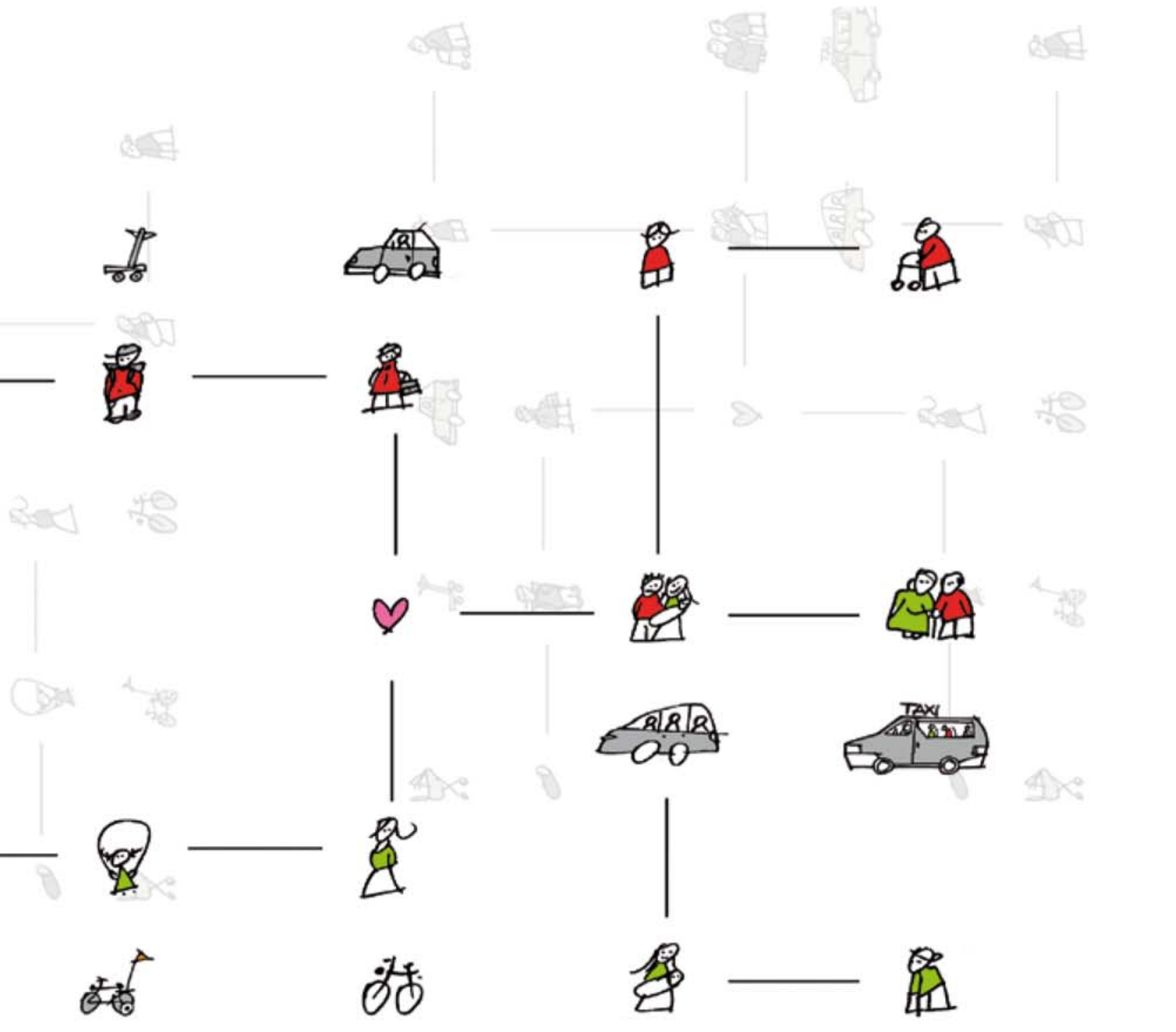
Modelling life trajectories and mode choice using Bayesian belief networks

DOI:
10.6100/IR667904

Document status and date:
Published: 01/01/2010

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)



Modelling life trajectories and mode choice using Bayesian Belief Networks

Marloes Verhoeven

/ faculty of architecture building and planning tu eindhoven

bouwstenen 143

Modelling life trajectories and transport mode choice using Bayesian Belief Networks

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van de
rector magnificus, prof.dr.ir. C.J. van Duijn, voor een
commissie aangewezen door het College voor
Promoties in het openbaar te verdedigen
op dinsdag 23 februari 2010 om 16.00 uur

door

Marloes Verhoeven

geboren te Oss

Dit proefschrift is goedgekeurd door de promotor:

prof.dr. H.J.P. Timmermans

Copromotor:
dr. T.A. Arentze

Technische Universiteit Eindhoven,
Faculteit Bouwkunde, Urban Planning Group

Cover illustration: Manon Grond
Cover design: Ton van Gennip, Tekenstudio Faculteit Bouwkunde
Printed by the Eindhoven University of Technology Press Facilities

BOUWSTENEN NR 143

ISBN 978-90-6814-625-7
NUR-code 955: Bouwkunde

Preface

This thesis is the result of the five years that I spent on my PhD research project at the Urban Planning Group of the Eindhoven University of Technology. First, I would like to give you some insight into my personal life trajectory and the life course events that I experienced during my PhD. This is related to the topic of my thesis. Next, I would like to thank everyone who has supported and motivated me during this part of my life. Without their help, contribution and support I would not have been able to complete the research project and this thesis.

My life trajectory started with my birth on 11th of April 1979 in Oss, where I lived with my parents Annie and Wim Verhoeven. In 1980 we moved to our new house <housing event> and soon after my brother Peter was born <household event>. I will not bother you with all the details and life course events during my childhood and adolescent life. Therefore, we skip to the year 2004. In this year I received my Master of Science degree at the Eindhoven University of Technology <study event>. After my graduation I had to make a decision, either leave the University and start a job in the “real world”, or stay at the University and start as a PhD student. That way I would be able to finish what I started as a Master student. Obviously, I chose the latter option <work event>. In the same year I had to hand in my student PT pass and I switched to a discount public transport (PT) pass <PT pass event>. During the year 2005 experienced several life course events. First, I moved to an apartment building in Eindhoven <housing event> and started living together with Mattijs <household event>. Next, we got our first car <car availability event> and I cancelled my discount pass for public transport <PT pass event>. In 2007, our household income increased, because Mattijs started his first job <household income event>. Mattijs and I got married in the summer of 2008 <household event> and soon after we moved into our newly bought house <housing event>. In the same year I purchased a discount pass for the public transport again <PT pass event>. In the beginning of 2009, I started a new job before this thesis was completely finished <work event>. With this short summary of my life trajectory I hope that I have given you a preview of the relevant life course events which will be discussed in this thesis.

I would like to thank a few people at the university for their support and help for making this possible. First of all, I would like to thank Professor Harry Timmermans, my promotor, for the chance he gave me to finish what I started with my Master project. It was a great honour to work with him and during our discussions he was always able to point me into the right direction. I really appreciate the opportunity he gave me to make a trip around the world during my PhD project. I also would like to thank Theo Arentze, my copromotor, for the endless support and patience. Theo inspired my during dicussions and was always there when I needed help. I really enjoyed our brainstorm sessions after which I always had plenty of new ideas and was highly motivated to explore these ideas. Harry and Theo thanks for correcting my English and for your endless help and support - without this it would not have been possible to finish my PhD research.

I would also like to thank my collegueas of the Urban Planning Group. In particular, I want to thank Peter van der Waerden for motivating me during my PhD journey. I really enjoyed our conversations and discussions during our many lunch-time walks. Sometimes we dicussed things about work, but most of the time we spoke about personal things. These discussions always came back to human (group) behaviour and decision making processes. Peter, I have got two things to say. Thanks for the good time at the University and I hope that we keep in touch or work together in the future. Of course I would like to thank “the girls” from the *secretary* for their support, but also their chats at the coffee machine. Mandy, Anja, Annemiek and Ingrid thanks for the fun during my time at the university, you made my time more pleasant. From the other Urban Planning Group staff members I would like to thanks Astrid Kemperman and Aloys Borgers for their company during the many lunch meetings at the 8th floor. I enjoyed our discussions and conversations, also on many other occasions. Leo van Veghel thanks for the help with the papers and books and I owe you an apology for the fact that I went for tea with Peter before you arrived at the university in the morning. From the DDSS group I would like to thank Joran Jessurun for writing the necessary programs for my research. You were always very understandinga and I really appreciate the help - thanks!

It is impossible to name all the other members of the Urban Planning Group. I would like to thank a few people for the interesting conversations, discussions and the good times after work: Erik, Nicole, Pauline, Anastasia, Gustavo, Oliver, Linda, Caspar and Oswald, thanks guys and girls! A special word of thanks to Nicole for reading my draft and final version and correcting my Dutch - English. I would also like to thank the initiators and active members of the PhD network: Ana, Christina, Paul, Jakob, Christian, Marija, Vincent, Daniel, William and Bart.

I would like to thanks my new bosses Bart van Hussen and Paul van Loon for their patience and for the possibility they gave me to finish my thesis during my new job. Without this opportunity I would probably still be struggling.

There are a few friends in my life who I would like to thank for their support, interest and understanding that my social life was sometimes on a break: Floortje, Hanneke, Marieke, Esther, Manon, Saskia, Loes, Tamara, Sagitta and Claudia. I would especially like to thank Manon for designing the cover of my thesis.

Besides my friends special thanks goes to my family for their interest, support and understanding: Peter and Mayke, my parents in law Eef and Nelleke, Annemarie and Ian and Janneke and Jeroen. I am very grateful to my parents, Wim and Annie, who were always there for me. They supported me during difficult moments and motivated me to finish this thesis. Thanks for making me feel proud of my research. Last but not least, I would like to thank my husband Mattijs for his everlasting support during this journey and the trust he had in me, even when I was completely broken down. Thanks for understanding and I hope we spent the rest of my life trajectory together. I love you.

Contents

List of figures

List of tables

List of equations

1 | Introduction

1 | Motivation

The so-called trip-based approach of transport demand forecasting has been frequently used in both academic and applied research in urban and transportation planning. It predicts transport demand in independent, sequential steps: traffic generation, destination choice, transport mode choice and traffic assignment. It has been criticised from different perspectives, the most important of which is the understanding that trips are derived from people's need to conduct activities.

Although seminal work on activity-based modelling can be traced back to the 1970s, the activity-based approach in urban planning and transportation research truly gained momentum in the early 1990s. This approach views travel patterns as a manifestation of the organisation of activities in time and space. Compared to previous approaches, including the trip-based approach, the activity-based approach added complexity to the modelling of transport demand by incorporating dependencies between the various choice facets making up an activity-travel pattern (transport mode, destination, departure time, etc.) at a higher spatial and temporal resolution. Such added complexity was felt crucial for better understanding and assessing the impact of urban and transport management programs and to better forecast environmental impacts of transportation. Since the early 2000s, several operational activity-based models have been developed (e.g. Albatross (Arentze and Timmermans, 2000, 2004), CEMDAP (Bhat *et al*, 2004), Famos (Pendyala, Kitamura, Kikuchi, Yamamoto, and Fujii, 2005), and the Daily Activity Model (Bowman *et al*., 2007)). The activity-based approach has become dominant in academic research.

There is now also evidence of dissemination to planning practice (e.g. Vovsha, Bradley, Bowman, 2005; Arentze, Timmermans, Jorritsma and Olde Kalter, 2008). Regardless of the progress made, operational activity-based models still have their limitations. Perhaps the most important of these is that existing models simulate the activity-travel patterns of a population for a single day. This does not only mean that some bias may be introduced in the simulations and forecasts, but also that the models do not allow simulating explicitly how individuals and households react to changes in factors influencing their organisation and implementation of activity-travel agendas and possibly adapt their activity-travel schedules. Acknowledging this limitation, the international research community has articulated the need to explore and model dynamics in activity-travel patterns along various time horizons. In that context, a distinction is made between long-term, mid-term and short-term dynamics. Long-term dynamics refer to events, such as moving house and changing jobs that may have a long-term impact on and involve a dramatic change in particular aspects of activity-travel patterns. In contrast, short-term dynamics relate to non-structural shifts in planned activity-travel schedules due to unforeseen events. Mid-term dynamics are in-between these two extremes and relate to incremental adaptations of activity-travel patterns.

This PhD thesis contributes to this emerging, but still scarce literature. The focus of attention is on long-term dynamics. In particular, it will be analysed whether life course events are associated with changes in activity-travel patterns and how these dynamics can be modelled, using changes in transport mode choice as an example.

2 | Outline

To that end, the thesis is divided into different parts. In the next chapter a brief overview of existing models of transport demand will be provided. It should be emphasized from the very beginning that this chapter is not meant to be a comprehensive, detailed review of transport choice models. Rather, this chapter will discuss in some detail a selection of previous research that has direct relevance to the problems that are addressed in this thesis.

Chapter three then continues by developing the conceptual framework underlying this study. This framework consists of two main components. First, it conceptualizes the factors influencing life trajectories. Second, it depicts the influence of life trajectories on behaviour, in particular transport mode choice.

Chapter four motivates the approach that is used to model these complex dynamics between life course events and transport mode choice decisions. The approach is based on Bayesian Belief Networks (BBN), a modelling approach that allows estimating and representing the direct and indirect influences between a set of categorical variables. The chapter describes the key principles underlying Bayesian Belief Networks.

The different techniques for data collection are described in the fifth chapter. An Internet-based survey was used to collect data on current behaviour and past events. Information about events was collected using a retrospective survey. About 700 respondents participated in the survey. The procedure for sampling respondents, details about the Internet-based survey, and sample characteristics are also discussed in this chapter.

Chapter six discusses the results of the analyses and model estimations. First, in order to examine whether there is evidence of time effects of occurrences of events, a multinomial logit model with time as an explanatory variable is estimated. Because the results supported the basic assumptions of time-related effects, next the two Bayesian networks are extracted from the data using structural and parameter learning algorithms. The life trajectory network captures the relations between the life course events, current states and the history of life course events, while the mode choice network considers the link of mode choice with life course events and the states.

The goodness-of-fit of the learned Bayesian Belief Networks is discussed in chapter seven. This chapter also gives an overview of validation tests of the learned Bayesian Belief Networks. The predicted life trajectories are compared with the observed life trajectories based on four criteria to assess whether the structural characteristics of the life trajectories are predicted correctly. The modal split (Car, Public Transport and Slow Transport) of the predicted mode choice is compared with the observed mode choice.

Chapter eight shows how the learned networks can be used in a micro-simulation to simulate the interdependencies between life course events and their impact on transport mode choice. A scenario is described to illustrate the simulation of life trajectories and mode choice. This will give further insight in the dynamics of the learned network.

The final chapter of this thesis discusses insights gained by this project, reflects on limitations and provides recommendations for further research.

2 | Literature

1 | Introduction

As indicated in the introduction, this study will examine and model transport mode choice in the context of life trajectories. To appreciate the contribution of this thesis and to discuss its foundations, this chapter will give a brief overview of existing transport choice models and research on life course events.

Previous research on transport mode choice can be classified into three more or less separate lines of research: (1) traditional transport mode choice models; (2) activity-based models; and (3) more comprehensive models, like dynamic activity-based models. Traditionally, transport mode choice was primarily examined as a stand-alone problem. Given the purpose and destination, the choice of transport mode was modelled as a function of the various attributes of the transport mode alternatives.

Later, when the activity-based approach became increasingly popular in urban planning and transportation research, transport mode choice decisions were modelled as part of more comprehensive models. Most of these models were cross-sectional in nature. More recently, some authors, in an attempt to build dynamic activity-based models or in better understanding behavioural change, have investigated changes in transport mode choice decisions (Dargay and Vythoulkas, 1999; Mohammadian and Miller, 2003)

In this discussion, it will be argued that a life course perspective offers potential advantages in understanding and modelling activity-travel decisions, including mode choice. In this chapter, some examples of a life course approach applied in domains other than transportation will therefore also be summarised and discussed. These examples serve to illustrate the potential of the life course perspective.

The chapter is organised as follows. First, some examples of stand-alone transport mode choice models will be discussed. This is followed by a brief description of the development of activity-based modelling, including an indication how transport mode choice decisions were modelled in some of the best-known activity-based models of transport demand. The fourth section then discusses previous research on transport mode decisions and change, triggered by particular life course events. After discussing the transportation research literature, some examples of a life course approach in other domains will be discussed. The chapter is completed by drawing some conclusions for the design of this study.

2 | Transport mode choice models

Many different models of transport mode choice have been developed in the past; attitudinal models (Fishbein and Ajzen, 1975) and random utility models (Ben-Akiva and Lerman, 1985) being the most commonly used approaches. In these models, mode choice is typically conceptualised as a function of the characteristics of alternative travel modes and a set of personal and household characteristics. Previous studies, based on the latter two approaches, assumed that these attributes generate some utility and that individuals maximize their utility when choosing between alternative transport modes, subject to budget constraints. Attitudinal models are an exception in that they do not involve maximizing utility, but rather assume that transport mode choice is based on a set of attitudes.

1 | Attitude models

The best-known attitude model in transportation research is the Fishbein and Ajzen model (1975). The model predicts behavioural intentions, but often it is assumed that travellers act on these intentions. The basic structure of the attitude model is represented in Figure 2 | 1 with three boxes (1) evaluative beliefs, (2) attitudes toward travel alternatives and (3) behavioural intentions.

The input of the model consists of evaluative beliefs. These beliefs are related to the existing choice alternatives known by the respondent. Evaluative beliefs are usually measured on point scales. A linear additive combination rule is used in the model to combine the separate evaluative beliefs into attitude scores using weights. Next, a choice rule is applied to select one alternative from the choice set. For example, the alternative with the highest overall attitude score is chosen. The model produces subjective output which is closely related to behaviour.

In the original attitude model (Fishbein and Ajzen, 1975) a person's attitude toward any object *j* is a function of his beliefs about the object and the evaluation of those beliefs. The expectancy-value formulation can be expressed as follows:

$$A_j = \sum_{k=1}^{K} b_{jk} e_{jk} \qquad \text{Equation 2 | 1}$$

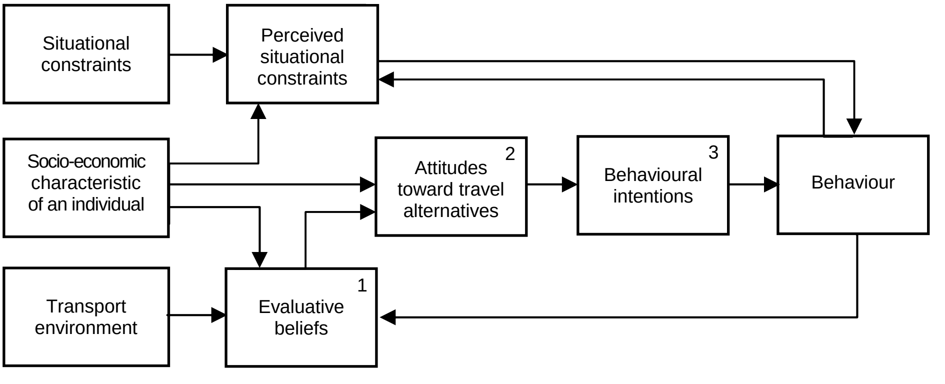


Figure 2 | 1: Psychological framework for individual's mode choice (after Golob, 1980)

where

A_j = attitude toward object *j*

b_{jk} = belief *k* about *j*

e_{jk} = evaluation of belief *k* about *j*

K = the number of beliefs

Attitudes, beliefs and evaluations are usually measured using rating scales. First, belief strength is assessed by means of a 7-point scale (e.g., *likely-unlikely)*. Respondents are asked to indicate how likely it is that an alternative possesses the characteristic. Next, respondents are asked to evaluate the attributes/characteristics, using a 7-point evaluative scale (e.g., *good-bad).* It is uncertain whether these scales should be scored in a *unipolar* fashion (e.g., from 1 to 7, or from 0 to 6) or in a *bipolar* fashion (e.g., from -3 to + 3). It is best to use equal-interval measures. In that case it is permissible to apply any linear transformation to the respondents' ratings without altering the measure's scale properties.

In the theory of reasoned action (Fishbein and Ajzen, 1975; Ajzen and Fishbein, 1980) the original model of attitude (A) was extended with behavioural intention (BI), and subjective norm (SN). It is assumed that an intention to perform a behaviour (*I*) is related to the attitude toward performing the behaviour (*A*) and the subjective norm for performing the behaviour (*SN*). The relationship is specified by the equation:

$$I = w_A A + w_{SN} SN$$ Equation 2 | 2

where the ws are weights determined empirically by means of linear regression.

Subjective norms indicate whether the behaviour is approved by important others (parents, partner, friends, authorities, etc.). The subjective norm (*SN*) is calculated as the sum across referents of the multiplication of the strength of each normative belief (N) approved by referent *r* and the person's motivation (M) to comply with the referent *r.* Thus:

$$SN = \sum_r N_r M_r$$ Equation 2 | 3

The more social pressure a person experiences to perform the behaviour, the higher *SN* is. To measure SN, respondents rate, with respect to each referent, the degree to which the referent would approve or disapprove a given behaviour using a 7-point scale. The respondents also rate how much they care whether the referent approves or disapproves their behaviour.

Ajzen (1985, 1991) revised and extended the theory of reasoned action into the theory of planned behaviour. “This extension involves the addition of one major predictor, perceived behavioural control, to the model. This addition was made to account for times when people have the intention of carrying out a particular behaviour, but the actual behaviour is thwarted because they lack confidence or control over behaviour” (Miller, 2005, p. 127). Three concepts of the theory of planned behaviour are described here: (1) attitude toward behaviour, (2) subjective norms, and (3) degree of perceived behavioural control. A general rule is: the more favourable the attitude and subjective norm with respect to a behaviour, and the greater the perceived behavioural control, the stronger an individual’s intention to perform the behaviour under consideration (Ajzen, 1991). The relative importance of these three concepts in the prediction of intention is expected to vary across behaviours and situations. Sometimes only attitudes may have a significant impact on intentions. In other situations attitudes and perceived control explain intentions, while in other applications all three predictors make independent contributions. The theory of planned behaviour is expressed in the following equation:

$$I = w_A A + w_{SN} SN + w_{PBC} PBC \qquad \text{Equation 2 | 4}$$

Control beliefs are added to the set of beliefs which, according to the theory of planned behaviour, determine intention and action. Control beliefs may be based in part on past experience with the behaviour or is influenced by second-hand information about the behaviour. For example by experiences of friends and relatives, and by other factors that increase or reduce the perceived difficulty of performing the behaviour in question (Ajzen, 1991). The perception of behavioural control (PBC) is calculated given equation 2 | 5. Each control belief *(c)* is multiplied by the perceived power (*p*) of the particular control factor to facilitate or inhibit performance of the behaviour. Beliefs about resources and opportunities are viewed as underlying perceived behavioural control. The inclusion of perceived behavioural control has been found to increase the accuracy of predicting behaviour not under volitional control (e.g., Fredericks and Dosett, 1983; Schifter and Ajzen, 1985; Ajzen and Madden, 1986; Netemeyer, Burton, and Johnston, 1991; Gärling, 1992). The perception of behavioural control (PCB) is calculated as follows:

$$PBC = \sum_k p_k c_k \qquad \text{Equation 2 | 5}$$

where:

c_k = control belief *k*

p_k = perceived power of particular control factor about belief *k*

The theory of planned behaviour provides a useful framework for dealing with the complexities of human behaviour. The expectancy-value formulation (attitude model) is not able to adequately describe the process of individual beliefs and produce the global response. Alternative models were developed to describe (1) the relations between beliefs and (2) the global constructs.

A representative, recent example of the application of attitude theory to transport mode choice decision is Wall, Devine-Wright and Mill (2007). Their focus is on drivers' motivations for switching travel modes. Multiple scales were used to measure attitudes with respect to transport modes. Principal components analysis was used to extract the underlying dimensions. This resulted in a five-factor solution with factors representing two norm activating constructs and three planned behaviour constructs. Behavioural intention was measured as the intention to maintain or reduce car use. The model linking behavioural intention to attitude toward the behaviour, subjective norms and perceived behavioural control was estimated using logistic regression. The model performed reasonable well.

The strengths and relevance of these attitudinal models are related to those choices where especially social norms play an important role. Compared to other modelling approaches, the measurement of attitudes and the estimation of these models, ignoring attitudinal differences, is generally weak.

2 | Random utility models

Random utility models are based on the assumption that choice behaviour is the outcome of a decision process in which individuals maximize their utility (McFadden, 1978; Ben-Akiva and Lerman, 1985)). Utility is assumed to consist of a deterministic part and an error component described in the following equation:

$$U_{ij} = V_{ij} + \varepsilon_{ij}$$ Equation 2 | 6

where:

U_{ij} = the utility of alternative *j* evaluated by individual *i*

V_{ij} = observable utility part from the researcher's perspective

ε_{ij} = random utility part (non observable by the researcher)

Similar to attitude models, often a linear function is assumed to capture utility:

$$V_{ij} = \sum_k \beta_k X_{ijk}$$ Equation 2 | 7

where:

β_k = estimated weight parameters

X_{ijk} = explanatory attributes (variables) of alternative *j* perceived by individual *i*

Different model specifications can be derived based on the assumptions about the distribution of these error terms. Most studies have applied a multinomial logit model, which can be derived by assuming that the error terms are independently and identically Gumbel distributed. (The model can also be derived from other theories, but that is beyond the current discussion.)

The multinomial choice model is given according to the following equation:

$$P_{ij} = \exp(V_{ij}) / \sum_{j'=1}^{J} \exp(V_{ij'})$$ Equation 2 | 8

where P_{ij} is the probability that alternative *j* is chosen by individual *i*, and V_{ij} (*j* =1, …., *J*) is the systematic component of the utility of alternative *j* to individual *i*. For each alternative *j*, V_{ij} is assumed to be a linear function of appropriate explanatory variables. Thus:

Equation 2 | 9

$$V_{ij} = \sum_{k=1}^{K} \beta_k X_{ijk}$$

where:

K = the number of explanatory variables

X_{ijk} = the value of the *k*-th explanatory variable for alternative *j* and individual *i*

β_k = a coefficient or parameter of explanatory variable *k*.

Traditionally, these models (equation 2 | 7) were estimated on the basis of revealed choice behaviour (Louviere, 1988). However, because revealed choice behaviour may not necessarily reflect underlying preferences, in the 1970s conjoint measurement models were developed. These models were also an answer to the weakness of the attitudinal models in terms of measurement. Conjoint measurement models are based on stated preferences or choice of respondents for hypothetical choice alternatives. First, the set of attributes influencing choice behaviour is selected and each attribute is defined in terms of attribute levels. Next, attributes levels are combined into profiles according to the principles of experimental design, and respondents are asked to rate the profiles on some preference scale or choose from a series of constructed choice sets the one they like best.

A recent example of the multinomial logit model for transport mode choice is Yagi and Mohammedian (2007). The utility function was specified in terms of attributes related to the travel (travel cost, travel time, and travel distance), household-related variables (household income, and vehicle ownership) and individual variables (employment status (e.g., full-time, part-time, and student), school type, personal income, gender, age, vehicle availability, work/school location, and various types of commuting allowance provided by the employer). In addition, some composite variables such as travel cost divided by the household income were chosen. The model performed satisfactorily.

In the standard applications of these models, it is assumed that purpose and destination are given. However, it is well-known that destination and transport mode choice are strongly interrelated. If the distance is beyond some threshold, the probability of choosing slow modes is dramatically reduced. Some authors have therefore formulated more advanced (e.g. nested logit) models to capture this interdependency. Examples include sequential choice model (Fujii, Kitamura and Monma, 1998; Borgers, Timmermans and van der Waerden, 2002), linked model methodology (Wen and Koppelman, 2000), flexible frameworks where decision structures are estimated simultaneously with the utility functions of choice alternatives (Train, 2003), and the co-evolutionary logit model (Krygsman, Arentze and Timmermans, 2007). In the next section, more comprehensive models of transport are described.

3 | Modelling transport mode choice in activity-based models

The models discussed in the previous section were also used as a component of more comprehensive models of transport demand, predicting not only transport mode choice, but also destination and route choice. The so-called four-step modelling approach has been dominant in this field of study and is still in practice. In the first stage, trips are generated as a function of land use and household characteristics. Second, destination choice is modelled to predict where the trip will terminate. In the third stage, given the destination, mode choice for a specific trip is predicted with a transport mode model. These three separate steps generate an origin-destination table, specifying the number of trips between a set of origins and a set of destinations. In the last step, these trips are assigned to the network using some route assignment algorithm. Because these models are independent, in principle any model of transport mode choice can be used in this four-step process.

Over the years, criticism about this approach increased in academic research and gradually this led to the development of so-called activity-based models of transport demand. The limitations of the four-step approach may be briefly summarised as follows (quoted from McNally and Rindt, 2008, p. 58):

1. Ignorance of travel as a demand derived from activity participation decisions.
2. A focus on individual trips, ignoring the spatial and temporal interrelationship between all trips and activities comprising an individual's activity pattern.
3. Misrepresentation of overall behaviour as an outcome of a true choice process, rather than as defined by a range of complex constraints which delimit choice.
4. Inadequate specification of the interrelationships between travel and activity participation and scheduling, including activity linkages and interpersonal constraints.
5. Misspecification of individual choice sets, resulting from the inability to establish choice alternatives available to the decision maker in a constrained environment.
6. The construction of models based strictly on the concept of utility maximization, neglecting substantial evidence relative to alternate decision strategies involving household dynamics, information levels, choice complexity, discontinuous specifications and habit formation.

Not mentioned before, but equally important, was the lack of interdependencies between transport mode choice and other decisions underlying the organisation of activities of individuals and households in time and space. The attempts of combined modelling of destination and mode choice is a step in this direction but other mechanisms such as car allocation in car-deficient households, and the complexity of the activity schedule are equally important.

"The activity approach began as a natural evolution of research of human behaviour, in general, and travel behaviour, in particular." (McNally and Rindt, 2008, p. 59). The fundamental idea of the activity approach is that travel decisions are driven by a collection of activities that form an agenda for participation. This means that travel decisions cannot be analysed on an individual trip basis. Specific travel decisions and the choice process associated with travel decisions can be understood and modelled only in the context of the entire agenda. "The collection of activities and trips actually performed comprise an individual's activity pattern, and the decision processes, behavioural rules, and the environment in which they are valid, which together constrain the formation of these patterns, characterize complex travel behaviour." (McNally and Rindt, 2008, p. 59). Over the years, several different modelling approaches have been suggested. Arentze and Timmermans (2002) classify these into constraints-based models, utility maximizing models, computational process models and micro-simulation models. These four modelling approaches are briefly discussed.

1 | Constraint-based models

Constraint-based models examine whether particular activity patterns can be realised within a specific time-space environment. As input these models require activity programs: a set of activities of certain duration which are performed at certain times. Different models have been developed in the past, like PESASP (Lenntorp, 1976), CARLA (Jones, Dix, Clarke, and Heggie, 1983), BSP (Huigen, 1986), and MASTIC (Dijst, 1995). These constraints-based models are not able to predict adjustment behaviour of individuals. Individuals are likely to change or adjust their activities when they are faced with a changing time-space environment. Transport mode choice decisions are not explicitly measured in this approach, but rather serve as input to assess the feasibility of activity-travel patterns, given a particular transport mode of a combination of different modes, and given a set of constraints.

2 | Utility maximizing models

Following the popularity of *utility maximizing* theory, discrete choice models were extended to include multiple choice facets. These models therefore represent a second approach in the development and application of activity-based models of transport demand. Utility maximizing theory is based on the assumption that choice alternatives can be represented as bundles of attribute values. The part-worth utilities are combined into some overall measure of utility according to a simple mathematical rule, such as a linear additive rule described in section 2 | 2 and 2 | 3. Often, the multinomial logit model is used. This model has a well-known limitation: the so-called interdependence from irrelevant alternative property, which states that the odds of choosing a particular alternative over another are independent of the size and composition of the choice set. It implies that the introduction of a new alternative will extract market share from the existing alternatives in direct proportional of their utility. In reality, however, one would expect that similar choice alternatives compete more with each other than dissimilar alternatives. The nested-logit model is a solution for this problem. Nested-logit models require grouping of similar choice alternatives in nests. Choice probabilities are predicted conditionally on the next higher nest. The best known activity-based model is the daily activity model (Ben-Akiva, Bowman and Gopinath, 1996). Bowman developed a prototype for the Boston area (Bowman, 1995) and implemented it in Portland (Bowman, Bradley, Shiftan, Lawton, and Ben-Akiva, 1998). An overview is given in Bowman and Ben-Akiva (1999). Several similar models however have been suggested, such as the HCG model (Ettema, Daly, de Jong and Kroes, 1997), PETRA (Fosgerau, 1998), COBRA (Wang and Timmermans, 2000), and Tel-Aviv Metropolitan Area model (Shiftan, Kaplan, and Hakkert, 2003). Slightly different, but also based on principles of utility maximization is Prism-Constrained Activity Travel Simulator (PCATS) and PCASTS-RUM (Kitamura and Fujii, 1998) and its variants such as FAMOS (Pendyala, Kitamura and Kikuchi, 2004; Pendyala *et al.*, 2005).

