Transportation Research Part C 113 (2020) 314–331

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Personalized incentives for promoting sustainable travel behaviors☆

Xi Zhu[a], Feilong Wang[a], Cynthia Chen[a,*], Derek D. Reed[b]

[a] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA 98195, USA
[b] Department of Applied Behavioral Science, University of Kansas, Lawrence, KS 66045, USA

ARTICLE INFO

Keywords:
Sustainable travel behaviors
Personalized incentives and controls
Preference learning
Particle filter

ABSTRACT

We develop a personalized system to modify individual travel behaviors by offering personalized incentives. Individual preferences are learned to provide personalized incentives so that the promoted alternative is likely accepted. Using knowledge from control theories and state estimation, we model travelers' choice-making behaviors with the random utility theory and responses from the individuals are mined by a particle filter for learning individual preferences to promote sustainable behaviors. The discrete nature of travel behavior naturally leads to limited observability. We overcome this problem by designing a measurement function from which additional information can be solicited. Additionally, the inherent trade-offs between factors that affect travel choices result in an infinite set of solutions. We thus propose two solutions: (1) the divide and conquer strategy in which a multi-dimensional conditional probability function is proposed; and (2) use of domain knowledge to restrict that preference values fall in certain ranges and are consistent with certain distributions. The performance of preference learning with these two solutions applied is shown via simulation tests and an online experiment involving human participants. For departure time choices, we show an average acceptance ratio of 0.68 for all participants when being promoted with alternatives with personalized incentives. We also show that changes in individual departure time choices will lead to 48% reduction in total travel time on a simple transportation network.

1. Introduction

It is widely recognized that America's over-reliance on driving is closely related to a number of urban issues, such as congestion, air pollution, and sedentary lifestyles (Katzev, 2003). Transportation Demand Management (TDM) strategies, designed to modify travel behavior patterns by providing travelers generic incentives (or costs) to certain travel choices (e.g., promoting transit use by offering low-cost pass), have not seen much success (Giuliano, 1992; Stopher, 2004; Möser and Bamberg, 2008). Parallel to this, in the past decade, the rapid proliferation of smart, personal devices has not only generated enormous data that allow us observe people's travel trajectories in time and space at an unprecedented scale (Chen et al., 2016), but also provided a medium that revolutionized the interactions between people and devices. Today, personalized interactions with individuals are a reality.

Under this backdrop, we design a personalized system to learn an individual's preferences, such that we could trigger the desired

☆ This paper has been accepted for a poster presentation at the 23rd International Symposium on Transportation and Traffic Theory (ISTTT23) July 24–26, 2019 in Lausanne, CH.

* Corresponding author.
E-mail address: qzchen@uw.edu (C. Chen).

https://doi.org/10.1016/j.trc.2019.05.015
Received 1 December 2018; Received in revised form 2 April 2019; Accepted 16 May 2019
Available online 21 June 2019

behavioral change by providing personalized incentives, rather than generic incentives implemented in traditional TDM strategies. Personalized incentives refer to incentives designed to suit an individual's preferences and constraints, and the individual's preferences are captured by β_k ($k \in \{1, 2, \cdots, K\}$), each of which describes how the individual values an influential attribute x_k of an alternative (here K gives the total number of influential attributes of an alternative, e.g., trip cost, trip time). Given a sustainable alternative to be promoted and its attributes, the system uses the learned individual preferences to present her incentives based on the theory of Random Utility Maximization (RUM) (Hess et al., 2018; McFadden, 1975). According to RUM, the probability of choosing among multiple alternatives depends only on the difference in their respective utilities, and an individual will select the alternative that provides the maximum utility. Here, the utility is a metric quantifying the attractiveness of an alternative in a choice scenario, and is assumed to be indirectly related to the various characteristics of the alternative, the individual and the surrounding environment (so called "indirect utility") (Ben-Akiva et al., 1985; Hensher, 1994). Therefore, the incentive presented is expected to add utility to the promoted alternative so that for the user, the probability of accepting it is no lower than that of the default alternative (e.g., single driving). While individual preferences cannot be directly observed, the system is designed to learn preferences from interactions with the individual: (1) when an individual is to conduct a trip, the system presents an incentive for the promoted alternative relying on its previously learned preferences along with other trip information; (2) the individual responds by choosing either the promoted alternative or the default choice and her decision is captured by the system so that the preference estimates are updated. In the current work, a binary choice scenario is presented to the individual. In the future it could also be extended to a multi-choice scenario by regarding the latter as a combination or a sequence of several binary choices.

Two challenges exist in achieving personalized learning. The first one is the limited number of observations that are available to the system (since a single individual can only generate a limited number of observations), thus preventing the use of traditional preference-learning techniques such as advanced econometric models (e.g., discrete choice models) derived from RUM (Ben-Akiva et al., 1985). Additionally, preferences estimated from such models represent averages for a sample or a group, thus not satisfying the personalization goal. Other techniques such as machine learning also cannot be applied for similar reasons (Lops et al., 2011; O'Mahony et al., 2009; Park et al., 2012). Additionally, machine learning techniques do not provide explanations for the decision-making choices, even though their prediction accuracy may be high (Moore et al., 2013).

The second issue is the lack of observability: given two alternative choices one being the default choice (0) and the other being the promoted one (1), based on RUM, we can write the probability of choice (1) being selected as $p_1 = Pr(U_1 \geq U_0)$, where $U_i = V_i + \varepsilon_i (i \in \{0, 1\})$ and is the utility associated with choice i, with V_i as the measurable systematic utility ($V_i = \sum_k \beta_k x_k$), and ε_i as the random utility (Ben-Akiva et al., 1985). Unfortunately, none of the key terms in this setup (p_1, U_i, or V_i) can be directly observed and the only piece of information that is directly observed is whether the individual accepts or rejects the promoted choice. This observable choice contains limited information, since if choice (1) is selected, it means $p_1 \in (0.5, 1)$, which can be the result of many possibilities for β_k.

To address the issue of limited sample size (which is also the core idea of personalization, i.e., learning based on limited observations from a single individual), we propose a particle filter approach that views β_k as the underlying states to be learned. The use of the particle filter approach however does not address the second issue, lack of observability, which has two dimensions. The first one, as noted above, is reflected in that the acceptance or rejection of an alternative by the individual conveys little information on the relative attractiveness of the two alternatives. To deal with this, we design an interface to solicit additional information. We assume that though p_1, U_i, or V_i cannot be directly observed, we can observe a utility ratio (R) that reflects the relative attractiveness of two alternatives. This ratio function is consistent with RUM as choice (1) is selected if and only if $R > 0.5$. It also conveys more information related to the individual's preferences as it continuously changes with U_1.

The use of the ratio R, however, does not address the second dimension of the lack of observability issue, which is related to the inherent trade-off nature of human decision making. The utility function is the weighted sum of various factors that come into the decision-making process (e.g., travel time and travel cost), or $V_i = \sum \beta_k x_k$. In other words, an infinite set of possible values for $\beta_k s$ gives rise to a single R. It is thus important to devise mechanisms to identify a unique set of $\beta_k s$ that most likely give the observed R. Two solutions are proposed to address it. The first one is the "divide and conquer" strategy, which decomposes a multi-dimensional problem into multiple conditional one-dimensional problems by considering responses from a pair of choice scenarios sharing similar attribute values. The second one is to add domain knowledge on travel behavior as constraints while learning the preferences. We present some simulation results to demonstrate the promise of these two solutions.

In addition, an online experiment was carried out to investigate how our system performs when interacting with humans. Participants were recruited to interact with the system by responding to a sequence of hypothetical binary choice scenarios on departure times. More specifically, they gave responses to the promoted alternatives designed based on their previous responses. The online stated choice experiment focused on promoting off-peak-hour travels (i.e. departure time changes). The results show that participants are likely to accept the promoted alternatives with personalized incentives: the average acceptance ratio (i.e. AR, the fraction of all the promoted suggestions that are accepted) is 0.68 for all participants, with 65% of them having an AR larger than 0.5, 20% of them with an AR equal to 0.5, and 15% with an AR smaller than 0.5. An evaluation of the system impact was also carried out, demonstrating that individual departure time changes will lead to 48% reduction in total travel times on a simple network.

The remainder of the paper is organized as follows. In Section 2 the proposed personalized system is presented, including the three modules comprising the system, the design characteristics of each module and how these modules work with each other towards learning underlying preferences and providing personalized incentives. The details on how we address above-mentioned challenges are also provided. Section 3 lays out the test results from both simulations and the online experiment. In Section 4 we present a framework with system-level considerations such that the total amount of incentives is minimized subject to a minimum level of system performance improvement. Lastly in Section 5 we discuss related issues and future research.

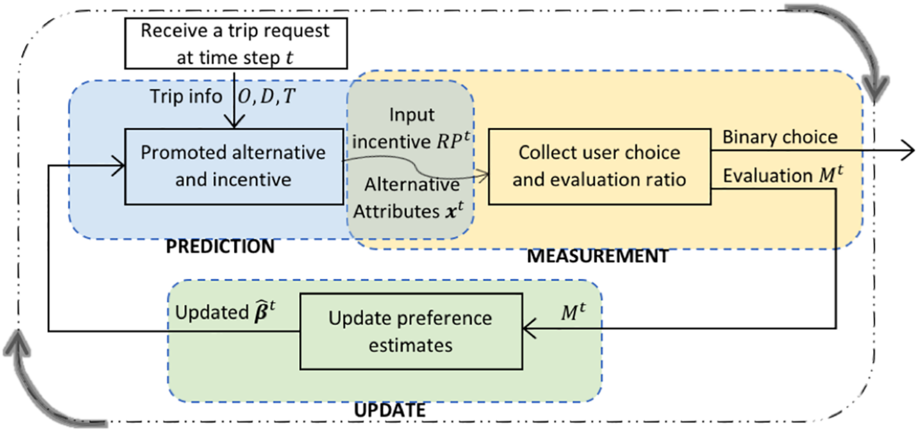

Fig. 1. Flowchart of the personalized system.

2. System description

2.1. System overview

Fig. 1 provides an overview on how an individual interacts with the personalized system in an iterative fashion. An individual can initiate by sending a trip request to the system. For clarity, an interaction triggered by the t^{th} trip request is noted as the interaction at time step *t*. A request consists of trip information such as origin *O,* destination *D*, and departure time *T*. In each interaction, three modules work sequentially to update preference estimates, including PREDICTION, MEASUREMENT and UPDATE. Based on the trip information contained in the trip request and the previous estimates on the individual's preferences, the PREDICTION module formulates an alternative that is to be promoted, and predicts the amount of the incentive needed for the individual to accept the promoted alternative. The incentive is represented by reward points, which are widely used (Richter et al. 2015; Hu et al. 2015). The reward points earned by accepting the promoted alternative can be exchanged for real rewards (e.g., a discount on transit fares). The MEASUREMENT module presents the individual the binary choice scenario (one being the default alternative and the other being the promoted alternative formulated in the PREDICTION module), along with attributes of each alternative such as travel cost and travel time as well as reward points. The individual makes the decision by either accepting or rejecting the promoted alternative. Besides observing the decision, the MEASUREMENT module also solicits additional information from the individual. Lastly, by utilizing the response collected, the UPDATE module updates preference estimates (β_ks) following a Bayesian estimation scheme in a particle filter. The updated preference estimates are then fed into the next interaction. Each module is described with more details in Sections 2.2–2.4.

