

## The experienced mode choice set and its determinants

### Commuting trips in the Netherlands

Ton, Danique; Bekhor, Shlomo; Cats, Oded; Duives, Dorine C.; Hoogendoorn-Lanser, Sascha; Hoogendoorn, Serge P.

**DOI**

[10.1016/j.tra.2019.12.027](https://doi.org/10.1016/j.tra.2019.12.027)

**Publication date**

2020

**Document Version**

Final published version

**Published in**

Transportation Research Part A: Policy and Practice

**Citation (APA)**

Ton, D., Bekhor, S., Cats, O., Duives, D. C., Hoogendoorn-Lanser, S., & Hoogendoorn, S. P. (2020). The experienced mode choice set and its determinants: Commuting trips in the Netherlands. *Transportation Research Part A: Policy and Practice*, 132, 744-758. <https://doi.org/10.1016/j.tra.2019.12.027>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

# Transportation Research Part A

journal homepage: [www.elsevier.com/locate/tra](http://www.elsevier.com/locate/tra)

## The experienced mode choice set and its determinants: Commuting trips in the Netherlands



Danique Ton<sup>a,\*</sup>, Shlomo Bekhor<sup>b</sup>, Oded Cats<sup>a</sup>, Dorine C. Duives<sup>a</sup>,  
Sascha Hoogendoorn-Lanser<sup>c</sup>, Serge P. Hoogendoorn<sup>a</sup>

<sup>a</sup> Department of Transport & Planning, Delft University of Technology, the Netherlands

<sup>b</sup> Transportation Research Institute, Technion – Israel Institute of Technology, Israel

<sup>c</sup> KiM Netherlands Institute for Transport Policy Analysis, the Netherlands

### ARTICLE INFO

#### Keywords:

Experienced choice set  
Mode use  
Mode choice set  
Commute trips  
Discrete choice analysis

### ABSTRACT

Active modes take up an increasingly important place on the global policy-making agenda. In the Netherlands, a country that is well-known for its high shares of walking and cycling, the government aims at achieving a modal shift among 200,000 commuting car drivers towards using the bicycle. To this end, policy measures need to be introduced. When the aim is to achieve a modal switch over an enduring period of time, it is more relevant to know the likelihood of including or excluding a mode in the mode choice set, compared to choosing a mode for a single trip. Therefore, we investigate the formation of the experienced choice set (set of modes used over a long period of time), where the aim is to identify determinants that influence the inclusion or exclusion of a mode in this set. We estimate discrete choice models, based on survey data from the Netherlands Mobility Panel (MPN) and a complementary survey, where individuals were asked to report the frequency of using certain modes of transport for commuting trips over the course of half a year. This study shows that the experienced choice set for commuting is unimodal for the majority of the individuals, and remains constant over time for most individuals. Reimbursement by the employer for using a certain mode is the most important determinant influencing the experienced choice set, followed by ownership characteristics and urban density. We show that the mode choice set formation depends on more determinants than previously assumed.

### 1. Introduction

Due to increasing urbanisation rates accompanied by growing transportation demands, governments worldwide have been increasingly interested in active modes of transport, i.e. walking and cycling. These modes can reduce congestion and emissions when replacing the car (standalone or in combination with public transport). Furthermore, these modes provide health benefits for the individual. The Netherlands has a very high share of active mode use, with 27% of trips made by bicycle and 17% on foot (CBS, 2018). Notwithstanding, the Dutch government aims at achieving a mode switch from the car to the bicycle for 200,000 commuters (Rijksoverheid, 2018). This aim is supported by the fact that about half of the commuting trips currently travelled by car are shorter than 7.5 km. Consequently, these could be travelled by bicycle. To ensure this aim can be reached, policy measures need to be

\* Corresponding author.

E-mail addresses: [d.ton@tudelft.nl](mailto:d.ton@tudelft.nl) (D. Ton), [sbekhor@technion.ac.il](mailto:sbekhor@technion.ac.il) (S. Bekhor), [o.cats@tudelft.nl](mailto:o.cats@tudelft.nl) (O. Cats), [d.c.duives@tudelft.nl](mailto:d.c.duives@tudelft.nl) (D.C. Duives), [sascha.hoogendoorn@mininm.nl](mailto:sascha.hoogendoorn@mininm.nl) (S. Hoogendoorn-Lanser), [s.p.hoogendoorn@tudelft.nl](mailto:s.p.hoogendoorn@tudelft.nl) (S.P. Hoogendoorn).

<https://doi.org/10.1016/j.tra.2019.12.027>

Received 5 December 2018; Received in revised form 19 December 2019; Accepted 26 December 2019  
0965-8564/ © 2019 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

developed that lead to mode switches among commuters.

Investigating mode use requires observing individuals over a period of time. [Kuhnimhof et al. \(2006\)](#) investigated mode use in Germany using a 7-day travel diary. They found that during the first three to four days the number of modes used increased significantly and thereafter stabilised. [Heinen and Chatterjee \(2015\)](#) investigated the mode use variability of individuals in the UK using a 7-day travel diary. They found that 44% of the respondents were unimodal, the other 56% used a variety of modes over the course of seven days. [Ralph \(2017\)](#) used three years of information from a 1-day travel diary in the USA and clustered young adults based on their mode use. She found that 86% were unimodal car user, of which 4% used the car for long distances. Only 3% of the individuals were multimodal, meaning that they used both car and public transport. A further 11% were car-less, meaning they relied on public transport when they travelled. [Ton et al. \(2019b\)](#) investigated the mode use of individuals in the Netherlands using a 3-day travel diary, where they clustered the individuals based on their frequency of using each mode per day into five clusters. Two clusters were unimodal, namely car only (27%) and bicycle only (9.5%). One cluster (12.3%) used all modes (car, public transport, bicycle, and walking) on a daily basis. The other two multimodal clusters were car and bicycle (27.5%) and car, bicycle, and walking (23.7%). In sum, the number of modes used increases with the number of days for which a travel diary is available, but generally only a selection of modes is used.

When zooming in to the commuting trip purpose, even less variability in the mode use over time has been observed. [Kuhnimhof \(2009\)](#) found that for commuting most individuals repeatedly uses the same mode (72%) over the course of 7-days This is supported by [Hensher and Ho \(2016\)](#), who found that use of a mode increases the likelihood of using that mode again. However, the variation in modes between individuals is higher for commuting compared to other trip purposes ([Kuhnimhof et al., 2006](#)). [Lavery et al. \(2013\)](#) investigated commuting trips to McMaster University in Canada. They asked individuals for their primary commuting mode and alternatively asked which modes the respondents considered available/feasible for their commute. When considering both used and available modes, the unimodality varies between 9% for active modes to 55% for public transport. A total of 51% of the individuals states that two modes are used and/or available, 37% mentions three modes, and 4% mentions four modes. Many of these results show routines or habitual behaviour regarding commuting trips.

When the aim is to identify individuals that might switch modes for their commute trips it is essential to understand the composition of their experienced choice set (modes used over a long term) and which determinants drive the inclusion or exclusion of modes from this set. In common mode choice research, discrete choice models are estimated to identify which determinants are associated with mode choice and predict the likelihood of using a mode for a certain trip. Here, the choice set (i.e. between which modes a commuter chooses) is often defined based on deterministic rules, related to travel time, distance, ownership or availability of modes ([De Jong et al., 2007](#); [Gehrke and Clifton, 2014](#); [Hamre and Buehler, 2014](#); [Kamargianni and Polydoropoulou, 2013](#)). However, for example [Kuhnimhof \(2009\)](#); [Lavery et al. \(2013\)](#) show that these assumptions in both science and practice, might not cover all observed behaviour. Consequently, we propose to use discrete choice models to estimate the determinants of the experienced choice set that predict the likelihood of including or excluding a mode for commuting purposes.

In this study we investigate the experienced choice set of individuals in the Netherlands, by observing their mode use for commuting trips over the course of half a year. The data used for identification of the experienced choice set is the Netherlands Mobility Panel (MPN) survey data featuring Dutch-speaking individuals from the Netherlands. This dataset contains personal and household data, and is enriched with a survey that investigates, among others, the modes used by the respondents for commuting trips over a time period of half a year. We identify determinants that are relevant for inclusion of a mode in the experienced choice set by applying discrete choice models. These determinants can be used to identify policy measures that aim to realise a mode switch from the car to the bicycle. Furthermore, we show that more determinants are relevant in the estimation and prediction of the mode choice set than often assumed in mode choice models.

The remainder of this paper is organised as follows. [Section 2](#) details the literature on determinants used for identifying the mode choice set. In [Section 3](#) the methodology is discussed, together with the modelling approaches applied. [Section 4](#) describes the data and shows an overview of the dataset. Then, in [Section 5](#) the experienced mode choice set is elaborated upon. [Section 6](#) covers the model estimation and validation results. [Section 7](#) discusses the experienced choice set in relation to past research and potential applications. Finally, the paper is concluded in [Section 8](#) and recommendations are provided.

