Towards a learning-based mode shift model: a conceptual framework

ABSTRACT: Random Utility Maximization (RUM)-based mode choice models are criticized for their poor characterization of several behavioral aspects including habit and inertia in mode choice. As a result, traditional models may end up with a misleading estimation of modal shift in cases where strong habits towards a specific mode exist. In addition, given the cross-sectional nature of traditional choice models, it is hard to address individuals’ sensitivity to various policy scenarios and to determine the period of time required to reap the benefits of the proposed policies. In order to contribute to this critical issue, this paper presents a review of existing literature on mode choice modeling focusing on capturing behavioral aspects related to habit and inertia formation. The paper then proposes a novel conceptual framework for a microsimulation learning-based mode shift model that compiles both reinforcement learning techniques with random utility maximization concepts. The proposed approach utilizes various elements of reinforcement learning methods to simultaneously integrate habit formation, level of information provision and awareness limitations. A numerical example is provided to illustrate the differences between traditional and learning-based modeling frameworks. The simulation results are in line with the prior expectations which can be considered a first step towards the model’s credibility.

KEYWORDS: Agent-based, Habitual Inertia, Limited Awareness, Mode Choice, Mode Shift, Reinforcement Learning.

1. INTRODUCTION

Mode choice models are traditionally based on the random utility maximization framework originating in microeconomics where utilities are assumed random to the modeler while choice strategies are considered deterministic from the decision maker’s viewpoint. Within such framework, a passenger assigns weights to the different attributes characterizing each of the available modes and accordingly selects the travel option that maximizes her/his utility with a higher choice probability (McFadden 1974; Ben-Akiva and Boccara 1995).

Numerous random utility maximization-based mode choice models have been developed. The simplest form of these models is the multinomial logit (MNL) model relying on the fundamental assumption of independently and identically distributed (iid) random component of utilities (McFadden 1978; Ben-Akiva and Lerman 1985). Further, more complex utility maximization-based models were developed with various mathematical formulations and different assumptions for the correlation between random residuals in order to overcome the simplified assumptions of the MNL models (McFadden 1980; McFadden and Train 2000).

The result is a set of models that can estimate the modal shares under a given set of attributes related to both the decision maker and the competing options. However, traditional models have been criticized for relaxing some behavioral aspects that may reduce their ability to predict accurately the change in passengers’ choices (i.e., mode shift). Furthermore, traditional models imply rational decision making, complete knowledge of the transportation system and full awareness of travelers of the changes occurring in the transport system once they occur (Barff et al. 1982; Verplanken et al. 1997; Verplanken and Aarts 1999; Verplanken et al. 2006). These implications are in conflict with research into bounded ratio-
nality which suggests that passengers’ rationality is bounded by the cognitive limitations of their minds, the information they could have, and the limited amount of time available to them to make decisions (Bogers et al. 2005; Molin et al. 2006; van Zuylen and Bogers 2008). Therefore, passengers lack the ability and resources to find an optimal solution, and they instead apply their rationality only after simplifying the available travel options. Hence, passengers may seek satisfactory solutions rather than the optimal ones (Bamberg et al. 2003; Chorus and Timmermans 2009). Interestingly, research has shown that passengers’ lack of both searching and processing of information can be attributed to the effect of several behavioral aspects of sub-optimal characteristics that can lead to the domination of a specific travel option even in cases where the rational choice favors another (Banister 1978; Johansson et al. 2006; Cantillo et al. 2007; Domarchi et al. 2008).

Over the decades, demand modelers claim that the random component of utility, within the traditional mode choice modeling framework, not only can explain the limitations of the observer, but also it considers the imperfect information and random variation in tastes on the part of the decision maker. However, a need for a learning-based mode shift model is supported by the following facts. First, the decision process a passenger has to undertake while shifting to an alternative mode of travel is informed and guided by information on the service levels of alternative modes. Such knowledge is usually gained through various means (including travel experience) over time. Second, mode shift decisions are affected by some behavioral factors in which passengers are more (less) inclined to choose (change) the modes they are already accustomed to. Third, the stochastic and time-dependent nature of the transportation system most likely gives rise to adaptive mode switching decisions by passengers, in which they may make their decisions over time by updating their expected utilities for each mode of travel based on previous experience (Wahba and Shalaby 2005). Given their cross-sectional nature, traditional approaches fail to address the time dimension involved in the decision making process. Therefore, a learning-based mode shift model is more desirable than the traditional method for choice modeling, especially in policy analysis where it is important to determine the period of time required to harvest the benefits of the proposed policies.