To illustrate how transport mode choice is modelled in these models, the daily activity schedule model is used as an example. The individual's demand for activity and travel is represented as a multidimensional choice in the daily activity schedule. This means all the combinations of activity and travel that an individual might choose during the day are listed. The daily activity pattern is based on tours, which are organised in schedules. The following parts can be distinguished in the daily activity pattern: (1) a primary activity, (2) the type of tour for the day's primary activity (including number, purpose, and sequence of stops), and (3) the number and purpose of secondary tours. The tour schedule consists of choices of destinations for activities, mode and timing of the travel. The number of secondary tours is determined by the choice of the daily activity pattern. Destination and mode of the secondary tours are conditioned upon the choice of a daily activity pattern. Choice of mode is modelled for the tour in the destination and mode choice model, instead of the usual choice of mode for a trip.

3 | Computational process models

Utility maximizing models have been criticized by some scholars who argued that individuals do not necessarily choose the alternative that generates the highest utility. Moreover, utilities are not invariant as implicitly assumed in the above models.

Individuals rather apply heuristics that may be context-dependent. These heuristics can be represented as “if...then...else” rules, which specify which decision will be made under a set of conditions. *Rule-based models,* sometimes also referred to as c*omputational process models*, conceptualize choices as outcome of heuristics. A large number of rules represent context-specific behaviour in these models. This often generates a black box feeling. An advantage of these models is the flexibility in defining complex interdependencies among facets of activity-travel patterns and other facets. Examples of computational process models are SCHEDULER (Gärling, Brännäs, Garvill, Golledge, Gopal, Holm and Lindberg, 1989), which however was never operationalised, SMASH (Ettema, Borgers and Timmermans, 1995), ALBATROSS (Arentze and Timmermans, 2000) and TASHA (Miller and Roorda, 2003).

Using ALBATROSS as an example, transport mode choice was modelled as follows. The schedule engine controls a sequence of steps, which intends to simulate the way individuals solve the problem or organizing their activities and associated travel in time and space. In each step, the schedule engine indentifies the condition information required for making principal scheduling decisions. Appropriate calls are sent to agents for the required analyses and the obtained information is passed on to the rule-based system, which translates returned decisions into appropriate operations on the current schedule. An initial schedule is derived based on the activity programme in terms of activities that need to be performed that day. Scheduled activities can be a result of long–term commitments, household constraints and other pre-scheduling decisions. Activities are selected and added to the skeleton as fixed activities. Next, the schedule position and profile are determined for each added activity.

The mode choice for primary, out-of-home-work activities is considered first in the decision sequence of the scheduling process. The sequence consists of six steps in total. In the first step is determined which person can use the car for that specific day. The choice set consists of the following options: car driver, car passenger, public transport (such as bus, train and taxi) and slow transport (walk and bike). The mode options public transport, slow transport and car passenger are always available for the system’s choice. The availability of the car depends on the presence of a car in the household and possession of a driver’s licence. Characteristics of the partner are included. This means that the system is able to consider implications of the choice in terms of who is going to use the car for which activity, in cases where there is only one car and more driver's licences. The next decision step handles the selection of activities, travel party and duration. Choice of time of the day is decided in the third step, and choice of trip-chaining in the next step. The fifth step in the decision sequence is the location choice. Choice of transport mode for each trip in the chain is last step in the decision sequence. In this step it is assumed that transport mode decisions are made at the trip-chain level, instead of the trip level. Two types of trip-chains are distinguished. The first one includes a primary work activity and the second one does not include a work activity, but includes other activities. Mode choice for a primary work activity, assigned in the first step of the decision sequence, is used as predicted mode for the other activities in the first trip-chain. For the second type of trip-chain the mode choice is considered in this step of the process. The same mode options as in the first step are available here. The availability of the option car driver is evaluated based on the following characteristics: driver's licence, number of cars in the household, and the use of the car by the partner. Only the primary work activity of the partner is taken into account, the other activities are not known at this stage. Mode choice is modelled in ALBATROSS separate for the home-work trips and for the other trips in the chain. Different constraints are taken into account when the decision for mode choice is taken.

4 | Micro simulation models

Although all of the above models may involve simulating the behaviour of individual travellers, in addition to these models which are based on certain theoretical concepts, other *micro simulation models* are more data-driven models. Examples include ORIENT (Sparmann, 1980), VISEM and PTV VISION (Fellendorf, Haupt, Heidl and Scherr, 1997), RAMBLAS (Veldhuisen, Kapoen and Timmermans, 2000), TRANSIMS (Wagner and Nagel, 1999) and MATSim (Balmer, Meister, Rieser, Nagel and Axhausen, 2008). The Transport Analysis and Simulation System (TRANSIMS) was the best-known micro simulation model. The underlying concepts and ideas have been transformed into Multi-Agent Transport Simulation (MATSim). Functionalities of activity-based travel demand generation, mode choice and route assignment and micro simulations are combined. The MATSim approach is iterative and the iterative approach is developed into an extension of the assignment procedure: The route adaptation process is extended towards other choice dimensions, such as time choice, mode choice, location choice. The modelling of transport mode choice in these models varies considerably, but the fundamental principles are not different from the treatment in the other modelling approaches.

This brief literature review shows that activity-based models typically simulate the activity-travel patterns of a population for a single day. Long-term decisions of individuals and households are not taken into account. The models are not able to simulate how individuals and households react to changes, for example, changes in their life. Traditional models are static and estimate behaviour in an equilibrium situation. In reality, behaviour may not be static but always in motion toward an equilibrium situation. These long-term dynamics are studied and modelled in this study.

4 | Transport mode choice and life course events

All currently fully operational activity-based models of transport demand are cross-sectional in nature. Future behaviour is predicted based on the relationships established at one point in time. Because this assumption obviously has some limitations, the development of dynamic activity-based models is one of the current research frontiers in transportation research. Arentze and Timmermans (2007, 2008) summarise recent development in modelling dynamics along various time horizons.

Zimmerman (1982) stressed the need to use the life cycle concept and its relation to household travel in travel research. This way the manner in which individuals and households live over time can be captured, and the question how in each life cycle stage their concerns are expressed in travel can be addressed. The theoretical framework of mobility biographies is also based on a life course approach (Salomon, 1983). Note that the words life cycle and life course are used interchangeable. Mobility biography refers to the total of an individual's longitudinal trajectories in the mobility domain. Salomon (1983) distinguished three domains: life style, accessibility, and mobility. The life style domain consists of three careers: demographic, professional and leisure career. Employment location, residential location, leisure and other locations careers are part of the accessibility domain. Four careers were distinguished in the mobility domain: car ownership, season ticket, holiday travel and daily travel careers. Events in these trajectories are assumed to have an impact on daily travel patterns, car ownership and other mobility characteristics. Salomon (1983) associated such events in certain life domains with daily mobility behaviour first. Salomon's model was extended by Lanzendorf (2003), who transferred the life course approach to individual travel decisions.

Part of this line of research is concerned with behavioural change. Several studies have addressed the questions whether changes in job location or house will trigger changes in activity-travel patterns. Residence and places of education and employment play an important role in the long-term and mid-term mobility of people (Beige, 2008). In line with this notion, several studies have focused on 'critical' or 'stressful' events. Van der Waerden and Timmermans (2003; van der Waerden, Borgers and Timmermans, 2003) argued that key events and critical incidents may be useful in better understanding the dynamics of travel decisions and resources. Key events were defined as planned events, like marriage, relocation etc. and critical incidents are referred to as unplanned events such as policies or accidents. In their studies, only key events were studied, which are referred to as life course events in this thesis. A life course event is defined in this study as a major event in a person's life such as marriage or a move that may trigger a process of reconsideration of current behaviour. Some events, such as a change in the place of residence, may dramatically change the space-time context within which travel decisions have to be made. Other life course events, such as a change in car availability, may reduce constraints and expand an individual's choice set. Moving house implies a shift in characteristics such as accessibility, distance/travel time relationships and perhaps also the utility an individual derives from alternative travel modes. A life course event such as changing jobs may also lead to changes in characteristics of travel modes. A final example is the birth or adoption of a child, which may induce new activities (e.g. day care) that are more difficult to complete using the currently used travel mode.

A number of significant key events have been described, such as acquisition of a driver's licence, residential relocation and job change (van der Waerden and Timmermans, 2003; van der Waerden *et al.*, 2003; Klöckner, 2004). In different studies the influence of one particular event, career or resource has been examined. For example, Lanzendorf studied the influence of child birth on mobility biographies (Lanzendorf, 2006), while Prillwitz and Lanzendorf (2006a) examined the impact of life course events on car ownership. Lanzendorf (2006) assumed that the maintenance tasks in the household needs to be rearranged after the birth of the first child. He was looking for typical patterns of change around this specific key event and the effects on the mobility biography in a long-term perspective. The analyses were based on qualitative retrospective interviews with 20 parents with young children. The study resulted in three conclusions. First, some typical patterns of change were found, although there was no clear indication of increased or decreased car use. The car ownership of mothers increased after child birth in comparison with time. Second, car use of mothers with more children increased, when the before situation (before the birth of the first child) is compared with the after situation (all children were older than one year old). Third, there was no difference found in the impact of the first or second child on travel patterns. No evidence was found in the data that people kept their travel behaviour with the first child more frequently than with the second child.

Prillwitz and Lanzendorf (2006a) studied the influence of four key events in a person's or household's life on car ownership (and ultimately travel behaviour). The four key events that were analysed were: (1) changing number of adults in the household, (2) birth of a first child, (3) changing weighted monthly income, and (4) residential move. The German Socio Economic Panel was used for binomial probit analysis. Empirical results suggest a strong influence of the four key events on car ownership growth. Also the household status variables age, number of cars per household and weighted monthly income had a strong impact according to both analyses. Residential relocation only showed a limited effect. Interactions between residential relocation and other key events might be relevant for travel behaviour and car ownership and they will study this in the future.

Stanbridge, Lyons and Farthing (2004) studied the effect of residential move on people's travel behaviour, in particular mode choice. The authors tried to better understand the experiential aspects of residential relocation. Their goal was to reveal the behavioural processes that took place. A set of qualitative interviews with recent home movers, 11 in total, was used. In many instances, people are consciously considering the travel mode implications during the course of the moving home. The study reported that some people change travel modes for particular journey purposes after the residential relocation.

In addition to this mainly qualitative, analytical work, attempts of modelling the impact of life course events on travel choice decisions are rare and to our knowledge, did not exist in transportation research at the start of this PhD project. More recently, Beige (2008) studied long-term and mid-term mobility decisions during the life course using Hazard models. A retrospective survey, which covered 20 years (1985 – 2004), was used for analyses over time and over the life course. Beige distinguished residential and occupational behaviour on the one hand and ownership of mobility tools on the other hand. The aim of her study was to explore the interrelationships between the two aspects of long-term and mid-term mobility, taking the personal and familial situation into account and how corresponding events affect long-term and mid-term mobility. Analyses over time and over the life course, as well as various durations and occurring changes were carried out. Event history analyses were applied to the retrospective data. Beige found that the ownership of mobility tools was relatively stable over time. Only a small percentage of the respondents (3%) varied mobility tool ownership every year. More respondents acquired a car during the observed period than abandoned one. Changes in residence, education and employment occurred more frequently in this period. She also concluded that changes in mobility tools and spatial changes are interconnected. A strong relationship between long-term and mid-term mobility was found in this study. Ownership of various mobility tools both influences and is influenced by residential mobility. Changes concerning locations (residence, education and employment) took place more frequently than changes in mobility tool ownership. Beige supports the statement that mobility tool ownership can be used as a proxy for the actual behaviour (Simma and Axhausen, 2003; Prillwitz and Lanzendorf, 2006b). Actual travel behaviour seems to be reconsidered and altered as spatial changes take place. No clear statements were made about causal relations between the various aspects of long-term and mid-term mobility behaviour as well as the influence of other life dimensions (personal and familial events). The impact of one event to another event cannot be automatically deduced from the chronological order. Individuals sometimes anticipate (future) changes. Age, gender, occupation, income, personal situation and familial situation are the most important influencing variables that Beige found for the long-term and mid-term mobility decision. Costs were not taken into account as explanatory variables. Beige suggests that travel behaviour can be influenced by the occurrence of key or life events. Habits and routines are broken or weakened at the time of an event. Individuals reconsider their behaviour and consciously reflect on their decisions.

Her study is conceptually very similar to the underpinnings of this study. The main difference is that the mobility tools will not be examined in as much detail. In addition, a different modelling approach will be applied. Hazard models often are used to examine the dynamics of one event and therefore are less appropriate to model a network of changes and how the impact of a single event is disseminated through the network. Beige used competing risk analysis in her study, where multiple events of different types are taken into account. However, this method does not allow indirect influence which can be modelled in Bayesian Belief Networks.

5 | Life course approach in other domains

Although the life course approach is relatively new in transportation research, it has a longer and stronger tradition in other domains. The life course approach has been developed since the early 1980s (Hareven, 1977; Elder, 1985; Willekens, 1991). The basic idea underlying this approach is that each human life history is a meaningful succession of individual life events within a specific historical and social time (Feijten, 2005). The importance of taken into account additional characteristics such as timing and order of events and the duration of the resulting state, besides the occurrence of a life event is stressed in the life course approach (Giele and Elder, 1998). If both current circumstances and past experiences are considered, this will lead to a better understanding. Earlier life transitions may have a cumulative effect on later life (Dykstra and Van Wissen, 1999).

At the micro-level, life events alter preferences and needs. The resources and restrictions of a household determine to what extent it can realise its preferences. At the macro-level, economic, social-cultural, and market circumstances determine the opportunities and constraints that influence the choice set of individuals. A disadvantage of macro or societal approaches is that these models do not permit translation into individual behaviour without the danger of ecological fallacy, while micro or individual approaches do not include contextual explanations of behaviour.

Mulder (1993) made four assumptions which were necessary to make the combination of a life course and cohort perspective a sensible, useful way of studying the behaviour of individuals. The first assumption is that individuals have goals in life. The goals are not specific. An example is the set of hierarchically ordered needs from

Maslow (1954). General goals become specific goals, which are also called preferences. General goals are assumed to be universal, while preferences may vary between individuals and during an individual's life course. The relationship between people's behaviour and their preferences is the basis for the second assumption. It is assumed that people behave rationally, in the sense of satisfying behaviour. On the one hand, people do not behave very differently from other people in their societal context for two reasons: they need social approval and they adopt procedures that society makes familiar to them. Individuals are capable of shaping their own procedures and re-defining their own preferences. Third, it is assumed that people's past behaviour determines and conditions future behaviour. It is also assumed that people sometimes act and think with a long-term perspective in mind. People try to shape their lives along reasonable consistent paths. Those paths are referred to as careers. Different careers can be distinguished: residential career, household career, occupational career, educational career etc.. A life course is defined as an individual's complex system of careers. The last assumption is about societal change. These four assumptions about human behaviour imply that people influence and are influenced by society through their preference formulation and behaviour. The shared (macro-level) societal context defines similar opportunity structures and common social norms concerning behaviour and careers for people within one birth cohort.

"The concept of life course is defined as the way in which an individual progresses through various stages or statuses in various careers in life without the normative connotations often associates with the concept of life cycle" (Mulder, 1993, p. 23). Elder (1985) defines life trajectories (similar to career), transitions and events as central concepts in the life course. The individual life course is composed of multiple, interdependent trajectories. According to Mayer and Tuma (1990) the concept of life course refers to the way in which social institutions shape and institutionalize individual lives in the interconnected domains of education, family and work. The concepts of timing, sequencing, duration and spacing are used to describe the life events, transitions and trajectories (Hagestad and Neugarten, 1985).

Mulder (1993) defines two types of dependence: event dependence and state dependence. Event dependence refers to the effect of the occurrence of an event in a parallel career such as changing jobs, marriage or divorce. The (long-term) effect of occupying a certain state in a career is denoted as state dependence. Careers have sometimes a causal relation with each other, for example an event in the household career may force someone to move and this move may in turn result in the abandonment of a job. Events in one career often take place at the same point in time as the event in the other career. The same point in time is not defined here and there may be a time lag in between the two related events. Individuals can react or anticipate to changes. Time ordering does not necessarily reflect causal ordering (Willekens, 1991). Willekens (1987) observed transitional periods in the life course, in which many decisions and changes occur. These periods are typically concentrated around adolescence and young adulthood. Parallel careers provide triggers for events in other careers and produce individual resources and constraints. Resources and constraints are career specific and preferences are individual specific, but they can vary during the life course.

It is necessary to observe a longer period over time to get the full picture of all possible effects. For instance, a triggering event may not be synchronized with the event and effects of live events on the situation may be either temporary or lasting over the life course. Feijten (2005) distinguished two time dimensions: macro-level (calendar time and historical time) and micro-level (individual time and timing of a life event as age or position within an order). Life events results in states, for example the birth of the first child of a married couple results in a household of three persons. There are disrupting events, like divorce or widowhood and unemployment, which often occur unintentionally and there are planned events, such as marriage, relocation etc. The duration of the resulting state can be short (temporary) or long (lasting). A state that lasts for a long time creates opportunities for making long-term investments that pay off in the long run.

The influence of personal and collective (cohort) past experiences on present behaviour has been studied by Mayer (1986). A commitment is seen as a choice in life that ties you to a situation for a long period and you can only leave this situation at high costs. On the other hand, an advantage of commitments is that it leads to stability. Many people prefer stability in their lives. Commitments in one career can also bring stability in another parallel career. This way, commitments can be seen as long-term investments that pay off in the long run. People who do not make commitments remain initially flexible and are able to react to attractive opportunities in several careers. At a certain moment, such people reach stability in their lives as well. Stability is attained when no disrupting events occur over a long period of time.

As mentioned before the timing of an effect can differ, cause and effect are almost synchronized, this means an immediate effect, or it takes a while to evolve, which is referred to as a lagged effect. Lagged effects run the risk of remaining unobserved in limited retrospective research. There could also be a conflict in causality; this happens when people anticipate to something they intend to do in the future or something they count on happening in the future. In this case the intention of the cause occurs in time before the effect; only the manifestation of the cause comes in time after the effect. A reversal of cause and effect may in fact be a case of anticipation.

Full life history data, methods and techniques are necessary for life course research. Different aspects can be observed, established or calculated based on life history data: the timing of each separate life event, the chronological order of events and the time lags between several events. Individuals can be compared concerning the timing or order of events. People can be grouped according to who experienced certain events and who have not. The ending times of past states can be observed in life history data. This means that the period the state lasted can be calculated. In cross-sectional data it is impossible to calculate the duration of states.

Feijten (2005) distinguished three dimensions: timing (immediately / lagged), duration (permanent / temporary), order (before: anticipation / after its cause) of events. The life course approach emphasizes the mutual influence of parallel life careers on the time aspects of life events and the cumulative impact of life experiences in the long term. Memory lapses are less of a problem in case of life course events (Van der Vaart, 1996). Life course events can be better recollected than other events.

6 | Conclusion

Many models of transport mode choice have been developed in the past. Mode choice is typically conceptualized as a function of the characteristics of alternative travel modes and a set of personal and household characteristics. Adjustment of individuals' behaviour is not included in models described in this chapter. There is a need to explore and model dynamics in activity-travel patterns along various time horizons from a transportation perspective. This will lead to dynamic models where behavioural change is included.

Life course may be a valuable approach as suggested by experiences in other domains. The context in which mode choice decisions take place is seen, in this study, as an individual's life course. Central concepts in the life course approach are life trajectories (similar to career), transitions and events. The individual life course is composed of multiple, interdependent careers (i.e. housing, household, education, occupational career) which develop over time in parallel. Earlier life transitions may have a cumulative effect on later life. The concepts of timing, sequencing, duration and spacing are used to describe the life events, transitions and trajectories.

A modelling approach that allows estimation of direct and indirect effects, inclusion of contextual and situation-specific variables, and specification and testing of causal mechanisms may offer some advantages. The primary aim of the study is to develop the suggested formalism in analysing and predicting dynamic travel mode choice in relation to life course events. This fits into the interest in developing dynamic integrated land use transportation models.

3 | Framework

1 | Introduction

As mentioned in the introduction, this thesis contributes to the literature on activity-based modelling. More specifically, the focus will be on the dynamics underlying activity-travel patterns. Various kinds of dynamics can be distinguished. For example, when an individual traveller notices that the actual travel time to a particular destination has been longer than expected, he has to decide whether or not to shorten the duration of that and / or subsequent activities, visit other destinations, drive faster, cancel activities, etc. This is an example of short-term dynamics that involves rescheduling of activities. At the other end of the spectrum, a different job location may imply that a habitual activity-travel pattern may no longer be very efficient and / or effective. Under such circumstances, individuals may need to explore new options and permanently reorganise their activities in time and space. Change of job is only one of the triggers that may induce individuals and households to reorganise their activities in time and space. The specific focus in this thesis is concerned with life course events in general that may cause a change in needs or preferences and / or influence the constraints that impact on activity-travel decisions. Thus, this thesis is based on the assumption that life course events may cause individuals and households to change their activity-travel patterns. We are interested in developing and testing a modelling approach that allows representing and simulating such dynamics. Our special focus is concerned with changes in transport mode choice. Before discussing the specific modelling approach, the key underlying conceptual considerations will be explained. In the next section the principles of adaptation and learning are discussed. These underlying principles form the foundation of dynamic activity-travel patterns, the influence of time is also discussed here. These considerations will be combined into a conceptual framework for this study.

2 | Adaptation and learning

Adaptation and learning are an integral part of daily life. People have certain needs, preferences and expectations and will try to achieve these preferences, dependent on the characteristics of their immediate environment. They may not know all the options in the beginning, but continued search and exploration of new alternatives may ultimately lead to equilibrium in the sense that the utility they derive from conducting their activities in time and space is sufficiently close to their preferences and expectations.

Many factors may however distort such equilibrium, leading to a discrepancy between needs / preferences and actual utilities, which in turn may lead to behavioural change. Van der Waerden *et al.* (2003) identified two important factors which may cause people to reconsider their habitual behaviour: critical incidents and life course events. Critical incidents are defined as unexpected events such as accidents or unexpected long delays. Life course events are for example the birth of a child, change of job, etc. The terms life course events and life trajectory events are used interchangeably in this thesis.

This research project concentrates on life course events and their influence on a person's life trajectory. The life trajectory is seen as the context in which behaviour takes place. The set of conditions influencing choice behaviour may change because of a life course event. People can react in different ways. For example, responsive behaviour occurs when individuals react to a context that has dramatically changed. People may however also change their behaviour in anticipation of an expected change. It is assumed that an individual is likely to reconsider his / her current choice behaviour after the occurrence of a life course event. The change may trigger learning processes so that the implications of an event for choice behaviour may materialize gradually over time after the occurrence of the event. The assumed effect of events on activity-travel decisions is thus captured in terms of the theory of learning and adaptation.

According to this theory, individuals develop and continuously adapt choice rules while interacting with their environment. Transport systems and urban environments are highly dynamic, non-stationary and uncertain (Arentze and Timmermans, 2003). Based on earlier work by van der Waerden and Timmermans (2003), van der Waerden *et al.* (2003) and Verhoeven, Arentze, Timmermans and van der Waerden (2005, 2006 and 2007) it is assumed that individuals adapt their behaviour such that, given a set of constraints, the utility derived from the outcomes of their behaviour meets at least a certain aspiration level. Under stationary conditions, after some period of time, individuals will show habitual behaviour (Han and Timmermans, 2006).

The link between this general theory of learning and adaptation and life course events is that by learning and adaptation individuals adapt to their environment to some extent, and are therefore in some state of equilibrium (that is, they exhibit habitual behaviour) until the occurrence of an event causes such an amount of change that an individual feels the need to start exploring alternative options, implying that they may reconsider their current choices and / or resources. Learning would imply that adaptation to new circumstances takes time and, hence, that current behaviour cannot be fully understood in terms of the current state of the individual and the environment.

Besides learning processes, other mechanisms may produce temporal effects such as delayed responses (people respond to a change only after some time), accumulation of stress (people avoid too many changes in a short amount of time), etc.. It is assumed that a life course event changes a certain personal situation, i.e., state**.** Figure 3 | 1 illustrates such influences. A state is defined in terms of a set of substates. In this example, the substates are different residential conditions, such as independent living, living in a dorm room or living with parents.

Figure 3 | 1 represents (a part of) the life trajectory of one individual: for instance, a boy who lives with his parents (substate C) and moves in the third quarter of 2003 to his first dorm room (substate B). Soon after this change he moves in with his girlfriend (e.g. live together (substate A)). This example illustrates transitions between different substates, which describe the residential condition of an individual; substate A = independent living (living on your own), substate B = student living (living in a dorm room) and substate C = parental living (living with parents or guardians). The event describes changes in the residential situation. Several different types of substates are possible; for example first dorm room, another dorm room, living on your own, renting a house, buying a house, etc.

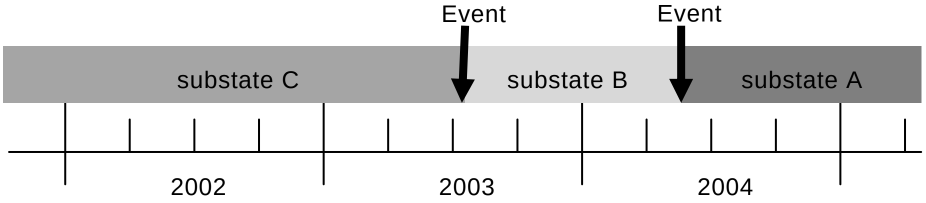


Figure 3 | 1: Effects of an occurrence on one state dimension

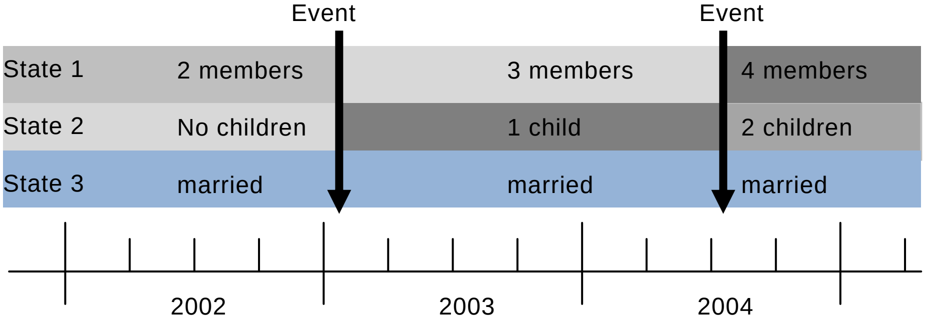


Figure 3 | 2: Effect of changes of an occurrence on more states

Figure 3 | 2 illustrates another example of changes in states caused by a life course event. In this case, the life course event describes changes in household composition. There are different aspects of household composition and, hence, dimensions of states, namely the number of household members (state 1), the number of children (state 2) and the marital state (state 3). Every state has certain substates. For example, the number of household members (state 1) has four substates: one member, two members, three members and four or more members of the household. Figure 3 | 2 illustrates the changes in household composition and the corresponding substates for one individual. In this example, the event in the first quarter of 2003 is a birth/adoption of a child. This result in a change in state 1: from two members to three members. State 2 changes from no children to one child, while state 3, the marital state stays the same, i.e. married. The second change, the event in the third quarter of 2004, is also a birth/adoption of a child. These two changes have no effect on state 3 (marital state), therefore this state stays the same. In case of a change such as getting married, it will only affect state 3 and states 1 and 2 will stay the same. Thus not every life course event changes all states. Figure 3 | 1 and Figure 3 | 2 represent parts of a life trajectory for a single individual. It illustrates the impact of occurrences of an event on different substates. This is only an example of substates for these two events. Events may however be interrelated and have complex relationships with (aspects of) choice behaviour and attributes of the person that are not influenced by occurrences of events (e.g., age, gender, etc.).

3 | Time influence

A basic assumption is that life course events not only change the (personal) conditions, but may also trigger adaptation and learning process. Behaviour is thus also influenced by time. Through search an individual explores choice opportunities in his or her environment and keeps a memory record of the varying rewards associated with his actions. Actions that produce positive rewards are reinforced and have a higher probability of being repeated in future choice situations under similar conditions, while actions with negative outcomes tend to be avoided. In stationary environments, reinforcement learning implies that random behaviour will ultimately evolve into habitual behaviour.

In non-stationary environments, a gradually changing environment or discrepancies between the changing environments and changing personal or household circumstances may imply that the behaviour of interest is no longer adequate to cope with the new situation. An individual may then have to change one or more facets of his habitual behaviour. Life course events may have a similar, but attenuated effect. A dilemma for any individual, who has limited knowledge about new circumstances, is the choice between exploration and exploiting current knowledge. Selecting actions that have not been tried before gives the opportunity of discovering new choices that may yield higher rewards than the currently best action. However, this comes with the risk of negative experiences. Individuals who wish to avoid such risks may stick to the current best choice. Thresholds for reconsidering current choices and the thoroughness of search will vary depending on the individual's tendency to take or avoid such risks.

Although these learning and adaptation process have been studied at some length in the transportation research community, especially in the context of uncertain travel times (Arentze and Timmermans, 2003; Ettema, Tamminga, Timmermans and Arentze, 2005; Ettema and Timmermans, 2006), in this study an attempt will be made to model these processes in an explicit manner. It is argued that the effects of life course events on behavioural dynamics may involve a time gap.

4 | Conceptual model

Figure 3 | 3 illustrates the assumption that behaviour is the result of the present state of a person and adaptation to a new state after a change caused by an event. An event, denoted by E in Figure 3 | 3, affects behaviour (B) in two different ways: firstly through the change of the present state (from S_1 to S_2) and secondly through learning, illustrated with the evolution from B_2 to B_2' to B_2" ending in the equilibrium situation indicated by B_2*. The present state is the actual environment or context of the person involved; including personal characteristics, possession and availability of transport modes, distances to different destinations and so on. This present state of a person influences behaviour, indicated in the figure with the arrow between S and B. The figure illustrates that behaviour can change even though the state stays the same. Life course events may change people's values and judgements, and may result in new behaviour. However, this may take time and after a while behaviour is in equilibrium again, illustrated in the figure with an asterisk.

The assumed influence of time, the adaptation and learning process, is illustrated in Figure 3 | 4. This example concerns car use. The event in this example is change in residential location: the person moved to a new city. There is a visual break in the behavioural curve, which is the moment, that event (E) occurred. The new situation (S_2), after the event, can result for example in less car use, which is a new behaviour (B_2). People need some time to adapt to the new circumstances. In Figure 3 | 4, this is illustrated with the time after the event until the equilibrium of the new behaviour (B_2*). At first the person still takes the car to work (habit). The person is also not very familiar with the environment and the possibilities for travelling to work.

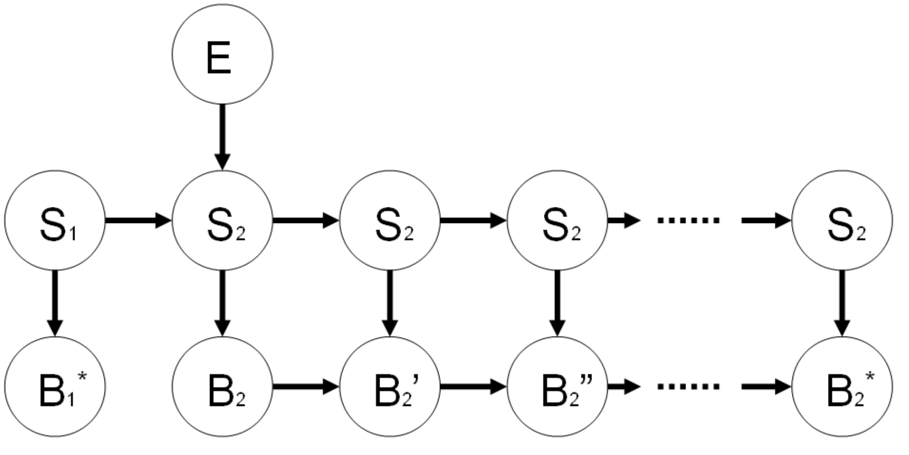


Figure 3 | 3: Conceptual model of the influence of an event

(E is event, S is state and B is behaviour)

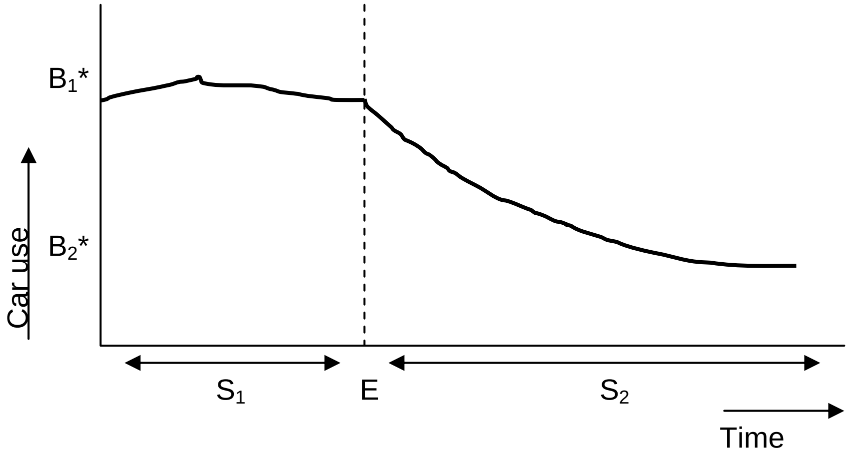


Figure 3 | 4: Example of time influence on car use

After a while the person, triggered to explore new options, discovers that there is a very good public transport connection from his home location to work and decides to choose that alternative option. This results in less car use.

5 | Conclusion

This thesis is based on the contention that life course events may trigger individuals and households to rethink their habitual activity-travel patterns. It may decide them to change one or more facets of their activity-travel patterns. A particular event may also lead to other life course events. Thus, life course events may have direct and indirect effects on other life course events and on activity-travel patterns.