## 2. Determinants of mode choice sets

In literature, various determinants are used for the specification of the mode choice set. Sometimes studies rely on self-reported availability of modes as perceived by respondents ([Lavery et al., 2013](#); [Whalen et al., 2013](#)), thus not relying on determinants for the formation of the choice set. [Table 1](#) presents an overview of the determinants that are used to specify the mode set and identifies the operationalisation of these determinants as mentioned in the literature.

[Table 1](#) shows that determinants can roughly be divided into four categories: availability of modes, trip characteristics, network characteristics, and individual characteristics. The first two categories are most common in the literature. Contrary to the trip characteristics, the determinants related to availability can be regarded both at the trip- and individual level. Travel time and distance, both trip characteristics, are operationalised in various ways, for example per mode, as aspect of the complete trip or in general. Some studies have incorporated individual characteristics to determine mode availability ([Calastri et al., 2017](#); [Vij et al., 2017, 2013](#)). These studies have applied latent class models, where the individual characteristics are used to determine mode availability based on class membership. [Calastri et al. \(2017\)](#) showed that including individual characteristics significantly improves the model fit. Consequently, we expect that also individual and household characteristics are relevant determinants of the experienced mode choice set.

**Table 1**  
Operationalisation of determinants used to specify the mode choice set for individuals.

Determinant	Operationalisation	Studies
Availability of modes	Car available	(Ben-Akiva and Boccara, 1995; Gehrke and Clifton, 2014; Hensher and Ho, 2016; Kamargianni, 2015; Kamargianni et al., 2015; Rodríguez and Joo, 2004; Ton et al., 2019a; Vij et al., 2017, 2013)
	Driver's license	(Gehrke and Clifton, 2014; Habib et al., 2011; Kamargianni et al., 2015; Ton et al., 2019a)
	Bicycle available	(Gehrke and Clifton, 2014; Kamargianni et al., 2015; Ton et al., 2019a; Vij et al., 2017)
	PT available	(Hensher and Ho, 2016)
	PT subscription (e.g. season pass)	(Vij et al., 2013)
Maximum distance	General maximum distance	(Bhat, 1995; Kamargianni and Polydoropoulou, 2013; Wardman et al., 2007)
	PT stop within walking distance	(Ben-Akiva and Boccara, 1995; Gehrke and Clifton, 2014)
	PT distance	(Rodríguez and Joo, 2004)
	Walking distance	(Calastri et al., 2017; Habib et al., 2011; Kamargianni, 2015; Kamargianni et al., 2015; Rodríguez and Joo, 2004)
	Bicycle distance	(Calastri et al., 2017; Habib et al., 2011; Kamargianni, 2015)
Maximum travel time	Walking time	(Gehrke and Clifton, 2014; Swait and Ben-Akiva, 1987a; Ton et al., 2019a)
	Bicycle time	(Gehrke and Clifton, 2014; Ton et al., 2019a)
	Car time	(Swait and Ben-Akiva, 1987a)
	PT time	(Swait and Ben-Akiva, 1987a)
	Threshold for inclusion of mode	(Cantillo and Ortúzar, 2005)
Cost	Threshold for inclusion of mode	(Cantillo and Ortúzar, 2005)
Network connection	Route available using a mode	(Swait and Ben-Akiva, 1987a; Ton et al., 2019a)
Individual characteristics as input for mode availability	PT stops in neighbourhood	(Habib et al., 2011; Kamargianni et al., 2015)
	Gender	(Calastri et al., 2017; Vij et al., 2017, 2013)
	Age	(Calastri et al., 2017; Vij et al., 2017)
	Education level	(Calastri et al., 2017)
	Income	(Vij et al., 2017, 2013)
	Employment	(Vij et al., 2017, 2013)
	Household size	(Vij et al., 2017, 2013)
	Children in household	(Vij et al., 2017)
	Marital status	(Vij et al., 2017, 2013)
	Parenthood status	(Vij et al., 2017, 2013)
Home ownership	(Vij et al., 2017, 2013)	

### 3. Methodology

The experienced choice set is defined as the set of modes used over an enduring period of time. In [Section 3.1](#) we present the methodology for retrieving the experienced choice set and identifying the relevant determinants. The model structures used for estimation and validation of the experienced choice set are discussed in [Section 3.2](#).

#### 3.1. The experienced mode choice set and its determinants

The experienced choice set can be retrieved in different ways. One can, for example, observe an individual over a long period of time using a GPS device or a travel diary, which is time consuming and largely impacts the privacy of the individual. This has been done before to study mobility patterns of individuals, where the duration of these data collection efforts range between one day ([Ralph, 2017](#)) and six weeks ([Vij et al., 2017, 2013](#)). Another method, which is less demanding on the individual, is to use a survey to ask questions related to the mode use of an individual over a long period of time. This method has previously been applied to study mobility patterns of individuals ([Lavery et al., 2013; Molin et al., 2016](#)). We apply the latter method and use a survey to collect data (see [Section 4](#)). The question posed to the respondents is: *which modes have you used at least once in the last half year for commute trips?* Where they could choose multiple modes from a list of the most prominent modes in the Netherlands, namely car, train, bus/tram/metro (BTM), bicycle, and walking. Access and egress modes are not included here. This question provides insights in the modes used by the respondents over a long period of time. By focusing on commuting, i.e. one trip purpose, it is easier for individuals to retrieve their mode use. This question collects aggregated data, consequently the experienced choice set reflects the general experienced choice set and does not directly represent trip-level variations.

The experienced choice set reflects actual observed behaviour. Therefore, it provides a rich source of information, both in terms of choices made by individuals and the composition and size of their experienced mode choice set. We propose to apply discrete choice models to identify determinants that influence the experienced choice set of an individual. The alternatives of the experienced choice set are constructed by combining all historically observed modes into a single alternative. The respondent then chooses between sets

of modes, e.g. car-train-walk and bicycle-car. This results with 31 potential experienced choice sets. Given that, for example, individual and household determinants are previously applied and found relevant for choice set specification, we will test the relevance of a number of determinants belonging to ownership, socio-demographics, household characteristics, urban density, and work conditions.

### 3.2. Model structures used for estimation and validation

The determinants influencing the experienced choice set are revealed by estimating a number of discrete choice models. We start simple by estimating Multinomial Logit (MNL) models. Due to the way in which alternatives are constructed, we expect shared unobserved variables (captured in the error term). This cannot be captured in the MNL model, thus requiring the use of various more complex structures such as Nested Logit (NL), Cross-Nested Logit (CNL), and Mixed Logit (ML). The utility function for experienced choice set  $C$  and individual  $n$  is specified according to Eq. (1) (Ben-Akiva and Bierlaire, 1999):

$$U_{Cn} = V_{Cn} + \varepsilon_{Cn}, C \in G_n \tag{1}$$

where  $V_{Cn}$  is the deterministic utility for individual  $n$  and experienced choice set  $C$ , which is part of the feasible choice set of that individual  $G_n$ , and where  $\varepsilon_{Cn}$  represents the random error term capturing the uncertainty. The deterministic part of the utility is composed in the following way for the experienced choice set (the index of individual  $n$  is omitted for simplicity):

$$V_C = ASC_C + \sum_{m \in C} \beta_m x_m \tag{2}$$

where the alternative specific constant (ASC) is defined per experienced set, and where the parameters are estimated for each mode  $m$  that is a member of the experienced set  $C$ . As an example, if the alternative is bicycle-walk, this means that for each variable  $x$  two parameters are estimated, one for the bicycle and one for walking. Before, a model with alternative specific parameters was estimated. This model was optimised by starting with the most comprehensive model formulation, including all variables. Accordingly, the insignificant parameters (from the base) were iteratively, one-by-one, fixed to zero. The result of this iterative model estimation procedure yielded a final model with limited behavioural interpretation, as for some of the alternatives only few explanatory variables were retained. Alternatively, the mode-specific approach was applied. This meant that we started again with the most comprehensive model formulation. However, now variables are excluded per mode-variable combination. This method captures correlations between alternatives in the deterministic part of the utility function and results with a highly improved behaviourally interpretable model. The downside of this approach is that effects per alternative are not revealed.

As mentioned before, the MNL model is only able to capture similarities between alternatives via the observed variables. The NL allows correlations between the unobserved variables of some alternatives, by grouping (nesting) them. In case of the experienced mode choice set, several of these nesting structures can be identified, related to for example size and composition of choice sets. An example of a nesting structure based on composition is shown in Fig. 1A, where alternatives that contain active modes, motorised modes, or a mixture of these are distinguished (note that this representation does not imply hierarchy). All alternatives can be assigned to one nest. Alternatively, a nesting structure based on size represents the number of modes that are combined into the experienced choice set (Fig. 1B). A variety of different nesting structures are tested and judged on model fit and behavioural interpretation.