Before describing details of each module, we review another system *Tripod*[1] that is also funded under the ARPA-E's TRANSNET program[2] and aims to shift individuals' travel behaviors via personalized travel alternatives and incentives. We briefly discuss how *Tripod* differs from our system[3]. Although both systems have a personalized focus and learn an individual's preferences (e.g., toward travel time), *Tripod* differs from our system in three important aspects. The first aspect involves the meaning of personalization. For *Tripod*, personalization means identifying an alternative set unique to an individual according to her trip request and the utilities of alternatives are calculated with the preferences learned from her historical behaviors. In our system, personalization means providing personalized incentives by identifying an individual's preferences based on which, the minimum amount of incentives is calculated so that the individual will likely accept the promoted alternative. The second aspect relates to the determination of incentives provided to an individual. *Tripod* allocates incentives according to the contribution of the individual's behavioral change to the entire network that is predicted with micro-simulation of the network system (Atasoy et al. 2015; Lima Azevedo et al., 2018). Behavioral consideration on whether the incentive is attractive enough for the individual to accept the promoted alternative(s) is not accounted for. This is exactly what our personalized system aims to achieve by determining the right amount of incentives such that the probability of the individual's accepting it is greater than a set threshold (e.g., 60%). Ideally, both behavioral and system-level considerations shall be accounted for and we discuss this future direction in Section 5. The third aspect relates to the methodology used to achieve

[1] *Tripod*: http://its.mit.edu/tripod-sustainable-travel-incentives-prediction-optimization-and-personalization.
[2] TRANSNET: https://arpa-e.energy.gov/?q = arpa-e-programs/transnet.
[3] Besides *Tripod*, two other systems are also funded by TRANSNET: *iPretii* (http://www.ipretii.umd.edu/) and *Metropia* (https://metropia.com/). Similar to *Tripod*, *iPretii* provides an individual a personalized set of alternatives that could be attractive to the individual and offers incentives based on the contribution of a behaviour change to the network system. However, details on how these are achieved in *iPretii* are not published. *Metropia* simply classifies individuals into different bins with clustering, based on historical choices. No details can be found on how the classification results in personalization or what personalization means.

personalization. *Tripod* identifies the set of personalized travel alternatives by first learning individuals' preferences via a regression model, which are then used for calculating alternatives' utilities and ranking them (Song et al., 2019; Song et al., 2018; Lima Azevedo et al., 2018). As noted earlier, the use of the regression model requires a sufficiently large sample size and tracking of an individual's preferences over time is also not possible. In our personalized system, we develop a particle filter approach that addresses three issues: limited sample size, lack of observability and multi-dimensional problem and allows tracking of an individual's preferences over time.

2.2. Prediction

As a response to an individual's trip request, the PREDICTION module determines the amount of reward points needed for the potential acceptance of the promoted alternative (departure time change in this paper). In real-world applications, multiple alternatives are likely and which one to promote depends on the day-to-day traffic dynamics, which are affected by various factors (Liu et al., 2017; Xiao and Lo, 2016; Ye et al., 2017). The decision on which alternative to promote requires a network model capable of simulating the traffic dynamics given different alternatives. This is outside of the scope of the current paper and discussed in Section 5 as future directions.

The estimation of individual a's evaluation toward the promoted alternative is modelled via the predicted systematic utility $\widehat{V}_1^{a,t}$:

$$\widehat{V}_1^{a,t}(RP^{a,t}) = \sum_k \widehat{\beta}_k^{a,t-1}\Delta x_k^{a,t} + \widehat{\beta}_{RP}^{a,t-1}RP^{a,t} \tag{1}$$

Here, given that individual a's true preferences are not observable, $\widehat{V}_1^{a,t}(RP^{a,t})$ is estimated using previously learned preferences including $\widehat{\beta}_k^{a,t-1}$ and $\widehat{\beta}_{RP}^{a,t-1}$, which represent the learned preferences on trip attribute differences $\Delta x_k^{a,t}$ and on reward points $RP^{a,t}$ respectively. Since Sections 4 and 5 focus on individual level, hereafter we omit a in the superscript for simplicity.

According to RUM, the individual would select the alternative with the maximum utility. Therefore, with an expectation that she would accept the promoted choice, RP^t is calculated such that

$$p_1 = \frac{\exp(\widehat{V}_1^t(RP^t))}{1 + \exp(\widehat{V}_1^t(RP^t))} > 0.5. \tag{2}$$

Here, p_1 is the probability that an individual is expected to accept the promoted alternative given RP^t points. In our system, we could adjust our expected probability of the alternative to be accepted by pre-specifying p_1 ($p_1 \in (0.5,\ 1.0)$). Given a specific value of p_1 (e.g., $p_1 = 0.6$), RP^t can be solved as:

$$RP^t = \frac{\log\left(\frac{p_1}{1-p_1}\right) - \sum_k \widehat{\beta}_k^{t-1}\Delta x_k^t}{\widehat{\beta}_{RP}^{t-1}} \tag{3}$$

It is possible that Eq. (3) returns a negative value, suggesting no reward point is needed to promote the alternative. In such a case, we let $RP^t = 0$.

2.3. Measurement

MEASUREMENT module presents the individual the formulated binary choice scenario and in turn, the individual makes her choice. Fig. 2 provides an example of the interface with a binary choice scenario and an additional question on how the individual evaluates the relative attractiveness of the promoted choice compared with the default choice. By asking the additional question, the utility ratio R is assumed to be observed:

$$R = \frac{e^{U_1}}{e^{U_0} + e^{U_1}}, \tag{4}$$

The individual answers the additional question by adjusting the slider between two ends with one being "not attractive at all" and the other being "absolutely attractive" (see Fig. 2). Consequently, R can be measured based on the position of the slider.

There are errors accounting for omitted factors in U_1 (factors that are important to the individual's decision-making process but not accounted for in the utility function), and the error in adjusting the slider (which translates into R) by the individual. We term both types of error as measurement noise in U_1. The measurement obtained through the slider is denoted as M and is expressed as:

$$M^t = \frac{e^{\upsilon_m + \boldsymbol{\beta}^t\Delta\boldsymbol{x}}}{1 + e^{\upsilon_m + \boldsymbol{\beta}^t\Delta\boldsymbol{x}}} \tag{5}$$

where for simplicity, $\boldsymbol{\beta}^t$ is used as the preference vector containing both preferences on trip attributes $\beta_k^t (k \in \{1, 2, \cdots, K\})$ and the preference on reward points β_{RP}^t at time step t, $\Delta\boldsymbol{x}$ is used as the attribute vector containing both trip attribute differences and the number of points, and υ_m is referred as the measurement noise (accounting for both types of errors) following a certain distribution. For the simulation tests in the following sections, υ_m is simply assumed following a normal distribution, i.e. $\upsilon_m \sim N(0, m^2)$. Eq. (5) serves as the measurement function in the particle filter approach.

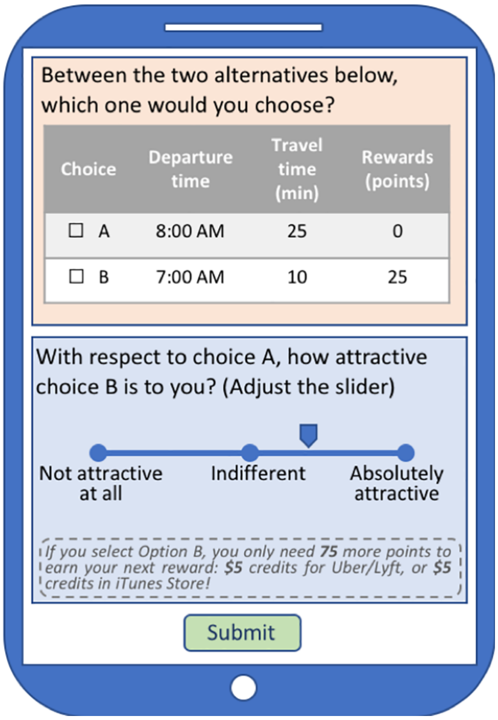

Fig. 2. An example interface with a choice scenario and an additional question.

2.4. Update

Following the measurement on the utility ratio, the UPDATE module works to update the preference estimates via a Bayesian estimation scheme. Given the non-linearity in our measurement function (Eq. (5)), the estimation is realized with a particle filter approach. The particle filter approach is a sequential Monte Carlo sampling method and designed to infer the state of a system from noisy measurements via a recursive predict-update scheme (Chen et al., 2011). In the following, we first describe how the UPDATE model works and then give a summary on its implementation in our system (Doucet and Johansen, 2011; Simon, 2006).

The preferences $\boldsymbol{\beta}^t$, although not observable, are estimated by utilizing the available information contained in the noisy measurements, such as the one given by Eq. (5). More specifically, Bayesian estimation calculates a conditional probability density function $p(\boldsymbol{\beta}^t|M^{1:t})$ to represent $\boldsymbol{\beta}^t$ by obtaining all the measurements up to time t (i.e. $M^1, \cdots, M^t$). The conditional pdf is updated according to Doucet and Johansen (2011):

$$p(\boldsymbol{\beta}^t\ |M^{1:t}) = \frac{p(M^t|\ \boldsymbol{\beta}^t)p(\boldsymbol{\beta}^t|M^{1:(t-1)})}{\int p(M^t|\ \boldsymbol{\beta}^t)p(\boldsymbol{\beta}^t|M^{1:(t-1)})d_{\boldsymbol{\beta}^t}} \propto p(M^t|\ \boldsymbol{\beta}^t)p(\boldsymbol{\beta}^t|M^{1:(t-1)}) \tag{6}$$

Essentially, to acquire the conditional pdf, we update the prior pdf $p(\boldsymbol{\beta}^t|M^{1:(t-1)})$ by taking the new measurement M^t into consideration, with the likelihood function $p(M^t|\ \boldsymbol{\beta}^t)$ following the measurement function (Eq. (5)) and rewritten as:

$$p(M^t|\boldsymbol{\beta}^t) = \frac{1}{\sqrt{2\pi}m}\exp\left(\frac{-\left[-\log\left(\frac{1-M^t}{M^t}\right) - \Delta\boldsymbol{x}^T\boldsymbol{\beta}^t\right]m^{-2}\left[-\log\left(\frac{1-M^t}{M^t}\right) - \Delta\boldsymbol{x}^T\boldsymbol{\beta}^t\right]}{2}\right) \tag{7}$$

And the prior $p(\boldsymbol{\beta}^t|M^{1:(t-1)})$ in Eq. (6) is

$$p(\boldsymbol{\beta}^t|M^{1:(t-1)}) = \int p(\boldsymbol{\beta}^t|\boldsymbol{\beta}^{t-1})p(\boldsymbol{\beta}^{t-1}|M^{1:(t-1)})d_{\boldsymbol{\beta}^{t-1}}, \tag{8}$$

where $p(\boldsymbol{\beta}^t|\boldsymbol{\beta}^{t-1})$ is the probability of preference evolution.

