the number of switches predicted by MAL models in the two halves ($t < -0.567$, $p > 0.5$). Furthermore, we conducted four Mann-Whitney tests (two-sided) to compare the simulated average proportions of switches by each of the learning models with the observed ones by game and by block. The null hypothesis of no difference between BL predictions and observed proportion of switches was significantly rejected ($p = 0.01$), while the other hypothesis was not ($p > 0.1$).

TABLE 3. The means and standard deviations of the simulated proportion of switches predicted by different learning models, compared with the observed ones, by game and block.

	Game 2R		Game 4R		Game 6R		Game 8R		Grand Mean
	1[b]	2	1	2	1	2	1	2	
Obs[a]	0.34	0.35	0.49	0.44	0.58	0.43	0.67	0.50	0.48
	(0.10)	(0.12)	(0.04)	(0.04)	(0.08)	(0.05)	(0.09)	(0.05)	
EWA	0.48	0.46	0.33	0.01	0.78	0.66	0.80	0.66	0.52
	(0.01)	(0.01)	(0.22)	(0.01)	(0.03)	(0.04)	(0.05)	(0.04)	
RL	0.40	0.33	0.31	0.01	0.33	0.03	0.82	0.71	0.37
	(0.03)	(0.01)	(0.20)	(0.01)	(0.21)	(0.02)	(0.03)	(0.04)	
BL	0.49	0.49	0.69	0.64	0.73	0.67	0.81	0.76	0.66^{**}
	(0.01)	(0.00)	(0.02)	(0.02)	(0.05)	(0.01)	(0.03)	(0.01)	
MAL	0.30	0.30	0.47	0.47	0.53	0.53	0.61	0.61	0.48
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Rad[c]	0.50	0.50	0.75	0.75	0.83	0.83	0.88	0.88	0.74^{**}

[a] represents the observed mean value of each dataset;
[b] this line represents block 1 and block 2 under each dataset;
[c] represents the predicted mean value of the Random benchmark model.

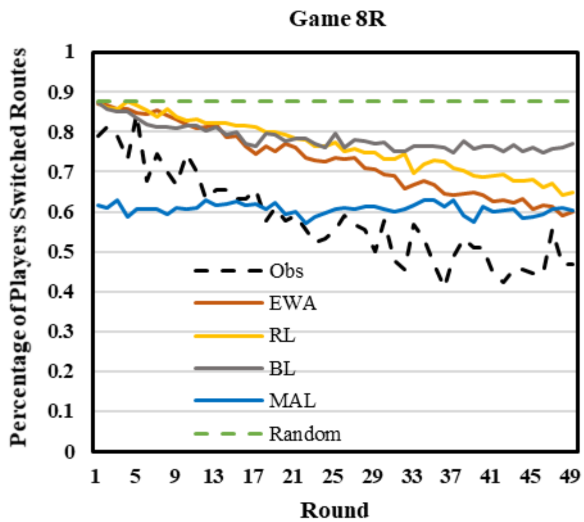


FIGURE 4. Percentage of players who switched routes, by round and model for Game 8R, with two benchmarks, i.e., Random Play (dotted line) and observed data (dashed line). Note that each point indicates the total number of switches out of 90 (= 18 players × 5 sessions) individual data points.

Our result is twofold. On the one hand, the EWA, RL and BL models, could predict a decreasing tendency among all simulated games, while the MAL model could only roughly predict a flat line in all cases. On the other hand, BL predicted a significantly higher proportion of switches than the observed proportion over all games, while others did not. To better illustrate this result, we took Game 8R as an example, and drew Fig. 4 to display the percentage of players who switched routes, by round and by model. A line-by-line comparison shows that the simulated results of EWA and RL models follow the observed decreasing tendency rather well, but that the BL model and Random play benchmark consistently predicted much higher switches. It appears that the MAL model captures the average level of the number of switches over time rather closely. However, it is worth mentioning that none of the tested learning models is capable to capture the mixed patterns in switches over time across different games; that is, switches over time decrease significantly on some networks but not significantly in other scenarios.

VI. CONCLUSION

In route choice games, a large group of self-interested users must to traverse through the same and congestible network, hoping for minimum delays. Given the high degree of strategic uncertainty and large-scale multiple equilibria, the recent experimental evidence of coordination success is rather surprising.

In this paper, we investigate how learning theories could contribute to explaining the consistently observed tacit coordination using a comprehensive horse race between the alterative learning models in four experimental games. To the best of our knowledge, this paper is the first systematic comparative study of learning theories in large-group route-choice games. We organize the main results to correspond to the order of our two research questions: which learning models could best capture the population decisions, and how successfully the learning models could reproduce the dynamics towards coordination at the aggregate level.