Due to the way the alternatives are bundled, i.e. combining all used modes into one experienced choice set, it seems plausible that alternatives that have modes in common share correlation in the unobserved variables. For example, the car-train and car-bicycle alternatives might exercise correlations due to the common car-mode. It is not possible to capture this structure in a NL model since nests are mutually exclusive. The CNL model relaxes this assumption by including alternatives as members of multiple nests (Vovsha, 1997). An example of a structure based on mode-nests is shown in Fig. 2. Often, the membership of a nest is predefined (Bierlaire, 2006), but can also be estimated together with the nesting parameters. This optimises the CNL model further as the degree of membership can vary between alternatives and nests. Again, a variety of different nesting structures was tested and judged on model

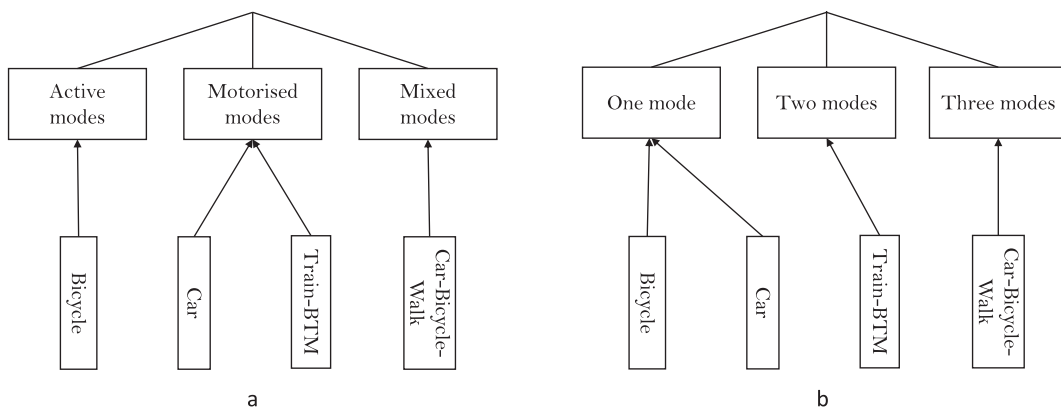


Fig. 1. Example of NL structures based on composition (a) and size (b) of the experienced choice set.

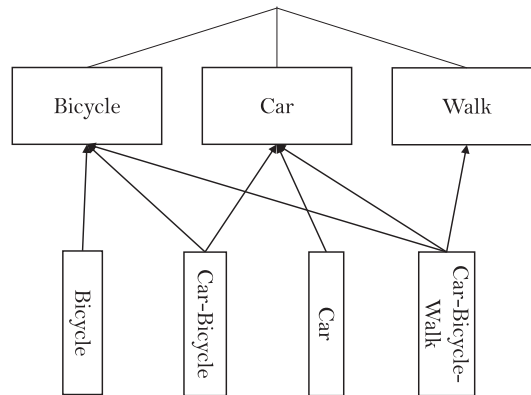


Fig. 2. Example of a CNL structure based on mode-nests.

fit and behavioural interpretation. ML with error component structure has a flexible error structure, and is theoretically able to reproduce the same structure as both the CNL and NL models (McFadden and Train, 2000). It has an extra advantage that it is able to incorporate heterogeneity and heteroscedasticity that can be present in the population.

In the model estimation process, first all variables and parameters are introduced for all alternatives. Afterwards, non-significant parameters were excluded iteratively, so that the model fit (adjusted rho-square compared to the equally likely model, log-likelihood, AIC, and BIC) is optimised. This optimisation is done for the MNL model, the other models are all based on this specification. All models are estimated using the Python Biogeme package (Bierlaire, 2016).

In order to test the predictive power of the best model for the experienced mode choice set, a k-fold cross-validation is performed with five groups. This means that the dataset is randomly distributed into five groups of 20% each. Accordingly, the model is estimated using 80% of the sample and the remaining 20% is used to predict the experienced choice set, given the estimated model. The stability of the parameters is tested, based on model estimations from the different samples. Furthermore, the predictive power is tested by calculating how often the model assigns the highest probability to the actual experienced choice set (hit rate) and the extent to which errors are made. Regarding the latter, if for example the experienced set is bicycle-walk and the predicted set is bicycle only, this is considered a less significant error than predicting train only as the experienced set.

#### 4. Data description

For this study, the data obtained via the Netherlands Mobility Panel (MPN) is used. This is a longitudinal household panel, which commenced in 2013, with the goal of investigating how travel patterns of individuals change over a long period of time. Two surveys focusing on personal and household characteristics and a three-day travel diary are distributed among panel members every autumn. This panel is to a large extent representative of the Dutch population. We refer the reader to Hoogendoorn-Lanser et al. (2015) for a detailed description of the MPN surveys and travel diary.

A companion survey on the perceptions, attitudes, and wayfinding styles towards active modes (coined PAW-AM) was distributed among the MPN panel members in June 2017. This survey addressed among other things the experienced mode choice set of individuals (see Section 3). To identify different determinants that influence the composition of the experienced choice set, we use data from the personal and household surveys. Consequently, the data from the MPN surveys (2016) and the PAW-AM survey (2017) are merged. This study focuses on commuter trips, as such respondents were required to have a job and commute towards their work location. 2775 respondents fulfilled these requirements. A total of 31 alternatives can be experienced, of which one was never chosen and 18 were rarely chosen (less than 20 times). Therefore, a final filtering was performed to include only experienced choice set alternatives that contain sufficient respondents, which leads to a dataset of 2652 respondents. A total of 12 experienced mode choice set alternatives are included for model estimation and cross-validation:

1. Car
2. Bicycle
3. Train
4. BTM
5. Walk
6. Bicycle-Train
7. Bicycle-BTM
8. Bicycle-Walk
9. Car-Train
10. Car-Bicycle
11. Car-Bicycle-Walk
12. Car-Bicycle-BTM

**Table 2**  
Potential determinants of the experienced mode choice set.

<b>Socio-demographics</b>	<b>Urban density</b>
Gender (m/f)	Level of urbanisation (low/medium/high)
Age (-49/50+)	
Education level (low/medium/high)	<b>Household characteristics</b>
	No. of individuals (1–2/3 + )
<b>Ownership characteristics</b>	Children (y/n)
Driver's license (y/n)	Household income (low/medium/high)
Car ownership (y/n)	
Bicycle ownership (y/n)	<b>Work conditions</b>
Public transport subscription (y/n)	Working hours (part-time/full-time)
	Reimbursement for car, public transport, or bicycle (y/n)

Based on the determinants used to specify mode choice sets in literature and the availability of data in the MPN, potential determinants of the experienced mode choice set are selected for this study. Table 2 shows an overview of the selected determinants and their operationalisation in the models. Due to the aggregated nature of the experienced mode choice set in this data collection effort, trip-level characteristics are not included. Five categories of determinants are identified, namely socio-demographics, ownership characteristics, work conditions, urban density, and household characteristics. This information is collected in the MPN surveys of the year 2016.

The characteristics of the respondents in the dataset are presented in Table 3. The surveys are only distributed to individuals of 12 years and older. The education level represents the highest completed level of education. Consequently, the younger population that have not finished studies yet, ends up in a lower level of education. The education levels represent the following: low (completion of secondary education), medium (completion of higher secondary education, pre-university education, or secondary vocational education), and high (completion of higher professional education or university education). Many respondents have a medium or high education level, potentially due to the focus on commuting trips. Furthermore, ownership and availability percentages are high. A large share of the respondents lives in a highly urban environment, which represents municipalities of 1500 inhabitants/km<sup>2</sup> or more. A moderate urban environment is defined as a municipality of 1000–1500 inhabitants/km<sup>2</sup> and a low urban environment is defined as less than 1000 inhabitants/km<sup>2</sup>. Most respondents have no children (under the age of 12) and live in a four or more-person household. This means that most households have children (over the age of 12) or other inhabitants. Finally, more than half of the respondents are reimbursed by their employer for travelling by a certain mode, where the largest share is reimbursed for the car (e.g. in the form of a company car or kilometre compensation).

## 5. The experienced mode choice set

This section investigates the experienced mode choice set for commuting trips, using the dataset described in the previous section as derived from the MPN and PAW-AM surveys. In Section 5.1 the size and composition of this set are discussed. Section 5.2 compares the experienced choice set with a deterministic choice set, based on ownership and availability. Finally, Section 5.3 investigates consistency in this experienced set over time.