We assume that an individual's preference evolution is in the following form:

$$\boldsymbol{\beta}^t = \boldsymbol{\beta}^{t-1} + \boldsymbol{u}, \quad \boldsymbol{u}\sim N(\mathbf{0}, \boldsymbol{\Sigma}) \tag{9}$$

Provided that preferences are stable in the very short term (Gärling and Axhausen, 2003; Moore et al., 2013), the evolution function (Eq. (9)) suggests that the system takes values from the previous time step t-1 for prediction but adds a process noise $\boldsymbol{u}$. The process noise $\boldsymbol{u}$ here is to account for possible fluctuations of individuals' preferences over time and is also essential in learning dynamic preferences (i.e. individuals' preferences change over time).

Eqs. (7) and (8) define a recursive way to update $\boldsymbol{\beta}^t$, with the initial pdf available (i.e. $p(\boldsymbol{\beta}^0|M^0)$, M^0 means no measurement available). Due to the nonlinearity in our measurement function, there is no analytical closed form solution. An approximate solution is obtained using the particle filter approach, with Monto Carlo sampling.

The key idea of the particle filter is to approximate the posterior pdf with a set of random samples with weights. These samples are particles. When the sample size is large enough, these particles approach an approximate representation to the posterior pdf. Suppose $\boldsymbol{\beta}^t\left\{\boldsymbol{\beta}_j^t, w_j^t\right\}$ denotes a collection of J particles, where $\boldsymbol{\beta}_j^t\,(j \in \{1, 2, \cdots, J\})$ is a preference sample and w_j^t is its corresponding weight. A weighted approximation to the posterior pdf is given in (10), with weights update given in (11) (Doucet and Johansen, 2011; Orhan, 2012).

$$p(\boldsymbol{\beta}^t|M^{1:t}) \approx \sum\nolimits_{j=1}^{J} w_j^t \delta(\boldsymbol{\beta}_j^t) \tag{10}$$

$$w_j^t \propto w_j^{t-1} \frac{p(M^t|\,\boldsymbol{\beta}_j^t) p\left(\boldsymbol{\beta}_j^t|\boldsymbol{\beta}_j^{t-1}\right)}{q\left(\boldsymbol{\beta}_j^t|\boldsymbol{\beta}_j^{t-1}; M^t\right)} \tag{11}$$

Here, $\delta(\boldsymbol{\beta}_j^t)$ is a delta function centered at $\boldsymbol{\beta}_j^t$ and $q\left(\boldsymbol{\beta}_j^t|\boldsymbol{\beta}_j^{t-1}; M^t\right)$ is the importance density, which is a known pdf to generate particles. Note here we assume that the importance density only depends on the state and measurement at the previous time step t-1. It is often convenient to set the importance density to be the same as the prior $p\left(\boldsymbol{\beta}_j^t|\boldsymbol{\beta}_j^{t-1}\right)$, such that the weight update Eq. (11) is further simplified as (Doucet and Johansen, 2011):

$$w_j^t \propto w_j^{t-1} p(M^t|\boldsymbol{\beta}_j^t) \tag{12}$$

2.4.1. A problem in multi-dimensional learning

As mentioned earlier, we face a problem when there are two or more parameters to be learned–an infinite set of parameters could give rise to a single M (Fig. 3). We propose two solutions to address this problem: one being the "divide and conquer" strategy to modify the particle filter as described above and the other imposing domain knowledge on travel behavior as constraints in Monte Carlo sampling.

2.4.2. Solution 1: Divide and conquer

In the usual setup of the particle filter approach, every update utilizes the current observation at time step t after the individual makes a choice on a scenario presented to her. The likelihood function used in the update process is $L = Pr(M^t|\boldsymbol{\beta}_j^t)$, representing the likelihood of observing M^t, given the preferences represented by particle $\boldsymbol{\beta}_j^t$, where $\boldsymbol{\beta}_j^t = \boldsymbol{\beta}_j^{t-1} + \boldsymbol{u}$, with $\boldsymbol{u}$ representing the process noise from t-1 to t. In the proposed divide and conquer strategy, a multi-dimensional problem is decomposed into multiple, conditional one-dimensional problems. The idea is that, when different scenarios possess alternatives sharing similar attribute values, the relative attractiveness across alternatives only hinges upon that attribute with different values. In summary, it takes two modifications in the previously introduced update process.

(1) Instead of using measurements in Eq. (5) directly, we define a new measurement function by: (a) identifying scenarios with similar attribute values, leaving only one attribute whose values are different across scenarios; for example, two choice scenarios could be $(\Delta x_1,\ \Delta x_2,\ M)$, $(x_1',\ \Delta x_2',\ M')$ where M and M' are two measurements on the individuals' responses to scenarios $(\Delta x_1,\ \Delta x_2)$ and $(\Delta x_1',\ \Delta x_2')$, with $\Delta x_1 = \Delta x_1'$; and (b) recalculating a modified measurement on the difference between the two measurements $\bar{M} = M - M'$.

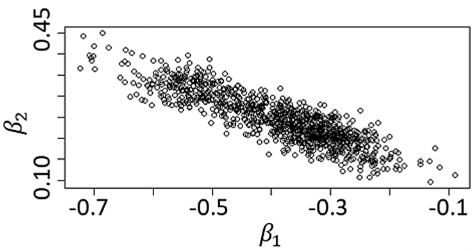

Fig. 3. Distribution of two-dimensional preference samples. Each sample gives rise to the same measurement M. True preferences are: $(\beta_1, \beta_2) = (-0.3, 0.2)$. Each point in the distribution is a sample in particle filter approach and there are 1000 samples in total, with distributions of attribute differences being $\Delta x_1 \sim N(30, 1^2)$ and $\Delta x_2 \sim N(-15, 0.5^2)$. t = 100.

(2) Accordingly, we modify the likelihood function (e.g., $L = Pr(M - M'|\Delta x_2 - \Delta x_2', \beta_2)$) such that the likelihood of obtaining the modified measurement is only conditioned upon a single parameter (e.g., β_2), and the attribute it associates with, or the difference between two different attribute values from different scenarios (e.g., $\Delta\bar{x} = \Delta x_2 - \Delta x_2'$). Specifically, the likelihood function in (12) is modified as:

$$p(M^t|\beta_j^t) = \frac{1}{\sqrt{2\pi}m}\exp\left(\frac{-\left[-\log\left(\frac{1-\bar{M}}{\bar{M}}\right) - \Delta\bar{x}\beta_j^t\right]m^{-2}\left[-\log\left(\frac{1-\bar{M}}{\bar{M}}\right) - \Delta\bar{x}\beta_j^t\right]}{2}\right) \tag{13}$$

Given that attribute values are available, the one-dimension parameter is thus solvable.

2.4.3. Solution 2: Sampling with domain knowledge

Domain knowledge refers to existing knowledge about travel behavior in existing studies. Past studies in travel behavior have revealed many insights on (1) what factors tend to matter in what types of behavior scenarios? For example, in departure time choice scenarios, only a limited number of factors (e.g., scheduled delay and travel time) are found to matter (Gaver Jr, 1968; Mahmassani and Chang,1986; Noland and Small, 1995; De Palma et al., 1997); (2) how various factors may affect travel behavior choices? For example, the parameters for travel time and travel cost must be negative and a rough range could be identified for each parameter from previous studies. In some cases, not only ranges are available but also their approximated distribution forms. Such knowledge can be useful in the resampling scheme within the particle filter process in providing us the information on what types of distributions to be resampled from.

To summarize, the proposed particle filtering approach in our study is implemented in the following steps. For each dimension β_k of the preference vector $\boldsymbol{\beta}$, we seek to use the divide and conquer strategy which translates multi-dimensional learning into a sequence of conditional one-dimensional learning.

a. Retrieve preference samples $\beta_{k,j}^{t-1}$ with corresponding weights $w_{k,j}^{t-1}$ from the previous update;
b. Perform the time evolution step using the process function (Eq. (9)):

$$\beta_{k,j}^t = \beta_{k,j}^{t-1} + u_j, \quad u_j \sim N(0, \sigma^2)$$

c. Update each sample weight by evaluating how likely the sample represents the individual's preferences β_k:

$$w_{k,j}^t \propto w_{k,j}^{t-1} \times L$$

Here $L = Pr(M - M'|\Delta x_k - \Delta x_k', \beta_{k,j}^t)$ is the likelihood of observing $\bar{M} = M - M'$ given sample j (see Eq. (13));

d. Normalize the weights following the update;
e. Generate a set of posterior particles via importance resampling based on the distribution of particle weights. Domain knowledge is added as constraints in the resampling process; and
f. Update preference estimate $\hat{\beta}_k^t = \sum_{j=1}^{J} w_{k,j}^t \beta_{k,j}^t$.

In the beginning, preference samples are initialized using a broad uniform distribution centered at values found in the literature and the weight of each sample is set as $1/J$. After updating preference estimates, they are then stored for future use, i.e. calculating incentives in the PREDICTION module for the next interaction with the individual. In Appendix B, we provided a convergence analysis of our preference learning algorithm.

3. System testing

System testing in this section includes tests with simulations and an online experiment. In simulation tests, an individual is characterized by a presumed preference vector, which could, for example, decide to switch to the suggested departure time. The elements in the preference vector represent the individual's preferences toward attributes including travel time, travel cost, and reward points. The performance on one-dimensional learning has been extensively studied in the authors' previous work (Zhu et al., 2018), showing that the system can learn and track individual preferences with both high accuracy and efficiency under various levels of measurement noise. This gives support to our proposed divide-and-conquer strategy. In this section, we show simulation results of multi-dimensional learning, with the divide-and-conquer strategy and domain knowledge applied.