In addition, it is argued that any modelling attempt should consider possible time gaps between life course events and behavioural change. An event does not necessarily lead to immediate changes in particular facets of activity-travel patterns. Change may increase the pressure on current, habitual patterns. However, it may take some time until accumulated pressure induces the consideration and exploration of new alternatives. Habitual patterns may even take longer to form due to learning and adaptation. Behavioural change may also occur in anticipation of life course events. In this case, the actual experiencing of pressure is not the trigger, but rather mental simulation of the consequences of possible or likely future scenarios.

Given this conceptualization, the question then becomes which approach should be adopted to model the direct and indirect effects of life course effects on travel mode change. This question will be addressed in the next chapter.

4 | Bayesian Belief Networks

1 | Introduction

The conceptualisation of a set of mutually dependent life course events directly and indirectly influencing travel mode choice requires a modelling approach that is consistent with these premises. Bayesian Belief Networks (BBN), which have been developed in overlapping fields such as Artificial Intelligence and Machine Learning, represent such a powerful approach. A Bayesian network is a network representation of the interrelationships and conditional dependencies between a set of variables (Neapolitan, 1990). This modelling approach is used in this research project to represent and simulate life trajectories and the direct and indirect effects of life course events on choice behaviour decisions, in particular transport mode choice. The potential advantage of BBNs over other techniques is that more complex causation patterns can be included and that the results can be directly interpreted in terms of the classified events.

This chapter is organized as follows. First the formulation and application of Bayesian Belief Networks are given. Next, an example of Bayesian Belief Networks is illustrated. In the fourth section two algorithms for identifying the causal relationships between variables (structure learning) are described: search & scoring-based algorithms and dependency-analysis based algorithms. The PC algorithm and the NPC algorithm (dependency-analysis based algorithms) are described in more detail. Given an identified structure, the parameters of the network need to be estimated. The EM algorithm for parameter learning, used in this study, is explained in the fifth section. The final section draws some conclusions related to our specific approach.

2 | Definition of Bayesian Belief Network

Bayesian networks are directed acyclic graphs (DAGs) in which the nodes represent variables, the arcs signify the existence of direct causal influences between the linked variables, and the strengths of these influences are expressed by the forward conditional probabilities (Pearl, 1988). The directed acyclic graphs do not allow undirected or bidirectional arcs, nor cyclic feedback loops. An arc between two nodes represents a causal relation: the node from which the arc originates is called the parent node and the other node is called the child node. A Bayesian network consists of two components: 1) a structure component (i.e., the DAG), which specifies the structure of cause-effect relationships between the variables; and 2) a parameter component, which consists of a set of conditional probability distributions that provide the statistical interpretation of the cause-effect dependence relationships depicted by the graphical structure.

The representation of variables (nodes) has several states, which correspond with the classes or options of the concerning variable. Each node has an underlying conditional probability table (CPT) that describes the probability distribution across the states of that specific node for each possible combination of states of the parent nodes.

The CPT of a node that has no parents is simple: it only contains the states of the node itself and the probability distribution across the states, where the sum of the probabilities is equal to 1 or to 100% (for each row in the CPT). The CPT of a child node is more complicated; the conditional probability table expands with the possible state configurations of the involved parent node(s). The CPT describes the probability distribution across the states of that specific child node for each (combined) state of the parent node(s). Each node has a certain probability distribution, which represents probabilities about the likelihood of possible outcomes for each node. *Prior* (unconditional) probability is the likelihood that some input parameter will be in a particular state; a *conditional* probability is the likelihood of the state of a parameter given the states of input parameters affecting it, and *posterior* probability is the likelihood that some variable will be in particular state. The initial probabilities for a node that has no parents correspond exactly with the probability distribution across the states in the CPT. Probabilistic reasoning makes use of an elementary relationship in probability theory that can be stated as:

$$P(A|B) = \frac{P(A,B)}{P(B)} \qquad \text{Equation 4 | 1}$$

where:

P(A | B) is the conditional probability of A, given B. It is also called the posterior probability because it is derived from or depends upon the specified value of B.

P(A, B) is the joint probability distribution of A and B.

P(B) is the prior probability or marginal probability of B. It is ‘prior’ in the sense that it does not take into account any information about A.

There are two main methods of reasoning (i.e. updating probabilities when something changes in the network - new evidence) in Bayesian Belief Networks: forward reasoning and backward reasoning. Evidence spreads through the network by these two methods. Forward reasoning is triggered by new evidence for one or more parent nodes. The probabilities of the involved child nodes (i.e. related nodes) are updated, the probabilities are made consistent with the new information (i.e. hard evidence entered into the network). Backward reasoning is triggered when new evidence is entered into child nodes. Probabilities can be calculated using the logic of the well-known Bayes rule:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \qquad \text{Equation 4 | 2}$$

where:

P(A) is the prior probability or marginal probability of A.

Once a BBN has been compiled the effects of hard evidence entered into one or more nodes can be propagated throughout the net, in any direction, and the marginal distributions of all nodes are updated. Efficient algorithms for updating Bayesian Belief Networks exist. Propagation algorithms were discovered in the 1980s by researchers. These algorithms are effective for large classes of BBNs and are implemented in software tools. This makes it now possible to use BBNs to solve complex problems without doing any of the Bayesian calculations by hand. At any moment in time all probabilities are consistent with all evidence available and states of the CPTs.

3 | Illustration of a Bayesian Belief Network

In this section, the use of Bayesian Belief models will be illustrated with a simple BBN, using Netica, a software tool for learning and analysing Bayesian Belief Networks (Norsys Software Corp, 1997). Note that other BBN software tools may have a different layout, which would result in different illustrations. The purpose of this example is to illustrate that Bayesian Belief Networks allow us to specify and analyze the direct and indirect effects of a series of variables on a particular node (choice behaviour of interest).

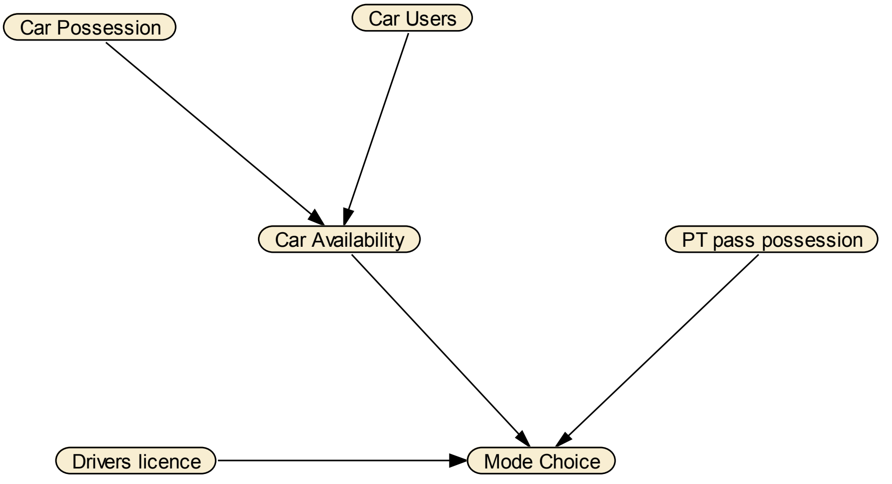


Figure 4 | 1: Simple Bayesian Belief Network

A hypothetical network is shown in Figure 4 | 1. It represents the direct influence of car availability, public transport (PT) pass possession and drivers licence on mode choice and the indirect influence of car possession and the number car users (through car availability) on mode choice. Car availability depends on the possession of car(s) in the household and the number of car users within the household. This network consists of six nodes, with the following states: Car Possession (no car, 1 car, 2 cars, > 2 cars), Car Users (1 user, 2 users, > 2 users), Car Availability (low, high), PT pass possession (yes, no), Drivers Licence (yes, no) and Mode Choice (car driver, car passenger, public transport, slow transport).

The nodes Car Possession and Car Users are the root nodes (no parents). They are the parent nodes for the Car Availability node, which is also called child node of the previously mentioned nodes. The parent nodes of Mode Choice are Car Availability, PT pass possession and Drivers Licence. Mode Choice is also called a leaf node (no children).

The CPT of a node that has no parents (thus root node) only contains the states of the node and the probability distribution across the states. The CPT's of the root nodes Car Possession, Car Users, PT pass possession and Drivers Licence are illustrated in Table 4 | 1. In this example, the a-priori probability or chance that someone does not owns a car is 8 percent, the chance of owning one car is 67 percent, two cars 21 percent, and more than two cars 4 percent chance. The a-priori probability or chance that there is one car user is 31 percent, the chance of two car users is 47 percent, and more than two car users 22 percent chance. For PT pass possession the a-priori probability of the possession of a pass is 31 percent and respectively 69 percent chance of not owning a pass. For drivers licence the a-priori probability of possession is 71 percent chance and for not owning a drivers licence 29 percent chance.

The CPT of a child node is more complicated. The conditional probability table expands with the states of the involved parent node(s). The CPT describes the probability distribution across the states of that specific child node for each (combined) state of the parent node(s). Table 4 | 2 and Table 4 | 3 show respectively the CPT of the child nodes Car Availability and Mode Choice.

The sum of all probabilities in each row is 100 percent. The higher the value, the higher the probability that this state occurs. If there are parent nodes involved, the actual probability depends on the probabilities of the parent states. Each node has a certain probability. The initial probabilities for a node that has no parents are simple; they exactly correspond with the probability distribution across the states as shown in Table 4 | 3. Probabilities of a child node depend on the probabilities of the parents.

Table 4 | 1: All CPT's of the root nodes (probabilities are percentages)

Root nodes	States	Probabilities (%)
Car Possession	No car	8
	1 car	67
	2 cars	21
	> 2 cars	4
Car Users	1 user	31
	2 users	47
	> 2 users	22
PT pass Possession	Yes	31
	No	69
Drivers Licence	Yes	71
	No	29

Table 4 | 2: CPT of the node car availability (probabilities are percentages)

Car Possession	Car Users	Low	High
no car	1 user	100	0
no car	2 users	100	0
no car	> 2 users	100	0
1 car	1 user	0	100
1 car	2 users	50	50
1 car	> 2 users	100	0
2 cars	1 user	0	100
2 cars	2 users	0	100
2 cars	> 2 users	50	50
> 2 cars	1 user	0	100
> 2 cars	2 users	0	100
> 2 cars	> 2 users	0	100

Table 4 | 3: CPT of the node mode choice (probabilities are percentages)

Car Availability	PT pass possession	Drivers Licence	Car Driver	Car Passenger	Public transport	Slow transport
Low	Yes	Yes	0	25	50	25
Low	Yes	No	0	25	50	25
Low	No	Yes	0	33	33	34
Low	No	No	0	33	33	34
High	Yes	Yes	30	30	30	10
High	Yes	No	0	40	40	20
High	No	Yes	40	40	10	10
High	No	No	0	60	20	20

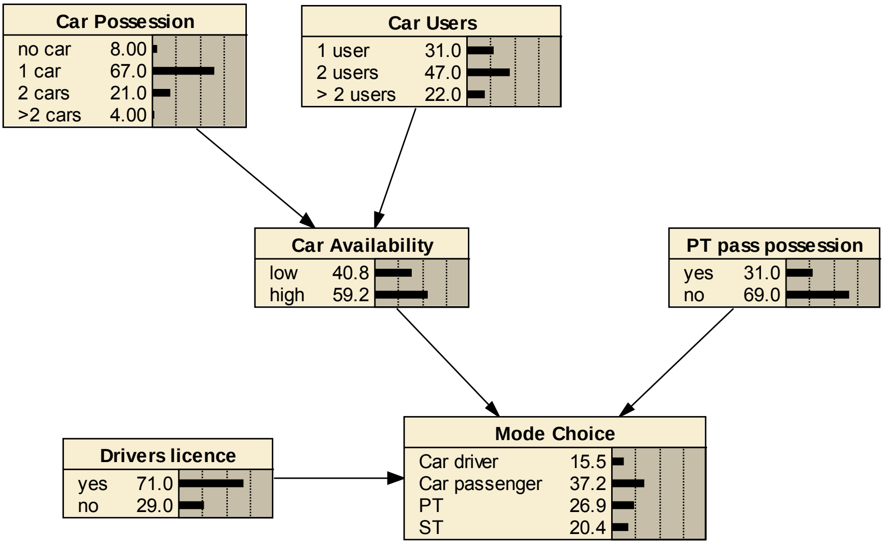


Figure 4 | 2: Compiled Bayesian Belief Network

Figure 4 | 2 gives the probabilities of the compiled network. The probability for the state 'low' of the car availability node can be calculated given the probability distribution of CPT's in Table 4 | 1, Table 4 | 2 and Table 4 | 3:

Probability state 'low' = (0.08 x 0.31 x 1.0) + (0.08 x 0.47 x 1.0) + (0.08 x 0.22 x 1) + (0.67 x 0.31 x 0) + (0.67 x 0.47 x 0.5) + (0.67 x 0.22 x 1.0) + (0.21 x 0.31 x 0) + (0.21 x 0.47 x 0) + (0.21 x 0.22 x 0.5) + (0.04 x 0.31 x 0) + (0.04 x 0.47 x 0) + (0.04 x 0.22 x 0) = **0.408**.

The probabilities for the state 'high' can be calculated in a similar way. However, because the probabilities sum to one, the probability for the state 'high' is equal to: 1 – 0.408 = **0.592**.

When hard evidence is entered into the network the probabilities for the related parent and child nodes are updated according to the Bayesian principle defined by Equation 4 | 1 and Equation 4 | 2 . Figure 4 | 3 illustrates the probabilities of the BBN when the evidence is entered into the network. The evidence indicates that there is one car in the household. The probabilities of the Car Availability node can be calculated in the same way, the changed numbers are in bold:

Probability state 'low' = (**0.00** x 0.31 x 1.0) + (**0.00** x 0.47 x 1.0) + (**0.00** x 0.22 x 1) + (**1.00** x 0.31 x 0) + (**1.00** x 0.47 x 0.5) + (**1.00** x 0.22 x 1.0) + (**0.00** x 0.31 x 0) + (**0.00** x 0.47 x 0) + (**0.00** x 0.22 x 0.5) + (**0.00** x 0.31 x 0) + (**0.00** x 0.47 x 0) + (**0.00** x 0.22 x 0) = **0.455.** In addition the probability for the state 'high' is equal to: 1 – 0.455 = **0.545**

Figure 4 | 3 shows the updated probabilities after entering hard evidence into the network. The updated probabilities can be compared with the probabilities in Figure 4 | 2. The updated probabilities of the nodes Car Users, Drivers Licence and PT pass possession are the same.

Those nodes are not directly related to the node Car Possession, where the hard evidence is entered. The probabilities of the other nodes, Car Availability and Mode Choice, are updated after the hard evidence '1 car' was entered into the network. The probability of the state 'low' of the node Car Availability increased from 40.8 to 45.5 in the new situation. For the Mode Choice node the probability of the option 'car driver' and 'car passenger' decreased from respectively 15.5 and 37.2 to 14.3 and 36.7.

.

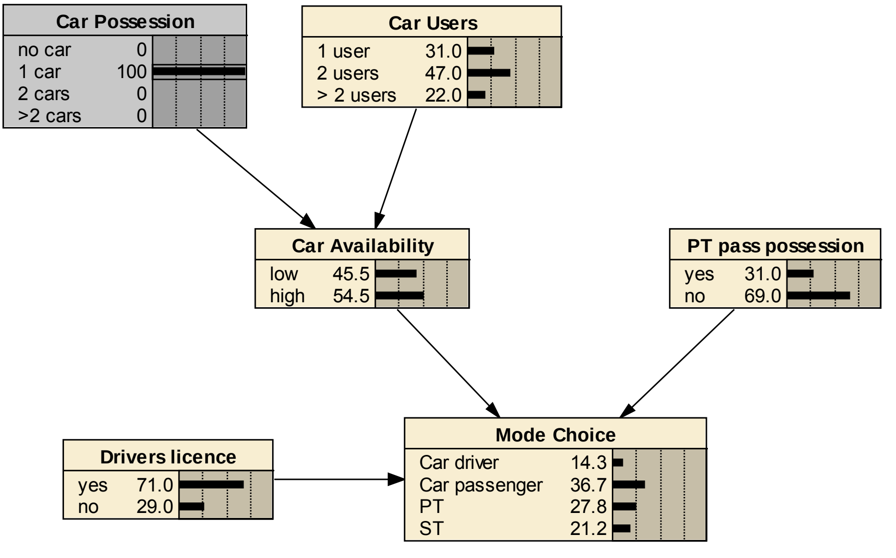


Figure 4 | 3: BBN with entered evidence for node car possession

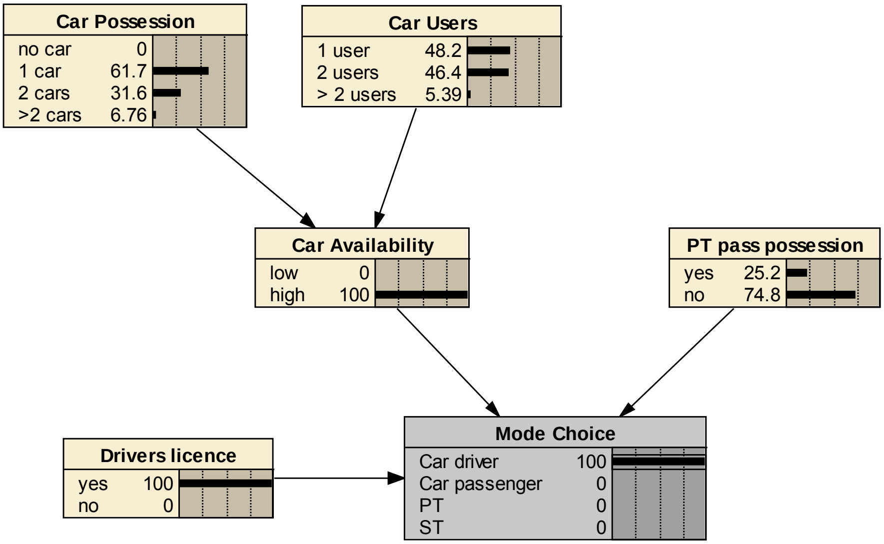


Figure 4 | 4: BBN with entered hard evidence for node mode choice

The probabilities for the other options ‘Public Transport (PT)’ and ‘Slow Transport (ST)’ increased from respectively 26.9 and 20.4 to 27.8 and 21.2. The percentage of increase or decrease (updated probability / former probability) is the highest for car driver (decrease) and slow transport (increase).

Backward reasoning is illustrated in Figure 4 | 4. The result of the entered evidence ‘car driver’ for the child node Mode Choice is portrayed. The probabilities of the parent nodes PT pass possession, Drivers Licence and Car Availability, and their parent nodes Car Possession and Car Users are automatically updated. Some probabilities increase, like the possession of 1 car, 2 cars and 2 cars or more in the household, while other probabilities decrease, like possession of a public transport pass.

This section illustrated an example of Bayesian Belief Networks. Parent and child nodes in the network are explained and the calculation of probabilities is illustrated given the underlying conditional probability tables. It is not always easy to construct the structure of a BBN. Fortunately algorithms for structure learning exist. These algorithms are described in the next section.

4 | Structure learning algorithm

The first step in estimating a Bayesian Belief Network is to identify the structural relationships among the variables (structure learning). Over the years, several structure learning algorithms have been developed. Two groups of algorithms or approaches can be distinguished: search and scoring-based algorithms and dependency analysis-based algorithms (also called constraint-based). These different methods are explained in the next sections.

1 | Search and scoring algorithms

Scoring-based methods view a Bayesian Network as a structure defining a joint probability distribution across the variables included in the network. These methods search for the structure that maximizes a goodness-of-fit on the observed joint probability distribution in the data. The joint probability distribution can be calculated with the following product (e.g. chain) rule:

Equation 4 | 3

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_j \mid \Pi_j)$$

Where X_1, X_2....X_n are variables in the network. Π_j is the set of parent variables of X_j.

These algorithms start with a graph without any edges between the variables. Then, first an edge is added using some search method. Next, a scoring method is used to check if the new structure with the added edge is better than the old graph. If the new graph is better, a new edge is added and this process continues until no new structure is better than the previous structure. To evaluate the structure, different scoring criteria have been applied in these algorithms, including the Bayesian scoring method. Examples of these algorithms are the K2 Algorithm (Cooper and Herskovits, 1992), the HGC algorithm (Heckerman, Geiger and Chickering, 1995; Ramoni and Sebastiani, 1997), the Kutato algorithm (Herskovits and Cooper, 1990) based on entropy-based methods, Suzuki's algorithm (Suzuki, 1996), the Lam-Bacchus algorithm

(Lam and Bacchus, 1994) and the WKD algorithm (Wallace, Korb and Dai, 1996) both based on minimum message length methods. Most of these algorithms apply heuristic search methods. Node ordering is required to reduce the search space.

2 | Dependency analysis algorithms

These algorithms rely on conditional independence (CI) tests to measure the dependency relationships, in contrast with the heuristic search methods used on the search and scoring algorithms that optimize a network structure as a whole. Examples are the Wermuth-Lauritzen algorithm (Wermuth and Lauritzen, 1983), the Boundary DAG algorithm (Pearl, 1988), Algorithm A and B (Cheng, 1998), the Constructor algorithm (Fung and Crawford, 1990), the SGS algorithm (Spirtes, Glymour and Scheines, 1990), the SRA algorithm (Srinivas, Russell and Agogino, 1990), and the PC algorithm (Spirtes, Glymour and Scheines, 1991). A basic concept in many algorithms is the mutual information between two given nodes, which is defined as:

$$I(A,B) = \sum_{a,b} P(a,b) \log \frac{P(a,b)}{P(a)P(b)} \qquad \text{Equation 4 | 4}$$

where $I(A, B)$ is the mutual information between nodes A and B; a and b represent possible states of A and B; $P(a, b)$ is the joint probability of $A = a$ and $B = b$; and $P(a)$ and $P(b)$ are the (marginal) probabilities of these states. Existence of mutual information is not a sufficient condition for a link between two nodes, as the influence may also run through other nodes. Constraint-based algorithms use the concept of *d*-separation: two nodes are *d*-separated when, loosely speaking, they are conditionally independent given possible paths through other nodes. The problem of finding the correct structure for a given set of variables is a NP-hard problem and therefore existing algorithms use heuristic search.

In this thesis, a dependency analysis algorithm for structure learning is used, more specifically, the PC algorithm. This algorithm and the NPC algorithm are described in more detail as a representative example of dependency analysis algorithms. The PC algorithm, used in this study, is a variant of the original PC algorithm (Sprites, Glymour, Scheines, Heckerman, Meek and Cooper, 2000). The basic idea of these algorithms is to derive a set of conditional independence and dependence statements by statistical tests. The PC algorithm consists of the following steps (Hugin Expert, 1995; Madsen, Lang, Kjærulff and Jensen, 2004):

1. Statistical tests for conditional independence (CI tests) are performed for all pairs of variables (except for those pairs for which a structural constraint has been specified).

2. An undirected link is added between each pair of variables for which no conditional independences were found. The resulting undirected graph is referred to as the *skeleton* of the learned structure.

3. Colliders are then identified, ensuring that no directed cycles occur. (A *collider* is a pair of links directed such that they meet in a node.) For example, if A and B are dependent, B and C are dependent, but A and C are conditionally independent given S (e.g. set of variables), not containing B, then this can be represented by the structure A --> B <-- C. Collider structures have the unique property that parent nodes A and C are mutually dependent only under given states of B. On that basis collider structures can be identified.

4. Next, directions are enforced for those links whose direction can be derived from the conditional independences found and the colliders identified.

5. Finally, the remaining undirected links are directed randomly, ensuring that no directed cycles occur.

The first two steps are repeated until all possible pairs have been tested. First, only pairs are tested (no conditional relations). After that pairs are tested conditionally on a third variable. This continues until all pairs have been tested conditionally. The following independence testing procedure is used for the first step in the PC algorithm: if variable X causes variable Y, it implies a probabilistic dependency, $P(Y \mid X) \neq P(Y)$. Thus, if the null hypothesis of marginal independence of X and Y, $H_0 : P(Y \mid X) = P(Y)$ or $P(X, Y) = P(X)P(Y)$

is rejected, the directed dependence $X \rightarrow Y$ is supported. The structure learning algorithms are based on making dependence tests that calculate a test statistic which is asymptotically chi-squared distributed assuming (conditional) independence.

One important thing to note about the PC algorithm is that, in general, it will not be able to derive the direction of all the links from the data, and thus some links will be directed randomly, ensuring that no directed cycles occur. This means that the learned structure should be inspected, and if any links seem counterintuitive the NPC algorithm (Necessary Path Condition (Steck and Tresp, 1999)), which allows the user to interactively decide on the directionality of undirected links, could be used instead. Another possibility is to import constraints in the structure learning process. Constraints describe where a directed or undirected link can not appear.

The NPC algorithm seeks to repair the deficiencies of the PC algorithm. The solution provided by the NPC algorithm is based on the inclusion of a criterion known as the Necessary Path Condition. This criterion forms the basis for introducing the notion of ambiguous regions, which in turn provide a language for selecting among sets of interdependent uncertain links. The resolution of ambiguous regions is performed in interaction with the user. In constraint-based learning algorithms, the skeleton of the graph is constructed by not including a link in the induced graph whenever the corresponding nodes are found to be conditionally independent. There can, however, be inconsistencies among the set of conditional independence and dependence statements (CIDs) derived from limited data sets. That is, not all CIDs can be represented simultaneously. The inconsistencies are assumed to stem solely from sampling noise. The number of inconsistencies in the set of CIDs reflects structural model uncertainty. Thus, the number of uncertainties is a confidence measure for the learned structure and can as such be used as an indication of whether or not sufficient data has been used to perform the learning. The inconsistent CIDs produce multiple solutions when inducing a directed acyclic graph (DAG) from them. These solutions differ with respect to the set of links included. To resolve the inconsistencies, the NPC algorithm relies on user interaction where the user gets the opportunity to decide on the directionality of undirected links and to resolve the ambiguous regions. An *ambiguous region* is a maximal set of interdependent links, or uncertain links. When the absence of a link (*a*) depends on the presence of another link (*b*), and vice versa, *a* and *b* are defined as interdependent. Both *a* and *b* are constitute what is called *uncertain links*. The main goal is to obtain as few and small ambiguous regions as possible. It should be noted that deterministic relations between variables will also produce ambiguous regions. The user is offered the possibility of providing information as to how the ambiguous regions should be resolved.

The first step in constructing a working BBN is structure learning. The second construction step is learning the parameters of the network. The estimation of the parameters is described in the next section.

5 | Parameter learning algorithm

It is necessary to specify the probability distribution for each node in the Bayesian network. If there are no missing values in the data, parameter learning is rather straight-forward. Under that condition it simply reduces to determine observed conditional frequencies for each child node and its parent nodes in the data. However, if there are missing values, estimation methods come into play. In the type of BBNs used in this thesis, all variables have discrete distributions, which simplify the calculations. As said, when there are missing values, the unknown parameters of these conditional distributions must be estimated from data. Different approaches can be used, such as maximum entropy, and maximum likelihood. When there are unobserved variables the direct maximization of the likelihood is often complex. The expectation-maximization (EM) algorithm is an approach to deal with this problem (Dempster, Laird and Rubin, 1977; Cowell and Dawid, 1992; Lauritzen, 1995). The EM algorithm tries to find the model parameters (conditional probability distributions) of the network from observed (but often not complete) data that maximizes the log likelihood of the current joint probability distribution on the case data. The estimated conditional probabilities are improved in each iteration. For each iteration, the conditional probabilities will match the data increasingly better until no further improvement can be made. It has proven a flexible tool for problems involving missing data or incomplete information.

The algorithm performs two steps: an expectation (E) step, which computes an expectation of the likelihood, and a maximization (M) step, which computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found in the E step. The parameters found in the M step are then used to begin another E step, and the process is repeated. It is an iterative process which continuously improves the resulting conditional probabilities. The learning process will terminate when the convergence, defined as the difference in log likelihood between two successive iterations, reaches a value smaller than the convergence threshold (or earlier, if the maximum number of iterations is reached). The algorithm has a few disadvantages: the choice of initial parameters may converge to a local optimum (poor choices can lead to bad estimations) and the convergence may take a long time.

If there are no data available for a specific combination of states, the probabilities are evenly distributed as this corresponds to probabilities when no information is available. For example, Table 4 | 2 shows that the probability distribution of car availability will be 50% for every state (low and high) if there are no data available for the combination of the following states of the parent nodes car possession and car users: one car and two car users or two cars and more than two car users. One may argue that the assumption of an even distribution when no information on the specific state of the parents is available in the dataset is too rigorous. Instead of an even distribution, an overall distribution across all cases in the dataset would already be a better estimate of the distribution of the actual probabilities. For example, when the overall distribution is very skewed, our a-priori expectation would also be that the distribution is skewed in the same way when there is no information about the parents. Note that this issue is not irrelevant. Unless the dataset is very large compared to the number of parents for a node, missing cases for specific configurations of a parent are likely. Assuming an even distribution in cases where the a-priori distribution is skewed, will lead to biased assessments of posterior probabilities in those cases. Therefore, to circumvent this bias, a more refined rule is used, i.e. rather than an even distribution, the overall distribution of a best a-priori estimate of the actual distribution is preferable.

6 | Conclusion

In this chapter, it is argued and shown that in principle Bayesian Belief Networks are an appropriate approach to modelling the direct and indirect influence of events on other events and mode choice. Structure and parameter learning can support the detection and representation of interdependencies in collected data in a network. Causal effects can be learned and temporal relations can be captured in constraints or domain knowledge, like node ordering.

A Bayesian Belief Network can be empirically derived and used to model and simulate life course events and their effects on transport mode choice and resource decisions. These networks offer an alternative to familiar hazard models, where the duration of some phenomenon (i.e. events) is modelled as a function of a set of explanatory variables. BBNs have the potential advantage that relatively complex direct and indirect relationships among life course events, and between these events and transport-related decisions can be captured.

In this chapter the tools for constructing a Bayesian Belief Network are described. The input necessary for structure and parameter learning is data and constraints. Data for this study is collected using a retrospective Internet-based survey. In the next chapter the data collection is described. Constraints for the network learning are given in the sixth chapter.

5 | Retrospective Internet-based survey

1 | Introduction

In the previous chapter it was argued that Bayesian Belief Networks represent a potentially powerful approach to modelling the impact of life trajectory events on dynamic activity-travel patterns. Such networks need as input empirical data to learn the structure of the network and the conditional probability tables of the variables that are identified to be relevant. Ideally, one would prefer to have a panel. However, in practice such samples are difficult to organise and are highly demanding in terms of time, money and other resources.

In the present study, it was therefore decided to use a retrospective survey. This means that respondents are invited to recall events they have experienced in their past and report relevant details of these events, including their timing. Retrospective surveys can be administered in various ways. In the present study, an Internet-based survey was designed and implemented.

This chapter starts by articulating the data needs. The following section discusses the pros and cons of retrospective surveys and the third section outlines the features of the Internet-based survey. The procedure, design and routing of the survey, response rates and the composition of the sample are described in the fourth section. Next, the cleaning of the data is reported. The results of a performance test of the retrospective data are presented in the sixth section and the chapter ends with the conclusion.

2 | Retrospective surveys

Data needs and appropriate data collection depend on the specific purpose of the study. In this context, it is important to understand which life trajectory events trigger behavioural change. To conduct such analyses, data are required about the life trajectory events experienced by individuals and the chronological order of these events. Panel data or pseudo-panel data are not relevant to collect such data when they are typically not administered to the same respondents for a large number of years. In addition to the costs of administering a panel, attrition rates of 40% are quite common. Retrospective surveys offer an alternative.

Behrens and Del Mistro (2006) define retrospective surveys as once-off surveys of individuals in which respondents are asked to recall past behavioural changes and the events and circumstances surrounding these changes. The potential advantages of the retrospective surveys method are that they do not present great administrative complexity and time delay in data collection of panel surveys. On the other hand, because respondents are invited to recall phenomena in the past, it goes without saying that the reliability of retrospective surveys depends fundamentally on the nature of the phenomena about which questions are asked. In general, one may argue that individuals build up memory traces about their experiences. Such memory traces will be stronger for those experiences that are more important to them, e.g., experiences that are unique, dramatic, etc. Vice versa, memory traces of insignificant experiences will be weak. Moreover, potentially incomplete or inaccurate responses are likely to occur if the event recalled and the time of recollection are far apart. Assuming that data quality, that is memory recall, is monotonically related to the strength of the memory trace, the quality of retrospective surveys may be sufficient if the retrospective questions are concerned with special, memorable events or phenomena, especially when the time elapsed between the occurring of the event and the time of the survey is not too far apart.

This logic was supported by Behrens and Del Mistro (2006). They concluded based on their retrospective travel survey experiment that even when considerable time had elapsed since making a behavioural change, respondents did not report uncertainty in their recollection of the number of years that had passed since the change. Especially the follow-up telephone qualitative interview to explore the reliability of the answers from the surveys and to establish how confident respondents were with their answers indicated this. An explanation for this finding is that all recalled travel behaviour changes were associated with a form of 'life shock' or trauma, which are memorable events. On the other hand, Baddeley (1979) argued that collecting reliable information about experienced events is difficult; respondents often cannot recall the events accurately. In an earlier study, Baddeley (1979) concluded that forgetting is not uniform. The information about the most recent experience of an event is likely to be more accurate and reliable than information about earlier occasions. Especially details of a given event are difficult to recall accurately. Some memorable personal events are easier to remember.