### 5.1. Size and composition of the experienced mode choice set

In the Netherlands, the car and bicycle are the most commonly used modes (CBS, 2018). When analysing the occurrence of different experienced mode choice sets for commuting purposes (Table 4), we see that car and bicycle are dominating.

**Table 3**  
Characteristics of the respondents (N = 2,652).

		Percentage (%)		Percentage (%)	
Gender	Male	46.8	Children in household	no	76.4
	Female	53.2		yes	23.6
Age	< = 34 years	41.0	Persons in household	1	14.7
	35–49 years	29.3		2	23.2
	50 < = years	29.8		3	18.5
				4+	43.7
Education level	Low	19.9	Reimbursement	Car	35.0
	Medium	40.8		PT	7.5
	High	39.3		Bicycle	10.5
Ownership	Drivers' license	89.7	Urban density	High	52.2
	Car	75.5		Moderate	18.6
	Bicycle	87.9		Low	29.3
	PT – subscription	30.5			

**Table 4**  
Ranking of the experienced mode choice sets.

No.	Rank	Alternative	Share (%)	Frequency
1	1	Car	47.1	1248
2	2	Bicycle	26.6	706
3	3	Car-Bicycle	9.8	260
4	4	Train	5.4	142
5	5	Walk	2.2	59
6	5	BTM	2.2	59
7	6	Car-Train	1.8	49
8	7	Bicycle-BTM	1.3	34
9	8	Bicycle-Walk	1.1	29
10	9	Car-Bicycle-Walk	1.0	26
11	10	Bicycle-Train	0.8	20
12	10	Car-Bicycle-BTM	0.8	20

Note that access and egress modes are excluded here. The most common sets consist of single-mode alternatives, with the exception of the car-bicycle choice set. Thus, individuals have a relatively small choice set for commuting trips, where most individuals only use one mode for their commute over a period of half a year. This was also found by [Kuhnimhof \(2009\)](#); [Kuhnimhof et al. \(2006\)](#), however they explored the mode use behaviour over only seven days. Our findings suggest that this unimodality is still largely present over a period of half a year, providing a first indication that individuals are habitual in their mode use for commuting trips.

[Table 5](#) visualises the modal shares per mode and shows how each mode is part of choice sets of different sizes. When investigating the experienced mode choice sets from this perspective, several observations can be made. First of all, it is confirmed that car and bicycle are the most common modes for commuting trips among individuals in the sample. The other modes; train, BTM, and walking, are used much less for commuting trips. Furthermore, BTM and walking are, relatively, more often part of multimodal choice sets (about 50% of the occurrences) compared to the other modes. Conversely, the car is most often used unimodally. Finally, the majority of the respondents have reported using a single mode for commuting trips in the last half year (83.5%), compared to 14.8% of the respondents that used two modes and 1.7% of the respondents that used three modes. The share of unimodal commuters is higher than the 72% found by [Kuhnimhof \(2009\)](#). One might expect that this percentage decreases when the observation period increases, however our findings show the opposite. Potentially, this is related to the context, i.e. Germany versus the Netherlands. The unimodality that we observe for commuting trips, does not necessarily mean unimodality in general, as for other trip purposes more or different modes can be used.

## 5.2. Comparison between choice set definitions

Both in research and practice, the mode choice set is often defined using deterministic rules ([de Jong et al., 2007](#); [Hamre and Buehler, 2014](#); [Gehrke and Clifton, 2014](#); [Kamargianni and Polydoropoulou, 2013](#)). These deterministic show large variations in their rigorosity between studies, yielding different choice sets when applying them on the same data. As an example, for the inclusion of public transport in the choice set, [Gehrke and Clifton \(2014\)](#) state that a bus or train stop should be present within respectively 0.5 and 1.0 mile from the home location and [Habib et al. \(2011\)](#) state that a stop should be available within the neighbourhood. On the other hand, [Ton et al. \(2019a\)](#) use the Google Directions API to identify whether a public transport route is available from home to destination. The nearest stop is not necessarily the best suitable stop for the entire trip, therefore one can argue whether including only the nearest stop will be accurate for a trip. Consequently, these rules will result with different choice sets. In this section we compare the experienced choice set with a choice set defined based on deterministic rules, in this case based on availability and ownership, to identify differences and similarities between rule-based and behaviour-based choice sets. The deterministic method defined in this paper should be regarded as an example and is not considered representative of the contributions in existing literature, however it is deemed reasonable in terms of the assumptions made. For more solid conclusions on this comparison, future research is advised to compare the experienced choice set with a variety of the state-of-the-art choice set generation methods ([Ortúzar and Willumsen, 2011](#)).

**Table 5**  
The experienced modes in the past half year depicted against the choice set size.

Mode	No. of modes in choice set			Total
	1	2	3	
Car	47.1%	11.7%	1.7%	60.4%
Train	5.4%	2.6%	0.0%	8.0%
BTM	2.2%	1.3%	0.8%	4.3%
Bicycle	26.6%	12.9%	1.7%	41.3%
Walk	2.2%	1.1%	1.0%	4.3%



**Table 6**

Availability/ownership choice set compared to the experienced choice set. dotted = exact match, diagonal stripes = subset, horizontal stripes = mismatch, white = mixture. w = walk, l = local transit (btm), t = train, b = bicycle, c = car, pt = public transport.

		Availability/Ownership Choice set								
		w	pt-w	bw	b-pt-w	cw	c-pt-w	cbw	cb-pt-w	Tot.
Experienced choice set	w	0	0	10	11	7	2	25	4	59
	l	0	4	6	31	0	1	7	10	59
	t	1	1	2	74	0	11	14	39	142
	b	4	5	134	184	27	9	273	70	706
	bw	1	0	6	10	2	1	7	2	29
	bl	0	1	4	22	0	1	1	5	34
	bt	0	0	1	5	1	2	1	10	20
	c	17	1	38	37	162	20	832	141	1,248
	ct	0	0	0	4	3	5	14	23	49
	cb	2	0	17	9	21	3	172	36	260
	cbw	0	0	1	2	5	0	16	2	26
	cbl	0	0	2	4	0	0	8	6	20
Tot.		25	12	221	393	228	55	1,370	348	2,652

To define the commuting availability/ownership choice set, we assume that the mode needs to be available to an individual on a daily basis. Therefore, the car and bicycle are included only if the respondent owns the mode. This means that no distinction is made between driver and passenger. Furthermore, public transport is only included if the individual has a subscription (e.g. discount, ticket for a specific line). This seems a plausible assumption for daily commuting trips, because not having this subscription and using it on a daily basis is expensive in the Netherlands (either for the individual or employer). Regarding walking we define no availability assumptions. This will not always hold for commuting trips (see Table 5), because people might not be able to walk or consider it too far to walk. The same applies for the inclusion of bicycle. However, because of the aggregated nature of the data no information on distance or travel time is available for many individuals (47%). Based on the above definition eight different choice set combinations can be identified. Four modes are distinguished in the deterministic choice set: car, public transport, bicycle, and walk. In the experienced choice set public transport is divided into BTM and train, therefore the comparison is not completely one-to-one.

Table 6 shows the comparison between the deterministic choice set and experienced choice set, with three exact matches between both sets: walk only, bicycle-walk, and car-bicycle-walk. The total of exact matches (dotted) is 22 out of 2,652 (0.8%). Consequently, when defining a choice set based on availability and ownership many differences are found in excluding and including relevant modes compared to the observed behaviour. The horizontal stripes show the mismatches between the two sets. We found a total of 171 mismatches (6.4%). The largest mismatch in number of respondents is between the car only experienced set and bicycle-walk ownership/availability set (38 out of 2,652). These individuals do not own a car (only a bicycle), but solely use the car as commuting mode. This means that these individuals borrow the car from someone else, are a passenger, or use it via a sharing system. For the majority of the respondents the experienced set is a subset of the ownership/availability set or vice versa (diagonal stripes). For example, in case of the car only experienced choice set and the car-bicycle-walk ownership/availability set (832 out of 2,652), individuals also own a bicycle, but do not use it. Some respondents show a mixture of the ownership/availability and experienced set (white), for example car-bicycle is experienced, whereas car-walk is owned/available. In that case the bicycle was borrowed from someone else or used via a sharing system. Consequently, ownership and availability are not the only explanatory variables for the experienced choice set. As mentioned before, different deterministic rules will result with different choice sets. As these rules are all based on logic and network information, they are likely to mismatch to a certain extent with observed behaviour (experienced choice set).