The online experiment is a stated choice experiment, wherein participants were asked to make hypothetical departure time choices on their commute. Participants were recruited from the AMT (Amazon Mechanical Turk) platform, which is widely used for conducting online experiments (Mason and Suri, 2012; Paolacci et al., 2010). Our system iteratively provided participants with recommended alternatives and participants responded to the system by indicating whether or not they would accept the recommendations. Through the interactions with the respondents, the system learns their preferences in order to provide personalized incentives associated with recommended travel choices which the participant is likely to accept. Different from the simulations, when interacting with an individual (in this AMT experiment), the true values of her preferences are unknown. Thus, we evaluate the acceptance ratio, which is defined as the fraction of times that a participant accepted the recommended alternatives.

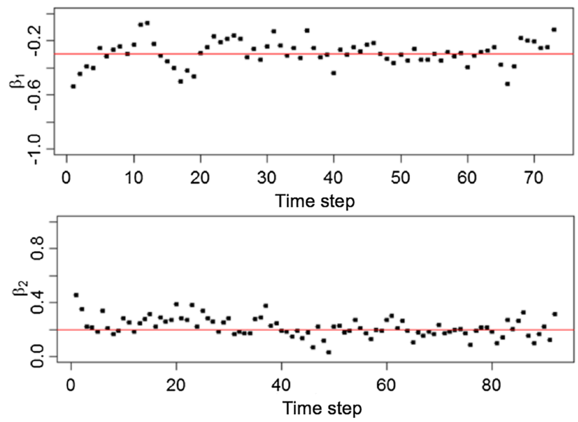

Fig. 4. Learning two-dimensional preferences using the "divide and conquer" strategy. (Single simulation run with true preference $\beta_1 = -0.3$, $\beta_2 = 0.2$ shown by red lines in the figure.)

3.1. Simulation tests

3.1.1. Performance of preference learning

3.1.1.1. Preference learning performance of "divide and conquer" strategy. As described earlier, the modified measurement function and the likelihood function between two scenarios sharing common attributes and one attribute with different values are $M - M'$ and $Pr(M - M'|\Delta x_2 - \Delta x_2', \beta_2)$, respectively. Here, we evaluate the performance of our proposed strategy in learning a two-dimensional preference vector $\boldsymbol{\beta} = (\beta_1, \beta_2)$. Fig. 4 shows an example of the preference learning with divide and conquer strategy applied at each time step. The result suggests that with the strategy, the system could estimate the multi-dimensional preferences quickly: starting from the initial values, the estimates approach true values within several steps, though fluctuating due to the measurement noises (Fig. 4).

3.1.1.2. Preference learning performance when adding domain knowledge as constraints. Fig. 5 shows between 40% and 50% reduction in error in the estimation of the preferences after we add domain knowledge as constraints in the resampling process. More specifically, in this test with true preferences being $\beta_1 = -0.3$ and $\beta_2 = 0.2$, when sampling from the posterior distribution in the particle filter, we abandon those samples having opposite signs from the true β_1 and β_2, based on the domain knowledge (e.g., the value of travel cost is negative).

3.1.2. Simulation tests on acceptance ratio

When interacting with humans, preferences are not known. Therefore, we evaluate the performance of the system by using the acceptance ratio (*AR*).

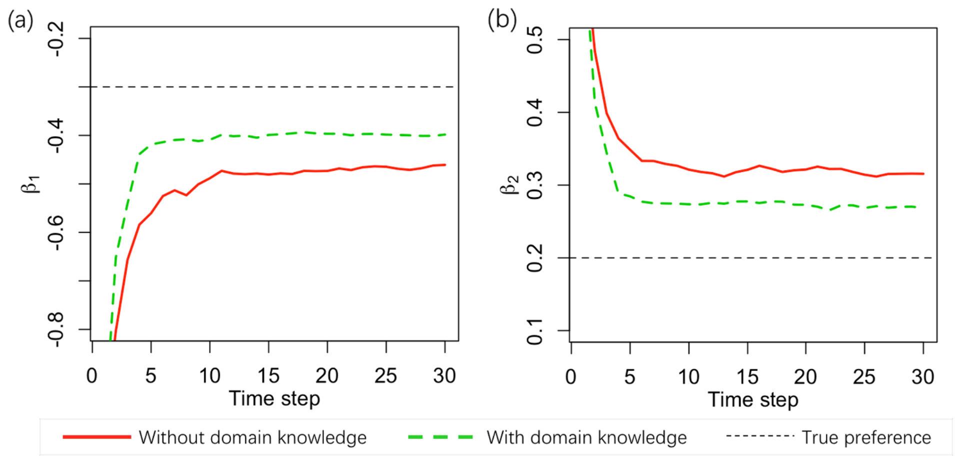

Fig. 5. Learned preferences vs time steps using two-dimensional particle filter with and without domain knowledge as constraints. (a) Learning β_1, and (b) learning β_2. (50 simulation runs; true preferences are $\beta_1 = -0.3$ and $\beta_2 = 0.2$).

Table 1
Effects of the setting of p_1. (η = 10%; results are the average of 100 independent runs).

p_1	0.6	0.7	0.8	0.9
AR	0.64	0.72	0.76	0.85
Avg_reward_point	39.5	41.6	42.3	46.7

$$AR = \frac{N_{times\ \ of\ \ the\ \ promoted\ \ alternatives\ \ being\ \ accepted}}{N_{total\ \ times\ \ of\ \ choices\ \ made}} \tag{14}$$

As a preliminary test in simulation, we investigate how *AR* could be affected by measurement noise and the probability threshold p_1 at which we expect an alternative to be accepted. We vary the measurement noise υ_m by changing the standard deviation *m* of the normal distribution (Eq. (5)). To make the noise comparable with U_1, m is defined by $m = \eta \times \mathrm{E}(V_1)$, where η is a tuning variable referring to the level of measurement noise with respect to $\mathrm{E}(V_1) = \mathrm{E}(\Sigma_k \beta_k x_k)$, the expectation of the systematic utility of an alternative in the simulation. p_1 is the probability that an individual will accept the promoted alternative given RP^t points as the incentive, and, according to Eqs. (2) and (3), directly influences the number of reward points used for each trip. For example, p_1 equals 0.5 means that the amount of reward point RP^t is calculated such that the attractiveness (utility) of the promoted alternative (with the incentives) is the same as that of the default choice.

3.1.2.1. Effects of measurement noise. In this simulation we set $p_1 = 0.6$. We observe that when the level of measurement noise η increases in simulation, *AR* decreases. More specifically, given noise levels at 10%, 50% and 100%, the *AR*s obtained are 0.64, 0.59 and 0.56, respectively. This may be because that under larger noise, the accuracy of preferences learned by the system is lower (Zhu et al., 2018), which leads to lower acceptance ratios.

3.1.2.2. Effects of p_1. As mentioned above, p_1 is an important parameter of our system which is the expected probability of the promoted alternative being accepted. By setting a larger p_1, we increase the expected probability of the promoted alternative being selected by providing more reward points, leading to a higher acceptance ratio by an individual. Table 1 shows the average reward points required per trip (*Avg_reward_point*) to yield an acceptance ratio that is close to p_1: *Avg_reward_point* increases from about 39.5 points per trip when $p_1 = 0.6$ to 46.7 points per trip when $p_1 = 0.9$.

3.2. Online experiment

The core of the experiment is a series of hypothetical choice scenarios (Fig. 6), mimicking sequential interactions between the system and participants. A total of 13 choice scenarios are presented to each participant. Each choice scenario contains a binary choice question, including a default choice associated with the participant and a recommended alternative that varies in departure time, travel time saving, and reward points. Some basic information on the participants was also collected, including individuals' travel behaviors and some socio-demographic information.

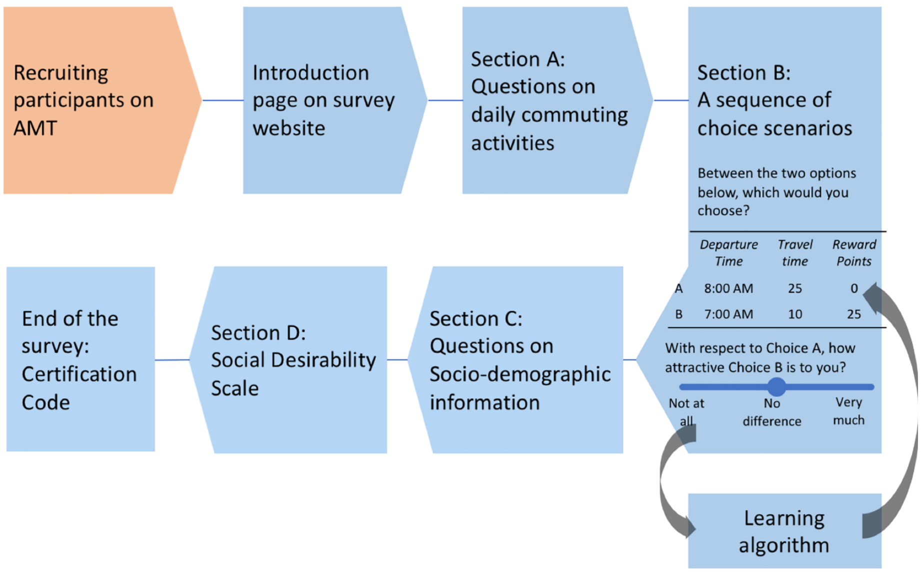

Fig. 6. Flowchart of the experiment.

Scenarios for this online experiment focus on departure time changes. To encourage an individual to depart at a time different from her default one, we provide reward points so that the deviation from the default choice is compensated. Note that the departure time changing model is used for the purpose of testing our methodology, and our methodology would be applicable to promoting other travel behavioral changes such as route changes or mode changes.

Since preference estimates are learned using data from one individual only, only alternative-based attributes contribute to the utility model[4]. Travel time and schedule delay (early or late) are selected as relevant alternative-based attributes based on existing studies, which report that these two factors are the primary ones in departure time choices for commuting trips (e.g., Gaver Jr, 1968; Small, 1982; Mahmassani and Chang,1986; Noland and Small, 1995; De Palma et al., 1997; Jou et al., 2008). Consequently, our utility model involving travel time, schedule delay and reward points can be expressed as:

$$U_1 = \beta_{TTS} \times TTS + \beta_{SDE} \times SDE + \beta_{SDL} \times SDL + \beta_{RP} \times RP + \varepsilon \tag{15}$$

Here, *SDE* and *SDL* are minutes scheduled earlier and later than the desired arrival time respectively, and for any individual only one of them is used. *TTS* is travel time savings in minutes and *RP* is the amount of reward points. Coefficients of *SDE* and *SDL* are expected to be negative while those of *TTS* and *RP* are positive.