Selective memory can play a role when retrospective questions are used. It may happen that respondents in fact 'do-not-know' the answer. The respondent can be stimulated to give an adequate response in several ways (de Leeuw, 2001): (1) encourage the respondent to use personal records and (2) stimulate a more thorough question-answer sequence by using a longer introduction. The researcher should be careful to ensure that the respondent understands the introduction and the question. Special techniques, like 'time-line follow-back methodology' and 'domain-dependent encoding' of memory, are used to probe the memory of respondents and improve their recall. The latter technique uses extra introductory questions to bring the respondent back to the situation in which the researcher is interested.

Experiences from Hollingworth and Miller (1996) showed that a retrospective survey proved to be a favourable alternative to a panel survey. Van der Vaart (1996) also argued that memory lapses are less of a problem in case of life events. Beige and Axhausen (2008) had good experiences with a retrospective survey covering a period of 20 years from 1985 to 2004. Thus, this brief summary of the literature suggests that retrospective surveys can be a useful and adequate tool to collect information about life trajectory events.

Examples of small scale applications of such surveys focus on non-standardized interviews, e.g., Baddely (1979) and Lanzendorf (2003). The method does not appear to have been applied extensively in practice. For model building purposes, however, one needs larger samples. The question then becomes whether retrospective surveys can be successfully designed and administrated through the Internet. Self-completion mail surveys seem less adequate as the protocol for completing the survey is relatively complex. Internet-based surveys have the potential advantage that some help and consistency checking is possible.

3 | Internet-based surveys

In principle, retrospective data can be collected through self-administered surveys (mail or internet), face-to-face interviews or telephone interviews. Interviews are more time consuming than self-administered surveys and have certain disadvantages for collecting retrospective data. Recently, Internet-based surveys have become quite popular, compared to surveys distributed by ordinary mail. Besides the low costs compared to mail and allowing faster processing of data and distribution by e-mail, an Internet-based survey has some technical advantages which are helpful with respect to collecting (retrospective) data.

As mentioned by De Leeuw (2001), it is important for the designer of the survey to fully understand what happens in a question-answer process to stimulate the respondent to give an adequate response (Tourangeau, 1984; Strack and

Martin, 1987; Schwarz, 1997). First, the respondent has to understand the question that determines the intended meaning. After that, the respondent has to recall relevant information from their memory. Particularly, with retrospective surveys this may be a difficult cognitive task. After retrieval from memory, a judgement is 'computed'. After a personal judgement is formed in the mind of the respondent the answer has to be communicated to the researcher. In the last step, the respondent may want to edit the response before it is finally given, especially with sensitive topics.

Some technical features of Internet-based surveys make them potentially advantageous in this context. These include: (1) extra information can be presented in pop-ups to support better understanding of a question, (2) dynamic routing, i.e., a dynamic sequence of questions depending on previous answers, skip irrelevant questions and different question phrasing, is possible, (3) drop down lists can be used to effectively handle pre-coded answers, to decrease the process time for respondents, and (4) checking possibilities can be used to make sure that answers are in a required range or format (numeric, number of characters) or to reduce item non-response if the question is mandatory. A disadvantage of these checking possibilities is that warning or error messages, where respondents have to return to the question and correct their "error" before being allowed to proceed, have been shown to increase respondent frustration. Consequently, a number of respondents terminate the survey before the end (Best and Krueger, 2004).

Furthermore, as Christian, Dillman and Smyth (2005) note, "The design and visual presentation of survey information, through the manipulation of verbal, numeric, symbolic, and graphical languages, can facilitate respondents' answering of survey questions and help them "get it right the first time". The likelihood of receiving error messages, that web survey designers may use to decrease item non-response and to verify that answers are in an acceptable format, can be reduced by helping respondents. Visual design techniques to reduce respondent frustration, increase response efficiency and improve the overall survey experience for respondents, can be applied by web survey designers. The quality of each answer will be improved, because the respondent has more time to understand the question and retrieve and compose an answer. Although these advantages of Internet-based surveys hold more in general, they arguably are particularly relevant for retrospective surveys given the memory demands these surveys impose. For that reason an Internet-based survey was designed to collect data for this study.

4 | Design and application

In this section, the details of the Internet-based survey to collect retrospective data about life course events are described. The design of the data collection instrument required several operational decisions, which are subsequently described and motivated in the following subsections.

1 | Life course events

Based on previous work by van der Waerden and Timmermans (2003) and van der Waerden *et al*. (2003), a list of seven life course events was compiled. These seven events are defined as changes in residential location, household composition, work location, study location, car availability, possession of a public transport pass, and household income. Changes in household composition, for example, are marked by births, marriage, divorce, etc. In this study, it is believed that these processes are the true markers of an unfolding life trajectory. Be it positive or negative, these events are unique and generally are the hallmark of individual's most important life course decisions. Similarly, even though perhaps less intense, changes in residential location, a new car, change of job, etc. also represent unique events in one's life that may have implications for travel behaviour and mobility. In this study, it is assumed that if respondents can be sufficiently motivated to actively participate in the survey and can handle the technical challenges induced by Internet surveys, there should be no reason to expect any less reliable results for a retrospective survey than for any other type of survey.

2 | Procedure

In the Netherlands, people are not always willing to participate in surveys. Of course, their willingness depends on several things, such as the topic of the survey, the way of approaching the respondent, implications of the results, benefits etc. A special procedure was designed to collect as many data as possible. Figure 5 | 1 illustrates the four steps of the procedure.

Step 1: a special TU/e e-mail account was opened for correspondence about the survey. The alias of the e-mail (vervoermiddelkeuze@bwk.tue.nl) was related to the topic of investigation (vervoermiddelkeuze = transport mode choice). This e-mail address makes the project more official for respondents and therefore they may feel more obligated to participate. It is more difficult to collect e-mail addresses of potential respondents than postal addresses. Many people also have a temporal e-mail address for work or private use and change them regularly to prevent spam. Nevertheless, e-mail addresses were collected in the Netherlands based on the participants of former surveys and connections of the University. Approximately 2400 emails were sent with a request to participate and to send the mail to at least three other persons within their social network. The latter request was our way to reach more potential respondents. Two rewards were offered, or at least a chance of a reward. The respondents who completed the survey could win one of 50 gift vouchers of 25 euro's. If people invited at least three other respondents, who completed the survey, they could also win one of 25 gift vouchers of 50 euro's. These rewards were given to motivate potential respondents to encourage more respondents to participate. In the first invitation e-mail the procedure was explained, as well as the rewards, and the goal of the research. The invitation e-mail included a link to a special online form to subscribe to the survey.

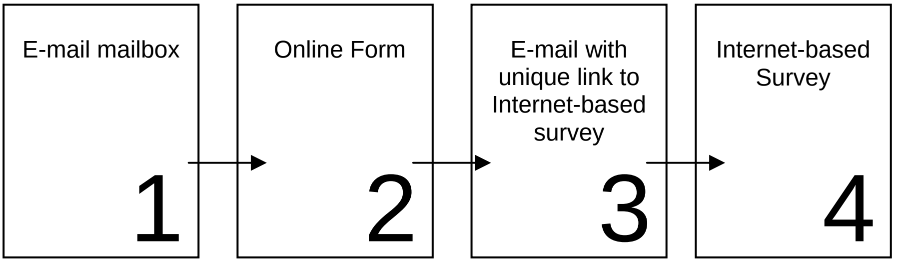


Figure 5 | 1: Four step procedure

It will be evident that the resulting sample is not a random sample. It means that if the data are used beyond the sample, appropriate weighting schemes need to be applied. The main interest of this PhD study, however, is to explore the suggested approach of modelling dynamics. Results are therefore based on this sample.

Step 2: This online form contained the same information as the invitation e-mail. The form consisted of questions about personal characteristics such as first and last name, gender, postal code, year of birth, e-mail address and their occupation. Some questions were mandatory, like name and e-mail address. People who received an invitation from another respondent could indicate that at the bottom of the form. This way the recruitment of other respondents could be checked and traced. In total, 939 people agreed to participate.

Step 3: the 939 respondents received a personal e-mail with a unique link to the Internet-based survey. Their first and last names from the online form were used to make the invitation e-mail more personal. The advantages of a unique link are (1) the survey is linked to a specific and known person, (2) respondents could stop and continue at another time (or at another computer), and (3) the link was only available once. In that way the respondents could not forward the link and there was control over the respondents.

Step 4: the Internet-based survey consisted of the following components: (i) personal and household characteristics; (ii) availability and possession of transport modes; (iii) occurrence of life course events, (iv) current travel behaviour, (v) perception of trip conditions and (vi) evaluation of the selected trip conditions.

3 | Design of the survey

Because the focus of the study concerns part 3 of the survey, occurrence of life course events, only that part will be described in more detail here. All questions of the survey are listed in Appendix 1A (in Dutch).

Respondents were prompted about seven life course events in their life. They were asked to indicate whether they experienced each of these events, and, if so, to provide additional information in a matrix about the timing of the event (month and year), the cause of the change (i.e., the specific type of event) that took place and the before and after situation for every change to a maximum of ten changes. Different types of changes involved in an event were defined (Figure 5 | 2) so that the respondents could understand what they had to recall. It was assumed that the description helped the respondent to recall those specific changes of an event.

First, respondents indicated whether they experienced a change or not. If they experienced at least one change, the respondent automatically received the matrix question about that particular event. As an example, Figure 5 | 3 represents the matrix question of *change in residential location*. All matrices of the seven life course events were identically structured. There was one exception: the matrix of *change in residential location* had seven topics/questions, whereas the other matrices had only five.

TU/e technische universiteit eindhoven

CHANGE IN RESIDENTIAL LOCATION

How many changes occured since you left the parental house?

Changes in residential location are defined as follows:

- moving to the first student room (house in the city where you study, which is not independent)
- move to another studentroom (not within the same house)
- independent living
- living together (move into your partner's house or move together to a different location)
- rent a (different) house
- buy a (different) house
- moving in with parents

Please indicate here whether you experienced this never or at least once.

- never
- at least once

back next

powered by NetQuestionnaires

Figure 5 | 2: Introduction question on occurrence of an event related to residential location

The five topics in the matrices referred to: the timing of the event, specifically the (1) month and (2) year, (3) the before and (4) after situation and (5) the cause of the change (i.e., the specific type of event). The two extra topics in this matrix (Figure 5 | 3) were included to collect more specific information about the housing situation: (6) housing type and (7) bought / rented residence. All questions of the Internet-based survey are illustrated with print screens in Appendix 1A, the predefined choice options and all variables are listed in Appendix 1B.

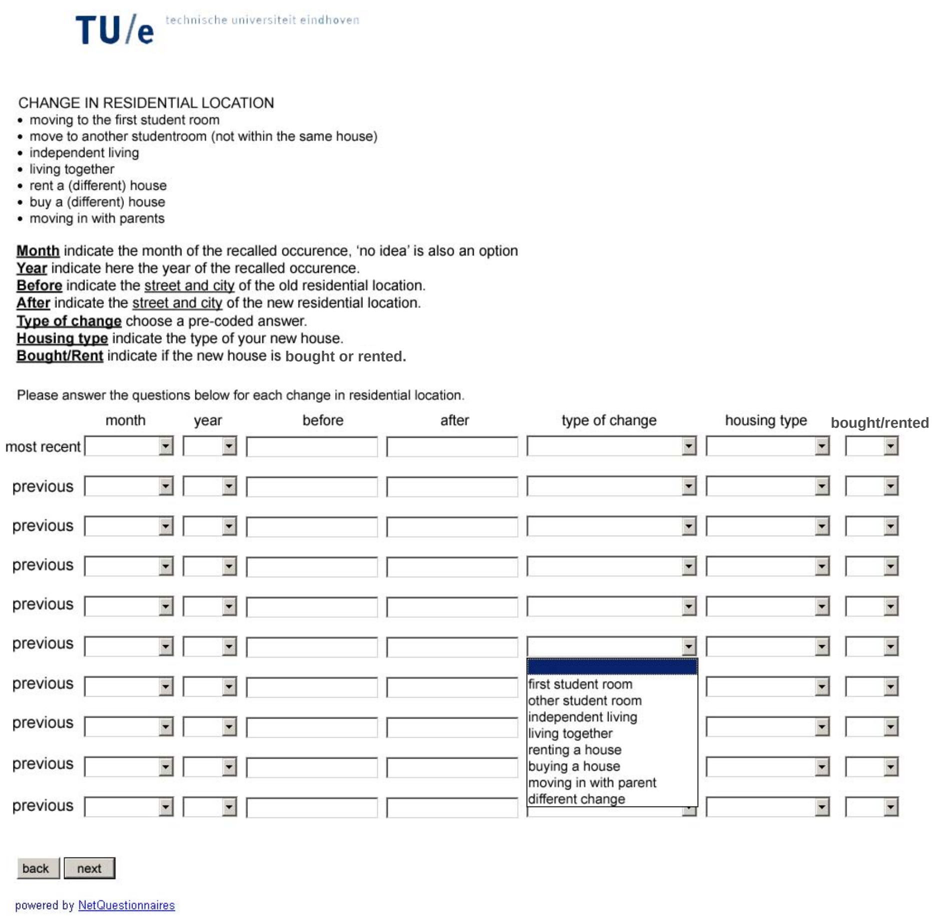


Figure 5 | 3: Matrix question event related to residential location

Respondents could indicate a maximum of ten changes, from the most recent change to changes in the past. In this study, it is assumed that the information about more than ten changes in the past would be less accurate and reliable. Moreover, it is unlikely that changes far back in the past influence a person's current travel behaviour. To stimulate the recall process, the matrix started with the most recent change. A mix of open ended and pre-coded questions was used in the retrospective survey. The question about the occurrence of an event was a pre-coded question (Figure 5 | 2), which was mandatory. Some questions/columns in the matrix question (Figure 5 | 3) were open ended questions. Like before and after situation, the other answers were pre-coded (drop down menu). Only the matrices for the life course events housing, work and study had open ended answer space in the columns before and after situation. The matrices for the other life course events had only predefined answer categories, thus also for the before and after situation.

The matrix questions were not mandatory. If these questions were made mandatory, the respondents should indicate ten changes (rows) and should answer all questions in each column. This would probably irritate the respondent and result in a lower response rate. An option would be to add a predefined answer 'no change' in each row, but this will require a lot of extra time of the respondent to complete each matrix question, which could result in prematurely termination of the survey. Figure 5 | 2 and Figure 5 | 3 illustrate the questions of change in residential location. The other life course events have similar questions.

4 | Routing

Each part of the survey had a header or an introduction page to give extra information about that specific part of the survey. The survey was designed with skip-logic or a specific routing. This means that certain irrelevant questions were automatically skipped.

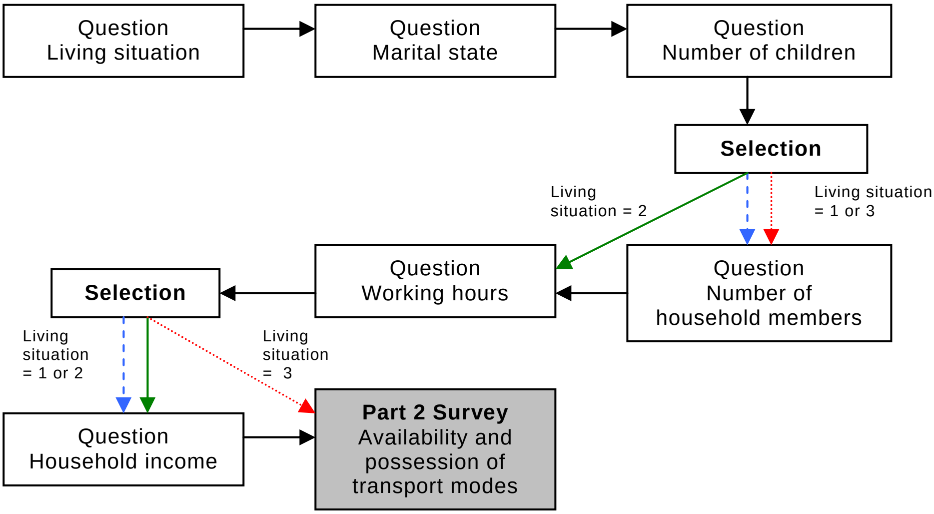


Figure 5 | 4: Routing part one (personal and household characteristics)

Figure 5 | 4 illustrate the general routing of the first part of the survey: personal and household characteristics. The routing depends on the answer of a specific question: living situation. The question about the number of household members is only relevant for respondents who live on their own (answer 1) or live with their parents (answer 3). The question about household income is only relevant for respondents who live on their own (answer 1) or occupy a student room (answer 2).

Figure 5 | 5 illustrate the routing of the second part of the survey: availability and possession of transport modes. This routing scheme illustrates that certain questions were skipped if the answer to the previous question was ‘no’. The first question was possession of driver’s licence. If the answer was ‘no’, the question about the number of year driver’s licence is irrelevant and was skipped. The answer to the third question about possession of one or more car within the household determined if the rest of the questions regarding car possession and availability were skipped or not.

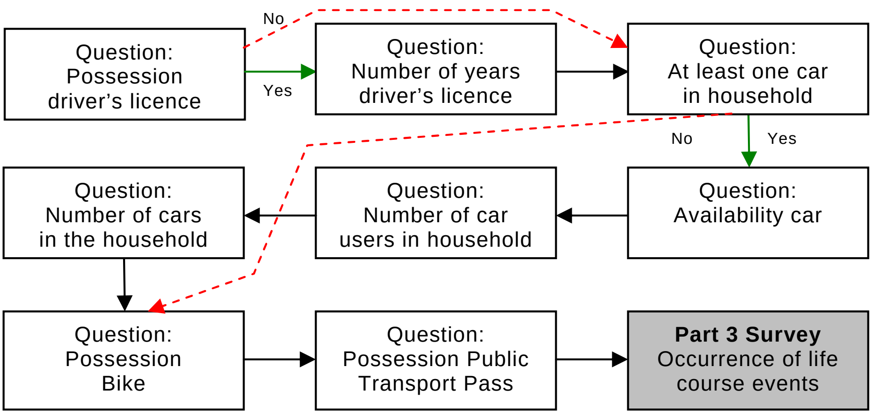


Figure 5 | 5: Routing part two (availability and possession of transport modes)

Figure 5 | 6 illustrates the routing of the third survey component, the occurrence of life course events. This routing is more complicated and consists of three columns in the figure. The first column illustrates the selection criteria, the second column shows the last part of the introduction question (since …, see Figure 5 | 2) and the possible answers, and the third column indicates the event matrix questions. Only if the answer was 'at least once' to the event question, respondents filled in the corresponding matrix question (third column). Otherwise respondents skipped the matrix question and answered the next life course introduction question. Sometimes an event matrix was automatically skipped, depending on the answer of the question about their living situation (part 1, personal and household characteristics). For example, respondents who lived with their parents (living situation = 3) or occupied a student room (living situation = 2) skipped the event matrix question about change in household composition. It is difficult to remember all changes in your parental household or student house and it is not relevant in this study. The skip logic for the life course event 'change in study location' was based on age (asked in part 1, personal and household characteristics, of the survey). Respondents born before 1969 skipped the life course question about changes in study. The skip logic for the life course event 'change in car availability' was based on the answer of the question about driver's licence (part 1, personal and household characteristics). The last part of the question, since.., was based on the living condition (part 1, personal and household characteristics). The routing was not visible for the respondents.

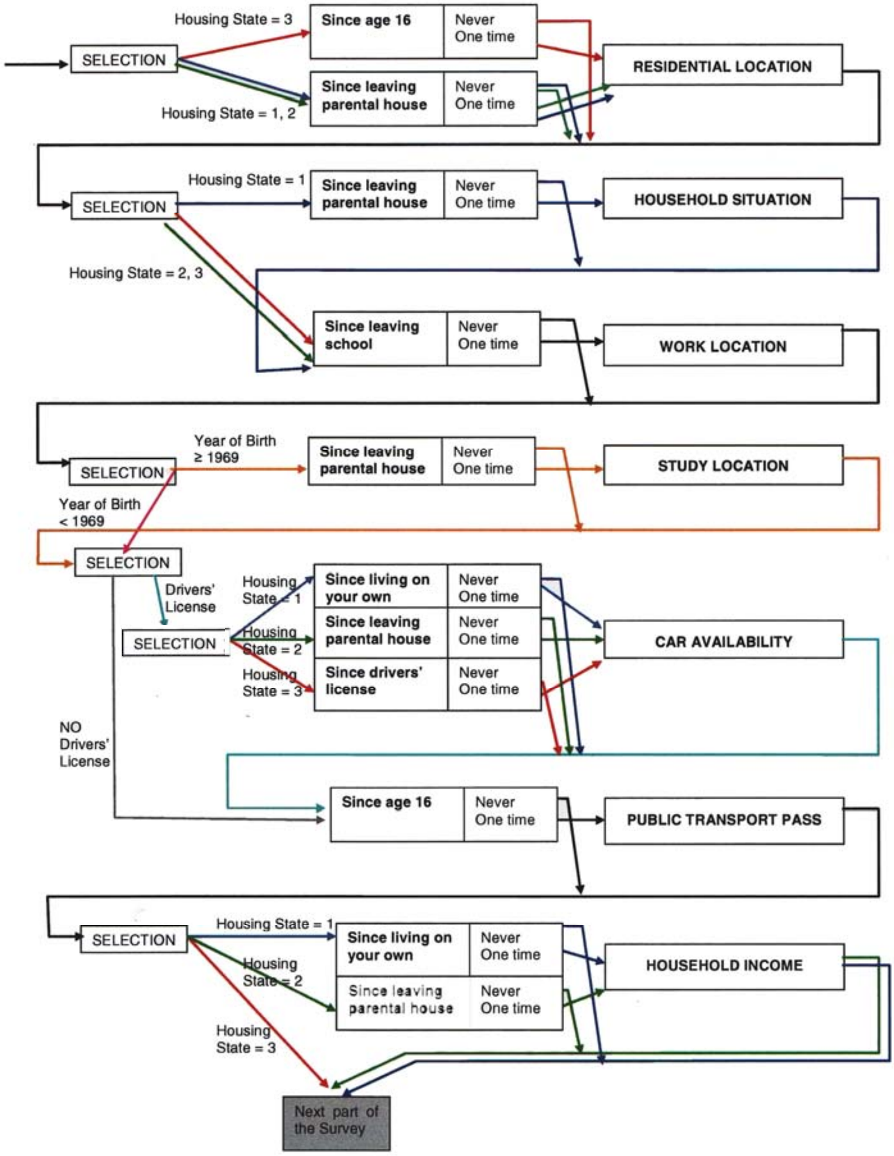


Figure 5 | 6 : Routing part three (occurrence of life course events)

5 | Response rates and sample characteristics

As mentioned before, a total of 939 respondents registered for the Internet-based survey. 807 respondents started the survey, while 710 respondents completed the whole survey. At two stages the respondents dropped out: 132 respondents did not start the Internet-based survey, and 97 respondents did not complete the survey. Therefore, two levels of nonresponse rate can be calculated: nonresponse at the first level, i.e. the percentage of respondents who registered for the survey but did not start. The first-level nonresponse rate was approximately 14%. The second-level nonresponse rate indicates the percentage of respondents who terminated the survey before it was completed. This rate is approximately 12%. In total, 75% of the registered respondents completed the survey. This is a relatively high percentage. Only the completed surveys (710) are used in the analyses.

As for sample composition, 59 percent were males, while 41 percent were females. This is in line with previous findings that Internet-based samples tend to be biased in the sense that males are overrepresented in the group of internet users. In total, 51.1 percent had a full time job (> 35 hours a week), 9.7 percent had no job, while the remainder had a part-time job. Of all respondents, 59.4 percent has no children, 6.8 percent has one, and 23.1 percent has two children, while the remainder has more than 2 children. 59.9 percent of the respondents is married or lives together with a partner. Almost everybody owns a driver's licence (93 percent) and 78 percent of the respondents owns at least one car. The possession of a bike is also very high (97.2%), and half the sample owns a particular public transport pass.

The sample is not completely representative of the Dutch population (Centraal Bureau voor Statistiek, 2004). The frequency of the personal characteristics age, education and household income differ. Younger respondents are overrepresented in the sample and the elderly are underrepresented. The use of an Internet-based survey in addition to the sampling procedure may have contributed to this finding, since not all elderly people posses a personal computer and have access to Internet in contrast to most younger people. Respondents with a higher education are overrepresented and respondents with a lower education are underrepresented. There is also a slight difference in car possession between the sample (78%) and the Dutch population (76%)

in 1998, but that might be explained by the fact that car possession has been increasing.

5 | Cleaning data

The data cleaning process consisted of the following steps. First, the chronological order of the life course events was checked, and the missing data in the life course part were corrected where possible. If it was not possible to complete the information, the occurrence was deleted. In addition, the consistency within the life course events and the consistency between the personal characteristics and the life course event were checked. The data in all cases with inconsistency was corrected. There were no cases deleted.

1 | Chronological order events

The first occurrence (first row) in the life course matrix was the most recent occurrence (see Figure 5 | 3). The second row was the occurrence before the most recent one, and so on. Not all respondents answered the questions about the occurrences of life course events in this particular sequence. Some respondents organised their occurrences the other way around. In some cases, two occurrences in the same year were entered in the wrong sequential order, based on month. The chronological order was checked for all respondents and all errors were corrected.

2 | Missing data

As mentioned before most of the questions were mandatory, except the life course matrices (see Figure 5 | 3). If the life course question matrix was made mandatory, respondents must indicate ten occurrences and must answer all questions for each topic and occurrence. A disadvantage of the setting that the life course matrix question was not mandatory is that respondent could skip one or more topics of an occurrence. For example, the respondent answered the question about month, year and type of the occurrence, but the respondent left the question about before situation and after situation blank. Unfortunately, this resulted in missing data.

There were no serious problems in most cases. Sometimes respondents did however not provide all information. The before and after situation should be consistent, for example before situation A (most recent occurrence) should match after situation B (next occurrence) and so on. These errors can easily be corrected and that was done for all incorrect cases.

Correcting missing data in the column month, year or type is more difficult and sometimes impossible. The missing data was corrected where possible to complete one occurrence and row in the life course matrix. If that was not possible, the occurrence was deleted.

3 | Consistency of data

The respondents answered some questions about their personal characteristics in the first part of the survey. For example, respondents indicated their living situation: (1) independent living, (2) student living, of (3) parental living. The answer to this question should of course be consistent with the last and thus most recent, registered life course occurrence of the corresponding event in part three of the survey. For example, the housing event had eight types of occurrences which can be recoded into the three living situations. All transformations from answer categories into states of life course events are described in Appendix 2.

To check the consistency, the personal characteristics were compared with the most recent occurrence of the corresponding life course event. In principle, there are two solutions if there was an inconsistency between the personal characteristics and the most recent occurrence:

1. correct the personal characteristics according to the most recent occurrence (routing inconsistent)
2. correct the most recent occurrence according to the personal characteristics (sequence occurrence)

The first solution was chosen for all seven life course events except for the housing event. If the data for the housing event was corrected according to solution one, the routing of the survey was not correct, i.e., inconsistent. For example, a respondent answered the question about his living situation with the answer student living. According to the most recent occurrence in the housing event his living situation was independent. The respondent skipped the question about the life course event household because of his answer to the question living situation at the personal characteristic part of the survey (see Figure 5 | 6). Correcting the personal characteristics in this case will change the routing of the respondent and this will result in more problems, like missing data for other life course events.

6 | Quality of the retrospective data

The performance of the retrospective survey instrument and administration procedure was assessed in terms of item non-response, error checking and data cleaning, and by applying a binary logit model of the effect of memory on reporting of events.

1 | Item non-response

When interpreting item non-response in the context of retrospective surveys, one should realise that respondents may skip questions for a number of reasons: (1) by mistake, (2) refuse to answer or (3) unable to provide a correct answer. This may be caused by a problem in the question-answer process (e.g., by not understanding the question or being unable to retrieve the necessary information), the lack of motivation of the respondent, the topic of the question (e.g., sensitive issues), or badly designed surveys. Note that missing data that is caused by the inability to recall the relevant information is of concern here. The matrix questions were not mandatory (see Figure 5 | 3), so there exists item non-response in the data.

To analyse item non-response, a rate was calculated for every event type. For example, 602 respondents reported information about a recent change in their housing situation. In the matrix (Figure 5 | 3) they had to answer questions about month, year, before situation, after situation and type of change of that most recent occurrence. For example, 125 respondents left the question about the most recent occurrence of the bought or rented house blank for different reasons (e.g., forgot or could not recall etc.). This resulted in an item non-response rate of over 20 percent indicated in the left column of Figure 5 | 7. For event type, every question in the matrix and for every occurrence the item non-response rate was calculated. The rates are shown in the different graphs, Figure 5 | 7 - Figure 5 | 13. The categories displayed on the x-axis are related to different questions that were posed for each reported event. The block of columns associated with each question, represents the ten possible occurrences of a change in every event type.

Note that the graphs show a percentage. The item non-response was calculated as the ratio of the number of cases with item non-response and the total number of cases for that change. In one extreme situation, related to the tenth occurrence for an event type, a single respondent caused the high percentage of item non-response. This respondent left blank the question about the type of change for the tenth occurrence of the car availability event. Because only one respondent reported ten occurrences for this event, the item non-response rate is here 100%. This extreme case is not shown in Figure 5 | 11.