5.3. Consistency in the experienced mode choice set over time

In the PAW-AM survey, the respondents were asked to recall which modes they have used over the past half year. To investigate consistency over time we compare the experienced choice set from the PAW-AM survey with the choice set containing all reported commuting modes in the three-day travel diary (MPN). In the travel diary, individuals were asked to report all the trips (and modes) made in the course of three days. By filtering the commuting trips from this diary, the experienced choice set based on the three-day travel diary is composed. It is uncertain whether this three-day period captures the whole spectrum of modes used. Notwithstanding, it is expected to help in identifying (in)consistency over time. The travel diary data was collected in Autumn 2016, whereas the PAW-AM survey covers the first half year of 2017.

Of the 2652 individuals that filled in the PAW-AM survey, only 1280 filled in the three-day travel diary and made at least one commuting trip. Approximately two-thirds (67.3%) of these respondents report the same choice set during both periods and are thus considered consistent in their experienced choice set over time. The choice sets that show consistency are the unimodal choice sets and two-mode sets, such as bicycle-walking and car-bicycle. The lack of consistency in the three-mode choice sets, might stem from the fact that only three days were observed. A total of 22.7% of the population reports a subset, either of the experienced set (13.6%) or of the three-day diary set (9.1%). Furthermore, a total of 9.3% of the respondents report a different choice set during the three days

**Table 7**

Experienced choice sets of the survey (2017) and the three-day travel diary (2016) compared. dotted = exact match, diagonal stripes = subset, horizontal stripes = mismatch, white = mixture. w = walk, l = local transit (btm), t = train, b = bicycle, c = car.

2017	2016																			Tot.		
	w	l	wl	t	tw	tl	b	bw	blw	bt	btlw	c	cw	cl	ct	ctw	ctl	cb	cbw		cbl	cblt
w	11	0	0	1	0	1	1	0	0	0	0	2	2	0	0	0	0	0	1	0	0	19
l	0	14	4	0	0	1	4	0	0	0	0	0	0	4	0	0	0	1	0	0	1	29
t	0	1	0	39	4	1	8	0	0	3	0	8	2	0	2	0	1	1	0	0	0	70
b	3	1	1	2	0	0	213	10	2	2	0	43	1	1	0	0	0	34	3	3	1	320
bw	1	1	0	0	0	0	7	6	0	0	0	0	0	0	0	0	0	1	0	0	0	16
bl	0	2	1	0	0	0	11	0	1	1	0	0	0	0	0	0	0	0	0	0	0	16
bt	0	0	0	10	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	13
c	3	4	1	5	0	2	16	0	0	2	0	553	12	5	4	1	0	13	0	1	0	622
ct	0	0	0	8	1	0	0	0	0	0	1	5	1	1	2	0	0	0	0	0	0	19
cb	0	0	0	2	0	0	44	0	0	0	0	62	1	0	0	0	0	23	0	1	0	133
cbw	0	0	0	0	0	0	5	0	0	0	0	6	1	0	0	0	0	2	0	0	0	14
cbl	0	3	0	0	0	0	4	0	0	0	0	2	0	0	0	0	0	0	0	0	0	9
Tot.	18	26	7	67	6	5	314	17	3	8	1	681	20	11	8	1	1	74	5	5	2	1,280

compared to the half year. Noticeably, 22.2% of this group has experienced a life event related to moving jobs or moving homes. One other reason of this shift in behaviour could be seasonality: autumn versus winter and spring.

When investigating the (in)consistency over time more thoroughly for the 1280 individuals that are part of both datasets, several observations can be made (see Table 7). First, consistency over time occurs most for the unimodal choice sets (96% of the matches). Consequently, the patterns that were uncovered in the half year survey were already present in the year before, confirming the habitual behaviour. The individuals that have a wider spectrum of modes in their experienced set are often partially consistent over time (e.g. expanding the set or shrinking the set). Second, several respondents have (partially) shifted from motorised modes to active modes, which might (again) be due to change in season. For example, 43 individuals shifted from car to bicycle, 34 shifted from car-bicycle to bicycle only, 62 shifted from car to car-bicycle, and 2 shifted from car-walk to walk only. Finally, when a subset of the travel diary set is reported in the survey, the data often suggests habit formation. For example, in case of bicycle-walk, 10 out of 17 individuals shift to bicycle only, for the car-bicycle-walk alternative all individuals become unimodal users, and the car-bicycle alternative shows that the majority of the individuals become either unimodal car or bicycle users.

**6. Modelling results**

The determinants that are relevant for the experienced choice set are uncovered using discrete choice models. This section details the results of this exercise. Section 6.1 describes the overall results of the estimated models. Section 6.2 discusses the determinants that are relevant for the experienced mode choice set. Finally, Section 6.3 reflects on the suitability of choice sets based on historical data for prediction purposes.

**6.1. Overall model estimation results**

In this study four different model structures are tested: MNL, NL, CNL, and ML. The NL, CNL, and ML models do not produce significantly superior results compared to the MNL model. All the nesting parameters that are estimated for each of the NL models (based on both size and composition) are not significantly different from one, consequently suggesting that the alternatives do not share unobserved variables. The CNL model (based on mode-nests) with variable membership did not converge properly, consequently an iterative process was used to find the best membership for each alternative to the nests. This iterative process consisted of alternately estimating the nesting parameters and attributes, while fixing the alphas, and estimating the alphas, while using the results of the previous iteration as fixed input. This model significantly improved the log-likelihood compared to the MNL, however no significant nesting parameters were found. The ML model reproduces the CNL model with fixed membership, thus performing worse than the CNL model with flexible membership. Consequently, we have to conclude that the MNL model produces the best results. This seems plausible, because we include a relationship in the utility function between alternatives that contain the same modes via the estimation of mode-specific parameters (Eq. (2)). Consequently, the car-bicycle and unimodal bicycle alternatives contain partially the same parameters. The other model structures do not find a significant effect of shared unobserved variables between alternatives. Table 8 shows the overall model fit results. The MNL model is estimated using a random draw of 80% of the data. It has a model fit of 0.542.

**6.2. Determinants of the experienced choice set**

In this section we discuss the different determinants that are relevant for explaining the experienced mode choice set according to the MNL model (see Table 9). The utility function consists of alternative specific constants and mode specific parameters. Regarding the first, we have fixed the parameter for car to zero, so that a comparison based on the relative utility can be made. Regarding the second, the model specification implies that if the alternative is car-bicycle, the parameter values for car and bicycle need to be

**Table 8**  
Model fit results.

	MNL
Initial log-likelihood (equally likely)	–5307.761
Final log-likelihood	–2383.285
# parameters	45
AIC	4856.6
BIC	5111.5
Adjusted rho-square (compared to the initial model)	0.543
# observations	2,136

**Table 9**  
Determinants of the experienced mode choice set for the mnl model.

Determinant	Level	Mode	Coefficient	Robust t-test		
<i>Alternative-specific parameters</i>						
Constant	Car		0	–		
	Bicycle		2.60	7.65*		
	Bicycle-BTM		–0.81	–1.71		
	Bicycle-Train		–2.48	–5.33*		
	Bicycle-Walk		–0.16	–0.33		
	BTM		–0.23	–0.52		
	Car-Bicycle		–1.21	–6.10*		
	Car-Bicycle-BTM		–3.62	–8.05*		
	Car-Bicycle-Walk		–2.96	–7.15*		
	Car-Train		–4.69	–13.64*		
	Train		–0.62	–1.52		
<i>Mode-specific parameters</i>	Walk		0.79	1.84		
	Education level (ref = Low)	Medium	Car	0.42	2.50*	
		High	Bicycle	0.28	2.12*	
			Car	0.49	2.63*	
			Train	0.59	2.70*	
	Age (ref = < 50 years)	≥50 years	BTM	–0.64	–2.09*	
			Car	–0.24	–1.71	
			Train	–0.87	–3.27*	
	Gender (ref = Female)	Male	Car	0.42	3.31*	
	Household size (ref = 3 + pers.)	1–2 pers.	Walk	0.51	2.08*	
	Children (ref = No)	Yes	Car	0.37	2.20*	
Walk			0.76	2.39*		
Urban density (ref = High)	Low	Bicycle	–0.44	–3.57*		
		BTM	–0.75	–2.32*		
		Train	–0.88	–3.18*		
	Moderate	BTM	–0.70	–1.84		
Driver's License (ref = No)		Car	1.71	5.59*		
Ownership (ref = No)	Bicycle	Bicycle	0.31	1.83		
		Walk	–0.77	–2.48*		
		BTM	–0.61	–2.09*		
	Car	Car	1.10	6.72*		
		Train	0.46	1.95		
		BTM	0.97	3.35*		
	PT sub.	Car	–0.54	–3.74*		
		Train	1.63	6.49*		
		Working hours (ref = Parttime)	Fulltime	Bicycle	–0.58	–5.12*
			Reimbursement (ref = No)	Bicycle	Bicycle	2.04
BTM	0.67	1.79				
Car	–0.78	–3.77*				
Car	Bicycle	–0.89		–6.07*		
	Car	1.67		7.96*		
	Walk	–1.55		–2.99*		
PT	Bicycle	–0.48		–1.75		
	BTM	2.93		8.44*		
	Train	3.08		9.95*		

\* = significant at the 5% level. ref = variable fixed to zero.

summed up (linear in parameters) to find the combined parameter. We discuss the parameter coefficients per category of variables (6.2.1 - 6.2.6).