Participants are randomly grouped into 9 context groups, defined by default commuting time (3 levels) and flexibility of work starting time (3 levels). Prior to presenting hypothetical choices, each participant is assigned to a particular group where the context information is given. For example, one participant may be told at the beginning of the series of choices that "imagining your company requires arrival at work before 8:00 am and your typical commuting time is 10 min". Following a national report on commuting trips (American Association of State Highway and Transportation Officials, 2013), for each factor we have three levels: 10 min, 25 min and 60 min for commute time; fixed, encouraged, flexible schedule for flexibility[5].

3.2.1. Response quality control and data cleaning

Before we move to presenting results, we discuss the quality of the data collected. AMT has been applied to a variety of research studies and experiments across disciplines. Existing studies using AMT samples support the validity of using AMT as a platform for conducting decision making research by comparing the AMT data and the results with those from consumer panel or community/college participants (Zhou et al., 2016; Necka et al., 2016; Kees et al., 2017; Hauser & Schwarz, 2016; Garrow et al., 2018). Qualified respondents are those who are over 18 years old and with approval rates higher than 99% (i.e. 99% of the tasks done by an individual are approved by the task-providers)[6]. Similar to traditional surveys, data collected from AMT may contain data/responses of low quality (e.g., careless responses). To control data quality, we apply established quality-control techniques to identify and filter out data of low quality (Iarossi, 2006; Meade and Craig, 2012). In our experiment, low-quality data may result from several concerns. First, since the experiment, a stated preference survey, relies on participants' understanding in the hypothetical background setting, those who do not pay sufficient attention to the introduction of our background setting will produce unqualified responses. Second, motivated by primarily receiving payment for participating in the experiment, some participants may give responses in a random and speedy manner. Third, some participants may respond in a way that is socially desirable. With all these concerns, we design five checks to examine the data quality of each of the 926 completed surveys (Berinsky et al., 2013; Maniaci and Rogge, 2014; Meade and Craig, 2012).

Check 1. Participants' understandings in background settings

After the introduction of the background information and before proceeding to scenario choice questions, participants were required to answer three simple and straightforward questions. Those who fail any one of those three questions were removed from the experiment. Our database shows that besides those 926 participants who completed the experiment, about 300 participants failed the check questions and were removed from the experiment. This suggests that these three questions effectively identified participants who were not paying attention in understanding the background settings.

Check 2. Consistency check

We assume those who provide inconsistent responses were not paying attention to the experiment and those responses should be removed. The inconsistency was checked in two ways: (1) a response was considered inconsistent if a participant rejected (accepted) a recommendation while indicating that it was attractive (not attractive) using the slider; (2) a response to the current recommendation was considered inconsistent, if, compared with the historical response to a previous recommendation, the current one was of higher utility but given a lower evaluation on the slider (we assume a higher *RP*, higher *TTS*, or lower *SDE/SDL* gives a higher utility). We observe that more than 60% of participants have fewer than 5 inconsistent responses.

[4] Other factors such as socio-demographic factors are irrelevant here.

[5] For participants assigned with the fixed level, they are told to imagine "your company requires all employees to arrive no later than 8:00 am."; correspondingly it is "your company encourages all employees to arrive no later than 8:00 am" for the encouraged level and "your company has no requirements for arrival time, but most of the staff arrive between 7:00 am and 9:00 am" for the flexible level.

[6] Approval rate is an index on AMT representing the "reputation" of an AMT user. The denominator is the total number of tasks done by the user and the numerator is the number of tasks that are approved/accepted as qualified answers by the task-providers.

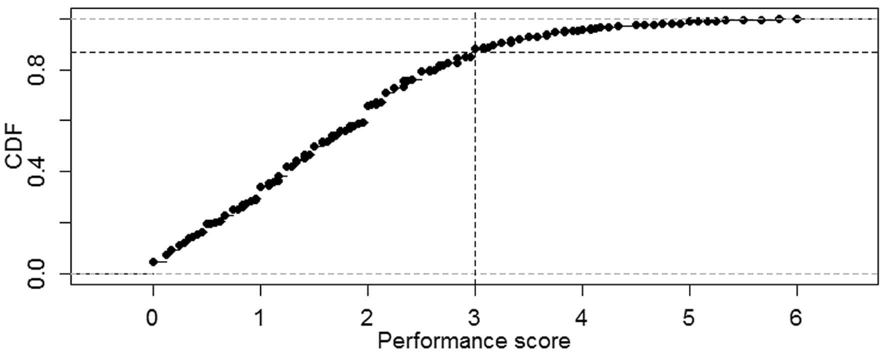

Fig. 7. Cumulative distribution for the response quality score.

Check 3. Repeated question

We repeat the first scenario at the end of Section B (i.e. the section with a sequence of departure time scenario choices as shown in Fig. 6), and the difference of responses to the two scenarios is used to check whether a participant was paying attention throughout the experiment. In our sample, 84.6% of participants made the same selections (i.e. if a participant chose the default choice in her first scenario, she chose the default choice in her last scenario as well). For 80% of participants, the difference between the two evaluations (slider readings in the two scenarios) is smaller than 0.2 ($|\Delta M| \leq 0.2$).

Check 4. Response time

For each participant, time spent responding to each choice scenario is recorded. We assume that too long or too short time taken indicates that participants did not respond to the presented scenarios carefully (Malhotra, 2008; Börger, 2016). Sixty percent of the participants finished all scenarios between 110 and 260 seconds (or from 8.46 to 20 seconds per scenario), with a median of 195 seconds.

Check 5. Social Desirability Scale

Social desirability is a common type of bias that impacts the validity of survey data (Nederhof, 1985). We use Social Desirability Scale (SDS) to test whether participants' responses result from their tendency of behaving in a socially desirable way, rather than reflecting their own preferences. Participants were asked at the end of the experiment to answer a short SDS survey with 10 items (Fischer and Fick, 1993; Strahan and Gerbasi, 1972).

In the final data cleaning process, each questionnaire is assigned with a score as an indicator of its response quality integrating all the aforementioned checks. Specifically, for each participant, a rank is given for each of the five checks and its total provides a participant score on the quality of her responses. Fig. 7 gives the cumulative distribution of the response quality scores for all participants, with a higher score indicating lower quality. Based on the Elbow rule, it is shown that a score at 3 is a good threshold to classifying questionnaires into either of good or low quality (Fig. 7). With this threshold, we remove 98 participants (about 10.6%) whose score are higher than 3.

Finally, we have 828 qualified participants in total, 387 of whom are female and rest (441) are male. About 70% of the participants are 40 or younger. The median annual income falls between \$50,000 and \$75,000. According to their stated commuting time in their daily lives, the average commuting time is about 31.58 min.[7] The correlation between the SDS score and *AR* is −0.003, which indicates that social desirability does not play a role in affecting participants' different choice making in our survey.

3.2.2. Results

Table 2 shows the median of learned βs in each group. Coefficients associated with *SDE*, *SDL*, *TTS* and *RP* (Table 2) are expected: deviation from scheduled departure time (either earlier or later) are disliked while travel time saving and reward points are favored. We further observe that with more flexibility in the working schedule, participants value less on schedule delay and more on the travel time saving and rewards. This observation is consistent with findings in previous studies (Small, 1982, Hurdle et al, 1983; De Palma et al., 1997). Another finding is that given a flexible schedule, participants prefer to arrive later rather than earlier, which is in contrast to the preference observed in scenarios with a fixed schedule. This reversal of preferences under different flexibility levels is also supported by existing studies such as He (2013).

We calculate *AR* for each individual. Distribution-wise, 15% of them tend to reject recommendations (*AR* is smaller than 0.5),

[7] The average commuting time is close to the National Report on Commuting Patterns and Trends from AASHTO in 2013, in which the average travel time for people who drive alone is 24.19 min. For carpooling the average travel time is 38.89 min.

Table 2
The median of individual preferences learned from the online experiment.

Flexibility Level	β_{SDE}	β_{SDL}	β_{TTS}	β_{RP}
Fixed	−0.092	−0.099	0.010	0.053
Encouraged	−0.075	−0.061	0.081	0.056
Flexible	−0.061	−0.055	0.129	0.068

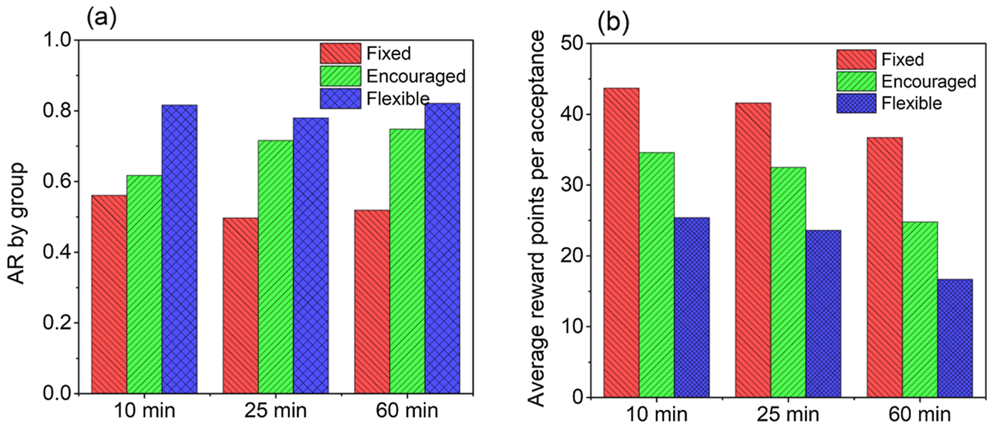

Fig. 8. Acceptance ratio and reward point use in each group. (a) average *AR* in each group; (b) points rewarded per acceptance in each group.

20% of them have equal chances for acceptance and rejection (*AR* equals to 0.5), while the majority (65%) are more likely to accept them (*AR* is larger than 0.5). The average *AR* is 0.68.

Fig. 8(a) gives *ARs* in all nine groups, which shows that a flexible working schedule leads to a higher *AR* regardless of commuting time. Fig. 8(b) examines the rewards needed for each accepted scenario. We observe that despite the commuting distance, with personalized incentives, it costs the system less to persuade commuters with flexible working schedules to change their behaviors. On the other hand, compared with short-distance commuters, it takes fewer reward points to persuade long-distance commuters to change their departure times. This observation is consistent with the previous finding (Chin, 1990) that with the same value of incentives, long-distance commuters are more likely to change their departure times.

Equity issues can potentially arise when a small portion of individuals obtain the disproportional amount of reward points, as the result of rewards manipulation[8] or simply lower values associated with rewards. We examined the existence of this issue with the AMT dataset involving 828 participants, each with 13 scenarios. About 11% of the total reward points are given to the top 5% of individuals on the average reward points given, and 34% of the total reward points are given to the top 20% of individuals. Among those top 5% of the individuals, 74% of them have an annual income over $50,000, which is higher than the 65% for the rest of the group. These statistics do not suggest grave equity concerns. In other words, rewards manipulation is not evident in this sample. In real-world implementation when rewards are of actual monetary value and the sample is a truly random sample of the target population, such issues could potentially arise.