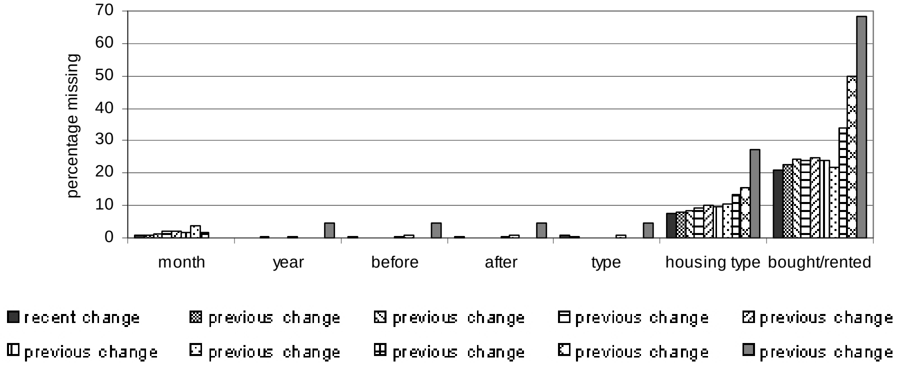


Figure 5 | 7: Item non-response housing event

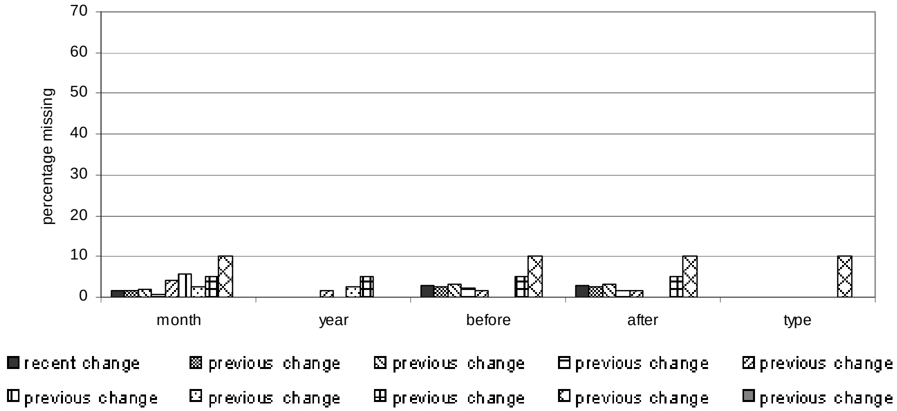


Figure 5 | 8: Item non-response household event

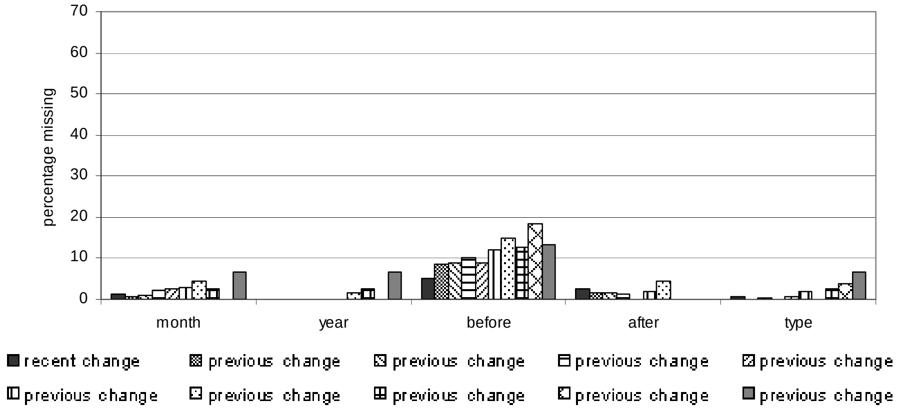


Figure 5 | 9: Item non-response of work event

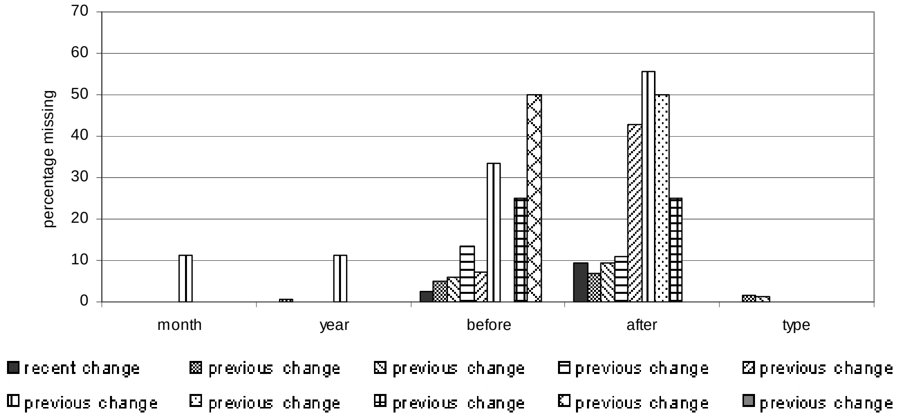


Figure 5 | 10: Item non-response of study event

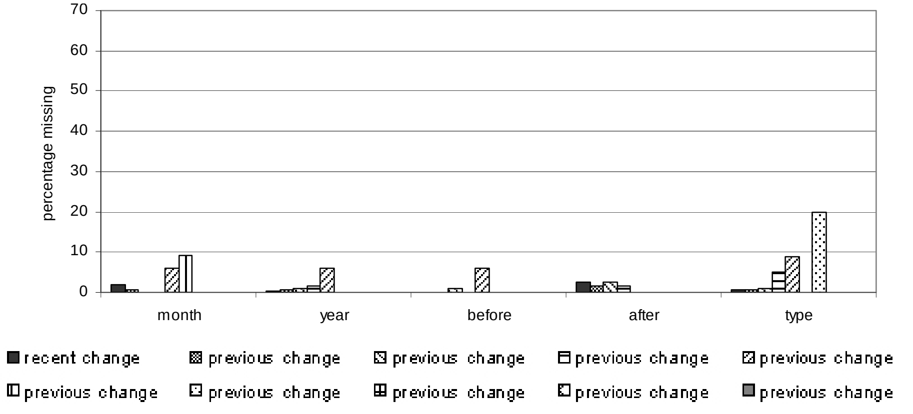


Figure 5 | 11: Item non-response car availability event

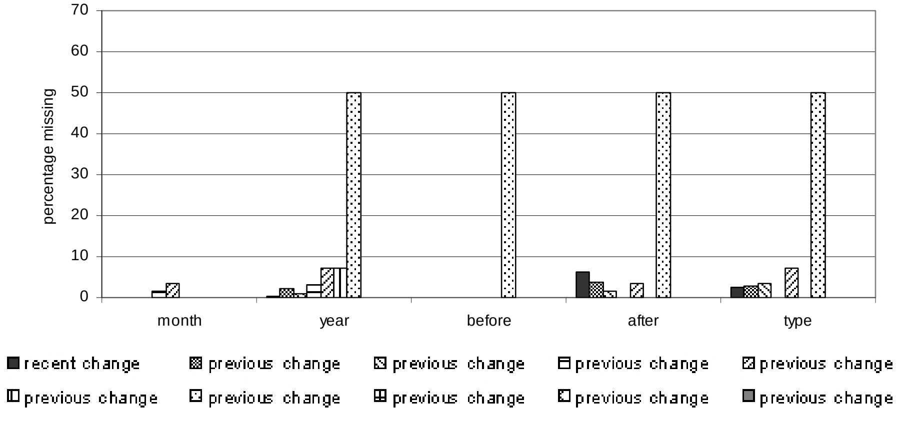


Figure 5 | 12: Item non-response PT pass event

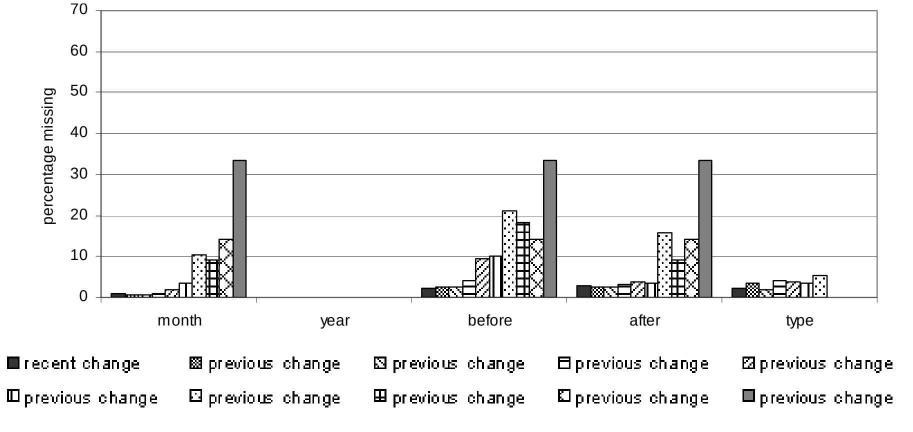


Figure 5 | 13: Item non-response household income event

All graphs indicate that in general the item non-response is low. However, the first graph (Figure 5 | 7) suggests some substantial variation in item non-response, depending on level of detail. For example, the nonresponse rate for the questions about housing type and owned/rented house is higher than the nonresponse rate for the other questions month, year, before and after situation, and type of change. It suggests that the ease of recalling the various aspects of events differs within and between different types of events and displays a tendency to increase over the number of events reported from most recent to the oldest. It seems that respondents had somewhat more problems recalling the specific information about some events that one would assume less memorable. Apart from ease of recall, perceived burden of providing the answers could have played a role. The housing, work and study event had open questions about the before and after situation instead of the drop down lists.

Thus, item non-response exists in our retrospective data, but overall, the percentage is relatively low, and more importantly, item non-response tends to decrease with increased salience of the event. Based on this performance criterion, one can conclude that indeed one does not need to be over concerned about respondents' ability and willingness to recall information about major life course events.

2 | Error checking and cleaning

A second aspect concerns the quality of the response. Because errors and inconsistencies exist in all kinds of data, the key question here is whether the survey displayed any abnormal amount of inconsistency. To address that question, several error checking possibilities were considered. In the matrix question, errors possibly may appear as: (1) inconsistencies between the before situation of the most recent change (first row) and the after situation of the previous change (second row), and (2) changes were not indicated in the right order, from recent to previous changes. Possible causes for these errors are given and a way to fix these errors is described.

Table 5 | 1: Incomplete data (year and type of change)

	Housing	Household	Work	Study	Car	PT	Income
Cases corrected	2	3	8	4	16	28	16
Usable cases	4	2	5	0	1	0	0
Unusable cases	6	0	2	0	4	12	2
Incomplete cases (total)	**12**	**5**	**15**	**4**	**21**	**40**	**18**

Findings related to item non-response have already been reported above. This further analysis is aimed at checking whether the missing information could be corrected based on other information available in the data. Of course, such consistency checks not only apply for respondents with missing information, but for all respondents.

Table 5 | 1 indicates the number of incomplete cases for each event based on the answers to the questions about year and type of change, which is the sum of cases that could be corrected, cases that are usable for analysis and cases that are unusable. Ultimately, the maximum number of unusable cases was only 12, which is a small percentage of the total number of cases, 710. It suggests that overall the quality of the surveys seemed at least satisfactory.

3 | Memory and recording of events

As mentioned before, a critical factor influencing the usefulness and reliability of retrospective surveys concerns the question how well respondents are able to retrieve events they experienced in the past from their memory. It is hypothesized that it will depend on the nature of the event and on time elapsed between moment of recollection (i.e., the moment of completing the survey) and the moment a particular event happened. More specifically, it is expected that the probability of a reported event will decrease with increasing time elapsed between the moment of occurrence and the moment of the survey.

To test this hypothesis, a binary logit model was estimated to predict the probability of reporting an event at a certain year in the past (the queried year). For each respondent and each event category, each year in the past constitutes an observation of whether or not the respondent reported an event of that category for that year. This means, for example, if a respondent could look 15 years back in time regarding a particular event, the data consist of 15 observations for that respondent and event. A univariate analysis where elapsed time is the only explanatory variable in the model is, however, not adequate for testing our hypothesis. With varying elapsed time (history) also the age of the respondent at the queried year varies simultaneously and, obviously, age may have an effect on occurrence probability of an event as well. Therefore, to correct for an age effect, the age of the respondent at the moment of the queried period (age-event) was also included as an independent variable in the model. Finally, education and gender were included to correct for demographic attributes.

Although in this way a correction was made for an ‘age-at-the-queried-period’ effect, it should be noted that there is another possible effect of age. That is, if the elapsed time is varied, while keeping age at the queried period constant, the current age of the respondent is simultaneously varied with the cohort group to which the respondent belongs. For example, if older people of today experienced an event less often when they were young compared to younger people when they were young, then the elapsed-time variable would have a negative effect on the probability of the event being reported. Clearly, this would not be a memory effect but a cohort group effect. If it were possible to include current age as an explanatory variable in the model as well, then this cohort effect could be corrected. Obviously, however, this would give estimation problems since a linear relationship exists between elapsed time, age at queried period and current age. Thus, a correction can be made for age at the queried period, which may confound a memory and cohort effect or a correction can be made for current age, which may confound a memory effect with an age-at-queried-period effect in the estimated coefficient for elapsed time. This should be kept in mind when interpreting the results.

Table 5 | 2 represents the results of the binary logit model. The first row of each life course event lists the parameters estimated in the model. Each second row of an event shows the *p*-values.

Table 5 | 2: Results binary logit model

	History	Age-event	Education	Gender	Constant	Chi square	Log likelihood
Housing	-0.002	-0.064	0.049	0.103	0.457	776.15	11111.610
	0.343	0.000	0.058	0.000	0.000		
HH	-0.014	-0.022	-0.017	0.003	-0.949	56.25	8725.90
	0.000	0.000	0.583	0.933	0.000		
Work	0.003	-0.046	0.027	0.203	-0.124	301.59	8972.87
	0.231	0.000	0.361	0.000	0.266		
Study	0.011	-0.124	0.068	0.120	0.757	148.16	3526.20
	0.368	0.000	0.122	0.013	0.030		
Car	-0.014	-0.039	0.001	0.136	-0.637	120.76	5444.28
	0.001	0.000	0.985	0.001	0.000		
PT	-0.090	-0.041	-0.057	0.184	-0.482	607.27	7219.10
	0.000	0.000	0.099	0.000	0.000		
Income	-0.010	-0.040	0.104	0.188	-0.429	161.26	6125.54
	0.009	0.000	0.005	0.000	0.001		

As it appears, the history parameter for the events; household composition (household), changes in car possession and availability (car), possession of public transport pass (PT), and household income (income) are all negative and significantly different from zero. For the events housing (i.e., change in residential location), work and study the history parameter is not significantly different from zero.

These results indicate that memory may have an effect in reporting of events in case of the household composition, income and transport mode related events, whereas it does not seem to play a role in case of housing, work and study related events. The negative effect on reporting probability seems to be largest for the public transport pass events (parameter = -0.090). In case of household, car and income the effect of history is substantially lower than the effect of age-event. Thus, this analysis suggests that changes in public transport (PT) pass are not that memorable. An alternative cohort explanation would be that, over the years, availability and use of public transport passes have increased (in the Netherlands) and, with that the probability of events related to these PT passes also increased. Keeping in mind the possible confounding of cohort and memory effects, it can be noted that nevertheless these findings are consistent with the assumption that many life course events are relatively easy to retrieve from memory, supporting the potential value of retrospective surveys.

7 | Conclusion

If one wishes to include life course or life trajectory events in any analysis or model of activity-travel patterns, data on such events should be collected. Conventional data collection approaches such as (quasi-)longitudinal personal travel data collection methods, including panel surveys, repeated cross sectional surveys, and cohort pseudo-panel surveys are typically not collected as part of national surveys and moreover require substantial financial resources to administer. Potentially, therefore, retrospective surveys, especially when administered through the Internet, are a good alternative. One would expect that the quality of data coming from a retrospective survey depends on the nature of the event about which information is collected and on the time elapsed between the occurring of the events and the time of the retrospective survey.

This chapter has reported the operational decisions made with respect to the design and administration of an Internet-based retrospective survey, which was used to collect data about life course events. In addition, results of analyses conducted to assess the reliability and validity of the collected data are reported. The results showed that item non-response in general was relatively low, especially for those life course events that serve as markers unfolding one's life. Moreover, inconsistencies in the data were not very different from experiences with traditional travel surveys and most problems could be relatively easy fixed using other pieces of information in the survey. Finally, a statistical analysis, albeit not capable of avoiding confounding cohort and memory effects, indicated that memory / cohort effects were not found for the more salient life course events, such as housing, work and study related events.

The data of the retrospective Internet-based survey is used as input for the learning of the two Bayesian Belief Networks, life trajectory and mode choice. The modelling framework, the constraints and the data preparation are described in the next chapter. The results of learned Bayesian Belief Networks will be reported as well.

6 | Learned networks

1 | Introduction

In the previous two chapters, it has been argued that Bayesian Belief Networks constitute a potentially relevant approach for modelling the complex direct and indirect relationships between one or more life course events and transport mode choice decisions. In addition, it has been argued that instead of detailed panel data, a retrospective survey may have sufficient validity and reliability to collect data on life course events. Based on such data, collected through the Internet, the results of learned Bayesian Belief Networks will be reported in this chapter.

The chapter is organised into two parts. In the first part, attention is paid to the influence of time. This analysis serves as a primary test to examine whether temporal effects do play a role. The time influence, registered as time elapsed after an occurrence of an event, is examined using a multinomial logit model (MNL). The different variables used in the MNL model are described first, followed by a description of the results. The operationalisation of the Bayesian

Belief Networks and their estimating results are described in the second part. The second part consists of five sections. First, the modelling frameworks for respectively a life trajectory and mode choice network are outlined. The life trajectory network captures the relations between the life course events, current states and the history of life course events, while the mode choice model considers the link of mode choice with life course events and the states. The variables and the constraints used in structure and parameter learning are explained given the modelling frameworks. The principles of learning are described in chapter four of this thesis. In the next section the data preparation is described. The data of the Internet-based survey had to be restructured before learning of the networks could take place. The learned networks for life trajectory and mode choice are discussed in the next sections. The last section summarises the most important conclusions of this chapter.

2 | Time effects

To analyse the influence of time on mode choice, a multinomial logit model was estimated. Three types of independent variables can be distinguished. They are referred to as variables *X*, variables *D* and variables *Z.* The independent variables consisted of appropriately effect-coded personal characteristics and availability and possession of transport modes (variables X), distances to different destinations (variables *D)*, and time elapsed since a person experienced an event (if any) of each type distinguished (variables *Z*). The importance of using the current state (variables *X*) as independent variables is emphasized in this analysis. These variables capture the effects of a current state so that the parameters for time variables (*Z*) purely represent a time effect (e.g., inertia in adaptation to a new state). In sum, the variables *X* and *D* measure the influence of the existing state on mode choice (from S or S_2 on B, see Figure 3 | 3), whereas the estimated parameters of *Z* variables represent the time influence, the adaptation process.

The independent variables *X* (personal characteristics and possession and availability of transport) that are included into the choice model include gender, age, education, income, number of household members, availability of driver's licence, car possession, bike possession, and public transport pass possession. In addition to these characteristics, the distance from home to work, study, stores, shopping mall and sport centre are used as independent variables (D). The respondents estimated the distance to those different locations in a matrix question. The time effects (variables *Z)* were operationalised in terms of the following equation:

$$Z = E * Ln(T) \quad \text{Equation 6 | 1}$$

where *E* indicates whether an individual experienced a certain event. If someone experienced an event, the value is equal to 1, and if someone did not experience that event, the value equals 0. The variable *T* corresponds to the time elapsed (in months) since the last occurrence or change. In other words, it corresponds to the number of months that have passed since a respondent experienced that specific event. The log transformation improved the goodness-of-fit of the model.

The dependent variable in the multinomial logit model is transport mode choice. This variable has three different choice options: car, slow transport and public transport. The respondents provided information with respect to their current mode choice behaviour for five different trip purposes. In particular, they indicated for each purpose the frequency, travel mode, alternative travel mode, destination, departure time from home, arrival time at home, estimated travel distance from their home to the destination and estimated travel time. Respondents could indicate the trip frequency using their own 'scale' (day / week / month), and all frequencies were rescaled into a monthly frequency (the most used category) to calculate total frequency. The total frequency is represented by the sum of frequencies given a particular transport mode across different purposes. This total frequency is used as the dependent variable in the multinomial logit model.

Recall that the third part of the survey provided the event-history data for this part of the study. For the set of seven predefined events, respondents were requested to indicate whether they experienced the event, and, if so, how many times it took place, the timing of the event (month and year), what exactly changed by the event (before and after situation, for example two people and three people in the household) and the nature of the change that took place (for example adoption or birth). Based on this information, the independent variable *Z* was calculated using Equation 6 | 1.

3 | Time effect results

The multinomial logit model was estimated using LIMDEP (Econometric Software, 2003). Slow mode was used as the base mode for this multinomial logit model. The estimated MNL model includes 76 variables and 709 observations. After 7 iterations, no further improvement in goodness-of-fit could be obtained. The log likelihood of the model is -188768.7 and the Rho-Square is 0.487.

Table 6 | 1: Part worth utilities variables *X* and *D*

Variables *X*	**Utility Car**	**Utility PT**
Constant	**-2.3417**	**-2.2376**
Males	**0.0588**	0.0412
Females	***-0.0588***	*-0.0412*
Age (17-24 year)	0.0491	-0.0003
Age (25-34 years)	0.0416	**0.1732**
Age (35-49 years)	0.0393	-0.1111
Age (50-79 years)	*-0.1300*	*-0.0651*
Education (high)	**0.1761**	**0.2969**
Education (low)	*-0.1761*	*-0.2969*
Alone, divorced, widow	-0.0562	**-0.7331**
Married/living together	*0.0562*	*0.7331*
Independent housing	**0.3430**	**-0.4578**
Student housing	**-0.5563**	**-0.7418**
Living with parents	*0.2133*	*1.1996*

Table 6 | 1 continued

Household of 1 person	-0.0745	**0.6532**
Household of 2 persons	0.0103	**-0.4198**
Household of 3 persons or more	*0.0642*	*-0.2334*
No employed work	**-0.1712**	-0.0111
Part-time work	0.0173	-0.0257
Fulltime work	*0.1539*	*0.0368*
Income (medium)	**-0.0773**	**-0.0799**
Income (high)	*0.0773*	*0.0799*
Driver's licence	**0.5198**	**-0.3426**
No driver's licence	*-0.5198*	*0.3426*
Car in possession	**1.2185**	**0.1277**
No car in possession	*-1.2185*	*-0.1277*
Car always available	**0.7015**	**0.1375**
Car sometimes available	**-0.2603**	**-0.1199**
No car available	*-0.4412*	*-0.0176*
No car sharing	**0.3223**	**-0.2568**
Car sharing with 1 person	**-0.1183**	0.0300
Car sharing with 2 persons or more	*-0.2040*	*0.2268*
One car in possession	**-0.3683**	**0.2328**
Two cars in possession	*0.3683*	*-0.2328*
Owner of a bike	**-1.0477**	0.0448
No owner of a bike	*1.0477*	*-0.0448*
No Public Transport (PT) pass	**0.7660**	**-0.5798**
Student PT pass	-0.1256	**0.4947**
PT subscription	**-0.3967**	**1.5600**
Benefit hour pass	*-0.2437*	*-1.4749*
Variables *D*	**Utility Car**	**Utility PT**
Distance to work	**0.0170**	**0.0214**
Distance to education	**-0.0030**	**0.0107**
Distance to the store	**0.1834**	**0.1041**
Distance to shopping area	**0.0092**	0.0011
Distance to sport centre	**0.0032**	-0.0009

1 | Personal characteristics and distances

Table 6 | 1 shows the estimated effects for the variables included in the model. The effects of the *X* and *D* variables are listed as part worth utilities for mode choice car and public transport (PT). Significant variables (*alpha* of 5%) are in **bold** and the calculated effects are in *italic*.

Almost all variables are significant. Both constants, for car and public transport (PT), are negative, implying that, all other variables being equal, the modes car and public transport are less attractive than the slow transport mode (ST) after correcting for distance. All estimated effects are in anticipated direction.

Table 6 | 1 shows that the utility of the car decreases very slowly with increasing age. Figure 6 | 1 illustrates for example the utilities for the different transport modes for the variable age. Above 50 years of age the utility for the car becomes negative relative to that of slow transport. Public transport has for people between 25 and 34 years old a positive utility compared with slow transport. This effect can be explained by the fact that Dutch students own a student PT pass which allows them to travel for free on PT during weekdays or weekends. After 34 years of age, the utility of PT decreases with increasing age. Males have an above average preference for both car and PT, while females have a below average preference for these modes.

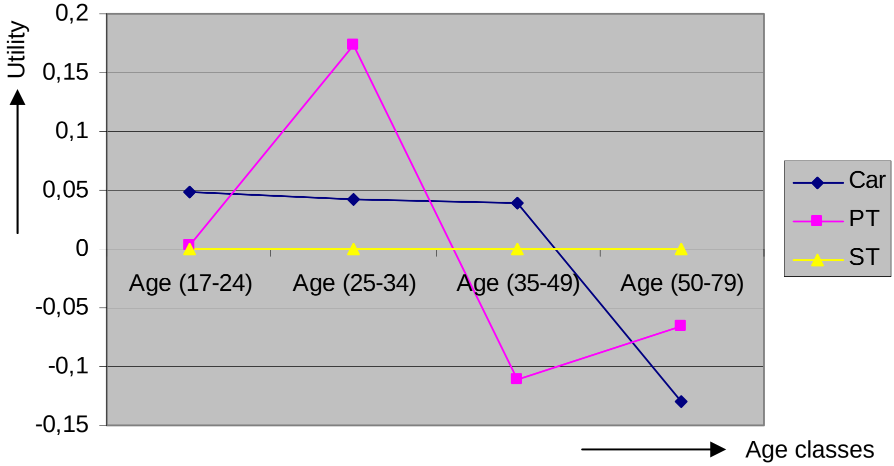


Figure 6 | 1: Utilities for the variable age

The effects of education are also as expected in the sense that higher education results in a higher utility for PT compared to the other modes. Three housing situations are distinguished: independent and student housing and living with parents. A person who lives independent has a higher than average utility for car mode and a lower than average utility of PT. A student has a negative utility for the car, while someone who lives with his parent has a positive car utility.

Table 6 | 1 also lists the utilities for the distance variables, almost all utilities are significant. A positive utility means that when distance to a destination (work, education, store, shopping area or sport centre) increases, the utility of the corresponding transport mode also increases. A negative utility indicates a negative influence of distance on the utility of that transport mode. With increasing distance to education the utility for the car decreases and for increasing distance to the sport centre the utility of PT decreases. These negative utilities can be explained by the fact that students often do not travel by car to school and the sport centre is often more accessible by car or slow transport compared to PT. With increasing distance to the store and shopping area the utility of the car and PT both increase. Similarly, with increasing distance to work the utility of car and PT also both increase. The estimated effects for distance are also in the anticipated direction.

2 | Time influence of events

The events considered were defined as follows: a change in residential location means that the respondent moves to a different residential location; a change in household composition means in this case an increase in the number of household members; a change in work location only included respondents with a job; a change in study location included students who started a new education or changed school/university; a change in car availability is an increase of number of cars or decrease in car users; a change in PT pass possession included respondents who possess a PT pass; and a change in household income is an increase in income. The results for the event variables are listed in Table 6 | 2.

Table 6 | 2: Part worth utilities variables *Z*

Variable	Utility Car	Utility PT
Change in residential location	-0.0427	-0.0415
Change in household	0.0331	-0.0586
Change in work location	0.0587	0.0801
Change in study location	0.0287	-0.0449
Change in car availability	-0.0232	0.0362
Change in PT pass possession	-0.0454	0.0962
Change in household income	-0.1230	-0.1129

The estimated effects are all significant. Figure 6 | 2 - Figure 6 | 8 display the effects of the different events. Slow Transport (ST) is the base in this logit model and its utility is always zero. For interpreting this graph, it is important to remember that a utility-line below the X-axis means that the more time that has elapsed since the most recent experience of event *i*, the lower the utility compared to the utility of ST. Similarly, when the utility-line is above the X-axis, an increase in elapsed time results in an increasing utility compared to that of ST. The result of experiencing the first life course event, change in residential location, is a new residence in a new environment. This change affects behaviour of the person involved. Figure 6 | 2 represents the time influence of this change.

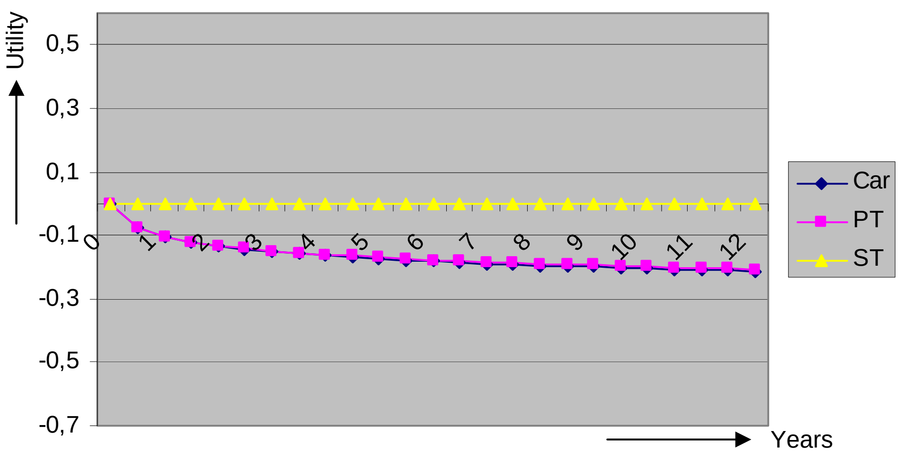


Figure 6 | 2: Change in residential location

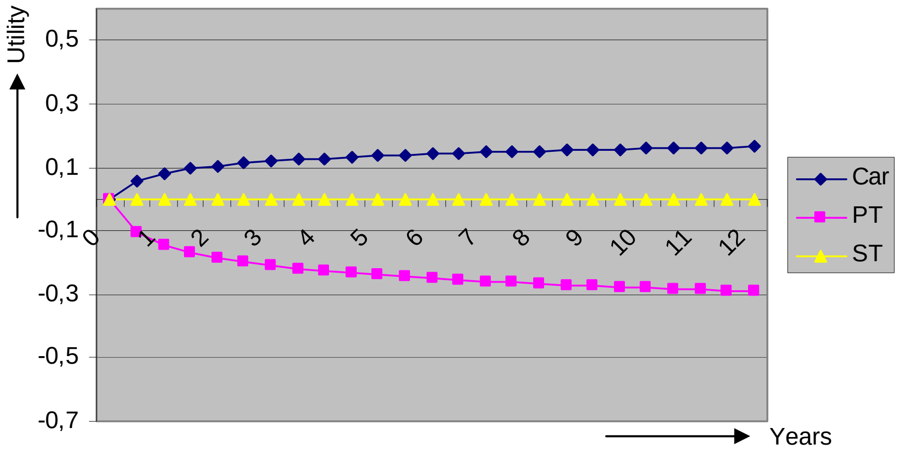


Figure 6 | 3: Change in household composition

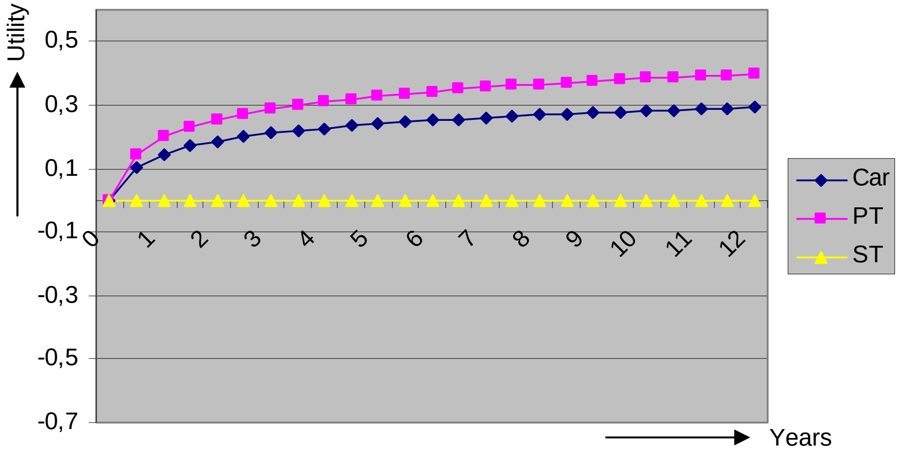


Figure 6 | 4: Change in work location

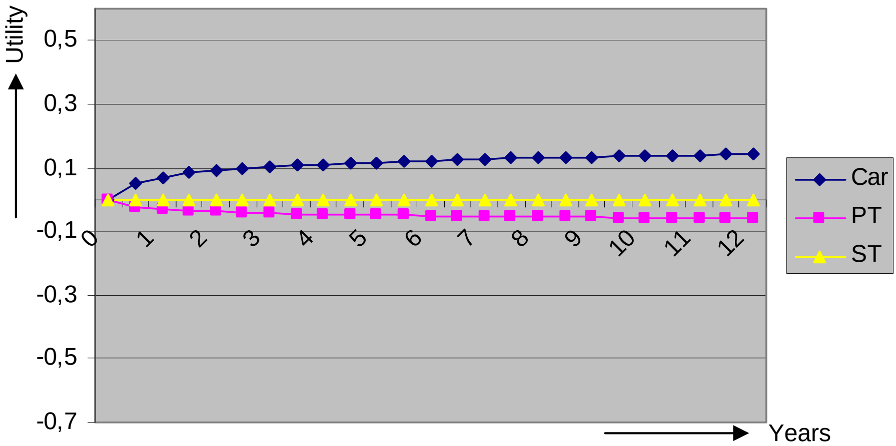


Figure 6 | 5: Change in study location

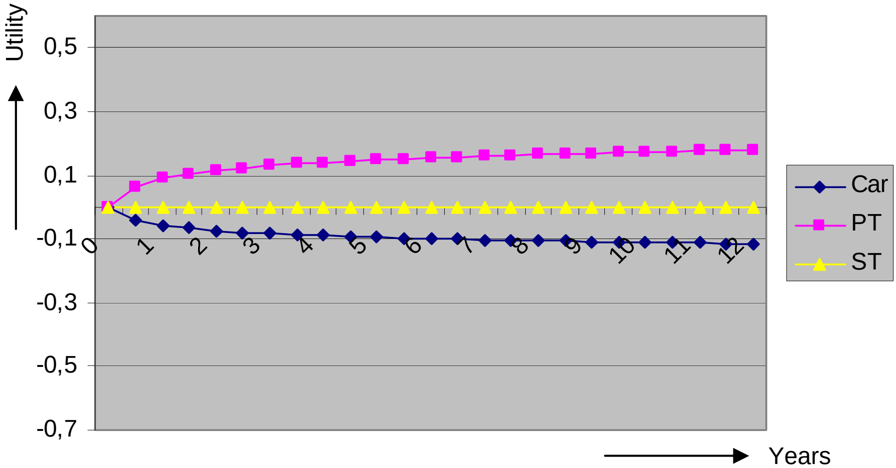


Figure 6 | 6: Change in car availability

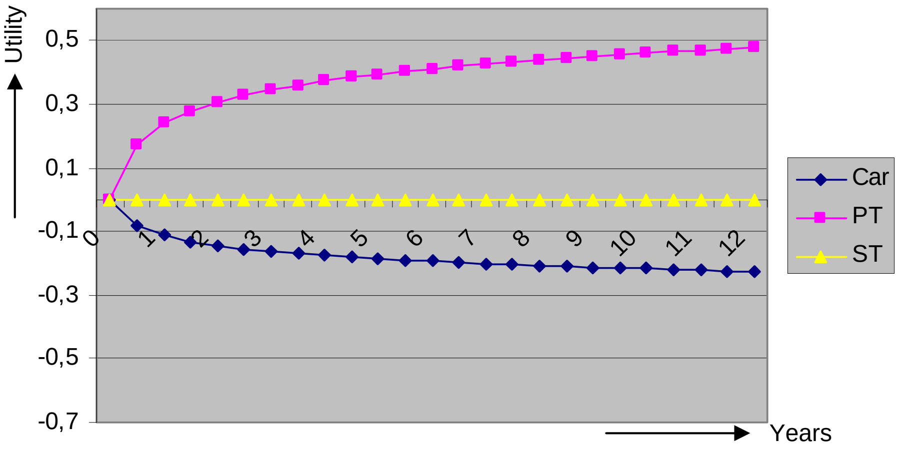


Figure 6 | 7: Change in public transport pass

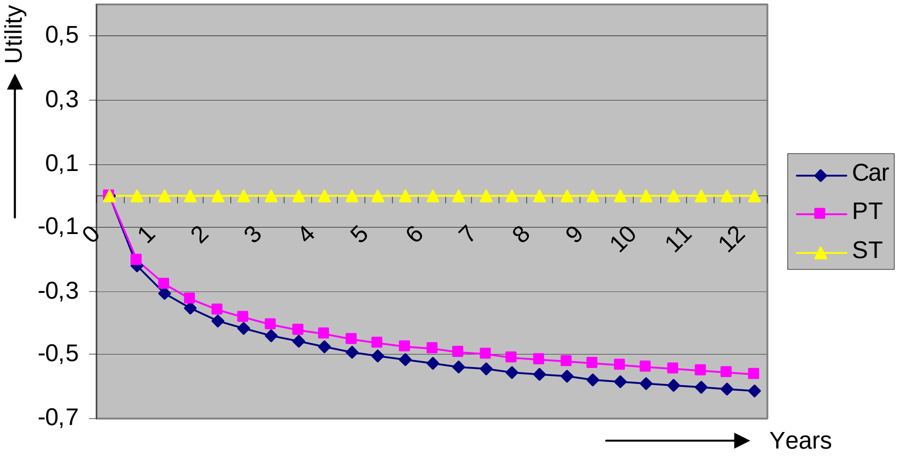


Figure 6 | 8: Change in household income

The transport modes car and public transport (PT) both have a decreasing utility compared to slow transport (ST). A possible explanation for this effect may be found in the adaptation process: first a person takes the car to every destination and after a while he/she slowly learns about the new environment and probably knows where everything is in the direct environment. After that learning process he/she will adapt his or her behaviour to the new circumstances and use slow transport more frequently.