In order to contribute to this critical issue, the main objective of this paper is first to highlight recent research efforts that account for behavioral factors in mode choice modeling. Second, the paper presents a conceptual learning-based mode shift model that integrates the random utility maximization concepts with an adaptive learning-based framework as an alternative way to mode shift modeling.

The presented model captures passengers’ experience and evaluation of service quality over time while considering the behavioral factors that affect the choice at the individual level. Unlike traditional models, the presented model simultaneously accounts for habits formation, level of information provision and awareness limitations in an attempt to comply with the assumptions of bounded rationality.

2. BEHAVIORAL ASPECTS AFFECTING MODE CHOICE DECISIONS

Mode choice has commonly been considered a rational decision making process that is related to some socioeconomic and demographic aspects of the decision maker and others representing the relative attractiveness of the available options (Eriksson et al. 2008). Of the decision maker characteristics, car ownership and availability are usually considered the major determinants of mode choice (Gwilliam and Banister 1977; Williams 1978; Barff et al. 1982). On the other hand, travel time and cost play a bigger role in determining mode choice than others that characterize the attractiveness of the competing modes (Quarmby 1967; Williams 1978).

In spite of using such attributes as explanatory variables, several research efforts have showed that mode choice should be characterized by some behavioral aspects that resist changing individuals’ previous choices (Ajzen 1991; Gärling et al. 1998; Fujii and Kitamura 2003; Gärling and Axhausen 2003). Interestingly, it was found that certain psychological and sociological factors can help to reinforce the relative attractiveness of a specific mode of travel relative to other options such that mode choice might not always be a result of cost and time savings but rather a result of paying psychological cost or overcoming negative emotions (Chorus et al. 2006; Cantillo et al. 2007; Chorus et al. 2009). Hence, traditional mode choice models can be criticized for focusing on rational decision making without paying attention to some behavioral aspects such as habits, beliefs, values, emotions and attitudes that interact while choosing a mode for travel (Ben-Akiva et al. 2002).

Although rational reasoning may have been the origin of many daily-based decisions, individuals do not go through such deliberate process when they repeat the same decision over a long period of time as it becomes habitual (Ronis et al. 1989; Aarts et al. 1997). As an early indication for the effect of habits formation on mode choice decisions, Seth (1976) stated that people tend to stay with the mode they are already accustomed even though other modes may be more appropriate for them. The previous argument was further supported by Goodwin (1977) which showed that habits may prevail even in cases where the more deliberate choice favors
another mode. In addition, Aarts et al. (1997) argued that mode choice decisions like many other routine behaviors are supposed to be often made in a habitual mindless fashion. In two studies about the moderating effect of habit on the intention-behavior relationship within established commuting contexts, Gardner (2009) showed that habits usually dominate behavioral outcomes. Furthermore, Chorus et al. (2009) found that even travelers that make rational decisions exhibit inertia during a series of risky choices such that choosing the same alternative from an initial set of equally risky alternatives repeatedly is a rewarding strategy under the essence of risk aversion.

In light of the above, routine-based choice can describe the reason behind the domination of car as a mode of travel which is hard to be altered even after a policy change which favors public transit. Therefore, it is suggested that mode choice modeling should be characterized by its underlying behavioral aspects in order to precisely describe the modal split (Banister 1978). Consequently, numerous research attempts have been made to capture the intricacies of the decision making process by including socio-psychological factors in terms of explanatory variables, either as dummy or latent variables, within the traditional mode choice models in addition to the conventional personal and modal service variables (Johansson et al. 2006; Cantillo et al. 2007; Domarchi et al. 2008; Habib et al. 2010).

In general, the previous research efforts provide evidence that mode choice is a complex process which not only involves socio-economic factors, but also socio-psychological variables that have shown to have strong influence on mode choice and improved the models in terms of fitness and statistical significance. However, latent class models can be criticized for incorporating the effect of psychological variables within a generalized function without giving proper attention to the individual travel experience. In addition, given their cross-sectional nature such models fail to address the time dimension involved in the decision making process.

In an effort to overcome the above mentioned gaps, we provide in this research a conceptual framework of a learning-based mode shift model capable to model the mode shifting mechanism while being consistent with the intuition behind bounded rationality. The proposed learning-based mode shift model is built on top of a base latent class model to take advantage of its strong theoretical foundation in modeling choice behavior. The learning process, however, ensures modeling personal behavior at the individual level based on personal experience and evaluation of the transportation system in a more dynamic fashion. Further, the learning process models the mode switching mechanism while simultaneously accounting for habitual inertia against shifting modes, different levels of information provision and awareness limitations. What is unique to the proposed approach is that it models the insights of the decision making process and the period of time required to reap the benefits of the proposed policy changes.