3.2.3. Evaluation on the transportation system

Studies have shown that the diversity of work starting hours (i.e. the preferred arrival time) could reduce the total travel cost (Small et al., 2007). This can be demonstrated by taking a simple dynamic queueing model—the bottleneck model (Arnott et al., 1990), which models the morning rush congestion as a vertical queuing process that occurs behind a bottleneck. The model defines a user equilibrium based on commuters' travel decisions: those prefer arriving closer to the work starting time are likely to experience longer travel delays in the queue, while those prefer fewer travel delays in the queue are more likely to arrive at work too early or too late. At the equilibrium[9], no commuter can reduce her travel cost by changing the arrival time and the total travel cost *TC* of all commuters is expressed as (de Palma and Lindsey, 2002):

$$TC = \left(1 - \frac{s}{2r}\right)\frac{N_D}{s}\frac{\beta_{SDE}\beta_{SDL}}{\beta_{SDE} + \beta_{SDL}} \tag{16}$$

[8] Some may purposefully reject recommendations with smaller rewards in order to obtain larger ones in the future.

[9] Assumptions to achieve the expressed equilibrium include: (1) desired arrival rate is larger than the service rate of the bottleneck (i.e., $r > s$); (2) homogeneous preferences among commuters; (3) the cost of a slight delay in departing from work when already late must be less than the cost of an equal amount of time spent queuing.

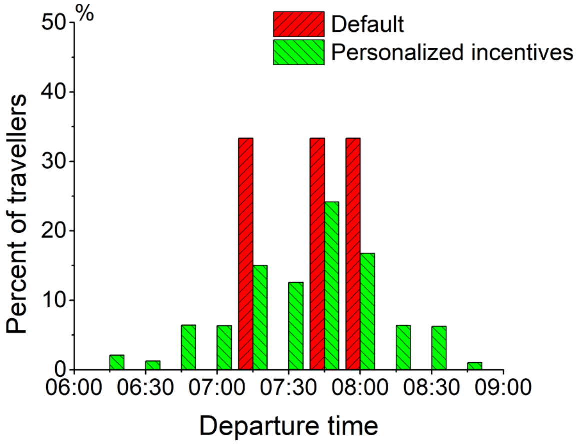

Fig. 9. Departure time distributions with and without personalized incentives.

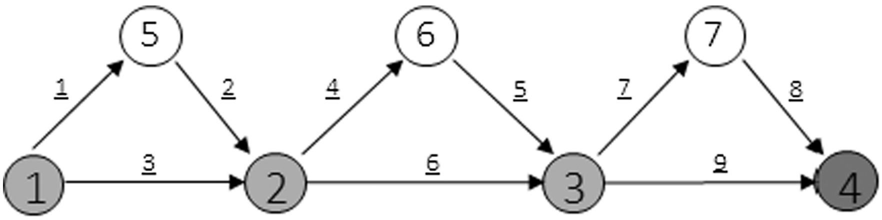

Fig. 10. Test network.

Here, r represents the desired arrival rate, by assuming a uniform arrival distribution. N_D is the total number of commuters who are delayed by the bottleneck with a service rate of s. β_{SDE} and β_{SDL} represent commuters' preferences on schedule delay early and schedule delay late, respectively. Note that a fixed work starting schedule means that all commuters arrive at the same time and the arrival rate r can be expressed as $r = \infty$. This suggests that a flexible work starting schedule with a r arriving rate ($r \neq \infty$) would reduce the total travel cost by a fraction of $s/2r$. Therefore, an incentive system promoting off-peak-hours travel would diversify the arrival profile and consequently lead to less travel cost.

Fig. 9 gives the departure time distribution collected from the responses of participants. It shows that with personalized incentives, the demand profile is significantly diversified, compared with the default departure distribution. In the following, we quantitatively evaluate how the diversified demand profile leads to the system-level impact using data collected from our online experiment, i.e. the influence of our proposed personalized system on total travel time in a simplified network.

For the study, we use a Cell Transmission Model (CTM) to load the dynamic demand (Fig. 9) in a simple network (Fig. 10). This network is designed with two important considerations: first, our online experiment involves three different commuting time groups in terms of commute time (10 min, 25 min, and 60 min). This suggests a network with three origins (Node 1, 2, and 3) and one destination (Node 4); second, to evaluate the travel time saving with the personalized incentive, we need to model the interaction between the demand and supply such that the formation and dissipation of congestions could be observed. The applied network together with the cell transmission model enables us to capture the interactions: (1) given the limited capacity of the links and the maximum passing rate for each node, queues will be built if the demand exceeds the supply and the alternative routes available allow for rerouting behavior; (2) the rerouting behavior would reshape the demand on each link and change the traffic state.

CTM discretizes traffic in both time and space (Daganzo, 1994) by splitting a road into sections (i.e. cells) and modelling the dynamics via updating traffic on cells at each time step. The update works iteratively following the equation:

$$occup_e(t+1) = occup_e(t) + infl_e(t) - infl_{e+1}(t) \tag{17}$$

where $occup_e(t+1)$ is the occupancy of cell e at time $t+1$, $occup_e(t)$ is the occupancy at time t, $infl_e(t)$ is the inflow to cell e at time t and $infl_{e+1}(t)$ is the outflow of cell e (i.e. the inflow to the downstream cell (cell $e+1$)). More details of the network configuration can be found in the Appendix A.

Fig. 11 shows the time-space diagrams on arc 2 with and without personalized incentives. With the personalized incentives, the spillback caused by the congestion at the head of arc 2 disappears. Fig. 12 gives the distribution of the travel time for commuters. It shows that the diversified demand profile brings a system-wide saving on total travel time. More specifically, compared with the default departure time profile, the diversified profile enabled by personalized incentives leads to a 48% reduction on total commuting time.

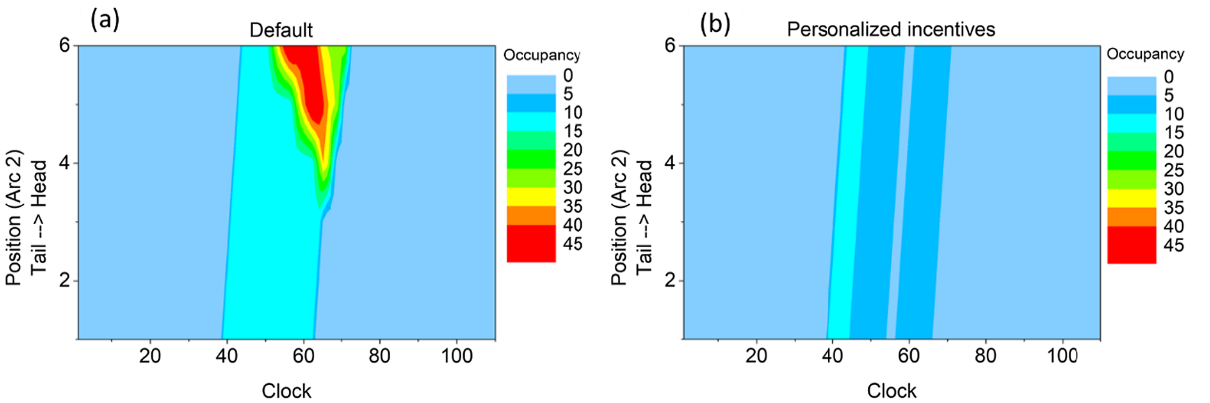

Fig. 11. Shockwaves (a) with and (b) without personalized incentives. The spillback in (a) caused by the congestion disappears in (b).

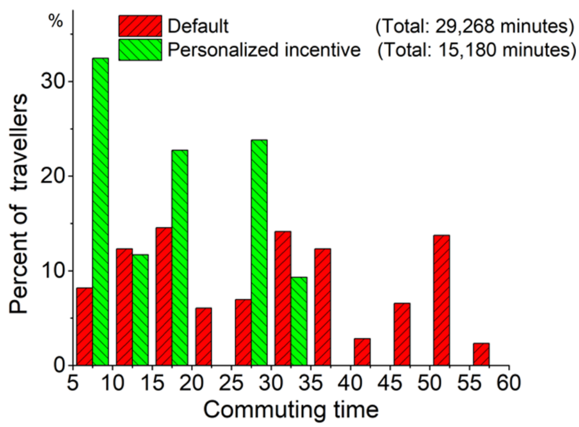

Fig. 12. Distributions of commuting time with and without personalized incentives.

4. System-level considerations

The personalized system introduced above focuses on the design of the incentive mechanism at the individual-level: for every individual, the system learns her preferences and tries to influence her travel choices by providing personalized incentives (reward points in this case). System-level considerations are not incorporated and they relate to, for example, deciding which promoted alternative(s) shall be presented to which individual, minimizing total reward points given subject to a minimum threshold improvement in system performance, or addressing equity concerns such that reward points are not doled out to favor those who may manipulate the system for more rewards. Achieving these system-level goals at the same time would require significant research that goes significantly beyond the current scope of the paper, which aims to address challenges at the individual level, or more specifically, learning individual preferences for personalized recommendations. In this section, we discuss some preliminary results with system-level considerations. More discussions are presented in Section 5.

Minimizing the total amount of rewards given subject to a minimum threshold improvement in system performance can be formulated as a Facility Location Problem. By it at time step t, we can determine to whom an incentive shall be provided. For the convenience of problem formulation, we present in the following a summary of notations used in the optimization framework (the notation of time step t in the superscript is omitted for simplicity):

Nomenclature	
N	total number of individuals to whom the personalized system could promote alternatives
a	index for individual a, and $a \in \{1, 2, \cdots, N\}$
$\boldsymbol{x}_0^a$	a vector of attribute values of the default alternative for individual a
$\boldsymbol{x}_1^a$	a vector of attribute values of the alternative to promote to individual a, which is assumed known
TT_0^a	individual a's default travel time in minutes
TTS^a	the travel time saving for individual a if she accepts the promoted alternative (in min)
y^a	$y^a \in \{0, 1\}$. $y^a = 1$ means that the system decides to promote the alternative to individual a and $y^a = 0$ otherwise
C	a pre-specified percentage (e.g., 5%) by which the transportation system performance is expected to improve

We denote $f\left[(\boldsymbol{x}_0^1, \boldsymbol{x}_1^1, y^1), \cdots, (\boldsymbol{x}_0^a, \boldsymbol{x}_1^a, y^a), \cdots, (\boldsymbol{x}_0^N, \boldsymbol{x}_1^N, y^N)\right]$ be a nonlinear function whose output is the total travel time reduction, and inputs include individual a's default travel choice $\boldsymbol{x}_0^a$, the alternative $\boldsymbol{x}_1^a$, and the binary variable y^a. Interactions exist among individuals at the network level—as more individuals change departures times, their choices affect the travel times of other individuals, resulting in changes in total travel time reductions. Considering the interactions, we expect a S-shape curve over time: (1) When promoting alternatives to a tiny fraction of individuals (i.e. $\sum_{a=1}^{N} y^a \ll N$), the impact on the traffic flow is insignificant and we could consider no interaction exists. The total travel time reduced could be calculated as a simple summation of travel time reductions by individuals who change their departure times; (2) With an increasing fraction of individuals being promoted alternatives, the interactions among individuals would emerge at a certain threshold and the total travel time reduced in the transportation system could only be estimated via an external traffic network model (which is outside the scope of this paper). This is supported by both simulations (Vinitsky et al. 2018) and field experiments (Stern et al. 2018) showing that changing behaviors for a small fraction of drivers (e.g., 1/20 in (Stern et al. 2018)) on the road could improve the overall traffic flow; (3) When the fraction gets larger and exceeds another threshold, the marginal benefits from more people changing their behaviors decrease, leading to flatter slope at the end of the curve.