### 6.2.1. Constants

The constants are alternative-specific and provide information on the average influence of the unobserved variables on the utility (relative to the reference alternative: car-only). The car-only alternative is most frequently chosen, which explains why the parameters values related to most other mode choice alternatives are negative. The constants for walk and cycle are positive, suggesting that these have unobserved variables that favour walking and cycling over the car. This might be due to the shorter distances for which these modes are used, which is not captured in the model (unobserved). Most parameters are significant, which indicates that the mode-specific parameters do not capture all the information in the data. Thus, other influences are present in the experienced mode choice set choice that are currently not observed.

### 6.2.2. Socio-demographics

Three socio-demographic variables are tested: age, gender, and education level. All are found to be significant. This is in line with the research of [Vij et al. \(2017, 2013\)](#); [Calastri et al. \(2017\)](#), who used among others, these three variables to identify the availability of several modes. The older commuting population ( $\geq 50$  years) is associated with a lower utility for the BTM, car, and train modes compared to the young population ( $< 50$  years). The active modes are thus more attractive to the 50+ population than the motorised modes, all else being equal. Men are more likely to have car included in their experienced mode choice set than women, which is also found by [Vij et al. \(2013\)](#). Furthermore, the education level parameters show that if an individual has a medium and high education level, they are more likely to include the car mode in the experienced choice set compared to an individual with a low education level. The bicycle and train also become more attractive for an individual with a high education level.

### 6.2.3. Household characteristics

The household characteristics exhibit a limited association with the experienced mode choice set. The household income does not yield a significant relationship. The salary of the combined household members does not make it more or less likely that a mode is included in the experienced choice set, contrary to [Witlox and Tindemans \(2004\)](#), who found a positive relationship between income and car use. The size and composition of the household do explain the inclusion of walking and using the car in the experienced mode choice set. The presence of children under the age of 12 in the household increases the probability of including walking and the car in the experienced choice set. This is in line with the results found by [Vij et al. \(2017\)](#), as they found that having children increases likelihood that one is dependent on a car. Furthermore, an individual living in a household consisting of one or two persons is more likely to include walking in the experienced choice set compared to an individual living in a larger household.

### 6.2.4. Urban density

Low urban density represents municipalities that contain mostly villages and rural areas, moderate density represents municipalities with medium-sized cities, such as Groningen, and high density represents municipalities with large cities, such as Amsterdam and Rotterdam. We took the high urban density as a reference point. With moderate density it is less likely that BTM is included in the choice set. Generally, the density and frequency of BTM is very high in high urban areas, but less so in moderately urban areas. In low urban areas, the bicycle, train, and BTM are less likely to be included in the experienced mode choice set. This finding shows that the experienced choice set is not only related to individual specific determinants, but also to where they live.

### 6.2.5. Ownership

The ownership variables are all significantly associated with the experienced mode choice set. This concurs with the use of ownership variables in many mode choice studies to identify the mode choice set ([Ben-Akiva and Boccara, 1995](#); [Gehrke and Clifton, 2014](#); [Habib et al., 2011](#); [Kamargianni, 2015](#); [Kamargianni et al., 2015](#); [Rodríguez and Joo, 2004](#); [Vij et al., 2017, 2013](#)).

Having a driver's license is positively associated with the inclusion of the car in the experienced choice set. It does not have a significant impact on the inclusion of other modes. In contrast, ownership of a certain mode positively relates to inclusion of that mode and at the same time negatively relates to inclusion of other modes in the experienced mode choice set. Owning a bicycle positively relates to including the bicycle in the choice set, which is in line with literature ([Heinen et al., 2010](#); [Muñoz et al., 2016](#)). Moreover, it also reduces the utility of walking for inclusion in the choice set. Owning a car increases the utility of inclusion of car in the choice set, furthermore it increases the utility for including train in the choice set. This suggests that train users often own a car. On the other hand, it reduces the utility of the BTM. Ownership of a public transport subscription results with the expected results: having a subscription increases the probability of including BTM and train in the choice set and reduces the probability of using the car. This suggests that train users are affected differently by car ownership, compared to car users and the ownership of a public transport subscription. The first yields a positive relation, whereas the second yields a negative relation.

### 6.2.6. Work conditions

For commuting trips, the work conditions are important determinants. Both the working hours and reimbursement are relevant for the experienced mode choice set, where the latter proves to be more important. Working full-time (more than 35 hours per week) decreases the probability of including the bicycle in the choice set, which is in line with literature ([Heinen et al., 2010](#)). This does not hold for the other modes of transport. Regarding the reimbursement, similarly to the trend observed for the ownership of modes: being reimbursed for using a mode to commute to work increases the probability of including that mode in the experienced choice

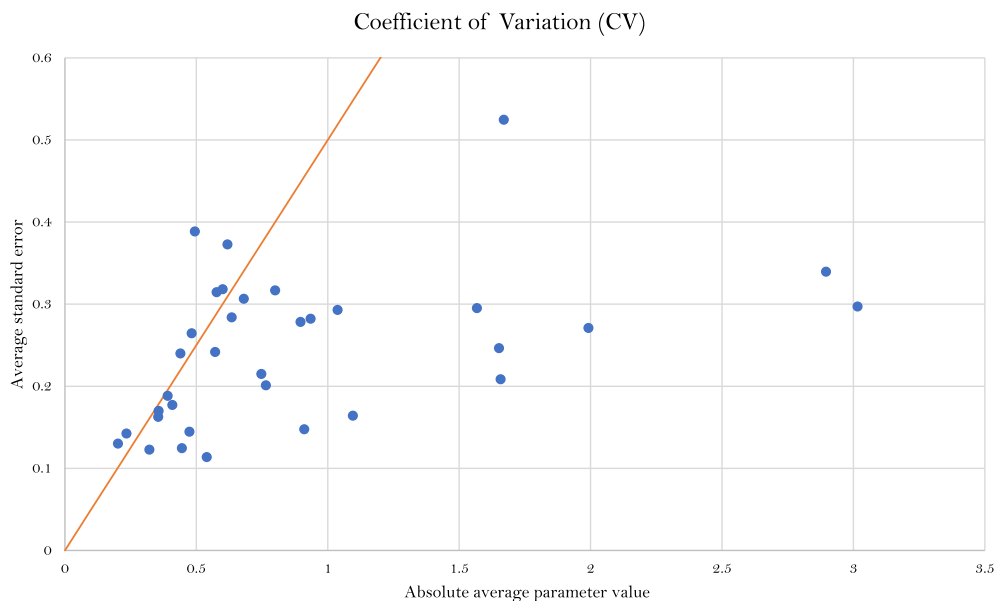


Fig. 3. Coefficient of variation (mean standard error divided by mean absolute parameter value).

set, whereas it decreases the probability of including another mode in the experienced choice set. One exception is the reimbursement for bicycle, which positively relates to inclusion of BTM in the choice set. Both the bicycle and BTM can be used for similar distance ranges and can thus be used as substitutes. Reimbursement of the car decreases the inclusion of walk and bicycle in the choice set. Finally, a public transport reimbursement increases the utility for BTM and train and reduces the utility of the bicycle.

### 6.3. Using the experienced mode choice set for out-of-sample predictions

To test the performance of the model for predictions, we have partitioned the data into five segments of approximately 20% each. One of the segments is used as the default case to calibrate the parameters involved. The other segments are used for validation. In this validation we investigate several aspects: model performance, stability of the parameters, and prediction accuracy for out-of-sample data.

The models perform well for the different segments. The default model produced an adjusted rho-square of 0.542, whereas the validation models show a model fit ranging from 0.526 to 0.541. This means that for all segments between 52% and 54% of the variation of the data can be explained using this model. The AIC and BIC values also show promising results, with some segments showing better results than the default model.

The stability of the parameters is investigated by calculating the coefficient of variation (CV) for each parameter. This is done by dividing the average standard error of the parameter by the average parameter value. If the CV is over 0.5, the parameter is considered less stable. Fig. 3 visualises the CV for all parameters. Left of the orange line the CV is over 0.5 and thus less stable. We exclude the constants in this comparison, because these capture the unobserved variables for each model and are therefore different per model. Eight parameters are considered less stable: high education – bicycle, medium education – car, 50+ years – BTM and car, own car – train, low urban – BTM, moderate urban – BTM, and reimbursement bicycle – BTM. These parameters show large variation in their parameter values and standard errors, where the last parameter is the least stable. The majority of the parameters, however, are stable over the different runs.