### 3. TOWARDS A LEARNING-BASED MODE SHIFT MODEL

In general, learning-based models have been widely used in a number of various fields of research. A significant contribution of this paper is modeling mode shift decisions as a learning-based process which involves learning by interaction with the transportation system in a dynamic context. The proposed approach can be conceptualized under both the Reinforcement Learning (RL) concepts and the Random Utility Maximization (RUM) framework. Within the proposed approach, passengers’ choices depend on examining the system and updating their perceptions of modal utilities while being influenced by some behavioral aspects that act against learning new knowledge (e.g., formation of habits). In particular, mode shift is modeled as a long term decision process which involves learning over a period of time until reaching a state of habit stabilization.

Within the reinforcement learning concepts, passengers are assumed to be goal-directed agents that apply an optimal policy to choose the best alternative. At each episode, agents perceive the state of the system and choose a mode of travel accordingly while considering their past experiences. Based on earning positive or negative rewards, agents adjust their choices while seeking to maximize the total return received in the long-term in terms of a value function considering travel time, cost, and other factors that affect the choice decision (Wahba and Shalaby 2009). For instance, according to Barto and Sutton (1998), if a state $(s_t)$ is visited at time $(t)$, the reinforcement learner updates its long-term estimate $V(s_{t} + 1)$ based on the immediate reward gained after that visit $R(s_t)$, in addition to what happened before that visit $V(s_t)$, using a simple reinforcement learning updating rule as follows:

$$V(s_{t} + 1) \leftarrow V(s_{t}) + \alpha [R(s_t) - V(s_{t})],$$

where:

- $\alpha$: Step size parameter

In an attempt to comply with the context of bounded rationality, the proposed approach employs the principles of reinforcement learning to account for the long-term accumulation of rewards while considering some behavioral aspects that affect the learning process. Specifically, the formation of habits is modeled in terms of the step size parameter ($\alpha$), the level of awareness of the changes in the transport...
system is considered in terms of the temperature parameter \( (\tau) \), and finally different updating rules will be used to represent different cases of information provision. On the other hand, the proposed approach also employs the principles of random utility theory to measure passengers’ satisfaction in an immediate sense in terms of the short-term reward within the learning process.

### 3.1. Modeling the Formation of Habits in terms of the Step Size Parameter (\( \alpha \))

The step size parameter \( (\alpha) \) is a small positive fraction \( (0 \leq \alpha \leq 1) \) which is commonly used in reinforcement learning methods to influence the learning rate, such that the higher the step size is, the more the agent learns from recent experience. The step size parameter is generally reduced over time within the learning process as the agent tends to rely more on what it has already learnt. From a behavioral perspective, this adaptive learning mechanism can logically describe the mode shift behavior where travelers (more specifically commuters) become more systematic with respect to their chosen mode and insensitive to changes in the transport system once habits are formed towards a specific mode of travel. Once the learning process has been completed, small scale economic policies may be of little effect due to habit formation, as indicated by Lucarotti (1977) and further supported by Banister (1978) who showed that the commuter may not change her/his chosen mode until a certain threshold of the corresponding utility has been reached. Consequently, future choices can be predicted with a high degree of accuracy if habits are identified, which is possible knowing that habits are characterized by their invariability, repetition and persistence (Golledge and Brown 1967).

The previous findings provide evidence that habits act against learning new knowledge, which is opposite to the function of the step size parameter in the reinforcement learning process. Thus, habits can be modeled in terms of the value of the step size parameter such that the strength of formed habits is inversely proportional with the step size towards learning new knowledge.

#### 3.1.1. Estimating the Step Size Parameter (\( \alpha \))

Over the past few decades, the conventional choice rule for modeling choice decisions has been the logit or exponential rule which is based on the maximum-likelihood parameter estimation (MLE) method. Logit models are discrete choice models, which attempt to explain the behavior of individuals making choice between finite number of alternatives. In the logit model, actions with higher propensities are chosen with higher probabilities (Hopkins 2007).

For a decision maker \( (i) \), the probability of choosing alternative \( (m) \) is

\[
P_{im} = \frac{e^{V_{im}}}{\sum_m e^{V_{im}}},
\]

where:

- \( P_{im} \): Probability that decision maker \( (i) \) selects alternative \( (m) \)
- \( V_{im} \): Utility that decision maker \( (i) \) obtains from alternative \( (m) \), where \( i = 1, \ldots, I; m = 1, \ldots, M \)

Research findings showed that traditional mode choice models do not only provide information about the probabilities of mode selection in a stochastic manner, but also the explanatory variables of the models imply some behavioral aspects. As an early indication of strong habits towards the auto mode, Banister (1978) found that the car is almost invariably preferred whenever available and argued that personal mobility is dependent on car ownership, license holding and car availability and less dependent upon the competitive attributes of alternative modes. This finding has been further supported by Domarchi et al. (2008) who showed that habitual frequency of car use is positively correlated with car availability, which means that the inclusion of auto ownership as an explanatory variable in traditional mode choice models can act as an indirect measure of car use habit.