Using maximum likelihood estimation to estimate the parameters in the sample and several statistic criteria to test models out the sample, we obtained the order of the explanatory power of the four models from best to worst, EWA $\approx$ RL $>$ BL $\approx$ MAL. Two hints about the human subjects' decisions suggested by the model comparison analysis are that, (1) past of experience is important to make present decision, (2) subjects' choices appear to be governed more by the "law of actual effect" than by the "law of simulated effects" [12].

By comparing the simulations based on parameter estimations with experimental data, all the learning models proved to have stable predictive power in replicating the dynamics of aggregate choice distribution, which did converge to Nash equilibrium (although not significantly). Regarding the dynamics of the proportion of subjects who switched from their last choices, the results are two twofold. EWA, RL and BL models could always predict decreasing tendency in any games, while MAL always a flat line regardless of the network topology. However, none of them were flexible enough to capture the mixed patterns in real experiments. In capturing the distribution of individual switching rates, learning theories face collective failure.

In summary, adaptive learning models such as EWA, in general, could both effectively account for the subjects' decisions and generate the dynamics of coordination similar to that in experimental games. Future research might test whether employing various types of learning, such as rule-based learning theory postulated [28], is a promising way of simultaneously capturing the dynamics towards coordination and substantial individual differences.

REFERENCES

[1] R. Selten, T. Chmura, T. Pitz, S. Kube, and M. Schreckenberg, "Commuters route choice behaviour," *Games Econ. Behav.*, vol. 58, no. 2, pp. 394–406, 2007.

[2] X. Yu and S. Gao, "Learning routing policies in a disrupted, congestible network with real-time information: An experimental approach," *Transp. Res. C, Emerg. Technol.*, vol. 106, pp. 205–219, Sep. 2019.

[3] N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, Eds., *Algorithmic Game Theory*. Cambridge, MA, USA: Cambridge Univ. Press, 2007.

[4] A. Rapoport and V. Mak, "Strategic interactions in transportation networks," in *The Handbook Behavioral Operations*, K. Donohue, E. Katok, S. Leider, Eds. Hoboken, NJ, USA: Wiley, 2018, ch. 16, pp. 557–586.

[5] W. Zhang, Z. He, W. Guan, and G. Qi, "Day-to-day rerouting modeling and analysis with absolute and relative bounded rationalities," *Transportmetrica A, Transp. Sci.*, vol. 14, no. 3, pp. 256–273, 2018.

[6] Z. He, L. Yang, and W. Guan, "A day-to-day route choice model based on travellers' behavioural characteristics?" *Procedia-Social Behavioral Sci.*, vol. 138, pp. 738–747, Jul. 2014.

[7] V. Mak, E. J. Gisches, and A. Rapoport, "Route vs. segment: An experiment on real-time travel information in congestible networks," *Prod. Oper. Manag.*, vol. 24, no. 6, pp. 947–960, 2015.

[8] E. J. Gisches and A. Rapoport, "Degrading network capacity may improve performance: Private versus public monitoring in the Braess Paradox," *Theory Decis.*, vol. 73, no. 2, pp. 267–293, 2012.

[9] T. Chmura, S. J. Goerg, and R. Selten, "Learning in experimental 2×2 games," *Game Econ. Behav.*, vol. 76, no. 1, pp. 44–73, 2012.

[10] C. Camerer and T. H. Ho, "Experience-weighted attraction learning in normal form games," *Econometrica*, vol. 67, no. 4, pp. 827–874, 1999.

[11] I. Erev, E. Ert, A. E. Roth, E. Haruvy, S. M. Herzog, R. Hau, R. Hertwig, T. Stewart, R. West, and C. Lebiere, "A choice prediction competition: Choices from experience and from description," *J. Behav. Decis. Making*, vol. 23, no. 1, pp. 15–47, 2010.

[12] C. F. Camerer, T.-H. Ho, and J.-K. Chong, "Sophisticated experience-weighted attraction learning and strategic teaching in repeated games," *J. Econ. Theory*, vol. 104, no. 1, pp. 137–188, 2002.

[13] Z. B. He, W. Guan, and S. F. Ma, "A traffic-condition-based route guidance strategy for a single destination road network," *Transp. Res. C, Emerg. Technol.*, vol. 32, pp. 89–102, Jul. 2013.

[14] Z. He, L. Zheng, W. Guan, and B. Mao, "A self-regulation traffic-condition-based route guidance strategy with realistic considerations: Overlapping routes, stochastic traffic, and signalized intersections," *J. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 545–558, Mar. 2016.