For investigating the prediction accuracy, we apply the model on out-of-sample data. This results with a probability for every alternative to be chosen. When summing these probabilities over individuals for each alternative, we can calculate what the probability is that the prediction is correct or not. In some cases, the model predicts a choice set that is too small, too large, or mixed. In this case, the prediction is considered better than when the wrong mode(s) are predicted, as it is able to identify part of the choice set. An example is if the prediction is bicycle and the actual choice set was bicycle-car. If the predicted choice set is a total mismatch, we consider it wrong. For example, the predicted set is BTM, whereas the actual choice set is bicycle-walk. Table 10 shows the prediction accuracy based on the above-mentioned situations for each of the model runs. The models predict the correct experienced choice set on average 49% of the time. Three alternatives are especially well predicted: car, bicycle, and train. These alternatives are observed most frequently in the dataset, consequently the model aims to predict these alternatives correctly. In 25% of the cases a choice set is predicted that is smaller, larger, or mixed compared to the observed choice set and in 26% the model produces a wrong choice set. All in all, around 74% of the observations the experienced mode choice set is predicted with sufficient accuracy. This suggests that the relevant determinants are, to a large extent, able to capture the experienced choice set of individuals for commuting purposes. Consequently, these determinants can also be used for the specification of mode choice sets in future mode choice studies.

**Table 10**  
Predictive power of the experienced mode choice set model on out-of-sample data.

	Correct	Too small	Too large	Mixed	Wrong
Run 1	48%	12%	12%	1%	27%
Run 2	50%	11%	12%	1%	27%
Run 3	50%	12%	12%	1%	26%
Run 4	49%	13%	12%	1%	25%
Run 5	49%	12%	11%	1%	27%
Average	49%	12%	12%	1%	26%

### 7. Discussion on the experienced mode choice set

This section discusses our findings of the experienced mode choice set in relation to past research. In particular, we discuss the determinants of the choice set, the unimodal choice sets, and potential applications of the experienced choice set.

In the literature, we identified that most studies specify the mode choice set based on ownership/availability and trip characteristics (e.g. Ben-Akiva and Boccara, 1995; Gehrke and Clifton, 2014; Hensher and Ho, 2016; Vij et al., 2013). Some studies have investigated the influence of individual characteristics on the availability/inclusion of modes in the choice set (Calastri et al., 2017; Vij et al., 2017, 2013). In this study, we tested both ownership and individual characteristics and found that both influenced the experienced choice set. Our comparison between the ownership/availability choice set and the experienced choice set shows that these sets are very different. A specification based on ownership would falsely exclude alternatives such as car and bicycle, which can also be used as shared modes. This discussion also showed that different deterministic rules result with different choice sets, and that these somewhat ‘arbitrary’ rules are different from observed behaviour. Furthermore, we found that the urban environment is also relevant for the experienced mode choice set. This means that regardless of the characteristics of an individual, certain modes have a higher or lower probability to be used depending on the type of environment in which one lives. This finding is confirmed by research on mode use and mobility patterns (Kuhnimhof et al., 2006; Ton et al., 2019b). Furthermore, the work conditions prove to be important explanatory variables of the experienced mode choice set for commuting trips. This means that the employer plays an important role in the mode use of its employees. Consequently, more individuals may start cycling when the system of reimbursement for commuting given by the employer is changed to benefit cyclists (Heinen et al., 2013). In sum, a wider variety of characteristics is relevant for the identification of the mode choice set compared to what has been previously assumed.

The majority of individuals has a unimodal experienced choice set for commuting trips. Many of these individuals show habitual behaviour that is consistent over time. This is in line with findings from Lavery et al. (2013); Kuhnimhof (2009). However, this study offers evidence to suggest that there might be ways to influence travellers in ways that will lead to an increase of the mode set or even a modal shift. The employer can, for example with support from the government, choose to provide reimbursement for sustainable or active modes instead of car use. This increases the probability of including the bicycle or public transport in the experienced choice set. Unimodality is higher in low density urban areas (86.8%) than in moderate (83.8%) or high density urban areas (81.5%). Stimulation of the bicycle use may increase the use of the bicycle in high density urban areas, potentially increasing the experienced choice set size. Consequently, this study shows that several determinants can be used to develop policy measures that might increase the number of modes used by commuters.

A potential application of the experienced choice set lies in the mode choice domain. When embedding this method in the probabilistic approach proposed by Manski (1977), a probability is assigned to each experienced choice set. In a simultaneous model the mode choice is then estimated such that it incorporates the probabilities for the experienced choice sets as shown in equation (3).

$$P(i|B, D, X_n) = \sum_{C \in G_n} P(i|C, B, X_n) \cdot P(C|G_n, D, X_n) \tag{3}$$

where  $P(i|B, D, X_n)$  is the probability that an individual  $n$  chooses alternative  $i$ , given the parameters  $B$  and  $D$  and explanatory variables  $X_n$ . This depends on the probability of a choice set being chosen by the individual and the choice from this choice set. The set  $G_n$  includes all non-empty deterministically feasible modes for individual  $n$ .  $G_n$  is a subset of the master choice set  $M$  that comprises all possible alternatives available for the choice context and population ( $G_n \subseteq M$ ). Eq. (3) thus consists of three parts (Swait and Ben-Akiva, 1987b); (1) a mode choice aspect given a choice set, (2) a deterministic choice set generation aspect to define  $M_n$ , and (3) a probabilistic choice set generation aspect that expresses the probability that choice set  $C$  is the actual consideration choice set. Among others Ben-Akiva and Boccara (1995); Cantillo and Ortúzar (2005) have applied variants of this method. This application has to be thoroughly tested with respect to potential endogeneity and bias, as the experienced choice set is based on mode use. Furthermore, this method needs to be benchmarked against commonly used mode choice set specification methods (such as those based on deterministic rules), to adequately identify the (potential) added value of this method.

### 8. Conclusions and recommendations

This paper presents the findings of an analysis of the experienced mode choice set and identifies the determinants that impact this set. The experienced choice set is the set of modes that is used over a long period of time. We focus on commuting trips in the Netherlands, using data from the Mobility Panel Netherlands (MPN) and a companion survey, in which we ask respondents for their

used modes for commuting trips in the past half year. We evaluate the experienced choice set by analysing the size and composition, comparing this set with a choice set based on ownership and availability, and by investigating its consistency over time. The determinants are identified by means of estimation of discrete choice models.

The analysis of the experienced mode choice set for commuting purposes shows that the size of this set is limited. The majority of the respondents only uses one mode in the course of half a year, which indicates habitual behaviour. We investigated which determinants are relevant for the formation of the experienced choice set. Determinants belonging to socio-demographics, household characteristics, urban density, work conditions, and ownership of modes are relevant for choice set formation. The work conditions, especially the reimbursement by the employer for using a specific mode, is particularly important for the experienced mode choice set of commuters. The second group in terms of importance is ownership or availability of modes. We show that more determinants are relevant in the choice set formation than previously assumed. While many studies specify the mode choice set based on ownership and trip characteristics, only some have extended this to including individual characteristics to identify availability of modes. The results are, to a large extent, transferable to out-of-sample data. According to our findings, future research into mode choice could benefit from including a wider variety of variables in the choice set specification.

New modes can potentially be added to the individual's experienced choice set given the right incentives. Policy measures could focus on the reimbursement provided by the employer as this can be used to increase the probability of including the bicycle in the choice set, if this mode is reimbursed and others are not.

Future research may aim to estimate a probabilistic integrated choice set and mode choice model using the experienced choice set. The method needs to be benchmarked against often-used methods to identify its added value in performance and computational effort. This study showed that habitual behaviour is present for the majority of the individuals regarding their commuting trips. When data is available on the experienced choice set over multiple years, the impact of habit formation and life cycle can be investigated in relation to the experienced choice set. Furthermore, it is interesting to investigate if habit formation also arises for other trip purposes or in general in the mobility pattern of the individuals. Next to that, given the increasing evidence that the social environment (i.e. household members and friends) influences mode choice behaviour of individuals (e.g. Pike and Lubell, 2018), additional research could investigate the impact of the social environment on the experienced mode choice set. Furthermore, this study focusses only on commuting. The null alternative (not commuting) would be an interesting addition to the model, as this would additionally increase the understanding of factors influencing the decision to work at home or commute to work. Finally, a potentially interesting addition to the experienced choice set is the way in which the modes are used, i.e. as private mode or via shared systems (and types thereof).