In light of the above, we hypothesize that the previous choice probability of a particular mode can be considered as an indicator for habitual inertia towards that mode. This hypothesis implies that the higher the previous choice probability, the stronger the formed habits towards that choice. In addition, knowing that habits act against learning new knowledge, the step size parameter is postulated to be an inversely proportional function of the previous dominating mode choice probability (e.g., \( \alpha_i = 1 - P_{im} \)),

Where:

- \( \alpha_i \): Step size parameter associated with decision maker \( (i) \)
- \( P_{im} \): Previous mode choice probability that decision maker \( (i) \) selects the dominating alternative \( (d) \)

"Within such treatment, previous dominating mode choice probabilities represent agents’ willingness to switch modes after a change in the system such that if the value of \( P_{im} \) is close to one (i.e., \( \alpha_i \) closer to zero) then there is a strong inertia towards the previous choice and the previous mode prevails (i.e., the agent will not learn much from recent experience)."
3.2. Modeling the Awareness Level in terms of the Temperature Parameter ($\tau$)

In general, balancing exploration and exploitation is an issue in reinforcement learning approaches. Exploiting the actions estimated (through agent learning) to be best is usually insufficient, because many relevant state-action pairs in the reinforcement learning framework may never be visited by the agent. Excessive exploration on the other hand will make it hard to learn the good actions to take at different states. Therefore, maintaining a balance between exploration and exploitation is necessary to ensure that the agent is really learning to take the optimal decisions (Barto and Sutton 1998).

One popular technique of exploration is the $\varepsilon$-greedy method, where a learner behaves greedily most of the time but every once in a while it selects an action at random with small probability $\varepsilon$. The disadvantage of this method is that it chooses among all actions with equal probability, irrespective of the estimated reward value of each action. An alternative is to use the softmax action selection method, where higher estimated rewards are chosen with greater priority than actions with lesser estimated rewards (Abdulhai and Kattan 2003). The most common softmax method relies on the Boltzmann distribution where an action ($a$) is selected using the following probability:

$$P_a = \frac{e^{Q_t(a)/\tau}}{\sum_{b=1}^n e^{Q_t(b)/\tau}},$$

where:

- $Q_t(a)$: Value of action ($a$)
- $\tau$: Temperature parameter

The temperature parameter ($0 \leq \tau \leq \infty$) is a parameter controlling the degree to which actions with higher values are favored in selection. In general, high temperatures cause the actions to be all nearly equiprobable whereas low temperatures cause greater difference in the action selection probability, in the limits as $\tau = 0$, softmax action selection becomes the same as greedy action selection.

3.2.1. Estimating the Temperature Parameter ($\tau$)

From a behavioral viewpoint, strong habit formation can act against exploring new alternatives and consequently against being aware of recent changes in the transport system. For example, travelers might be unaware of the changes in the transit service due to the lack of exploring as a result of strong habits toward driving. In other words, the formation of habits might put passengers in a state of limited awareness of system changes.

In such context, the exploration rate can be maintained to address passengers’ awareness of changes in the transport system which in turn is affected by the strength of the formed habits. In particular, the temperature parameter ($\tau$) is assumed to be inversely proportional with the previous dominating mode choice probability which acts as an indicator for habitual inertia (e.g., $\tau_i = 1 - P_{di}$),”.

Where:

- $\tau_i$: Temperature parameter associated with decision maker ($i$)
- $P_{di}$: Previous mode choice probability that decision maker ($i$) selects the dominating alternative ($d$)

This assumption implies that the higher the previous choice probability of a specific mode, the stronger the formed habits towards that mode, and consequently the lower the temperature parameter (i.e., the agent tend to exploit greedily). This treatment maintains the balance between exploration and exploitation as a function of the previous dominating mode choice probabilities such that if the value of $P_{di}$ is close to one (i.e., $\tau_i$ closer to zero) then there is a strong inertia towards the previous choice and the agent tends to exploit and vice versa.”.

3.3. Modeling the level of Information Provision in terms of the Updating Rules

One of the drawbacks of the traditional mode choice models is being unable to precisely address the effect of information provision on mode choice decisions, specifically the variation in the perceived transport service attributes and the way it could affect the choice behavior among passengers over time (Quentin and Hong 2005; Wahba and Shalaby 2009).