Given the objective function and constraints discussed above, a general optimization framework could be written as follows:

$$\begin{gathered}\min_{(y^1, \cdots, y^N)} \sum_{a=1}^{N} y^a \cdot RP^a \\ s.\ t. \\ \frac{f\left[(\boldsymbol{x}_0^1, \boldsymbol{x}_1^1, y^1), \cdots, (\boldsymbol{x}_0^a, \boldsymbol{x}_1^a, y^a), \cdots, (\boldsymbol{x}_0^N, \boldsymbol{x}_1^N, y^N)\right]}{\sum_{a=1}^{N} TT_0^a} \geq C \\ y^a \in \{0, 1\}, \quad a \in \{1, 2, \cdots, N\}\end{gathered} \tag{18}$$

We tested this framework by taking the last scenarios of all the respondents in our online experiment. We assume no interaction among individuals' travel choices and therefore the system-level constraint (i.e. the first constraint in (18)) is expressed as:

$$\frac{\sum_{a=1}^{N} y^a \cdot TTS^a}{\sum_{a=1}^{N} TT_0^a} \geq C \tag{19}$$

Given a pre-specified system-level constraint, we could then know how many individuals to whom the system should promote alternatives. Solving the optimization framework (18) with its system-level constraint expressed in (19) identifies 8.7% of the individuals (72 in 828 individuals) to whom personalized incentives shall be provided in order to meet a system-level constraint of $C = 5\%$.

5. Conclusion and discussions

Personalized incentives have the potential leading to higher acceptance for sustainable travel behavior changes, which has many benefits ranging from reduction in congestion to energy saving and air quality improvement. In this paper, to encourage sustainable travels, a personalized system is proposed to calculate the minimum amount of incentives needed for an individual to accept the promoted alternative with a pre-defined probability (> 0.5). With simulations, we show that the system could dynamically learn individuals' preferences, which are essential to provide personalized incentives. The online experiment involing humans has an average acceptance ratio of 0.68. For the system-level impacts, we show that the diversified demand profile resulting from individual behavioral changes leads to a 48% reduction in total travel time. Obviously, when implemented in the real world involving not only human participants but also real traffic conditions, we expect a reduction in acceptance ratio. It is, however, hard to speculate how much the reduction will be without actually conducting a real-world implementation. This will represent an important future direction.

One important future research direction to be pursued is the integration of person-level and system-level considerations, as noted in various places previously throughout the paper. For real-world implementations, system-level considerations must be taken into account as they will allow us to: (1) consider diverse alternatives (e.g., mode changes, departure time choices, and route choices, etc.) based on the state of the traffic; (2) capture interactions among individuals and their respective impacts; and (3) decide on the amount of personalized incentives considering not only behavioral factors (as this paper does) but also achieving a system-level goal. Research directions 1 and 2 above can be achieved by simply linking our personalized system to existing traffic and network models. This will enable a number of capabilities that are outside the scope of this paper but needed for real-world implementation, for example, offering diverse alternatives to an individual depending on the state of the traffic and more accurately capturing the emerging effects of individual choices on the system (compared to the no interaction assumption in Section 4).

Research direction 3 above requires new models to be developed, or a sequential connection with a network model (as suggested above for research directions 1 and 2) will not be satisfactory. Many network models have been developed to design an incentive/disincentive mechanism (Rey et al., 2016) such as a pricing scheme (Lou et al., 2010) and to evaluate the impact of the mechanism on network performance under pre-determined system-level goals. These system-level goals may include: pareto-optimal, revenue-neutral, balanced budgets, and/or equitable distribution of the reward points. These models typically assume some types of equilibriums (Xu et al. 2011; De Palma and Lindsey, 2004) and homogeneous preferences in the population (e.g., value of time) (Xu et al., 2011). Several designs have been proposed, such as lottery-based incentive mechanism (Rey et al., 2016), system-optimum tolling and second-best flat and step tolling (De Palma et al., 2005), and their impacts have been evaluated. Some models take into account the heterogeneity of travel behaviors among travelers (De Palma and Lindsey, 2002, 2004) by assuming a distribution for preferences, finding that the heterogeneity could influence the design of incentive/disincentive mechanisms (De Palma and Lindsey, 2004). These models are ensemble models as they devise an optimal pricing scheme in the form of a distribution. Individual-level behavioral factors are not considered—questions such as to whom and what amount of the incentive shall be provided and whether the individual will likely accept the incentivized alternative are unknown. Without answering these questions, there are substantial practical difficulties in implementing in the real world those incentive schemes. This means a new kind of model integrating system- and individual-level goals is needed so that the incentives doled out to individuals not only improve the attractiveness of the promoted alternative so that the individual will likely accept it but also meeting system-level goals. To the authors' knowledge, no such model exists.

This paper presents a promising approach and a case study that integrates knowledge from control theories (in particular state estimation techniques) and travel behavior to achieve a personalized incentive system. The system allows dynamic learning of individual preferences based on only a few observations, tolerates the binary nature of the observed choices, and uses the behaviorally-sound random utility maximization framework. Though much needs to be done as discussed in this Section, the work presented does point to a promising future where control theories and the associated tools may be leveraged for changing human behaviors.

Acknowledgments

The information, data, or work presented herein was funded in part by the Advanced Research Project Agency-Energy (APAR-E), U.S. Department of Energy, under Award Number DE-AR0000614. The study was also supported by USDOT Tier 1 TOMNET (Teaching Old Models New Tricks) UTC Center. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. The authors appreciate constructive conversations with a number of colleagues, including Dan Work of Vanderbilt University who invited the corresponding author to a IPAM workshop at UCLA in 2016 and had many subsequent discussions on the use of particle filter for travel behavior modeling, An-Tsu Chen of UW who helped develop the measurement function, and Jeff X. Ban of UW and Anthony Chen of Hong Kong Polytechnic University on discussions for system evaluation and the use of facility location models.

Appendix A. Network configuration

In the system evaluation presented in Section 4, we use CTM to load the dynamic demand in the network shown as Fig. 10. Taking links 3, 6 and 9 as freeways and the other links as arteries, the turning portion at nodes 2, 3 and 4 are 60% for freeways and 40% for arteries. We assume equal merging portions (i.e. 0.5 for both freeway and arteries) at all merging conjunctions. Further, the length of each time interval is set as 2.5 min in implementing CTM. More details on the properties of the network are provided in Table A.1.

Table A.1
Configuration of the test network.

Link	Length (min)	Flow capacity Q (vehicles per min)	Jam density N	Length (miles)	Speed (mph)
1	17.5	15	400	8.75	30
2	17.5	15	400	8.75	30
3	26	20	1050	17.5	40
4	7.5	15	169	3.75	30
5	7.5	15	169	3.75	30
6	11.25	20	450	7.5	40
7	5	15	113	2.5	30
8	5	15	113	2.5	30
9	7.5	20	300	5	40

Appendix B. Convergence of the learning algorithm

In our paper, we use a particle filter to estimate the posterior distribution of an individual's preferences. The results in (Crisan and Doucet 2002) show that the convergence of the particle filter is ensured under very loose assumptions, with convergence rate independent of the dimension of the state space $\boldsymbol{\beta}$.

Given a bounded importance weight and a standard sampling scheme (i.e. importance sampling), the samples converge to the posterior distribution, with the mean square error at time step t expressed as following (Crisan and Doucet 2002; Van Der Merwe et al. 2000).

> "*Let $B(\mathbb{R}^n)$ be the space of bounded, Borel measurable function on $\mathbb{R}^n$. We denote $||f|| \triangleq \sup_{x\in\mathbb{R}^n} |(x)|$. For all $t \geq 0$, there exists c_t independent of the number of samples J, such that for any $f_t \in B(\mathbb{R}^{n_x\times(t+1)})$*",

$$E\left[\left(\frac{1}{J}\sum_{j=1}^{J} f_t(\boldsymbol{\beta}_{0:t}^{(j)}) - \int f_t(\boldsymbol{\beta}_{0:t})p(d\boldsymbol{\beta}_{0:t}|M_{0:t})\right)^2\right] \le c_t\frac{||f_t||^2}{J}. \tag{20}$$

Here, the expectation E considers the randomness in particle filter realizations. Recall that we follow a standard sampling scheme which sets the importance density to be the same as the prior. Our importance weight (Eq. (12)) follows the likelihood function (Eq. (7)), which is bounded. Since we use the sample mean of particles as the estimation of the state $\boldsymbol{\beta}$, thus in our case, $f(\boldsymbol{\beta}^j) = \boldsymbol{\beta}^j$. With particles initialized with a bounded uniform distribution and domain knowledge applied, $||\boldsymbol{\beta}^j||$ is bounded to the maximal $\boldsymbol{\beta}^j$ allowed (i.e. the upper bound of the uniform distribution).

References

American Association of State Highway and Transportation Officials: Commuting in America, 2013. The National Report on Commuting Patterns and Trends. American Association of State Highway and Transportation Officials, Washington, DC.

Arnott, R., de Palma, A., Lindsey, R., 1990. Economics of a bottleneck. J. Urban Econ. 27, 111–130.

Atasoy, B., Ikeda, T., Song, X., Ben-Akiva, M.E., 2015. The concept and impact analysis of a flexible mobility on demand system. Transport. Res. Part C: Emer. Technol. 56, 373–392.

Ben-Akiva, M.E., Lerman, S.R., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press.

Berinsky, A.J., Margolis, M.F., Sances, M.W., 2013. Separating the shirkers from the workers? Making sure respondents pay attention on self-administered surveys. Am. J. Political Sci. 58, 739–753.