An increase in the number of household members results in a larger household. Again, the change in state itself is represented by the *X* variables (household of one person, household of two persons and household of three or more persons). Table 6 | 1 shows that an increase in household size results in an increase of utility for the car and a decrease (from one to two persons) or increase (from two to three or more persons) of utility for PT. The adaptation process for this change is shown in Figure 6 | 3. The utility of the car increases and the utility of PT decreases, when time since the experience of the event increases. This suggests that households only slowly change their behaviour towards the new equilibrium whereby the car is used more frequently compared to ST or PT.

The time effect of a change in work locations is illustrated in Figure 6 | 4. The utility of the car and PT increases when the time after experiencing a change in work location increases. Again, this suggests that people tend to stick to their (old) behaviour and only slowly adapt their behaviour to the new circumstances.

When a change in study location occurred longer ago the utility of the car increases and the utility of PT decreases, as represented by Figure 6 | 5. A possible explanation is as follows: a student in the Netherlands receives a student PT pass to travel for free. After starting an educational program or changing university, people may at first overreact to the change resulting in very frequent use of the PT. After a while they adapt and partly switch back to the use of the car.

Figure 6 | 6 display the results of a change in car availability. It indicates that the utility-curve for the car is below the X-axis, meaning that an increase in time after experiencing a change in car availability results in a decreasing utility compared to that of ST. The utility-curve for PT is above the X-axis, which means an increasing utility with time compared to ST. In this case a change in car availability means either an increase in the number of cars or a decrease in the number of car users. Again, the effect of the change in state (number of cars or car users) on transport mode choice behaviour is given by the parameters of the X variables in Table 6 | 1. The graph of Figure 6 | 6 represents the time effect of this change. The negative effect of time after the event on car use may indicate that people overreact to the increase in car availability and, then, partly return back to slow mode as time passes.

The adaptation process after a change in PT pass, in this case buying or receiving a Public Transport pass, is revealed in Figure 6 | 7. Time after the event has a positive effect on PT and a negative effect on the car compared to ST. The effect of the change in state (from no PT pass to student PT pass/PT subscription/benefit hour pass) on transport mode choice behaviour is given by the parameters of these *X* variables in Table 6 | 1. People possessing a PT pass use the PT more frequently and the car less frequently compared to ST. Thus, the time effect suggests that people only slowly adapt their behaviour to the new state.

Figure 6 | 8 illustrates the adaptation process after experiencing an increase in household income. Time has a negative effect on both car and PT use. This suggests that people respond with an immediate big increase in car or PT use to the state change and after a while partly return to slow transport.

To summarise, all figures show significant temporal effects for the life course events on mode choice, after controlling for the other variables. This analysis suggests that there are two different ways in which people may react to a change (new situation): first people can overreact and after a while they (partially) return to their old behaviour and, second, the adaptation takes time and people slowly react on the change. All seven life course events had significant temporal effects. Moreover, the effects are interpretable as particular patterns of adaptation provided, of course, that all co-variants were controlled for by the *X* variables and possible effects of aging are negligible within the time spans considered. This provides evidence for the hypothesis that transport mode choice behaviour is dynamic and that these dynamics are associated with life course events, at least in a statistical sense.

4 | Modelling framework life trajectory

As discussed in previous chapters, choice behaviour is assumed to be context dependent. A change in the living conditions of a person, like moving to another house, city or country, will result in a new context which could impact the transport mode choice for all/certain trips. A life course event changes a certain personal situation, i.e., state. Besides the assumption that a life course event changes a person or household-level state, it is also assumed that events may influence each other. When an individual starts with a new job in a different location he/she might consider moving to that new location sooner or later. In that case the occurrence of a change in work location could trigger a change in residential location. Another example of influences between life course events is a possible effect of a change in work, for example promotion on income. The modelling framework described in the third chapter is transformed into two formal Bayesian Belief Network models for network learning, illustrated in Figure 6 | 9 and Figure 6 | 10. The first network models life trajectories, while the second network is concerned with the impact of these events on transport mode choice. The reason for estimating two networks rather than a single network including all nodes is that different sets of observations were available for life trajectory events and mode choice. Mode choice is only available for the base year (2004, the year of data collection). Mode choice was not collected in retrospect, because the reliability of such data is doubted, especially when linked to the events.

Figure 6 | 9 shows that two types of variables are included in the life trajectory network: (1) personal characteristics, such as gender and age, and (2) variables related to life course events (set of nodes within the rectangles). The variable of interest is occurrence event. A year is chosen as unit for both networks, this refers to one year of a persons life trajectory. Bayesian Belief Networks have discrete variables, for this reason the time in the life trajectory had to be classified. Too much detail, for example a unit of one month, results into a few observations within each unit. A unit of more than one year results in a lot of observations within one unit. This will not improve the sensitivity of the network. Besides these statistical arguments, the item-nonresponse of year, reported in the Internet-based survey, is lower than the reported month (see chapter 5 | 6 | 1). This indicates that the respondents can better recall the year in which the event occurred than the month. The month is probably too detailed, especially for occurrence that respondents experienced longer ago.

The other variables are referred to as external variables. The life course events are defined with respect to a certain moment in time, *t*, a certain person and type of event (using some classification of events) for which the model intends to predict whether or not a change occurred as a particular instance of an event of that type. The node occurrence event defines whether there was a change at time *t* for that certain event and what kind of change it was. The existing state before time *t*, e.g., residential situation in case of a housing event, is represented by the node labelled state. Time ago (A) and time ago (B) define the time elapsed since the last occurrence of certain types of changes A and B, respectively before *t*, such as for an example a decrease (A) or increase (B) of the number of family members in a household.

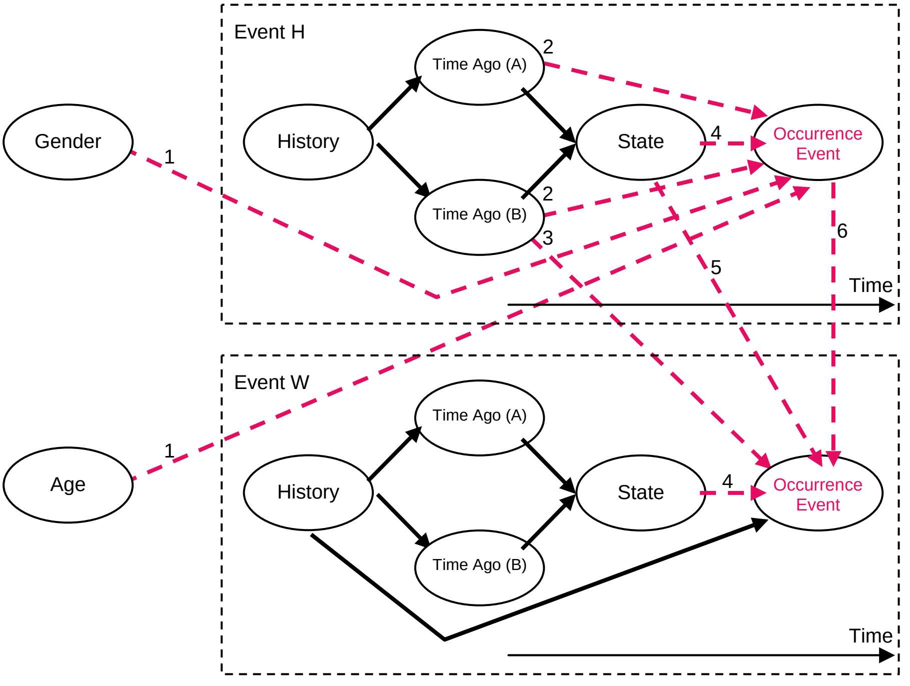


Figure 6 | 9: Basic structure of the network for life trajectories

Of course, the general model, where needed, can be extended with more classes, C, D, etc. The history node indicates for how many years ago information is available about that event for the person considered. This variable defines the range of the time ago variables in the sense that the time ago (from *t*) of the most recent event cannot be longer than the history. If no event occurred in this time frame then the state of time ago is defined as 'never' (meaning not in the time frame defined by history). Figure 6 | 9 illustrates two events, H and W. More events can be included and structured in the same way. The dashed arrows represent possible links that could emerge from the learning algorithm (structure learning). Note that this includes cross-relationships between events as well.

The purpose of the network is to predict for a person if an event will occur at a certain time *t* (node occurrence event) based on the other nodes in the Bayesian Belief Network. It is assumed that both the existing state before time *t* (state) and the last occurrence of an event (time ago) may influence the probability of the occurrence of an event on time *t*. If there is a (positive) effect of time, the probability of the occurrence of an event at time *t* increases. On the other hand a person can also remain in a certain state and a change becomes less likely, due to a stronger commitment to this situation (Elder, 2000).

Constraints were included in structure learning to prevent that directions of links are back-in-time (since links represent causal relationships they, logically, cannot run back in time). Furthermore, the following constraints are defined on the basis of a distinction between external variables (those that are observed when a prediction is to be generated) and target variables (those that are to be predicted). In this case, the target variables are the occurrence event nodes and the other nodes are the external variables. All external variables can only have a link with target variables occurrence event within an event or across an event. Links between the external variables are not allowed. The external variables age and gender can only have *outgoing* links to the target variables. This is done to prevent the creation of large underlying conditional probability tables, while the direction of links have no implications for predictions. The possible links, between external variables and the variable occurrence event, are represented in Figure 6 | 9 with dashed lines.

There are six possibilities for these links (numbers correspond with numbers in Figure 6 | 9 and the subscripts *H*, *W* refer to event *H* and event *W*):

(1) age / gender → occurrence event

(2) time ago^{H} → occurrence event^{H}

(3) time ago^{H} → occurrence event^{W}

(4) state^{H} → occurrence event^{H}

(5) state^{H} → occurrence event^{W}

(6) occurrence event^{H} → occurrence event^{W}

Furthermore, links that represent logical relationships are enforced. The logical links are represented as black lines in Figure 6 | 9. The logical relationships include the following: first, since history determines the range of the time ago variables, there is a logical relationship between history and each of the latter variables. Second, there is a logical relationship between history and occurrence of the event. If an event cannot possibly occur for a given person (for reasons described in Chapter 5), then the node history is undefined for that person and that specific event. 'Not applicable' is coded as a separate class (i.e., state) of the history node. This means that there exists a logical relationship in the sense that an event cannot occur if history is undefined. For each event type, the above variables and logical relationships hold.

The nodes are described in general terms in the schematic representation of the network (Figure 6 | 9). The learned Bayesian Belief Network included all seven life course events: change in residential location (in short, housing), change in household composition (household), change in work location (work), change in study location (study), change in car possession and availability (car availability), change in availability of public transport pass (PT pass), and change in household income (household income).

5 | Modelling framework mode choice

The second network is an extension of the first network, the variable mode choice is included in this network (Figure 6 | 10). As mentioned before, choice behaviour is context-dependent, this holds for mode choice behaviour as well. In this project, it is assumed that the context is represented by a person's life trajectory. For this second network the following constraints are imposed. The learned links in the first network are given and used as input for the structure learning of the second network. The variable mode choice is the variable of interest here and it can only have *incoming links*. The remaining variables are external variables. The node mode choice can have links with all external variables, except the history nodes. The possible links, between external variables and the variable mode choice, are represented in Figure 6 | 10 with dashed lines.

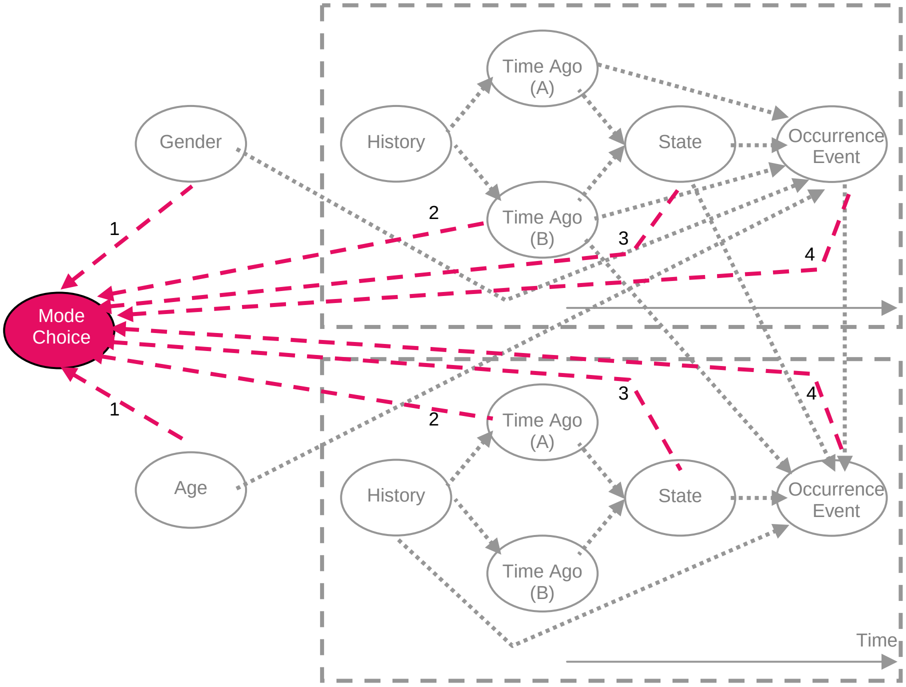


Figure 6 | 10: Basic structure of the network for mode choice

There are 4 possibilities (numbers correspond with numbers in Figure 6 | 10):

(1) age / gender → mode choice

(2) time ago → mode choice

(3) state → mode choice

(4) occurrence event → mode choice

6 | Data preparation

This section is inserted to explain the preparation of the data before it can be used for modelling purposes. Unfortunately, the data from the Internet-based survey cannot directly be used as input for the programs used for structure learning (HUGIN Expert, 1995). Sometimes the data had to be restructured in a different way compared to the data output of the Internet-based survey. The different steps in restructuring the data are explained given Figure 6 | 11. This figure illustrates the sequence of the various programs that were used. The input and output files will be briefly discussed. The most important part of the data preparation is restructuring the data into a life trajectory.

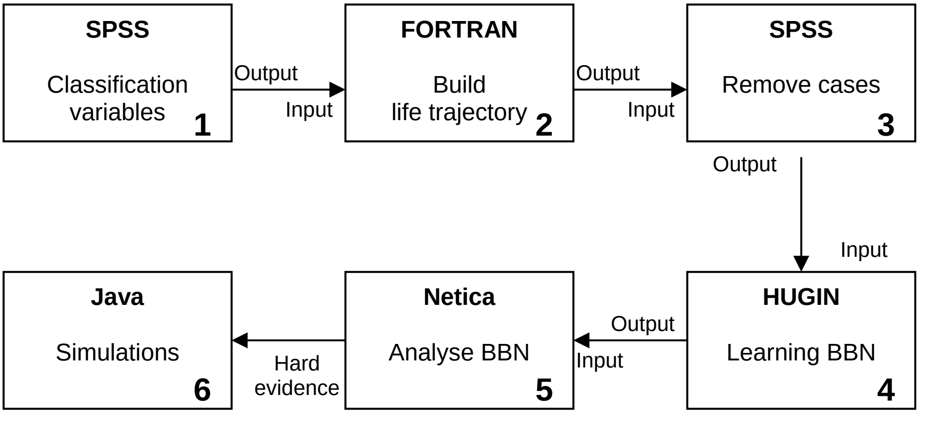


Figure 6 | 11: Sequence of the used programs

First, the cleaned data is classified into new variables. The classification of variables is described in Appendix 2. The output file consists of the following variables: age, gender, mode choice, life event (year, type, before and after), current state, eventmax, max.

The second step is building a life trajectory using FORTRAN (Silverfrost, 1999). The output consists of the following variables: age, gender, mode choice (only for 2004), occurrence events, state, time ago, and history. Using these variables, transitions probabilities for one year were derived and served as input for the BBN learning. There are seven input- and output files, one for every life course event.

The seven output files were merged into one file in step three. Besides merging the files the cases with missing values were deleted from the file. Sometimes there was no information available in a certain year causing missing data in the FORTRAN output file.

The fourth step is learning the Bayesian Belief Network (BBN) in HUGIN Expert (1995). This process of learning a network is described in the fourth chapter of this thesis. The input for BBN learning is a file with the variables age, gender, mode choice (only for the second network), occurrence events, state, time ago, history and a file with constraints. The constraints are described in the fourth and fifth section of the current chapter.

The learned Bayesian Belief Network is compiled in another program, called Netica (Norsys Software Corp, 1997). Netica is more user-friendly to analyse the network and to use the network for simulations. All coded nodes names and states are changed into text, which is self explanatory.

The last step in the sequence is the Life Walker program, which is built in Java. Two subprograms, basic and predict, are used for testing the network and simulation purposes. The ‘basic’ program enters the values of the input file as hard evidence into the network. No hard evidence is entered into the network for the variables that are predicted with this program (occurrence event nodes). The output of the ‘basic’ program is the probability distribution for the occurrence event nodes for any given event, person and year.

The ‘predict’ program can be used to build an individual life trajectory. The input is a file with values for all variables in a base year regarding each person of a simulated population, and the values are entered into the network as hard evidence. The probabilities of the occurrence event nodes are recalculated, and a random number is produced (Monte Carlo simulation) which defines which state (in general: no change; change type A; change type B; change type C) is chosen for the node occurrence event. Of course, the state with the largest probability has the highest chance to be chosen. The occurrence nodes can also influence each other directly or indirectly. For that reason, the decision sequence of these occurrence nodes is also entered. Before the probabilities for the next occurrence event node in the sequence are predicted the probabilities are automatically updated given this new hard evidence. In this case hard evidence is referred to as selected state of previous occurrence event node in the sequence. The sequence is predefined, as follows. The node with only outgoing arrows is the first one in the sequence. The input for the next year is based on the results of the occurrence node in the previous year. For instance, if someone lives in a student room (state) in 2010 and the program predicts an occurrence in that same year, for example independent living, the state in 2011 will be independent living. So the input for the next year is only altered if an occurrence was predicted. Otherwise, only the values for the history and time ago nodes are raised with one year. The number of years of life trajectory simulation is a variable in the input file.

Next, the different variables used in the structure learning step are explained. The output file in the first step and the input file in the second step contain the following variables: age, gender, mode choice, life event (year, type, before and after), current state, eventmax, max. Age is a simple translation from year of birth. Gender needs no additional preparation. The remainder of the variables need some preparation or classification from the survey data to the required data for the BBN. The steps will be described later.

The **mode choice** node in the Bayesian Belief Network has three states; car, slow transport and public transport. In the survey the respondents provided information with respect to their current mode choice behaviour for five different trip purposes. In particular, they indicated frequency, travel mode, alternative travel mode, destination, departure time at home, arrival time at home, estimated travel distance from their home to the destination and the estimated travel time. Respondents could indicate the frequency using their own ‘scale’ (day / week / month). The frequencies were rescaled into monthly frequency, which was the most frequent used category. The six mode options in the survey (see Appendix 1B) were classified into the three options, car driver and car passenger are recoded into car, bike and walking are recoded into slow transport and bus and train are recoded into public transport. The mode option with the highest frequencies, across the different trips purposes, is set as the overall mode choice.

The data of the **life events** is recoded into fewer classes. The seven events can be divided into two groups with the same matrix structure in the Internet-based survey. The first group is the one with the housing, work and study event, the other events, household, car availability, PT pass and household income are in the second group. In the first group, only the year and type of change are used. Appendix 2 represents the classifications of type of change. In the second group, the before and after situation are also used besides year and type of change. The before and after situation are recoded into new classes (see Appendix 2). There are three (new) types of changes in this group: decrease, increase and same. This new classification for type of changes is based on the transition of the before situation to the after situation. For example, if a child is born, the number of household members before is two persons and the number of household members after the occurrence is three. This means an increase of the number household members. A disadvantage of classification is that the original information reported in the Internet-based survey will be lost. For example the actual change, like a death in the household, a child moving out etc. are both coded as a decrease.

As explained in section five of the previous chapter, the personal characteristics should be consistent with most recent occurrence of the corresponding life course event. After cleaning the data no more inconsistencies existed. So it does not matter if the **current state** is based on the personal characteristics or the most recent occurrence in the event matrix, as in the end the result is the same.

In Fortran (Silverfrost, 1999) two variables are used: eventmax and max. **Eventmax** defines the number of years back in time with information available about that specific occurrence of the corresponding event. **Max** corresponds with the maximum of all eventmaxs. This variable is an extra variable to make sure that a person's life trajectory for one event has the same number of years as another life trajectory based on a different event for the same person.

The variable **eventmax** corresponds with the **history** node in the BBN and defines for how many years in the past there is information available (the recall period). The calculation of the variable eventmax is described here. First, the maximum recall period for one event is calculated, based on the year of the last occurrence. For example, if a respondent registered seven changes and the last one (oldest occurrence) was in 1985, the maximum is 19 years (given the data collection in 2004). The recall period for which the respondent should indicate changes/occurrences is not always the same (due to routing and questions in the Internet-based survey). This recall period is defined by the last part of the question "how many changes occurred, since…". There were five different options: since you left the parental house (*Tx*), since you live on your own (*Tz*), since you left school (*Ty*), since you got your driver's licence (*Ti*), and since your 16th birthday (*T16*). The age at which someone got his / her driver's licence (*Ti*) and turned 16 years old (*T16*) can be calculated given the personal characteristics (driver's licence and year of birth). There is no age marker for the first three periods (*Tx*, *Tz* and *Ty*). If the respondent registered occurrences for all seven events there is no problem. Only if a respondent indicated that no changes occurred, then the period is sometimes undefined, depending on the last part of the question "since…". To solve this problem events with an undefined recall period will be related to events with the same last part of the question (since…) with a defined recall period. This will be explained according to the routing of the survey (Chapter 5 | 4 | 4). There were eight possible routings through the survey, depending on the combination of the answers to three personal characteristics questions: living situation, age, and year of driver's licence). The first routing will be described here, while the other possible routings are illustrated in Appendix 4. A respondent who followed routing one has the following characteristics: living situation= independent housing, age= > 35 years old, and driver's licence= yes. Figure 6 | 12 illustrates the different time lines and recall periods. If a line is blank (white) this means that the respondent skipped that specific event question due to the predefined routing. The *T* in the figure illustrates the moment of the data collection (2004) and the other indicators (*Tx, Ty, Tz and T16*) correspond with the last part of the question "how many changes occurred, since…".

Housing			Tx							T
Household			Tx							T
Work				Ty						T
Study										
Car availability					Tz					T
PT pass	T16									T
Household Income					Tz					T
	past									present

Figure 6 | 12: Recall periods for routing 1

Tx = leave parental house

Ty = leave school

Tz = live on your own

Ti = driver's licence

T16 = since 16th birthday

In Figure 6 | 12 the recall period for the housing and the household event is the same (*Tx*). In case the respondent did not experience and thus did not report household events, the **eventmax** for the household event can be deduced from the **eventmax** of the housing event. If all related events were left blank and the recall period could not be deduced from other corresponding events, the history was calculated according to a 'general' age. This means that a mean age value was calculated based on all respondents. The following mean age value was calculated for the different events: Housing from age 21, Household from age 21, Work from age 23, Car from age 21 and Household Income from age 21. This was only used in a few cases (maximum of ten).

In the second step, a life trajectory was built for every respondent. The **max** value defines the number of rows, i.e. the number of years for the life trajectory of every respondent. The variables / nodes used in the BBN are literally configured in this second step. For every year in the life trajectory the following variables are calculated: age, gender, history, occurrence event, state, and time ago for every event type. The building of the life trajectory always starts with the year 2004 (moment of the data collection) and it ends when the **max** number of years equals zero. Table 6 | 3 illustrates an example of two life trajectory for the housing event. The variable gender is constant during the life trajectory, therefore it is not listed in this table. In this case there are three types of occurrences: independent housing (1), student housing (2) and parental housing (3). These three types correspond respectively with time ago A, B, and C. The rest of the variables are self explanatory.

Table 6 | 3: Example of database structure for life trajectory housing event

Person	Max	Age	Year	History	Occur-rence	State	Time ago A	Time ago B	Time ago C
1	15	32	2004	12	0	1	3	8	0
1	14	31	2003	11	0	1	2	7	0
1	13	30	2002	10	0	1	1	6	0
1	12	29	2001	9	1	2	0	5	0
1	11	28	2000	8	0	2	0	4	0
1	10	27	1999	7	0	2	0	3	0
1	9	26	1998	6	0	2	0	2	0
1	8	25	1997	5	0	2	0	1	0
1	7	24	1996	4	2	?	0	0	0
1	6	23	1995	3	0	?	0	0	0
1	5	22	1994	2	0	?	0	0	0
1	4	21	1993	1	0	?	0	0	0
1	3	20	1992	0	0	?	0	0	0
1	2	19	1991	-1	0	?	0	0	0
1	1	18	1990	-2	0	?	0	0	0
1	0	17	1989	-3	0	?	0	0	0
2	20	42	2004	20	0	1	10	19	17
2	19	41	2003	19	0	1	9	18	16
2	18	40	2002	18	0	1	8	17	15
2	17	39	2001	17	0	1	7	16	14
2	16	38	2000	16	0	1	6	15	13
2	15	37	1999	15	0	1	5	14	12
2	14	36	1998	14	0	1	4	13	11
2	13	35	1997	13	0	1	3	12	10
2	12	34	1996	12	0	1	2	11	9
2	11	33	1995	11	0	1	1	10	8
2	10	32	1994	10	1	1	4	9	7
2	9	31	1993	9	0	1	3	8	6
2	8	30	1992	8	0	1	2	7	5
2	7	29	1991	7	0	1	1	6	4
2	6	28	1990	6	1	3	0	5	3
2	5	27	1989	5	0	3	0	4	2
2	4	26	1988	4	0	3	0	3	1
2	3	25	1987	3	3	2	0	2	0
2	2	24	1986	2	0	2	0	1	0
2	1	23	1985	1	2	?	0	0	0
2	0	22	1984	0	0	?	0	0	0

Age and history of course decrease every year. In some cases the number of years of the history variable is not equal to the number of years in the max variable. In this case, the max variable is defined by the history value of another related event of this person, as explained before. In this case the value of history becomes negative at the end of the life trajectory. For example, in Table 6 | 3 the value of the history variable becomes negative in the year 1991 for the first person. The values of the max and history variable for the second person are the same.

When an occurrence took place in a certain year according to the event matrix, the type of change is registered in the column occurrence event with the same classification (1, 2 and 3) as described for the time ago A, B and C variables. Sometimes two occurrences of one event took place in the same year. In this case only the most recent occurrence of that event is registered in the life trajectory. The state in the first year (2004) is based on the current state in the input file. If an occurrence took place in a certain year, the state variable is changed according to the occurrence. For example, in Table 6 | 3 the first person experienced a change in housing in the year 2001. He started to live on his own, this is listed in the table as independent living which corresponds with 1. The state in this same year is based on the previous occurrence, if there is any. For example, the previous occurrence was in 1996, for the first respondent, when the respondent moved to a student room (occurrence is 2). The state in the years from 1996 to 2001 was therefore student housing, which correspond with 2 in the state variable. Before 1996 no occurrence happened, so the state up to 1996 is unknown. For the event in the first group (housing, work and study event) the state is always based on the previous occurrence. While for the other events, household, car availability, PT pass and household income, the state is based on the before situation of the same occurrence. In this case the state variable is never unknown or missing.

7 | Learned life trajectory network

As described before, in total 710 respondents completed the Internet-based survey. 701 surveys were used to build life trajectories for these respondents. Some respondents have a life trajectory of three years, while others have a life trajectory of 20 years, and everything in between. In total, the life trajectories resulted in 7649 cases, where the unit of observation is a year and a person. The learning of the network is based on these cases.

Hugin was used to build and estimate the Bayesian Belief Network for the life trajectory, using the input data and constraints described earlier in the fourth section of this chapter. The level of significance was set to the standard value of 0.01 for the learning process. Figure 6 | 13 illustrates the learned network with 42 nodes and 75 learned links. Note that 42 links were predefined as constraints within an event: 19 links history node → the time ago A/B/C nodes, 19 links time ago A / B / C nodes → state node, and 4 links history node → occurrence event node. Besides the predefined links 33 links were learned. These learned links can be divided into five different categories:

(1) links within an event : (time ago^{H} → occurrence event^{H} and state^{H} → occurrence event^{H})

(2) links across events (not including links between the event nodes): (time ago^{H} → occurrence event^{W} and state^{H} → occurrence event^{W})

(3) links between the event nodes: occurrence event^{H} → occurrence event^{W}

(4) links with the personal characteristics nodes: (age / gender → occurrence event)

(5) links between the personal characteristics nodes

H and *W* refer to event *H* and event *W*.

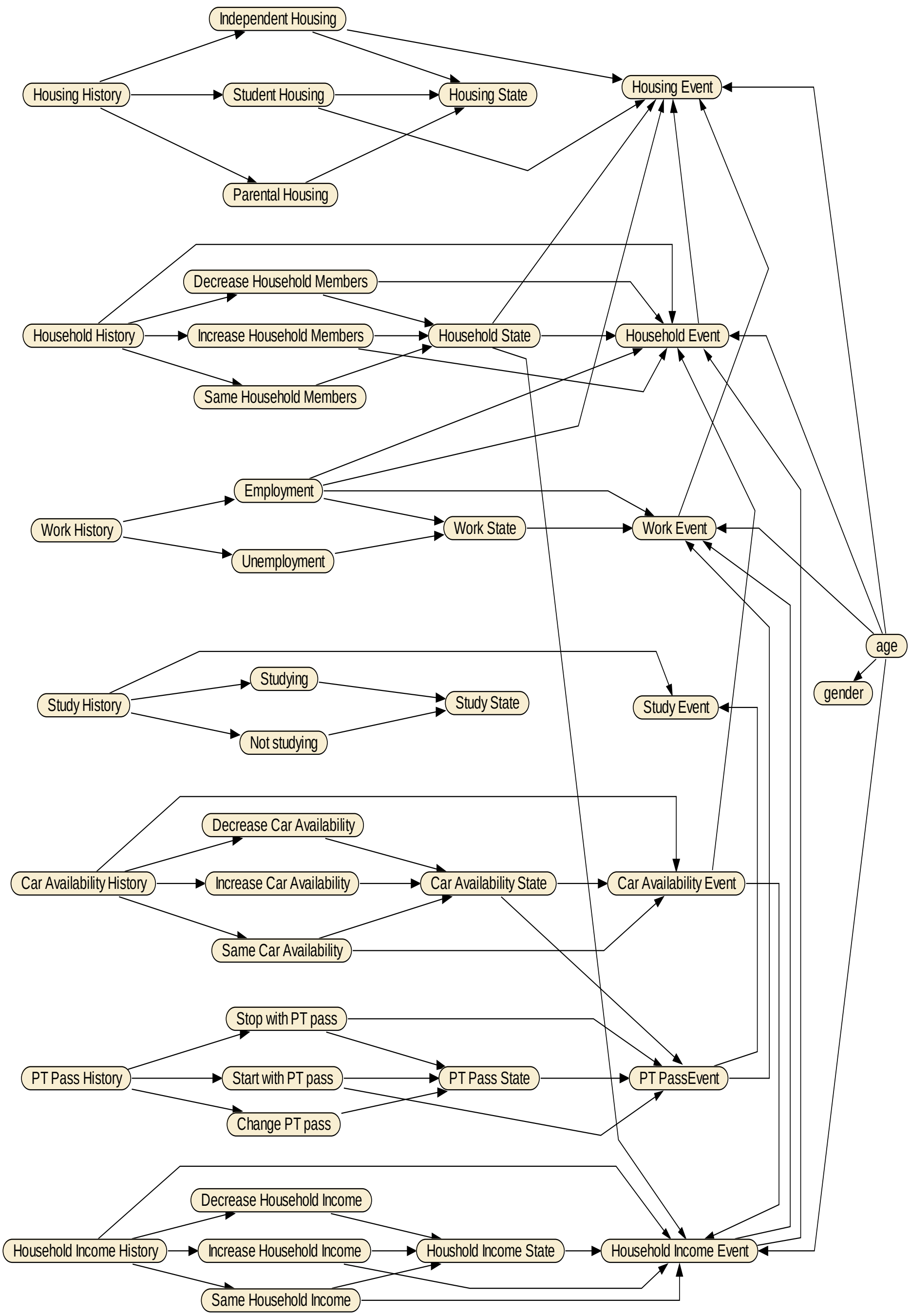


Figure 6 | 13: Learned life trajectory network

The 33 learned links were distributed across these five categories, as follows: 15 links were learned within an event, 5 links across events, 8 links between the event nodes, 4 links with the variable age, and 1 link was learned between the personal characteristics.