In order to describe the choice behavior, adaptive learning models assume that consumers learn about the relative quality of products adaptively using learning rules. In this context, Hopkins (2007) showed that small differences in the learning rules between belief-based and reinforcement learning behavior can have large effects on market outcomes. In addition, the results showed that even simple adaptive learning models can help explain actual choice behavior at the micro decision making level.

In general, two commonly used assumptions about available information can be identified while updating agent propensities, according to Hopkins (2007). The first corresponds to a state of partial information, in which an agent can only observe the reward resulting from the implemented action. The second corresponds to a state of full (perfect) information, in which an agent can observe the return of all possible actions including the rewards of actions that were not taken.
An important aspect of mode choice decisions under the assumptions of bounded rationality is that an agent is considered in a situation of partial information while choosing a travel option, as it (the agent) can only perceive the reward from the alternative that is actually chosen, while information about unselected modes is unavailable. Therefore, updating rules under partial information are more desirable while dealing with mode choice.

Two updating rules can be described under the states of partial information, a belief-based learning and a stimulus-response type learning (reinforcement learning) rule. The belief-based learning rule can be described as follows. At time \( t \), a passenger \( (i) \) perceives a utility \( \left( R_{im}^t \right) \) after choosing mode \( (m) \). He/she updates his/her utilities as follows:

\[
V_{im}^t(t+1) = V_{im}^t(t) + \alpha_i \left[ R_{im}^t(t) - V_{im}^t(t) \right], \quad V_{im}^t(t+1) = V_{im}^t(t), \quad \text{for all } n \neq m,
\]

Where:

- \( V_{im}^t(t) \): Utility that decision maker \( (i) \) obtained from mode \( (m) \) till time step \( (t) \)
- \( V_{in}^t(t) \): Utility that decision maker \( (i) \) obtained from any other mode \( (n \neq m) \) till time step \( (t) \)
- \( R_{im}^t(t) \): Immediate utility that decision maker \( (i) \) obtains from mode \( (m) \) at time step \( (t) \)
- \( V_{im}^t(t+1) \): Updated utility that decision maker \( (i) \) obtains from mode \( (m) \) at time step \( (t+1) \)
- \( V_{in}^t(t+1) \): Updated utility that decision maker \( (i) \) obtains from any mode \( (n \neq m) \) at time step \( (t+1) \)
- \( \alpha_i \): Step size parameter \( (0 \leq \alpha_i \leq 1) \), associated with decision maker \( (i) \)

If values of \( \alpha_i \) are closer to zero, then agent’s experience from long ago still have a significant effect on current beliefs, while values of \( \alpha_i \) closer to one means that only the very recent experience is remembered. Within the previous model, the propensity towards the selected mode potentially incorporates the reward of each action, while the utility of each unselected alternative remains unaltered, as there is no new information about the alternative with which to update its value. In this model, the agent is assumed to have adaptively formed beliefs about the quality of each of the competing modes.

The reinforcement type learning on the other hand could be conceptualized as follows. At time \( (t) \), a passenger \( (i) \) perceives a utility \( \left( R_{im}^t \right) \) after choosing mode \( (m) \). Upon perceiving the utility, the passenger updates his/her utilities as follows:

\[
V_{im}^t(t+1) = V_{im}^t(t) + \alpha_i \left[ R_{im}^t(t) - V_{im}^t(t) \right], \quad V_{im}^t(t+1) = \alpha_i V_{im}^t(t), \quad \text{for all } n \neq m,
\]

Within the previous model, the propensity towards the selected mode potentially incorporates an accumulation of positive feelings (e.g., familiarity or recognition) such that the utility of the unselected modes decreases naturally as familiarity with those alternatives declines relative to the selected mode.

Importantly, both updating rules respond only to the experienced utilities of the selected mode \( \left( R_{im}^t(t) \right) \), while information on utilities of the unselected modes is not utilized because they were not observed/experienced at that time. That can be interpreted as being in a state of partial information, or in other words, the agent is boundedly rational. However, such rules might be inadequate for evaluating the effects of the emerging information technologies in terms of the impacts of Intelligent Transportation Systems (ITS) deployments on service reliability and real-time information provision capabilities, which in turn affect passengers’ behavior.