#### CRedit authorship contribution statement

**Danique Ton:** Conceptualization, Methodology, Formal analysis, Investigation, Validation, Visualization, Writing - original draft. **Shlomo Bekhor:** Methodology, Formal analysis, Writing - review & editing. **Oded Cats:** Conceptualization, Supervision, Writing - review & editing. **Dorine C. Duives:** Investigation, Supervision, Writing - review & editing. **Sascha Hoogendoorn-Lanser:** Investigation, Writing - review & editing. **Serge P. Hoogendoorn:** Funding acquisition, Supervision, Writing - review & editing.

#### Acknowledgements

We thank the anonymous reviewers for their constructive feedback that helped improve the quality of the paper. This research was supported by the Allegro project (no. 669792), which is financed by the European Research Council and Amsterdam Institute for Advanced Metropolitan Solutions. The data was made available by the Netherlands Mobility Panel administered by KiM Netherlands Institute for Transport Policy Analysis.

#### References

- Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions. In: *Handbook of Transportation Science*. Springer, pp. 5–33.
- Ben-Akiva, M., Boccara, B., 1995. Discrete choice models with latent choice sets. *Intern. J. Res. Mark.* 12, 9–24.
- Bhat, C.R., 1995. A heteroscedastic extreme value model of intercity travel mode choice. *Transp. Res. Part B Methodol.* 29, 471–483.
- Bierlaire, M., 2016. PythonBiogeme: a short introduction.
- Bierlaire, M., 2006. A theoretical analysis of the cross-nested logit model. *Ann. Oper. Res.* 144, 287–300.
- Calastri, C., Hess, S., Choudhury, C., Daly, A., Gabrielli, L., 2017. Mode choice with latent availability and consideration: theory and a case study. *Transp. Res. Part B Methodol.*
- Cantillo, V., Ortúzar, J.de.D., 2005. A semi-compensatory discrete choice model with explicit attribute thresholds of perception. *Transp. Res. Part B Methodol.* 39, 641–657.
- CBS. 2018. Personenmobiliteit in Nederland [WWW Document]. Statline, <http://statline.cbs.nl> [in Dutch].
- De Jong, G., Daly, A., Pieters, M., Miller, S., Plasmeijer, R., Hofman, F., 2007. Uncertainty in traffic forecasts: literature review and new results for The Netherlands. *Transportation (Amst)*. 34, 375–395.
- Gehrke, S.R., Clifton, K.J., 2014. Operationalizing land use diversity at varying geographic scales and its connection to mode choice: Evidence from Portland, Oregon. *Transp. Res. Rec.* 2453, 128–136.
- Habib, K.M.N., Tian, Y., Zaman, H., 2011. Modelling commuting mode choice with explicit consideration of carpool in the choice set formation. *Transportation (Amst)*. 38, 587–604.
- Hamre, A., Buehler, R., 2014. Commuter mode choice and free car parking, public transportation benefits, showers/lockers, and bike parking at work: evidence from the Washington, DC region. *J. Public Transp.* 17, 4.
- Heinen, E., Chatterjee, K., 2015. The same mode again? An exploration of mode choice variability in Great Britain using the National Travel Survey. *Transp. Res. Part A Policy Pract.* 78, 266–282.
- Heinen, E., Maat, K., van Wee, B., 2013. The effect of work-related factors on the bicycle commute mode choice in the Netherlands. *Transportation (Amst)*. 40, 23–43.

- Heinen, E., Van Wee, B., Maat, K., 2010. Commuting by bicycle: an overview of the literature. *Transp. Res.* 30, 59–96.
- Hensher, D.A., Ho, C.Q., 2016. Experience conditioning in commuter modal choice modelling—Does it make a difference? *Transp. Res. Part E Logist. Transp. Res.* 95, 164–176.
- Hoogendoorn-Lanser, S., Schaap, N.T.W., OldeKalter, M.-J., 2015. The Netherlands Mobility Panel: An innovative design approach for web-based longitudinal travel data collection. *Transp. Res. Procedia* 11, 311–329.
- Kamargianni, M., 2015. Investigating next generation's cycling ridership to promote sustainable mobility in different types of cities. *Res. Transp. Econ.* 53, 45–55.
- Kamargianni, M., Dubey, S., Polydoropoulou, A., Bhat, C., 2015. Investigating the subjective and objective factors influencing teenagers' school travel mode choice—An integrated choice and latent variable model. *Transp. Res. Part A Policy Pract.* 78, 473–488.
- Kamargianni, M., Polydoropoulou, A., 2013. Hybrid choice model to investigate effects of teenagers' attitudes toward walking and cycling on mode choice behavior. *Transp. Res. Rec.* 2382, 151–161.
- Kuhnimhof, T., Chlond, B., Von Der Ruhren, S., 2006. Users of transport modes and multimodal travel behavior: Steps toward understanding travelers' options and choices. *Transp. Res. Rec.* 1985, 40–48.
- Kuhnimhof, T.G., 2009. Measuring and modeling multimodal mode use in the longitudinal section. *Proceedings of the 88th Annual TRB Meeting*.
- Lavery, T.A., Páez, A., Kanaroglou, P.S., 2013. Driving out of choices: An investigation of transport modality in a university sample. *Transp. Res. Part A Policy Pract.* 57, 37–46.
- Manski, C.F., 1977. The structure of random utility models. *Theory Decis.* 8, 229–254.
- McFadden, D., Train, K., 2000. Mixed MNL models for discrete response. *J. Appl. Econom.* 15, 447–470.
- Molin, E., Mokhtarian, P., Kroesen, M., 2016. Multimodal travel groups and attitudes: A latent class cluster analysis of Dutch travelers. *Transp. Res. Part A Policy Pract.* 83, 14–29.
- Muñoz, B., Monzon, A., Daziano, R.A., 2016. The increasing role of latent variables in modelling bicycle mode choice. *Transp. Res.* 36, 737–771.
- Ortúzar, J.de.D., Willumsen, L.G., 2011. *Modelling transport*. John Wiley & Sons.
- Pike, S., Lubell, M., 2018. The conditional effects of social influence in transportation mode choice. *Res. Transp. Econ.* 68, 2–10.
- Ralph, K.M., 2017. Multimodal millennials? The four traveler types of young people in the United States in 2009. *J. Plan. Educ. Res.* 37, 150–163.
- Rijksoverheid, 2018. Staatssecretaris Van Veldhoven: 200.000 mensen uit de file op de fiets [WWW Document]. <https://www.rijksoverheid.nl/actueel/nieuws/2018/06/12/staatssecretaris-van-velhoven-200.000-mensen-uit-de-file-op-de-fiets> [in Dutch].
- Rodríguez, D.A., Joo, J., 2004. The relationship between non-motorized mode choice and the local physical environment. *Transp. Res. Part D Transp. Environ.* 9, 151–173.
- Swait, J., Ben-Akiva, M., 1987a. Empirical test of a constrained choice discrete model: mode choice in Sao Paulo. Brazil. *Transp. Res. Part B Methodol.* 21, 103–115.
- Swait, J., Ben-Akiva, M., 1987b. Incorporating random constraints in discrete models of choice set generation. *Transp. Res. Part B Methodol.* 21, 91–102.
- Ton, D., Duives, D.C., Cats, O., Hoogendoorn-Lanser, S., Hoogendoorn, S.P., 2019a. Cycling or walking? Determinants of mode choice in the Netherlands. *Transp. Res. Part A Policy Pract.* 123, 7–23.
- Ton, D., Zomer, L.-B., Schneider, F., Hoogendoorn-Lanser, S., Duives, D., Cats, O., Hoogendoorn, S., 2019b. Latent classes of daily mobility patterns: the relationship with attitudes towards modes. *Transportation (Amst.)*. <https://doi.org/10.1007/s11116-019-09975-9>.
- Vij, A., Carrel, A., Walker, J.L., 2013. Incorporating the influence of latent modal preferences on travel mode choice behavior. *Transp. Res. Part A Policy Pract.* 54, 164–178.
- Vij, A., Gorripaty, S., Walker, J.L., 2017. From trend spotting to trend'splaining: Understanding modal preference shifts in the San Francisco Bay Area. *Transp. Res. Part A Policy Pract.* 95, 238–258.
- Vovsha, P., 1997. Application of cross-nested logit model to mode choice in Tel Aviv, Israel, metropolitan area. *Transp. Res. Rec. J. Transp. Res. Board* 6–15.
- Wardman, M., Tight, M., Page, M., 2007. Factors influencing the propensity to cycle to work. *Transp. Res. Part A Policy Pract.* 41, 339–350.
- Whalen, K.E., Páez, A., Carrasco, J.A., 2013. Mode choice of university students commuting to school and the role of active travel. *J. Transp. Geogr.* 31, 132–142.
- Witlox, F., Tindemans, H., 2004. Evaluating bicycle-car transport mode competitiveness in an urban environment. An activity-based approach. *World Transp. Policy Pract.* 10.