A link between time ago or state node and occurrence event node (category 1) means that there is a direct impact of a time ago or state variable on an occurrence event. For example, this may describe a higher probability for a housing occurrence (independent living) when a last occurrence (independent living) was more than three years ago. The absence of a link does not necessarily mean that there is no relation or influence. There can also be an indirect relation between nodes. For example, in the chain time ago → State → occurrence event, the state node is the intermediate node and there exists an indirect relation between the time ago node and the occurrence event node via the State node. Almost all state nodes have a link with the event node (within an event), except housing and study. This means that there is no direct link between the previous state and the occurrence of an event for the housing and study event. Most time ago nodes show a direct link with the occurrence event nodes. In total 10 links were learned of the 19 possible links.

Some state or time ago nodes of one event have a direct link with another Occurrence event node (Category 2). These links indicate that different events are also related. The learned network included the following links: household state is linked to occurrence event nodes of the housing and household income events. Time ago of work event (employed) has a direct link with the occurrence event nodes of housing and household event.

The links between occurrence event nodes (Category 3) indicate that the occurrence of one event influences the occurrence of another event. This relation should be investigated in more detail, because the absence of an occurrence (no change) is also a state of the occurrence event node. This means that the relation can also mean that the absence of an occurrence of one event can influence an occurrence or the absence of an occurrence of another event. The direction of the arc is not that important, as the child node can also influence the probabilities of the parent node. The following eight relations are learned in the network: 1 housing event – household event; 2 housing event – work event; 3 household event – car availability event; 4

household event – household Income event; 5 work event – public transport pass event; 6 work event – household income event; 7 study event – public transport pass event, and 8 car availability event – household income event. All relations between the occurrence event nodes seem logical.

As for category 4, the node age is linked to four occurrence event nodes, namely: housing event, household event, work event and household income event. This may mean that a specific event occurs at a specific age, for example, student housing before the age of 25 years old. All relations between the node age and the occurrence events nodes should be investigated further to describe the effect of age on the occurrence of an event.

Between the external variables (Category 5) there is a link between Age and Gender. This does not mean that gender influences age or the other way around, but rather it describes the association of these variables in the sample.

8 | Learned mode choice network

This second network examines if and how life trajectory events impact on transport mode choice. Life trajectory decisions, as discussed in the previous section, include socio-demographic events, long-term decisions such as residential and job choice and resource decisions, such as car purchasing behaviour, that are all assumed to influence transport mode choice decisions. In order to examine the impact of these life trajectory decisions, a network is derived that includes transport mode choice as a node in the network. This second network is based on 701 cases. As explained before, in this research project, there is only one observation per respondent on mode choice. The respondents only reported about their mode choice in the year of the survey, 2004.

Figure 6 | 14 illustrates the learned mode choice network with 43 nodes and 79 learned links. 75 links were already learned in the first network and 4 extra links were learned with the node mode choice.

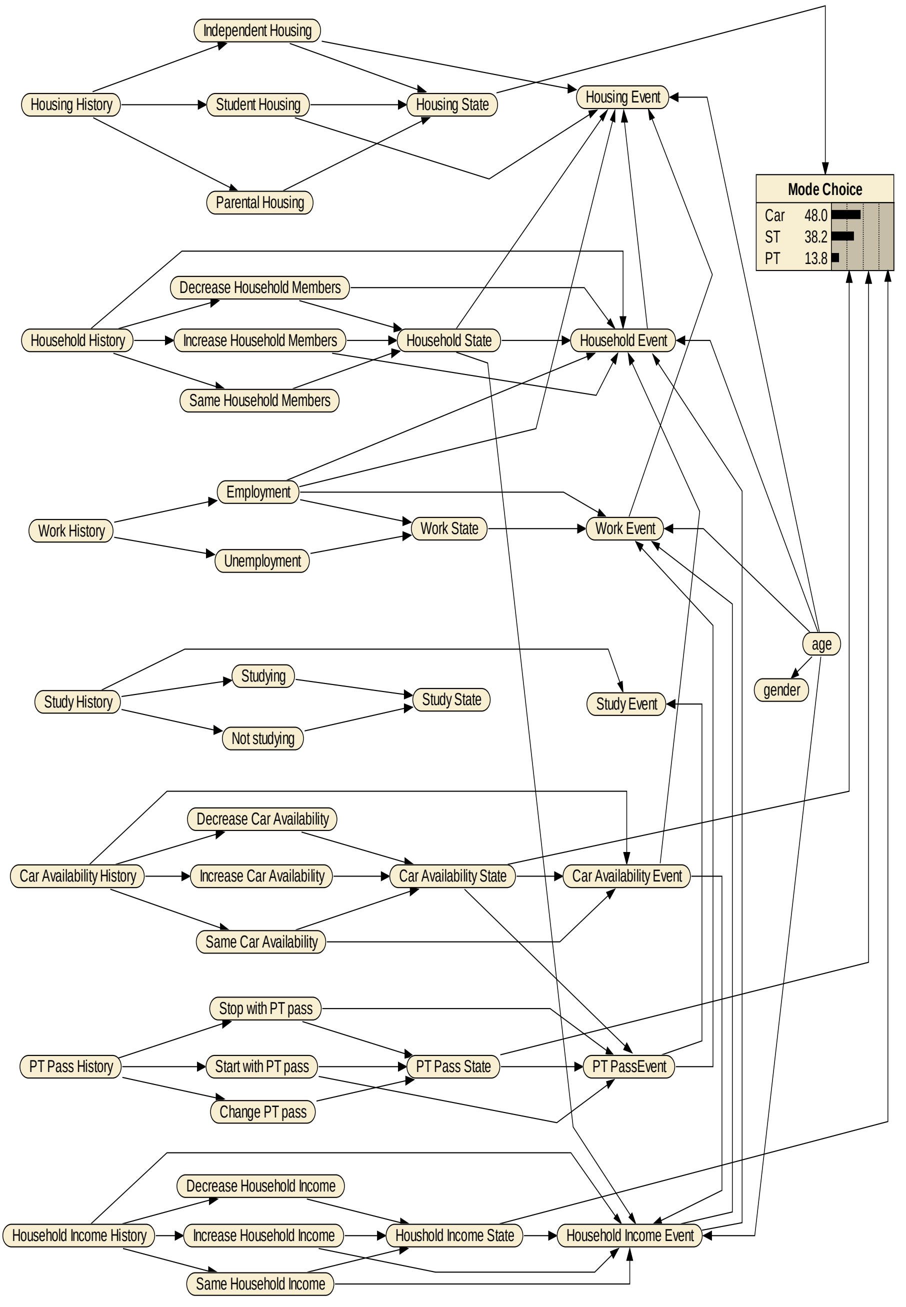


Figure 6 | 14: Learned mode choice network

All 4 links originate in state nodes: housing state, car availability state, public transport pass state and household income state. This means that the current state (of these events) is related to mode choice. It seems logical that car availability and public transport state relate to the choice set which has an impact on mode choice. The relation between housing state, household income state and mode choice is also explainable.

In order to use more data, all CPTs of the life trajectory network were copied into this network, except of course the CPT of mode choice which does not exist in the first network. The links with the variable/node mode choice are learned and it is assumed that these relations are the same in other years.

9 | Conclusion

This study seeks to apply Bayesian Belief Networks in estimating direct and indirect effects of life course events on transport mode choice. However, in order to test whether time has an effect, first a simple multinomial logit model was estimated. The results support the contention that a certain time influence exists. However, such models can only represent direct effects of a set of predictor variables on a target variable and, hence, are too restricted for the model development purpose of this thesis. To reveal direct and indirect relationships at the same time, Bayesian networks were learned from data.

Time is not modelled continuously, like in hazard models. A year is chosen as unit for these models, this is indicated with one year of a persons life trajectory. The respondents could better recall the year in which the event occurred than the month. The sequence of two or more events can not be determined within one year with these models. As mentioned before individuals can react or anticipate to changes. Time ordering does not necessarily reflect causal ordering.

Two networks were estimated: the first network can be used to simulate a person’s life trajectory and the second network can be used to predict mode choice for an individual at a certain time given his / her life trajectory. Both network models were successfully learned from the data and the learned links could be explained. It is important to emphasize that certain logical relationships had to be inserted as constraints in the modelling approach. These constraints are enforced and therefore the nature and significance of these links are not determined by the value of the relevant association measures.

7 | Validation

1 | Introduction

The previous chapter has reported both learned Bayesian Belief Networks and described the learned links. In this chapter, the goodness-of-fit of the learned Bayesian Belief Networks is discussed. A relatively satisfactory goodness-of-fit does not necessarily mean that the models are capable of representing a complete life trajectory and most importantly the effects of life trajectories on transport mode. Further validation along these lines is thus required.

This chapter describes the results of such validation tests. More specifically, the predicted life trajectories and mode choices are compared with the observed life trajectories and mode choice as stated in the Internet-based survey. The validation tests for the life trajectory model are based on the following set of criteria; number of occurrences (in short ‘count’), interval times between occurrences of events (in short ‘interval time’), simultaneous occurrences of events (in short ‘synchronic events’), and sequence of occurrences of events (in short ‘sequence’). These criteria correspond more or less with the three dimensions Feijten (2005) distinguished. Interval times and synchronic events refer to timing and duration of events. Sequence is related to the order of events. For the mode choice model the modal split (car, public transport and slow transport) is compared in the year 2004.

The chapter is organised as follows. First, the necessary input for a simulation is explained. In the next section the validation results of the life trajectory network are described. First, the log likelihood for this network is calculated. Next, an example life trajectory is given which is used for the other validation tests. This is followed by a discussion of the log likelihood of the mode choice network and the comparison of the modal split. The last section contains the conclusions and summarises the most important results.

2 | Input simulation

The Life Walker program, which was described briefly in Chapter 6 | 6, has been developed to simulate the trajectories. Figure 7 | 1 illustrates the simulation sequence. The learned Bayesian Belief Network is used to simulate transitions in a particular year. The input for the next year is based on these simulation results of the occurrence event nodes in the current year. Thus, the input for the next year is only altered if an occurrence was predicted. In that case the state of the corresponding event is updated according to the predicted occurrence. For example, if an increase of household members (occurrence event) is predicted in 2000, the number of household members (state event) in the year 2000 (before the occurrence) is increased for the year 2001.

Otherwise, if no occurrence is predicted in the current year, only the values for the history and time ago nodes are raised by one year. The values of the history nodes are of course raised by one year regardless of the outcome of the predicted occurrences. If an occurrence is predicted in 2000, the corresponding time ago node is set to one year for the following year (2001).

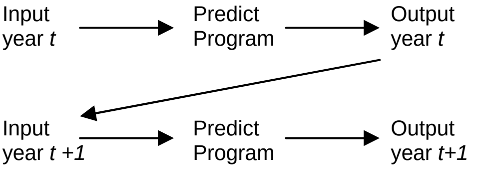


Figure 7 | 1: Subprogram 'predict'

The input of a simulation consists of the input of all variables/nodes in the network except the nodes that will be predicted during the simulation process. Occurrence event nodes will not be used as input when simulating a life trajectory. Because a life trajectory consists of several years, the simulation has to run for a number of years.

The occurrence of an event in year *t* depends on the predicted probabilities for that event given all variables/nodes in the network. Monte Carlo simulation is used to sample a specific state for the occurrence event node. Only one state of an event can be chosen in one year. Of course, there could be more than one occurrence predicted in one year for different events (indicated earlier with synchronism). For example, in one year a change in household and a change in housing can be predicted. A year is chosen as time measurement, within that year there is no order between occurrences.

Because the process is probabilistic, different simulation runs lead to different results. For validation purposes, five simulations were run. The input for every simulation run is the same: a set of 700 cases. Each case corresponds with one respondent in the sample. The values for every node in the life trajectory model, except the occurrence event nodes, are listed in the input file: age, gender, history, time ago and state variables. The number of years for the life trajectory simulation is a variable in the input file. All life trajectories are simulated until the year 2004. For example, for one person this results in a life trajectory of 18 years, while for another respondent the life trajectory may consist of only five years. In every simulation run the same number of cases is generated. The output file consists of 7648 cases, where one case corresponds with one year of a person's life trajectory.

3 | Validation life trajectory network

In this section different validation test are described. The total sample (i.e. all life trajectories from the 700 respondents) is used for structure and parameter learning of both networks. There is no hold-out sample used for validation purposes. The goal of this section is to illustrate a method of validation, in this case different tests for validation of the life trajectory network, and not to describe the definitive validation of the network. In this section, first the log likelihood is calculated for three models: null-model, overall model and prediction model. Next, an illustration is given of an example life trajectory, which is used for the explanation of the four validation tests (count, interval time, synchronism and sequence).

1 | Goodness-of-fit

To assess the goodness-of-fit of the learned network regarding the prediction of events, the log likelihood is calculated for each event separately, based on a prediction sequence. The sequence starts with an event with no *incoming links* from other event nodes and continues with events which only have links from the event nodes which precede in the sequence. If this rule is not applied the influence of occurrence event nodes cannot be taken into account. The occurrence event nodes of car availability and public transport pass event have no *incoming links* of the other occurrence event nodes (see Figure 6 | 13). These two events have to come first in the sequence, while the order of these events is interchangeable. Second in line are the occurrence nodes household income and study with only one *incoming link* from respectively the car availability event and the PT pass event. The household and work events are next. They have respectively *incoming links* from the car availability & household income events and the public transport pass & household income events. Last in the sequence is the occurrence node housing, which has *incoming links* of the household and work events. The chosen sequence is the public transport pass event, car availability event, study event, household income event, work event, household event and housing event.

To predict the probabilities for the first event in the sequence (public transport pass), hard evidence for all nodes was entered into the network, except for all nodes in the prediction sequence (occurrence event nodes). The probabilities for the second event in the sequence were predicted given hard evidence for all nodes of the network including the public transport pass event, except the other event nodes that follow later in the sequence, and so on.

Log likelihood values were calculated for three models: null-model, overall model and prediction model. In the null-model uniform distributions were chosen, implying that every option has the same chance. In the overall model the probabilities were distributed according to the overall probability distribution in the data set. The probabilities in the prediction model were based on the simulations with the network.

To illustrate the calculation of the log likelihood Table 7 | 1 illustrates the first three cases of the simulation results for the prediction for the study event. The first three columns list the simulated probabilities and the next three columns list the probabilities derived from the data collection. The simulation results listed in this table for all three options (A, B and C) have a value above zero. In the Internet-based survey only one occurrence was reported in one year. This means that the probabilities of the occurred event (A, B or C) are one and the other two probabilities have the value zero.

The log likelihood was calculated for every event and for every model (e.g., null model, overall model and prediction model). The calculated log likelihoods are listed in Table 7 | 2.

Table 7 | 1: Illustration part of the prediction model (study event)

Prediction (Simulation)			**Observed (Survey data)**			**Log likelihood**
Probability A	Probability B	Probability C	Probability A	Probability B	Probability C	
0.965986	0.0170068	0.017007	1	0	0	-0.03461
0.999998	0.000001	0.000001	1	0	0	-2E-06
0.965986	0.0170068	0.017007	1	0	0	-0.03461
….	….	….	….	…	…	…
Sum						**-273.936**

The first column lists the number in the prediction sequence. For every event the log likelihood value decreases if the values for the null model are compared to those of the overall model and the prediction model. The lowest log likelihood values are in the column of the prediction model. In general, a lower log likelihood indicates a better performance of the model. To express the improvement of the models into a value (goodness-of-fit), the Rho-Square statistic was calculated. This is the ratio of the log likelihood for the considered model and the log likelihood for the null model. The closer the value is to 1, the better the performance of the model. In case of a perfect goodness-of-fit, the log likelihood of the prediction model is equal to zero and Rho-Square is equal to 1.

Two Rho-Squares were calculated, the first one is the ratio between the log likelihood of the prediction and null-model, and the second one is the ratio between the log likelihood of the prediction and overall model. The values of the first Rho-Squares in Table 7 | 2 are relatively high, all above 0.68. This indicates that the Bayesian network model out performs the null-model. The values of the second Rho-Square are lower, but all values are above 0.20. This indicates, based on generally accepted norms, that the models perform well.

In the next subsections the four validation test are described. First, one example life trajectory is given. A part of this life trajectory is used to describe each particular validation test.

Table 7 | 2: Log likelihood values life trajectory network

	Event	Null-model	Overall model	Prediction model	First Rho-Square	Second Rho-Square
7	Housing	-10602.40	-3193.79	-1819.90	0.8284	0.4302
6	Household	-10602.40	-3767.82	-1883.04	0.6843	0.5002
5	Work	-8402.19	-3422.19	-2652.44	0.9674	0.2249
3	Study	-8402.19	-417.56	-273.94	0.8224	0.3440
2	Car Availability	-10602.40	-2714.49	-2162.44	0.7960	0.2034
1	PT pass	-10602.40	-1645.47	-1183.69	0.8884	0.2806
4	Household Income	-10602.40	-3184.56	-1749.95	0.8349	0.4505

2 | Example life trajectory

To compare the observed life trajectories from the Internet-based survey with predicted life trajectories five simulations were run. The number of years for the prediction of the life trajectory with the simulation is unique for each person in the sample. The number of years corresponds with the number of years of the observed life trajectory. For example, a person reported a life trajectory for twelve years, from 1992 until 2004. A life trajectory was also predicted for these same years. All years of the simulated life trajectory were analysed and compared with the observed life trajectory according to a set of validation tests: (1) number of occurrences, (2) interval times between occurrences of events, (3) simultaneous occurrences of events, and (4) sequence of occurrences of events.

Figure 7 | 2 illustrates an example of a life trajectory for one person over twelve years. Each event has a time line, and a year is visualized with a vertical line. The bold vertical lines indicate that an occurrence took place in that particular year.

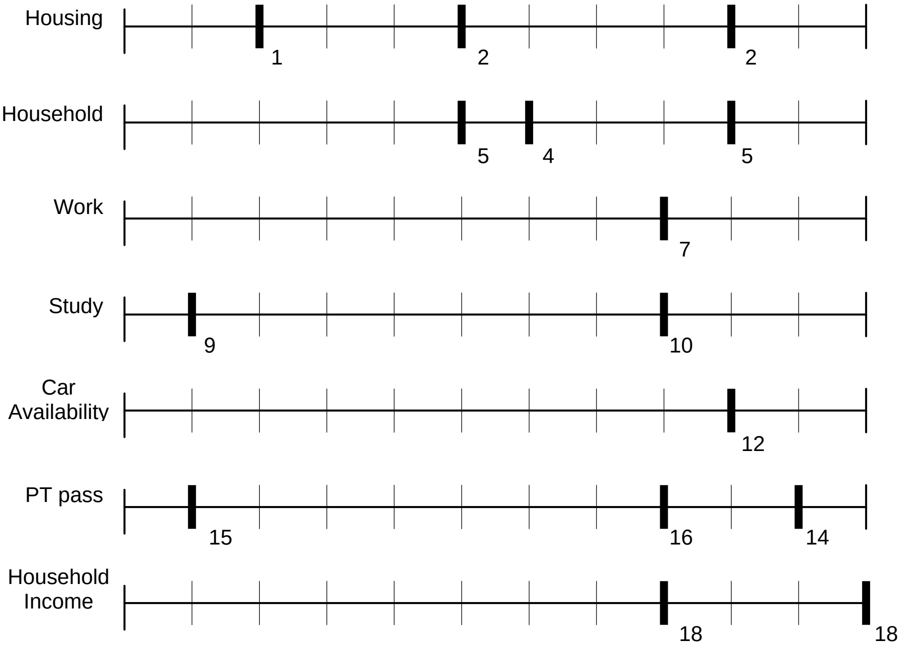


Figure 7 | 2: Life trajectory seven events

This person experienced in total 15 occurrences during these twelve years. The numbers in the figure correspond with the different event specific occurrences (i.e. subevents). The numbering starts at the first occurrence of the housing event and ends at the last occurrence of the household income event. In total there are 19 different subevents (see Appendix 3). This life trajectory is used in this section as an example to explain the different validation tests.

3 | Analysis simulations: Count

The first validation test is whether the total number of years with an occurrence is successfully reproduced by the network. The life trajectories of the 700 individuals in the sample produce 7648 years (i.e. person-years) in total. In every year an occurrence could happen or not. The frequency is the sum of all person-years when an occurrence happened across the (sub)events. The observed frequencies are compared with the frequencies of the simulations. Two classification levels are distinguished for the frequencies:

1. person-years with / without occurrences (including subevents)
2. person-years with / without occurrences (not including subevents)

All years are taken into account in both levels, including the person-years where no occurrence happened. At the first level all available information is taken into account. That is, the level of detail of the subevents is also included. For example, for the housing event this means the number of person-years is summed for the following subevents: no change; change, independent housing; change, student housing; change, parental housing. In the second level, there is no distinction between the subevents. For example, the number of person-years of the three states of the housing event (change, independent housing; change, student housing; change, parental housing) are aggregated into one value. This means that there is only a distinction between no change (e.g., no occurrence) and change (e.g., occurrence). Chi-Square was calculated to test whether the model is capable of reproducing the observations. The larger Chi-Square value, the greater the evidence against the null hypothesis (H_0): no difference between the predicted and observed observations. Table 7 | 3 provides the results for the first level: person-years with / without occurrences (including subevents). The first column lists the event with all states (no occurrence and all subevents). The second column shows the total number of counts in the observed life trajectories and the third column presents the percentage of the total count within each event. The total counts for the predicted life trajectory is shown in the fourth column and the percentage of the total count within an event is listed in the fifth column. The Chi-Square and its p-value are listed in the last two columns. Table 7 | 4 reports the results for the second level: person-years with / without occurrences (not including subevents).

Table 7 | 3: Results count level 1 (including subevents)

Housing	**observed**	**%**	**predicted**	**%**	**Chi Square**	**p-value**
no occurrence	6708	0.877	33731	0.882		
independent housing	850	0.111	3980	0.104		
student housing	71	0.009	420	0.011		
parental housing	19	0.002	109	0.003		
Total	7648	1.000	38240	1.000	**52.263**	0.156
Household	**observed**	**%**	**predicted**	**%**	**Chi Square**	**p-value**
no occurrence	6654	0.87	33691	0.881		
decrease	310	0.041	1334	0.035		
increase	627	0.082	2920	0.076		
same	57	0.007	295	0.008		
Total	7648	1.000	38240	1.000	**92.178**	0.027
Work	**observed**	**%**	**predicted**	**%**	**Chi Square**	**p-value**
no occurrence	6520	0.853	32501	0.85		
employed	1072	0.14	5445	0.142		
unemployed	56	0.007	294	0.008		
Total	7648	1.000	38240	1.000	**0.384**	0.944
Study	**observed**	**%**	**predicted**	**%**	**Chi Square**	**p-value**
no occurrence	7583	0.992	37931	0.992		
studying	24	0.003	115	0.003		
not studying	41	0.005	194	0.005		
Total	7648	1.000	38240	1.000	**0.1401**	0.987

Table 7 | 3 continued

Car availability	**observed**	**%**	**predicted**	**%**	**Chi Square**	**p-value**
no occurrence	7057	0.923	35181	0.92		
Decrease	150	0.02	701	0.018		
Increase	259	0.034	1377	0.036		
Same	182	0.024	981	0.026		
Total	7648	1.000	38240	1.000	**23.044**	0.512
PT pass	**observed**	**%**	**predicted**	**%**	**Chi Square**	**p-value**
no occurrence	7336	0.959	36458	0.953		
Stop	100	0.013	602	0.016		
Start	116	0.015	626	0.016		
Change	96	0.013	554	0.014		
Total	7648	1.000	38240	1.000	**54.444**	0.142
Household Income	**observed**	**%**	**predicted**	**%**	**Chi Square**	**p-value**
no occurrence	6847	0.895	34215	0.895		
Decrease	43	0.006	235	0.006		
Increase	215	0.028	1028	0.027		
Same	543	0.071	2762	0.072		
Total	7648	1.000	38240	1.000	**0.7788**	0.8545

A p-value < 0.05 (if *alpha* is 5% is applied) means that the null-hypothesis is rejected, and hence that the observed and simulation frequencies differ significantly. In Table 7 | 3 all significance values of the Chi-Square are above 0.05 except the one for the household event. The number of person-years with a decrease and increase in the number of household members is underestimated in the simulation. For all other events the simulation produces the same percentage of person-years with and without an occurrence. The occurrence is defined here as subevents, this means that the number of person-years with a specific subevent is predicted

In Table 7 | 4 the events household and PT pass event have a significance value below 0.05. In case of the household event, the person-years with occurrences are under predicted in the simulation and for the PT pass event the person-years with occurrences are over predicted. In general, the life trajectory model reproduced the number of person-years with and without occurrences in the life trajectories quite well.

Table 7 | 4: Results count level 2 (not including subevents)

Housing	**observed**	**%**	**predicted**	**%**	**Chi Square**	**p-value**
no occurrence	6708	0.877	33731	0.882		
occurrence	940	0.123	4509	0.118		
Total	7648	1.000	38240	1.000	**15.194**	0.218
Household	**observed**	**%**	**predicted**	**%**	**Chi Square**	**p-value**
no occurrence	6654	0.87	33691	0.881		
occurrence	994	0.13	4549	0.119		
Total	7648	1.000	38240	1.000	**72.738**	0.007
Work	**observed**	**%**	**predicted**	**%**	**Chi Square**	**p-value**
no occurrence	6520	0.853	32501	0.85		
occurrence	1128	0.147	5739	0.15		
Total	7648	1.000	38240	1.000	**0.3357**	0.562
Study	**observed**	**%**	**predicted**	**%**	**Chi Square**	**p-value**
no occurrence	7583	0.992	37931	0.992		
occurrence	65	0.008	309	0.008		
Total	7648	1.000	38240	1.000	**0.138**	0.710
Car availability	**observed**	**%**	**predicted**	**%**	**Chi Square**	**p-value**
no occurrence	7057	0.923	35181	0.92		
occurrence	591	0.077	3059	0.08		
Total	7648	1.000	38240	1.000	**0.6439**	0.422
PT pass	**observed**	**%**	**predicted**	**%**	**Chi Square**	**p-value**
no occurrence	7336	0.959	36458	0.953		
occurrence	312	0.041	1782	0.047		
Total	7648	1.000	38240	1.000	**49.322**	0.026
Household Income	**observed**	**%**	**predicted**	**%**	**Chi Square**	**p-value**
no occurrence	6847	0.895	34215	0.895		
occurrence	801	0.105	4025	0.105		
Total	7648	1.000	38240	1.000	**0.0185**	0.892

4 | Analysis simulations: Interval times

The first validation test only compared the number of occurrences in the observed and the predicted life trajectories. The moment in time was not taken into account. This second validation test analyses the interval times between occurrences within one event and between occurrences of two different events. It is important to know whether the network predicts the same interval times as were reported in the observed life trajectory. The timing of occurrences is analysed given two interval types. The different intervals are explained given the example life trajectory. Mean values are calculated for every interval type. An independent samples *t*-test is used to test whether the means of the two independent random samples (observed and predicted sample) are statistically different. It is important to note that there is no distinction between the type of occurrences, implying that all subevents as mentioned before are listed as occurrences without further qualification in this analysis.

1 | Interval times within an event

Figure 7 | 3 illustrates a part of the life trajectory graphed in Figure 7 | 2. This part shows that this person experienced three housing occurrences in the years 1995, 1998 and 2002. Three intervals are distinguished here: start interval, sequence interval and end interval. While only one interval (i.e. sequence interval) was used in the analysis between the observed and predicted life trajectories. The interval from the start (beginning of the life trajectory) to the first occurrence is named start interval. In case there was an occurrence in the first year of the life trajectory there is no start interval. The same holds for the end interval, which is the interval between the most recent occurrence and the end of the life trajectory. These two intervals are not taken into account in this analysis. The start and end moment of the life trajectory are completely random. Both are determined in the survey, the start interval can be defined by the occurrence of another (related) event, by routing of the survey, the end interval is established by the moment of the data collection. The first occurrence in the life trajectory is not predicted given the previous occurrence, because this occurrence is not known. For these reasons, the start and end interval are not taken into account here.

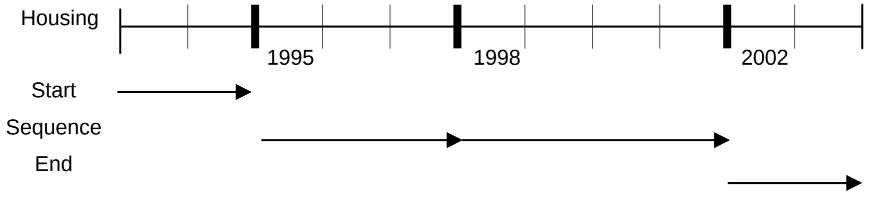


Figure 7 | 3: Interval time within an event

The interval between two occurrences of the same event is called sequence interval. If there are no occurrences in a person's life trajectory registered, no sequence intervals exist. In this example there are two sequence intervals: three and four years. The mean value of the sequence intervals in this example is 3.5 years.

<u>2 | Interval times between events</u>

Figure 7 | 4 illustrate a part of the total life trajectory portrayed in Figure 7 | 2. Two events are listed, the first one is the housing event and the second one is the household event. In total, four types of intervals are distinguished here: 1) start interval, 2) end interval, 3) sequence interval, and 4) interval between two events. The first three intervals are the same as in Figure 7 | 3. Only the interval between occurrences of two different events is new here.

The new interval between events is further explained given Figure 7 | 4. It is assumed that an occurrence of the household event in year *t* triggered an occurrence of the housing event in year *t* + 3. In this example the interval time between housing and household is three years. There are two ways to calculate the intervals between these two events. The different intervals are illustrated with solid and dashed lines in Figure 7 | 4. First, the housing event is set as a reference (solid lines) and second the household event is set as reference (dashed lines). This results in different intervals.

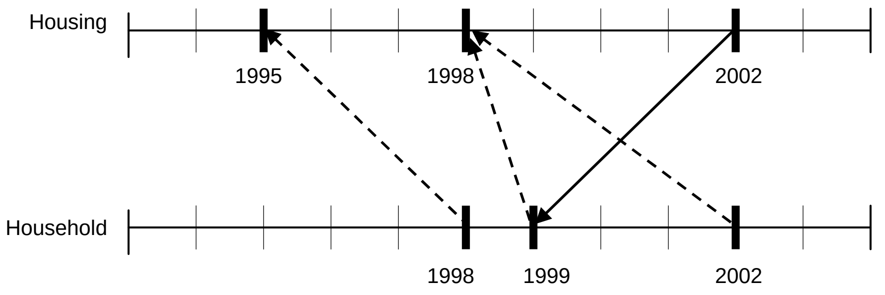


Figure 7 | 4: Interval time between events

In the example life trajectory in Figure 7 | 4, there are three occurrences of the housing event, but only one occurrence is preceded by an occurrence of the household event (solid line). The housing occurrence in the year 2002 is preceded by a household occurrence in the year 1999. Thus, the interval is three years. There are three intervals in this example for the household event (dashed line). The occurrence in 1998 was preceded by a housing occurrence in 1995 and the housing occurrence in 1998 happened prior to the household occurrences in 1999 and 2002. This results in intervals of respectively three, one and four years.

The mean value is calculated for all intervals and life trajectories. The mean value for the observed life trajectory is compared with the mean values of the predicted life trajectory using a *t*-test. Note that in the year 1998 and 2002 two occurrences took place within the same year. It is impossible to tell which occurrence happened first in that year. For this reason these relations, occurrences in the same year, are discussed separately in the next subsection.

3 | Results for all intervals

Only the mean sequence intervals and the mean interval between two events are analyses. Table 7 | 5 lists the means and number of the intervals of the observed life trajectories and Table 7 | 6 lists all information for the predicted life trajectories. The sequence intervals are in the diagonal of the table and the intervals between two events are in the other cells of the table. The upper part of the tables contains the mean values and the lower part lists the total number of intervals. The total of Table 7 | 6 is the sum of five simulations.

The null hypothesis (H_0) is that the model predicted the observed means of intervals well ($\mu_1 = \mu_2$). The independent sample *t*-test checks whether the predicted and observed means are equal. Table 7 | 7 lists the results of the *t*-test.