Although a state of perfect information might not practically exist, it could be argued that the new Advanced Traveler Information Systems (ATIS) can supply travelers with information on the alternative modes as well as the selected option. In other words, passengers may receive real-time information on the travel times of different modes through the network. Hence, an updating rule under the state of perfect information can be used to update simultaneously all utilities as follows:

\[
V_{im}^t(t+1) = V_{im}^t(t) + \alpha_i \left[ R_{im}^t(t) - V_{im}^t(t) \right], \quad \text{for } m = 1 \text{ to } M
\]

To our knowledge, the proposed approach is the first towards a learning-based mode shift model. The underlying hypothesis is that passengers are expected to adjust their choices according to their experience with the performance of the transport system and their previous valuation of the available alternatives, while being subject to awareness limitations and habit formation that might have been formed towards a specific mode of travel.

In this research, individual passengers are represented as agents that are endowed with different propensities associated with each of the possible choices in the choice set. As utility maximizers, the agents’ policy will be the choice of the travel option that maximizes their satisfaction on long-term basis. However, balancing exploration and exploitation will be used to ensure that agents can make different choices over time and thus learn which of these alternatives is more effective in achieving the desired goals. Further, agents will examine their choices by interacting with the transport system through a microsimulation model which represents the agents’ environment. The outlined learning mechanism will be iterated until the agents learn their choices and achieve a state of choice stabilization.
Towards a learning-based mode shift model: a conceptual framework

Under such framework, mode shift is modeled as a dynamic process of repetitively making decisions and updating perceptions according to a long-term adaptive learning process. This dynamic feedback, using learning and adaptation, is unique to the proposed framework and presents a more behaviorally sound mode shift approach that is suitable to perform mode shift analysis after a policy change.

4. NUMERICAL SIMULATION

This section provides numerical simulation results and compares different choice strategies under both the traditional and learning-based mode shift modeling frameworks. Various updating rules are examined under different states of information provision, specifically the states of partial and perfect information and how they could affect the mode switching behavior. Two updating rules are modeled under the state of partial information; namely a belief-based rule which considers the accumulation of former beliefs about the quality of each alternative and a reinforcement learning-based rule which considers natural decay of beliefs where familiarity with those alternatives decline. On the other hand, one updating rule is modeled under the state of perfect information which assumes the availability of information regarding the unselected modes to the decision maker.

The modeling scenario considers a hypothetical mode choice situation, in which 100 passengers face a daily mode choice between auto, transit and walk options. Under the traditional mode choice framework, a simple conventional logit model is used to estimate the choice probabilities based only on travel time and cost without dealing with the different transit travel time components (access/egress time, wait time, in vehicle time and transfer time), in order to simplify the calculations. The learning-based mode shift model on the other hand involves examining the system and updating long-term experience throughout twenty learning episodes during which the decision makers choose a travel alternative, earn an immediate reward and accordingly update long-term estimates. The specifications of the used logit model are as follows:

\[ V_{(Auto)} = 1.0 - 0.1 \times Auto\_TT\_\text{(min)} - 0.05 \times Auto\_Cost\_($) \]
\[ V_{(Transit)} = -0.1 \times Transit\_TT\_\text{(min)} - 0.05 \times Transit\_Cost\_($) \]
\[ V_{(Walk)} = -0.5 - 0.1 \times Walk\_TT\_\text{(min)} \]

The simulation scenario assumes that on episode one the attributes corresponding to each alternative are as follows: Auto\_TT\_\text{(min)} = 15, Transit\_TT\_\text{(min)} = 15, Walk\_TT\_\text{(min)} = 30, Auto\_Cost\_($) = 1.6 and Transit\_Cost\_($) = 1.5. Between episode one and episode two, the transit travel time is reduced due to a significant change that favors the transit option such that Transit\_TT\_\text{(min)} = 4.

In order to illustrate the evolution of habits with respect to the change in the awareness level, balancing exploration and exploitation will ensure continual exploration after the fifth episode at which the decision maker will become aware of the changes in transit mode by means of direct experience. In other words, the reinforcement learner will act greedily (i.e., exploit its choice) by choosing the most favorable mode based on what it has learnt at the beginning of the simulation. At the fifth episode, the agent will become aware of the changes in the system and the assumption of exploring starts (i.e., explore transit), and continues until the termination of the simulation. Importantly, note that the choice situations, the values of the attributes and the model parameters are chosen arbitrarily; hence, the outcomes presented in this section should be considered merely an illustration of the model, not a case study.

4.1. Simulation Results

4.1.1. Traditional Mode Choice Model

Based on the conventional mode choice framework, the auto mode was the most attractive alternative with the highest choice probability (70.2%) on episode one. However, after reaching the steady state conditions following the reduction in transit travel time on episode two, the transit option took the lead of the choice where 51.3% of the passengers use transit and 46% use auto.