[15] D. Braess, "Über ein Paradoxon aus der Verkehrsplanung," *Unternehmensforschung*, vol. 12, no. 1, pp. 258–268, 1968.

[16] A. Rapoport, T. Kugler, S. Dugar, and E. J. Gisches, "Choice of routes in congested traffic networks: Experimental tests of the Braess Paradox," *Game Econ. Behav.*, vol. 65, no. 2, pp. 538–571, 2009.

[17] A. Rapoport, H. Qi, V. Mak, and E. J. Gisches, "When a few undermine the whole: A class of social dilemmas in ridesharing," *J. Econ. Behav. Org.*, vol. 166, pp. 125–137, Oct. 2019.

[18] V. Mak, D. A. Seale, E. J. Gisches, R. Yang, M. Cheng, M. Moon, and A. Rapoport, "The Braess paradox and coordination failure in directed networks with mixed externalities," *Prod. Oper. Manag.*, vol. 27, no. 4, pp. 717–733, 2018.

[19] Y. Iida, T. Akiyama, and T. Uchida, "Experimental analysis of dynamic route choice behavior," *Transp. Res. B, Methodol.*, vol. 26, no. 1, pp. 17–32, 1992.

[20] A. Rapoport, E. J. Gisches, R. Lindsey, and T. Daniel, "Pre-trip information and route-choice decisions with stochastic travel conditions: Experiment," *Transp. Res. B, Methodol.*, vol. 68, pp. 154–172, Oct. 2014.

[21] W. Zhang, Z. He, W. Guan, and R. Ma, "Selfish routing equilibrium in stochastic traffic network: A probability-dominant description," *PLoS ONE*, vol. 12, no. 8, 2017, Art. no. e0183135.

[22] Z. He, B. Chen, N. Jia, W. Guan, B. Lin, and B. Wang, "Route guidance strategies revisited: Comparison and evaluation in an asymmetric two-route traffic network," *Int. J. Mod. Phys. C*, vol. 25, no. 4, 2014, Art. no. 1450005.

[23] H. Qi, S. Ma, N. Jia, and G. Wang, "Experiments on individual strategy updating in iterated snowdrift game under random rematching," *J. Theor. Biol.*, vol. 368, pp. 1–12, Mar. 2015.

[24] H. Qi, S. Ma, N. Jia, and G. Wang, "Individual response modes to pre-trip information in congestible networks: Laboratory experiment," *Transportmetrica A, Transp. Sci.*, vol. 15, no. 2, pp. 376–395, 2019.

[25] D. Hensher and P. R. Stopher, Eds., *Behavioural Travel Modelling*. New York, NY, USA: Taylor & Francis, 1979.
[26] I. Erev and A. E. Roth, ''Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria,'' *Amer. Econ. Rev.*, vol. 88, no. 4, pp. 848–881, 1998.
[27] R. Battalio, L. Samuelson, and J. Van Huyck, ''Optimization incentives and coordination failure in laboratory stag hunt games,'' *Econometrica*, vol. 69, no. 3, pp. 749–764, 2001.
[28] D. O. Stahl, ''Rule learning in symmetric normal-form games: Theory and evidence,'' *Game Econ. Behav.*, vol. 32, no. 1, pp. 105–138, 2000.

JIANGUO SONG received the LL.B. degree from Henan University, Henan, China, in 1989, and the LL.M. degree from Zhengzhou University, Zhengzhou, China, in 1992. He is currently pursuing the Ph.D. degree in management science and engineering with the School of Information Management, Wuhan University, China. His current research interests include traveler behavior and urban traffic control and management.

HANG QI was born in Nanyang, Henan, China, in 1990. She received the B.S., M.S., and Ph.D. degrees in management science and engineering from Tianjin University, Tianjin, China. Since 2018, she has been an Assistant Professor with the Institute for Advanced Studies in Finance and Economics, Hubei University of Economics, Wuhan, China. Her current research interest includes experimental study on traveler behavior, mixed behavior transportation system.

GUANGCHAO WANG received the B.S. degree in logistics management from the Shandong University, Shandong, China, in 2010. and the Ph.D. degree in management science and engineering from the Tianjin University, Tianjin, China, in 2016. His Ph.D. thesis was on traveler behavior modeling. Since 2016, he has been an Assistant Professor with the School of Information Management, Central China Normal University. His current research interests include the modeling of traveler behavior, the application of economic and optimization theories in transportation, sharing transportation, and connected and automated transportation.