Börger, T., 2016. Are fast responses more random? Testing the effect of response time on scale in an online choice experiment. Environ. Resour. Econ. 65 (2), 389–413.

Chen, C., Ma, J., Susilo, Y., Liu, Y., Wang, M., 2016. The promises of big data and small data for travel behavior (aka human mobility) analysis. Transport. Res. Part C: Emer. Technol. 68, 285–299.

Chen, H., Rakha, H.A., Sadek, S., 2011. Real-time freeway traffic state prediction: a particle filter approach. In: 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp. 626–631.

Chin, A.T.H., 1990. Influences on commuter trip departure time decisions in Singapore. Transport. Res. Part A: General 24, 321–333.

Crisan, D., Doucet, A., 2002. A survey of convergence results on particle filtering methods for practitioners. IEEE Trans. Signal Process. 50, 736–746.

Daganzo, C.F., 1994. The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. Transport. Res. Part B: Methodol. 28, 269–287.

De Palma, A., Khattak, A., Gupta, D., 1997. Commuters' departure time decisions in Brussels, Belgium. Transport. Res. Record: J. Transport. Res. Board 1607, 139–146.

De Palma, A., Kilani, M., Lindsey, R., 2005. Congestion pricing on a road network: a study using the dynamic equilibrium simulator METROPOLIS. Transport. Res. Part A: Policy Practice 39, 588–611.

De Palma, A., Lindsey, R., 2002. Comparison of morning and evening commutes in the Vickrey bottleneck model. Transport. Res. Record: J. Transport. Res. Board 1807, 26–33.

De Palma, A., Lindsey, R., 2004. congestion pricing with heterogeneous travelers: a general-equilibrium welfare analysis. Netw. Spatial Econ. 4, 135–160.

Doucet, A., Johansen, A.M., 2011. A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later.

Fischer, D.G., Fick, C., 1993. Measuring social desirability: short forms of the Marlowe-Crowne social desirability scale. Educ. Psychol. Measur. 53, 417–424.

Gärling, T., Axhausen, K.W., 2003. Introduction: habitual travel choice. Transportation 30, 1–11.

Garrow, L.A., Chen, Z., Ilbeigi, M., Lurkin, V., 2018. A new twist on the gig economy: conducting surveys on Amazon mechanical turk. Transportation 1–20.

Gaver Jr, D.P., 1968. Headstart strategies for combating congestion. Transport. Sci. 2 (2), 172–181.

Giuliano, G., 1992. Transportation demand management: promise or Panacea? J. Am. Plann. Assoc. 58, 327–335.

Hauser, D.J., Schwarz, N., 2016. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. Behav. Res. Methods 48 (1), 400–407.

He, S.Y., 2013. Does flexitime affect choice of departure time for morning home-based commuting trips? Evidence from two regions in California. Transp. Policy 25, 210–221.

Hensher, D.A., 1994. Stated preference analysis of travel choices: the state of practice. Transportation 21, 107–133.

Hess, S., Daly, A., Batley, R., 2018. Revisiting consistency with random utility maximisation: theory and implications for practical work. Theory Decis. 84, 181–204.

Hu, X., Chiu, Y.-C., Zhu, L., 2015. Behavior insights for an incentive-based active demand management platform. Int. J. Transp. Sci. Technol. 4, 119–133.

Hurdle, V., Hauser, E., Fargier, P., 1983. Effects of the choice of departure time on road traffic congestion. Theoret. Appr. Transport. Traffic Theory 8, 223–263.

Iarossi, G., 2006. The power of survey design: a user's guide for managing surveys, interpreting results, and influencing respondents. World Bank Public.

Jou, R.C., Kitamura, R., Weng, M.C., Chen, C.C., 2008. Dynamic commuter departure time choice under uncertainty. Transport. Res. Part A: Policy Pract. 42 (5), 774–783.

Katzev, R., 2003. Car sharing: a new approach to urban transportation problems. Anal. Social Issues Public Policy 3, 65–86.

Kees, J., Berry, C., Burton, S., Sheehan, K., 2017. An analysis of data quality: professional panels, student subject pools, and Amazon's Mechanical turk. J. Advert. 46 (1), 141–155.

Lima Azevedo, C., Seshadri, C., Gao, S., Atasoy, B., Akkinepally, A., Christofa, E., Zhao, F., Trancik, J., Ben-Akiva, M., 2018. Tripod: sustainable travel incentives with prediction, optimization and personalization. 97th Annual Meeting of the Transportation Research Board (TRB 2018), Washington D.C., US, January 2018.

Liu, W., Li, X., Zhang, F., Yang, H., 2017. Interactive travel choices and traffic forecast in a doubly dynamical system with user inertia and information provision. Transport. Res. Part C: Emer. Technol 85, 711–731.

Lops, P., de Gemmis, M., Semeraro, G., 2011. Content-based recommender systems: state of the art and trends. In: Recommender Systems Handbook. Springer, Boston, MA, pp. 73–105.

Lou, Y., Yin, Y., Lawphongpanich, S., 2010. Robust congestion pricing under boundedly rational user equilibrium. Transport. Res. Part B: Methodol. 44, 15–28.

Malhotra, N., 2008. Completion time and response order effects in web surveys. Public Opin. Quart. 72 (5), 914–934.

Maniaci, M.R., Rogge, R.D., 2014. Caring about carelessness: participant inattention and its effects on research. J. Res. Pers. 48, 61–83.

Mahmassani, H.S., Chang, G.L., 1986. Experiments with departure time choice dynamics of urban commuters. Transport. Res. Part B: Methodol. 20 (4), 297–320.

Mason, W., Suri, S., 2012. Conducting behavioral research on Amazon's mechanical turk. Behav Res. 44, 1–23.

McFadden, D., 1975. The revealed preferences of a government bureaucracy: theory. Bell J. Econ. 6, 401–416.

Meade, A.W., Craig, S.B., 2012. Identifying careless responses in survey data. Psychol. Methods 17, 437–455.

Moore, J.L., Chen, S., Joachims, T., Turnbull, D., 2013. Taste Over time. Temp. Dynam. User Prefer.

Möser, G., Bamberg, S., 2008. The effectiveness of soft transport policy measures: a critical assessment and meta-analysis of empirical evidence. J. Environ. Psychol. 28, 10–26.

Necka, E.A., Cacioppo, S., Norman, G.J., Cacioppo, J.T., 2016. Measuring the prevalence of problematic respondent behaviors among MTurk, campus, and community participants. PLoS One 11 (6), e0157732.

Nederhof, A.J., 1985. Methods of coping with social desirability bias: a review. Eur. J. Soc. Psychol. 15, 263–280.

Noland, R.B., Small, K.A., 1995. Travel-time uncertainty, departure time choice, and the cost of morning commutes (No. 1493).

O'Mahony, M.P., Cunningham, P., Smyth, B., 2009. An assessment of machine learning techniques for review recommendation. In: Artificial Intelligence and Cognitive

Science. Springer, Berlin, Heidelberg, pp. 241–250.
Orhan, E., 2012. Particle Filtering, vol. 8.
Paolacci, G., Chandler, J., Ipeirotis, P.G., 2010. Running Experiments on Amazon Mechanical Turk. Social Science Research Network, Rochester, NY.
Park, D.H., Kim, H.K., Choi, I.Y., Kim, J.K., 2012. A literature review and classification of recommender systems research. Expert Syst. Appl. 39, 10059–10072.
Rey, D., Dixit, V.V., Ygnace, J.-L., Waller, S.T., 2016. An endogenous lottery-based incentive mechanism to promote off-peak usage in congested transit systems. Transp. Policy 46, 46–55.
Richter, G., Raban, D.R., Rafaeli, S., 2015. Studying gamification: the effect of rewards and incentives on motivation. In: Gamification in Education and Business. Springer, Cham, pp. 21–46.
Simon, D., 2006. Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches. John Wiley & Sons.
Small, K.A., 1982. The scheduling of consumer activities: work trips. Am. Econ. Rev. 72 (3), 467–479.
Small, K.A., Verhoef, E.T., Lindsey, R., 2007. The Economics of Urban Transportation. Routledge.
Song, X., Atasoy, B., Ben-Akiva, M., 2019. Sequential personalized menu optimization through bandit learning. 98th Annual Meeting of the Transportation Research Board (TRB 2019), Washington D.C., US, January 2019.
Song, X., Danaf, M., Atasoy, B., Ben-Akiva, M., 2018. Personalized menu optimization with preference updater: a Boston case study. Transp. Res. Rec 0361198118758674.
Stern, R.E., Cui, S., Delle Monache, M.L., Bhadani, R., Bunting, M., Churchill, M., Hamilton, N., Haulcy, R., Pohlmann, H., Wu, F., Piccoli, B., Seibold, B., Sprinkle, J., Work, D.B., 2018. Dissipation of stop-and-go waves via control of autonomous vehicles: field experiments. Transport. Res. Part C: Emerg. Technol. 89, 205–221.
Stopher, P.R., 2004. Reducing road congestion: a reality check. Transp. Policy 11, 117–131.
Strahan, R., Gerbasi, K.C., 1972. Short, homogeneous versions of the Marlow-Crowne Social Desirability Scale. J. Clin. Psychol. 28, 191–193.
Van Der Merwe, R., Doucet, A., De Freitas, N., Wan, E., 2000. In: The Unscented Particle Filter. MIT Press, Cambridge, MA, USA, pp. 563–569.
Vinitsky, E., Kreidieh, A., Flem, L.L., Kheterpal, N., Jang, K., Wu, C., Wu, F., Liaw, R., Liang, E., Bayen, A.M., 2018. Benchmarks for reinforcement learning in mixed-autonomy traffic. In: Conference on Robot Learning, pp. 399–409.
Xiao, Y., Lo, H.K., 2016. Day-to-day departure time modeling under social network influence. Transport. Res. Part B: Methodol. 92, 54–72.
Xu, H., Lou, Y., Yin, Y., Zhou, J., 2011. A prospect-based user equilibrium model with endogenous reference points and its application in congestion pricing. Transp. Res. Part B: Methodol. 45 (2).
Ye, H., Xiao, F., Yang, H., 2017. Exploration of day-to-day route choice models by a virtual experiment. Transp. Res. Procedia 23, 679–699.
Zhou, H., Fishbach, A., Shaddy, F., Steinmetz, J., Bregant, J., Schroeder, J., Choshen-Hillel, S., 2016. The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. J. Pers. Soc. Psychol. 222, 493–504.
Zhu, Xi, Wang, Feilong, Chen, Cynthia, 2018. Personalized control for promoting sustainable travel behaviors. IFAC-PapersOnLine 51 (9), 31–36. https://doi.org/10.1016/j.ifacol.2018.07.006.