Obviously, conventional mode choice models have always been cross-sectional models under steady state conditions before and after the change. Hence, they might be useful to describe mode shares but they do not capture the time-dependent processes underlying a possible mode shift which involves breaking the previously formed habits and building new ones. In other words, conventional models cannot provide information about the period of time required to reap the benefits of a proposed policy scenario. This drawback of conventional mode choice models in policy analysis further supports the need for a learning-based mode shift model, which is presented in the following sections.

4.1.2. Learning-based Mode Shift Model

In order to utilize the effectiveness of traditional mode choice models in describing current mode split, the outcomes of the above mode choice model in terms of utilities and choice probabilities for each of the competing modes were used as the initial values of the learning process before the policy change. However, different updating rules were used to model the evolution of the agent’s experience throughout
the simulation with respect to different cases of information provision.

4.1.2.1. Partial Information (Belief-based Updating Rule)

\[
V_{im}(t + 1) = V_{im}(t) + \alpha_i [R_{im}(t) - V_{im}(t)], \quad V_{in}(t + 1) = V_{in}(t),
\]

for all \( n \neq m \)

Although the transit option had the higher choice probability after the reduction in its travel time, the superiority of transit is realized on episode ten after passing a period of unawareness which is followed by a period of reformation of habits. The simulation results based on the belief-based model are illustrated in Figure 1.

Initially, the same values of choice probabilities across modes were used such that the car was the superior mode. After the change on the second episode, an increase in the observed utility of transit has been achieved favoring transit over other modes in the choice set. However, knowing that the agent is still exploiting its favorite choice (i.e., auto mode), this increment in transit utility was not yet observed by the decision maker. In other words, the agent was unaware of the changes in the system till the fifth episode when it started to explore and directly examine the transit mode. Based on that, the values of choice probabilities remained unaltered during the first five episodes as shown in Figure 1, which could be interpreted as being in a state of unawareness.

During the unawareness zone, the previously formed habits remained stable. However, when the agent became aware of the changes at the fifth episode, habits started to reform according to the new experience with the system until reaching another stabilization zone on episode sixteen.

It can be also noticed that the superiority of transit has not been realized immediately after being aware of the changes, but rather after a transition period of habits reformation from the fifth episode till the tenth episode, in which the long-term value of transit utility exceeded the long-term value of the utility of auto and hence transit mode was more likely to be selected. In this scenario, the length of the modal shift period only depends on the impact of the policy change on the utility functions, regardless of how long the agent has been exploiting the auto mode before exploring the transit mode.

Obviously, the belief-based updating rule is in line with the assumptions of bounded rationality such that the propensity towards the selected mode potentially incorporates the reward of each action, while the utility of each unselected alternative remains unaltered, as there is no new information about the alternative with which to update its value.

Importantly, habits started to reform only after examining and being aware of the change in the service. In this
Towards a learning-based mode shift model: a conceptual framework

model, the agent is assumed to have adaptively formed beliefs about the quality of each of the competing modes.

4.1.2.2. Partial Information (Reinforcement Learning-based Updating Rule)

\[
V_{in}(t + 1) = V_{in}(t) + \alpha_i[R_{im}(t) - V_{in}(t)], \quad V_{in}(t + 1) = \alpha_i V_{in}(t),
\]

for all \( n \neq m \)

Although the transit option had the higher choice probability after the reduction in its travel time, the superiority of transit is realized on episode eleven after passing a period of unawareness which is followed by a period of reformation of habits. The simulation results based on the reinforcement learning-based model are illustrated in Figure 2.

Initially, the same values of choice probabilities across modes were used such that the car was the favorable mode. After the change on episode two, an increase in the observed utility of transit has been achieved promoting transit over other modes in the choice set. However, and knowing that the agent is still exploiting the auto mode, this increment in transit utility was not yet observed by the decision maker till the fifth episode as being in the unawareness state.

Unlike the belief-based rule, the previously formed habits in addition to the modal choice probabilities were always varying with respect to time such that the superiority of auto mode was increasing with familiarity while being decaying for the other modes with unfamiliarity. However, when the agent became aware of the changes and started to get familiar with the transit mode, the attractiveness of the car started to decline while that of transit started to develop until it became the superior mode which is more likely to be selected at the eleventh episode.

Similar to the belief-based updating rule, the superiority of transit has not been realized immediately after being aware of the changes, but rather after a transition period of habits reformation from the fifth episode till the eleventh episode. However, under the reinforcement learning-based updating rule, the length of the modal shift period is longer as it depends on both the impact of the change on the utility functions and the period of time the agent has exploited the auto mode (i.e., frequency of past use) before exploring the transit mode.

Obviously, the reinforcement learning-based updating rule is based on familiarity and accumulation of positive feelings such that the utility of the unselected modes decreases naturally as familiarity with those alternatives declines. Hence, habits were always varying depending on the frequency of past choice regardless of being aware of the change or not.

4.1.2.3. Perfect Information

\[
V_{in}(t + 1) = V_{in}(t) + \alpha_i[R_{im}(t) - V_{in}(t)], \text{ for } m = 1 \text{ to } M
\]
Although the transit option had the higher choice probability after the reduction in its travel time, and knowing that the decision maker is fully aware of the system characteristics, the superiority of transit is realized on episode six after passing a period of reformation of habits. The simulation results based on the assumption of perfect information are illustrated in Figure 3.

Initially, the same choice preferences across modes were used such that the car was the superior mode. After the change on the second episode, an increase in the observed utility of transit has been achieved favoring transit over other modes in the choice set.

Interestingly, and even though the agent was still exploiting its favorite choice (car option), the increment in transit utility was observed by the decision maker under the assumption of perfect information. In other words, the agent was fully aware of the changes in the system without exploring and directly examining the transit mode. Based on that, the previously formed habits and the modal choice probabilities started to evolve since episode two, until reaching another stabilization zone on episode twelve.

As illustrated in Figure 3, the superiority of transit has not been realized immediately after perceiving the reduction in transit travel time, but rather after a transition period of habits reformation from the second episode till the sixth episode. In this scenario, the length of the modal shift period only depends on the impact of the policy change on the utility functions, regardless of how long the agent has been exploiting the car option before exploring transit.

Practically, the assumption of perfect information requires receiving continuous information updates on each mode (selected as well as unselected). Adding to that the issue of information reliability and how much the passenger trusts the supplied information, it can be said that being in a state of perfect information might not practically exist. However the previous updating rule presents the upper bound of information provision which would speed up the learning process and the expected modal shift.

5. CONCLUSION

This paper reviewed some recent approaches that consider the effect of behavioral aspects on mode choice decisions. Research results showed that traditional mode choice models are based on static knowledge and lack to recognize that the interaction with the environment generally leads to adaptation of behavior through learning.

As an alternative way to mode shift modeling, we introduce a novel conceptual framework that models mode shift as an adaptive leaning process which involves learning by interaction with the transportation system in a dynamic context. Within the presented framework, passengers’ choices depend on updating their perceptions of choice utility while taking
Towards a learning-based mode shift model: a conceptual framework

into consideration some behavioral factors that affect the choice at the individual level. The underlying hypothesis is that passengers are expected to adjust their choices according to their experience with the performance of the transport system and their previous valuation of the available alternatives, while being subject to awareness limitations and habit formation that might have been formed towards a specific mode of travel. Further, the presented approach models the insights into understanding individuals’ sensitivity to various policy scenarios and how long it would take to reap the benefits of the proposed policies, which is typically not answered by traditional approaches owing to their cross-sectional nature.

The learning-based mode shift model as presented and numerically illustrated in this paper tackles some of the behavioral limitations of the traditional models under the contexts of bounded rationality and limited awareness. Importantly, the presented approach combines both perspectives of habitual inertia and awareness limitations rather than substituting one for the other as assumed by other models. Further, while traditional mode choice models implicitly assume rational decision making, perfect information availability and full awareness of the changes in the transport system, the presented model is considered more behaviorally realistic and incorporates a number of practical implications.

The presented approach implies that passengers’ awareness is limited and depends on their direct experience with the transport system and the available level of information provision. In addition, unlike other models considering only the formation of habits, this simulation presents the time-dependent processes of change in behavior that follows a change in the transport system, during which passengers update their formerly formed habits and change their choices accordingly.

In general, the illustrated learning behavioral patterns are in agreement with the prior expectations which can be considered a first step towards the model’s validity. Nevertheless, another strong point towards the model credibility is utilizing the effectiveness of the traditional mode choice models in describing the choice situation at the initiation of the learning process. In light of the above, this paper presents a promising approach that states the art for a more behaviorally sound mode shift model. What is unique to the proposed model is that it can explain the transitional process underlying the modal shift mechanism which is important from a service design point of view.

It is clear that further work is required to make these ideas practical and capable of implementation. Conducting controlled lab experiments of travel behavior is suggested for further work to specify and test the mode shift process and estimate the parameters of this process under various assumptions and levels of information provision. It is also suggested to collect travel data after policy implementation at regular time intervals (e.g., every six months) until the mode shares stabilize. The collected data can then be used to validate the proposed formulations and assumptions of habit formation, level of information provision and awareness limitations. In addition, future efforts are also suggested to test the forecasting performance of the model (i.e., temporal transferability) as well as testing its transferability across space.

REFERENCES