Table 7 | 5: Interval times (observed data)

Means	Housing	HH	Work	Study	Car	PT	Income
Housing	4.5	4.4	3.4	3.2	5.4	3.5	3.8
HH	5.2	4.5	4.3	3.8	5.7	4.2	5.1
Work	5.0	5.4	3.9	2.8	6.2	4.1	5.2
Study	1.6	2.1	1.7	2.3	1.7	1.9	1.9
Car	5.6	5.9	4.8	3.3	5.2	4.0	5.6
PT pass	5.5	4.8	4.1	3.5	6.7	3.8	5.1
Income	4.6	5.2	4.3	3.0	5.6	3.8	4.8
intervals	**Housing**	**HH**	**Work**	**Study**	**Car**	**PT**	**Income**
Housing	531	466	549	45	337	158	453
HH	615	629	621	29	445	178	524
Work	684	596	740	44	424	188	539
Study	20	14	12	7	14	10	14
Car	329	309	350	27	279	126	322
PT pass	177	149	185	31	112	120	165
Income	435	400	434	34	311	148	457

Table 7 | 6: Interval times (predicted data)

Means	Housing	HH	Work	Study	Car	PT	Income
Housing	4.7	4.5	3.7	2.4	5.7	4.7	4.6
HH	5.4	4.7	4.5	2.8	6.6	5.7	5.4
Work	5.4	5.2	4.4	2.2	6.4	5.1	5.3
Study	2.2	2.1	2.2	2.2	2.8	2.4	2.4
Car	5.9	5.6	5.0	1.9	5.5	5.6	5.9
PT	5.6	5.3	4.3	2.2	6.3	4.4	5.3
Income	5.0	4.9	4.4	2.6	6.1	4.9	4.9

Table 7 | 6 continued

intervals	Housing	HH	Work	Study	Car	PT	Income
Housing	2510	2117	2801	105	1816	727	2192
HH	2728	2817	3043	72	2068	833	2477
Work	3206	2949	3792	70	2363	1077	2889
Study	104	73	111	28	69	42	88
Car	1569	1513	1710	37	1381	558	1399
PT	1011	889	1143	53	715	827	929
Income	1958	1824	2083	69	1488	597	2175

Table 7 | 7: Results *t*-test for all interval times

	Housing	HH	Work	Study	Car	PT	Income
Housing	0.18	0.45	0.11	**0.03**	0.36	**0.00**	**0.00**
HH	0.46	0.28	0.22	0.08	**0.002**	**0.00**	0.32
Work	0.10	0.28	**0.006**	0.10	0.47	**0.001**	0.65
Study	0.07	0.91	0.21	0.91	**0.003**	0.29	0.29
Car	0.45	0.36	0.37	**0.006**	0.45	**0.00**	0.33
PT pass	0.89	0.26	0.53	**0.015**	0.55	0.11	0.65
Income	0.17	0.23	0.55	0.28	0.13	**0.001**	0.83

All significance values below 0.05 are bold in Table 7 | 7. In these cases the null hypothesis is rejected (*alpha* of 5% is applied), which means that the difference between the observed mean intervals and the predicted mean intervals is significant. The difference between the observed mean intervals and the predicted mean intervals is significant for 12 out of 49 intervals. In general, three interval times are underestimated and the other nine intervals are overestimated. The overestimations and underestimations are discussed in more detail for each interval group (sequence interval and interval between two events). In the sequence interval group there is one overestimation. The observed interval for the work event is 3.9 (Table 7 | 5) while the predicted interval is 4.4 (Table 7 | 6). A total of 11 out of 42 intervals in the last group (i.e. intervals between two events) are significantly different. The overestimations and underestimations of these intervals are described, without reference to the interval times in Table 7 | 5 and Table 7 | 6. The following intervals are underestimated: (1) housing and study, (2) car availability and PT pass, and (3) PT pass and study. The network overestimates the interval between (1) housing and PT pass, (2) housing and household income, (3) household and car availability, (4) household and PT pass, (5) work and PT pass, (6) study and car availability, (7) car availability and study, and (8) household income and PT pass.

There are more overestimated intervals than underestimated intervals. This indicates that the predicted time interval between two events is longer than reported in the survey. In general, the model can estimate more or less the same interval times for the other events. The most difficult interval to estimate is the interval between an event and the PT pass event. The model needs further improvement especially for this event.

5 | Analysis simulations: Synchronic events

As mentioned before, a time interval does not exist when two events occur in the same year. The time order within one year is no longer traceable due to recoding of the data according to the chosen perspective of one year. If occurrences were reported in the same year in the Internet-based survey, it is important that the network can reproduce this synchronism. Figure 7 | 5 illustrate a part of the example life trajectory graphed in Figure 7 | 2. Two events, housing and household, are illustrated here. In the year 1998 as well as in the year 2002 two occurrences happened in the same year. This is referred to as synchronic events.

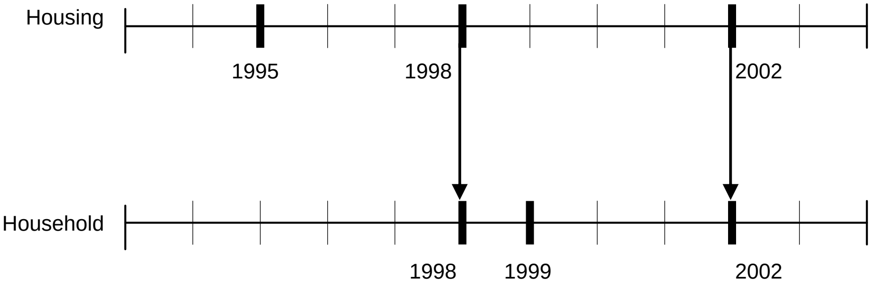


Figure 7 | 5: Events in the same year

Table 7 | 8 shows the number of synchronic events .Table 7 | 9 lists the number of observed occurrences for each event with no event occurrence as well as occurrences with one or more other occurrences. The total here is the sum of the previous rows; this number exceeds the total number of occurrences reported. The last row in the table indicates the number occurrence reported in the Internet-based survey (see Table 7 | 4). The probability of synchronic events depends on the total number of events. For example, when a housing event happened the probability of a household occurrence in the same year is the ratio between synchronic events (housing and household) and the total number of occurrence (housing). In this example this is 288 / 940 = 0.31. Table 7 | 10 lists all probabilities. In Table 7 | 11 – Table 7 | 13 the same information is listed for the predicted events: synchronic events, the number of synchronic events, and probabilities of synchronic events.

Table 7 | 8: Synchronic events frequencies (observed data)

Observed	Housing	HH	Work	Study	Car	PT	Income
Housing		288	271	32	178	87	240
HH	288		213	12	207	60	276
Work	271	213		27	160	145	359
Study	32	12	27		14	25	27
Car	178	207	160	14		66	186
PT pass	87	60	145	25	66		112
Income	240	276	359	27	186	112	

Table 7 | 9: Synchronic events more than two in one year (observed data)

Observed	Housing	HH	Work	Study	Car	PT	Income
no other event	384	426	457	13	196	75	186
1 event	236	272	361	8	155	92	252
2 events	168	165	180	20	120	77	206
3 events	96	81	78	13	74	34	104
4 events	45	40	41	6	37	24	42
5 events	10	9	10	4	8	9	10
6 events	1	1	1	1	1	1	1
Total	**1096**	**1056**	**1175**	**137**	**811**	**495**	**1200**
Occurrences	**940**	**994**	**1128**	**65**	**591**	**312**	**801**

Table 7 | 10: Synchronic events probabilities (observed data)

Observed	Housing	HH	Work	Study	Car	PT	Income
Housing	0.00	0.31	0.29	0.03	0.19	0.09	0.26
HH	0.29	0.00	0.21	0.01	0.21	0.06	0.28
Work	0.24	0.19	0.00	0.02	0.14	0.13	0.32
Study	0.49	0.18	0.42	0.00	0.22	0.38	0.42
Car	0.30	0.35	0.27	0.02	0.00	0.11	0.31
PT pass	0.28	0.19	0.46	0.08	0.21	0.00	0.36
Income	0.30	0.34	0.45	0.03	0.23	0.14	0.00

Table 7 | 11: Synchronic events frequencies (predicted data)

Simulations	Housing	HH	Work	Study	Car	PT	Income
Housing		1021	1134	59	529	252	820
HH	1021		838	50	615	208	966
Work	1134	838		53	668	588	1574
Study	59	50	53		32	85	39
Car	529	615	668	32		151	805
PT	252	208	588	85	151		211
Income	820	966	1574	39	805	211	

Table 7 | 12: Synchronic events more than two in one year (predicted data)

Observed	Housing	HH	Work	Study	Car	PT	Income
no other event	2110	2174	2523	118	1373	758	1303
1 event	1404	1435	2012	108	923	677	1475
2 events	649	630	847	51	484	244	875
3 events	276	242	284	22	212	86	302
4 events	65	63	68	8	62	13	66
5 events	5	5	5	2	5	4	4
6 events	0	0	0	0	0	0	0
Total	**4509**	**4549**	**5739**	**309**	**3059**	**1782**	**4025**
Occurrences	**3815**	**3698**	**4855**	**318**	**2800**	**1495**	**4415**

Table 7 | 13: Synchronic events probabilities (predicted data)

Observed	Housing	HH	Work	Study	Car	PT	Income
Housing	0.00	0.23	0.25	0.01	0.12	0.06	0.18
HH	0.22	0.00	0.18	0.01	0.14	0.05	0.21
Work	0.20	0.15	0.00	0.01	0.12	0.10	0.27
Study	0.19	0.16	0.17	0.00	0.10	0.28	0.13
Car	0.17	0.20	0.22	0.01	0.00	0.05	0.26
PT pass	0.14	0.12	0.33	0.05	0.08	0.00	0.12
Income	0.20	0.24	0.39	0.01	0.20	0.05	0.00

Table 7 | 14: Results p-values for the binomial test

Observed	Housing	HH	Work	Study	Car	PT	Income
Housing		0.000	0.000	0.000	0.000	0.000	0.000
HH	0.000		0.000	**0.270**	0.000	0.000	0.000
Work		0.000		0.000	0.000	0.000	0.000
Study	0.000	**0.226**	0.000		0.000	0.000	0.000
Car	0.000	0.000	0.000	0.000		0.000	0.000
PT pass	0.000	0.000	0.000	0.000	0.000		0.000
Income	0.000	0.000	0.000	0.000	0.000	0.000	

Table 7 | 14 illustrates the results of the binomial test. A p-value of 0.000 indicates a significant difference between the observed and predicted probability of synchronic events. In almost all cases there is a significant difference. The probability of the synchronic events household and study is an exception. The values 0.270 and 0.226 indicate that there is no significant difference. This means the model is successful in predicting these synchronic events. In general the model is less successful in predicting correctly the observed synchronic events.

6 | Analysis simulations: Sequence (SAM)

Occurrences in the life trajectory took place in a certain order or sequence. In order to compare the sequence of the observed life trajectory with the life trajectory of the simulations, the Sequence Alignment Method (in short SAM) was used. Different sequences can be configured from the example life trajectory (Figure 7 | 2). Three different sequences were constructed and analysed.

The first sequence option (labelled as OccurrenceOnly) is constructed with no reference to a year. This means that only the occurrences are listed in the sequence. There is no distinction in subevents in the sequence. The seven events: Housing, Household, Work, Study, Car availability, PT pass and Household Income correspond respectively with the numbers 1 - 7. All occurrences of events are listed in one sequence. When two or more events occur in one year the order of events that is listed above is used. In the second year (1994) two occurrences happened, a change in study and PT pass circumstances. The occurrence of the study events results in the '4' of the sequence followed by the '6' of the occurrence of the PT pass event. Next, two housing occurrences happened in year 1995 and 1998, which result in two '1''s in the sequence and so on.

OccurrenceOnly:	4	6	1	1	2	2	3	4	6	7	1	2	5	6	7

The next sequence (labelled as SubstatesOnly) is a more detailed description of the first sequence. The subevents are listed in this sequence. There is still no reference to the year of occurrence, just like in the OccurrenceOnly sequence. The occurrence of the study event in Figure 7 | 2 is referred to as '9', the first number in the sequence. '18' is the last number in the sequence, this number is related to the occurrence in household income in the last year of the life trajectory (2004).

SubstatesOnly:	9	15	1	2	5	4	7	10	16	18	2	5	12	14	18

The last sequence (labelled as Life Trajectory) is a numerical translation of the time lines in the life trajectory (Figure 7 | 2). This sequence consists of seven strings with twelve numbers. The bold lines in the life trajectory represent an occurrence and the type of occurrence (subevent) is given by a number. This number is copied into in the strings of the life trajectory. The other lines represent years were no occurrence happened; these years are represent in the sequence strings by zeros.

Life Trajectory:	0	0	1	0	0	2	0	0	0	2	0	0
	0	0	0	0	0	5	4	0	0	5	0	0
	0	0	0	0	0	0	0	0	7	0	0	0
	0	9	0	0	0	0	0	0	10	0	0	0
	0	0	0	0	0	0	0	0	0	12	0	0
	0	15	0	0	0	0	0	0	16	0	14	0
	0	0	0	0	0	0	0	0	18	0	0	18

Sequence alignment methods calculate the similarity between two sequences in terms of the number of operations that is required to equalize the two sequences. The operations involve adding, deleting, and swapping. In principle, different weights can be attached to these operations. Sequence alignment methods were originally developed for uni-dimensional sequences (Kruskal, 1983; States and Boguski, 1991). Joh, Arentze and Timmermans (2001) further expanded these methods to the case of multidimensional sequences. The Life Trajectory sequence can be viewed as a multidimensional profile. In this case, two different sequence alignment methods were applied: UDSAM (the sum of uni-dimensional sequence alignments) and DPSAM (Dynamic programming-based multidimensional sequence alignment method (Joh *et al*., 2001)). The UDSAM measure calculates the alignment costs of each uni-dimensional sequence and the sum of these measures across all dimensions is taken as a measure of distance or dissimilarity between two multidimensional patterns. The DPSAM measure compares sequences on multiple dimensions simultaneously, taking dependencies between the dimensions into account. To speed up computing time, the specific method applied searches only one trajectory for each event-type, which is closest to the diagonal of the two-dimensional comparison table of each event-type when integrating uni-dimensional operations into multidimensional ones, (Joh *et al.*, 2001).

In our analyses, a weight of one was used for additions and deletions, while swapping received a weight of two (it can be seen as a combined deletion and addition operation). Thus, the SAM measure for each sequence alignment is the weighted total number of additions, deletions and swapping operations required to make the two sequences identical. A higher number implies that the two sequences are less identical. The value of SAM (costs to align the sequences) will increase when the level of detail increases. It is difficult to analyse the SAM value if there is no comparison or range from the minimum to maximum SAM value. For this reason, a minimum and maximum were calculated and the mean of all SAM values was compared to this range.

Given the three sequences, four analyses were executed. The UDSAM measure is used for the OccurrenceOnly sequence and the SubstatesOnly sequence. Both sequences are uni-dimensional. The Life Trajectory sequence is analysed two times: first as an entity with the DPSAM method (multi-dimensional) and second for each event separately with UDSAM method (uni-dimensional). The Life Trajectory sequence is multi-dimensional given the seven strings which represent the different time lines in the life trajectory.

The calculation of the SAM value for sequence one and two is explained given an example (Figure 7 | 6). This figure illustrates an observed and predicted sequence for one person. Every vertical line in the figure indicates an occurrence. The observed sequence consist of eight occurrences and the predicted sequences has twelve occurrences in total.

The minimum SAM values means that the sequences are very similar (identical if they are of equal length). This means in this example that eight occurrences in both sequences (observed and predicted) are the same. Only four occurrences had to be inserted or deleted, this depends on the point of view. If the observed sequence should match the predicted sequence then occurrences are inserted and if the predicted sequence should match the observed sequence then occurrences are deleted. The minimum SAM value in this example is four. The maximum SAM value is obtained if the sequences are highly dissimilar. This means in this example that the occurrences in one

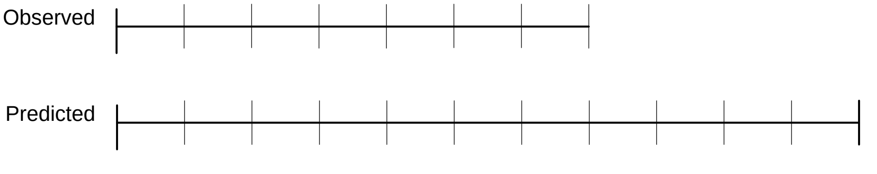


Figure 7 | 6: Two sequences without reference to years

sequence (observed or predicted) does not occur in the other sequence at all. In this case, all occurrences in the sequence are deleted and occurrences that correspond with the other sequence are inserted. If the observed sequence should match the predicted sequence first all eight occurrences are deleted and then twelve occurrences are inserted. The maximum SAM value is in this example is eight plus twelve, thus twenty. In the calculation of the minimum and maximum SAM value the prediction of the length of the sequence is not taken into account. The count validation test already checked this aspect (length of sequence). The SAM value is indicated with a percentage, calculated as: percentage SAM value = (SAM value - SAM value$_{min}$) / SAM value$_{range}$

The lower the percentage the closer the SAM value is to the minimum SAM value, thus the better the performance. In total 3495 cases were analysed. In some cases, the length of the observed or predicted sequence is zero. This means that no occurrences were observed or predicted for that person. These cases, in total 723, were not included in the analysis. In total, 2772 cases were analysed.

Table 7 | 15: Results OccurrenceOnly sequence (UDSAM)

	Minimum	Maximum	Mean	Std. Deviation
Length sequence	1	31	8.54	5.83
Minimum SAM	0	28	3.73	3.73
Maximum SAM	2	62	17.08	11.66
SAM range	2	60	13.35	10.14
SAM value	0	31	8.48	5.74
SAM percentage	0	1	0.35	0.26

Table 7 | 16: Results SubstatesOnly sequence (UDSAM)

	Minimum	Maximum	Mean	Std. Deviation
Length sequence	1	31	8.54	5.83
Minimum SAM	0	28	3.73	3.73
Maximum SAM	2	62	17.08	11.66
SAM range	2	60	13.35	10.14
SAM value	0	37	9.79	6.69
SAM percentage	0	1	0.45	0.28

The results of the analysis of the OccurrenceOnly sequence option are listed in Table 7 | 15, while Table 7 | 16 shows the results for the SubstatesOnly sequence option. Of course, the values for the length of the sequences, the minimum and maximum SAM value and the SAM range are the same for the OccurrenceOnly and SubstatesOnly sequence option. Those values are listed in the first four rows of the tables.

The percentage of the analysis of OccurrenceOnly and SubstatesOnly sequence are both below 0.5. The mean percentage of the SubstatesOnly sequence option (0.45 in Table 7 | 16) is higher than the mean percentage of the OccurrenceOnly sequence option (0.35 in Table 7 | 15). This can be explained by the difference in level of detail between the two sequence options. More detail in the sequence means a higher probability of differences between the sequences, which leads to a higher SAM (and SAM percentage) value.

The length of each string in the Life Trajectory sequence option is the same for the observed and predicted sequences. If a person reported a life trajectory of twelve years the predicted life trajectory also consists of twelve years. Therefore, the minimum SAM value is always zero. SAM values were calculated for 3495 cases. The total length among the 3495 cases varied from 1 year to 40 years. The maximum SAM values are calculated as follows: SAMmax = length sequence x 2 x 7. The multiplication with 2 is for replacing a value in the sequence and the multiplication with 7 is for every event. This is rather straightforward and a simple calculation, DPSAM methods is not taken into account in this calculation. The SAM percentage is calculated as described before. Table 7 | 17 reports the results of the Life Trajectory sequence.

The mean value of the total SAM value based on the uni-dimensional (UDSAM) method is higher (15.91) than the mean value based on the multidimensional method (10.35). In general, the DPSAM value is always below the sum of UDSAM or it can exactly match the sum. The mean value of the SAM percentage (DPSAM) is close to zero. The last column in Table 7 | 17 lists the number of cases with a SAM value of zero. The results indicate that the study event and the PT pass event were simulated most closely by the Bayesian Belief Network.

Table 7 | 17: Results Life Trajectory sequence

DPSAM analyses	**Minimum**	**Maximum**	**Mean**	**Std. Deviation**	**Zero SAM**
Length sequence	1	40	10.93	10.37	
Minimum SAM	0	0	0.00	0.00	
Maximum SAM	14	560	152.98	145.12	
DPSAM analyses					
SAM value	0	49	10.35	9.04	
SAM percentage	0	0.14	0.08	0.04	
UDSAM analyses					
SAM value Housing	0	20	2.75	2.83	1179
SAM value Household	0	18	3.00	3.32	1393
SAM value Work	0	20	3.24	3.59	1369
SAM value Study	0	6	0.32	0.83	2984
SAM value Car	0	26	2.36	2.73	1353
SAM value PT pass	0	16	1.54	2.33	2033
SAM value Income	0	20	2.70	3.11	1422
SAM value total	0	66	15.91	13.68	

Highest SAM measures were obtained for the work event and the household event, suggesting that the Bayesian network was relatively less successful in correctly simulating the timing of occurrences and/or sequence of these event types.

Based on the SAM measures, it is difficult to tell whether faulty predictions are caused by (1) predicting the wrong subevent or (2) predicting a wrong timing of the subevent. Overall, however, the results of the SAM analyses demonstrate that the learned Bayesian Belief Network predicts the sequence of the occurrences in the life trajectories relatively well.

4 | Validation mode choice network

In this section, first the log likelihood will be discussed for three models: null-model, overall model and prediction model. Mode choice was only registered in the year of the retrospective Internet-based survey (2004). This means that the observed data for mode choice is limited to one year. Beside the goodness-of-fit, one test is executed. The observed modal split in 2004 is compared with the predicted modal split.

1 | Goodness-of-fit

One way of assessing the validity of the learned network is to use the values for all variables/nodes in the network in the year 2004, except mode choice, and predict the posterior probabilities of a particular transport mode choice in the year 2004, given this hard evidence. In this case, the hard evidence corresponds with the situation in the corresponding year.

The network was used to calculate the posterior probabilities for every case given the hard evidence. To check the performance of this mode choice model the log likelihood value for three models (null-model, overall model and prediction model) was calculated. Table 7 | 18 lists the values for the different models. In the overall model, the probabilities were distributed according to the overall probability distribution. Figure 6 | 14 illustrates the following distribution: car = 0.480, ST = 0.382 and PT = 0.138. Two Rho-Squares were calculated, the first one is the ratio between the log likelihood of the prediction model and null-model, and the second one is the ratio between the log likelihood of the prediction model and overall model. The value of both Rho-Squares is reported in Table 7 | 18. Both Rho-Square values of the model are above 0.36, which indicate that the model performs relatively well.

Table 7 | 18: Log likelihood values mode choice network

	Null model	**Overall**	**Prediction**	**First Rho-Square**	**Second Rho-Square**
Mode Choice	-769.03	-697.57	-439.84	0.4281	0.3695

2 | Mode choice in 2004

Monte Carlo sampling is used to select an individual's mode choice based on the predicted posterior probabilities. The predicted modal split can be calculated given the individuals' mode choices. The observed modal split in 2004 is compared with predicted modal split to show how the model can be used for simulation. The results are shown in Table 7 | 19 and Table 7 | 20.

Only one Monte Carlo simulation was run. In 64 percent of the cases the predicted individual's mode choice corresponds with the observed individual's mode choice (not shown in any table). However, it can be calculated by the ratio of the sum of the values in the diagonal (199 + 220 + 29 = 449) and the total (700). Table 7 | 20 reports a relatively small over prediction of public transport at the expense of the other two mode choices car and slow transport.

Table 7 | 19: Mode choice in 2004 (observed and predicted)

		observed mode			
		Car	**PT**	**ST**	**Total**
predicted mode	**Car**	199	69	23	**291**
	PT	70	220	37	**327**
	ST	23	30	29	**82**
	Total	**292**	**319**	**89**	**700**

Table 7 | 20: Modal split in 2004 (observed and predicted)

Mode	observed	%	predicted	%	Difference	prediction
Car	292	0.417	291	0.416	-0.001	Under prediction
PT	319	0.456	327	0.467	0.011	Over prediction
ST	89	0.127	82	0.117	-0.010	Under prediction
Total	700	1.000	700	1.000		

5 | Conclusion

The log likelihood values regarding the prediction of the occurrence of events and mode choice indicate that both networks perform relatively well. In this chapter, it was investigated whether the life trajectory model was capable of representing a complete life trajectory. The predicted life trajectories are compared with the observed life trajectories as stated in the Internet-based survey. Aspects of the life trajectory that were used for validation were the number of occurrences, interval times between occurrences of events, simultaneous occurrences of events and sequence of occurrences of events. The modal split (car, Public Transport and Slow Transport) of the predicted mode choice was compared with the observed mode choice. The results of the four validation tests are summarised.

(1) Results indicate a slight under prediction of the number of occurrences for the household event and an over prediction of the PT pass event at one of the three levels of detail. Thus, the life trajectory model reproduced the number of occurrences in the life trajectories quite well.

(2) Two interval times were tested: the sequence interval and intervals between two events. In general, the model predicted more or less the same interval times for the events, except for the PT pass event.

(3) The results of the validation test showed that the network was less successful in predicting correctly the observed synchronic events.

(4) The sequence of events is another important aspect of a life trajectory. The results of the Sequence Alignment Method demonstrated that the Bayesian Belief Network more of less reproduced the sequence of events as registered in the observed life trajectory. Three levels of detail were tested, the first two have no reference to the year of the occurrence and the difference between the two levels is the level of detail (without or with subevents). The third level had a reference to the year of the occurrence and included all subevents. The results of the analyses of all three levels illustrate that the network predicts the sequence of the occurrences in the life trajectories relatively good.

For the mode choice network, the test of validity involved examining whether modal split was predicted correctly in the simulation. Results of this illustration indicated a small over prediction of public transport and under prediction of car and slow transport. This suggests that the mode choice network is able to simulate more or less the same mode choice as registered in the data.

This chapter illustrated the method of validation for both networks. As mentioned before, no hold-out sample was used. A new data collection is necessary for external validation.

8 | Application

1 | Introduction

In the previous chapter the goodness-of-fit of both networks is discussed as well as the results of validation tests for the life trajectory and the mode choice network. This chapter illustrates the use of both networks. The application of both networks is clarified in two ways. First, the network can be used to study direct and indirect effects of variables on other variables in the network. Second, the network can be used for simulation purposes. For example, life trajectory and mode choice dynamics of new inhabitants in a neighbourhood can be simulated using both networks together. Mode choice is predicted for every year of the predicted life trajectory. The influence of predicted mode choice in the previous year is also taken into account given a parameter.

This chapter starts with an example that illustrates how hard evidence can be inserted in the life trajectory network for the four states of the housing event (i.e. no change, independent housing, student housing and parental housing) to simulate direct and indirect effect on the other nodes. The evidence is discussed for the seven occurrence event nodes. Next, in the mode choice network, the influence of hard evidence is illustrated with the node car availability state as an example. Here the mode choice probabilities are given for two situations (i.e. before and after entering evidence). The use of the networks to simulate full life trajectories is discussed in section four. Details are illustrated in section five, which describes a scenario that is used for simulating life trajectories and mode choice until the year 2019. One personal life trajectory is discussed as an example and mode choice is predicted given three different parameter values. The chapter ends with a short summary and conclusions.

2 | Hard evidence life trajectory network

The indirect and direct influence of variables / nodes on other variables/nodes in the network can be analysed when hard evidence is entered into the network. When hard evidence for one variable / node is entered into the network the probabilities for all other variables/nodes will be updated. The effect of entering hard evidence into the network can give more insight in the influence and relationship between variables / nodes. As mentioned in the introduction hard evidence is entered in the life trajectory network for the four states of the housing event. Direct and indirect effect on the seven occurrence event nodes is registered with (changed) probabilities.

Figure 8 | 1 shows only a small part of the complete life trajectory network. The nodes of interest (i.e. the occurrence event nodes) are illustrated. This figure portrays the situation when no evidence is entered into the network. Figure 8 | 2 shows the updated probabilities for all seven nodes after the evidence “change, independent housing” is entered into the network. The node that contains the hard evidence is coloured grey and the probability for that particular state is set to 100.

Table 8 | 1 lists the probabilities for all occurrence event nodes without and with evidence entered into the network. The first column lists all seven occurrence event nodes with their classes (e.g. states). The second column consists of the probabilities for all nodes (and states) when no evidence is entered into the network. The next four columns illustrate the new and updated probabilities when hard evidence for different states of the housing event node is entered into the network, as an example. The state with value 1 (in the last four columns) corresponds with the hard evidence entered into the network. The other states of the same node have automatically the value 0. In this example, hard evidence was entered four times into the network, once for every state of the housing event node.

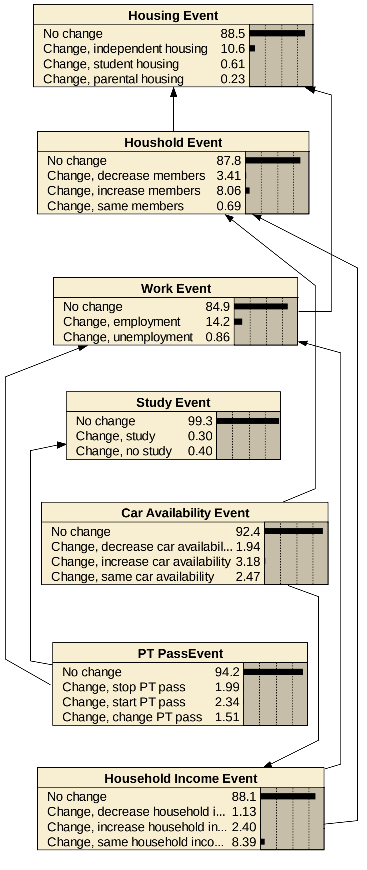


Figure 8 | 1: No evidence entered into the life trajectory network

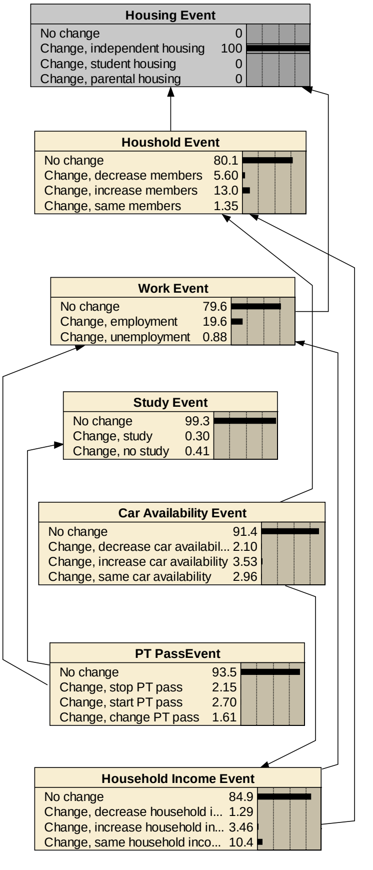


Figure 8 | 2: Evidence “independent housing” entered into the network

Table 8 | 1: Updated probabilities hard evidence Housing event

Events	no evidence	Evidence= no change	Evidence= independent housing	Evidence= student housing	Evidence= parental housing
Housing					
No change	0.885	1	0	0	0
Independent	0.106	0	1	0	0
Student housing	0.006	0	0	1	0
Parental housing	0.002	0	0	0	1
Household					
No change	0.878	0.887	0.801	0.911	0.862
Decrease	0.034	0.032	0.056	0.029	0.023
Increase	0.081	0.075	0.130	0.052	0.110
Same members	0.007	0.006	0.014	0.007	0.005
Work					
No change	0.849	0.856	0.796	0.830	0.761
Employment	0.142	0.136	0.196	0.145	0.232
Unemployment	0.009	0.009	0.009	0.025	0.007
Study					
No change	0.993	0.993	0.993	0.993	0.993
Study	0.003	0.003	0.003	0.003	0.003
No study	0.004	0.004	0.004	0.004	0.004
Car Availability					
No change	0.924	0.925	0.914	0.924	0.920
Decrease car	0.019	0.019	0.021	0.019	0.021
Increase car	0.032	0.031	0.035	0.032	0.033
Same car	0.025	0.024	0.030	0.025	0.026
PT Pass					
No change	0.942	0.943	0.935	0.938	0.933
Stop PT pass	0.020	0.020	0.022	0.021	0.023
Start PT pass	0.023	0.023	0.027	0.025	0.030
Change PT pass	0.015	0.015	0.016	0.016	0.015
HH Income					
No change	0.881	0.885	0.849	0.887	0.852
Decrease	0.011	0.011	0.013	0.012	0.012
Increase	0.024	0.023	0.035	0.021	0.031
Same household	0.084	0.082	0.104	0.081	0.104

The initial probabilities (e.g. before hard evidence is entered into the network) and the updated probabilities (e.g. after hard evidence is entered into the network) are compared to analyse the influence of the housing event node. If the probabilities for a certain state increase then the hard evidence has a positive influence on that particular state. A lower probability indicates a negative influence.

Overall, the probabilities do not change substantially. A few examples with positive and negative influence of hard evidence are described below. The hard evidence “independent housing” results in higher probabilities for the household states “decrease members”, “increase members” and “same members”. This means, in this example, that the occurrence of living independently influences the number of household members. These two events can happen at the same time (e.g. in the same year). On the one hand, the hard evidence “student housing” has a positive influence on the probabilities for the work state “unemployment”.

The hard evidence “student housing” has a negative influence on the probability for the household state “decrease members” and “increase members”. This means that the initial probability becomes smaller after hard evidence is entered into the network. This seems logical: when a person moves to a student house the number of household members is set to one. This implies that an increase in the number of members is less likely. The hard evidence “parental housing” results in a lower probability for the household state “decrease members” and at the same time a higher probability for the household state “increase members”. Moving to the parents’ place often means an increase in the number of household members; a decrease in the number of household members is less likely.

In this example, only the influence of the housing event node on other occurrence event nodes in the network is discussed. The housing event seems to influence the household and the work events. All other probabilities show no major influence of the housing event. The rest of the occurrence event nodes can be analysed the same way as described here.

3 | Hard evidence mode choice network

To illustrate the use of the model, in the same way as explained in the previous section, hard evidence is entered into the mode choice network. The node of interest here is mode choice. Four state nodes have a direct link with mode choice (see Figure 6 | 14): housing state, car availability state, PT pass state and household income state. As an illustration, hard evidence is entered into this network for the node car availability state.

Figure 8 | 3 illustrates a part of the mode choice network. In this situation no hard evidence is entered into the network. Figure 8 | 4 shows the same part of the network when hard evidence is entered into the network. In this example "no car" is entered as evidence. All probabilities are automatically updated and the probability for the mode choice option "car" is reduced from 48.0% to 9.93%. This newly updated probability does not equal 0%, while there is no car available. That seems strange, but in fact it is explainable. The mode choice option "car" consist of two options: car as driver and car as passenger. The latter option is always available even when there is no car available in that particular household. This results in a probability higher than 0% for the mode choice option "car".

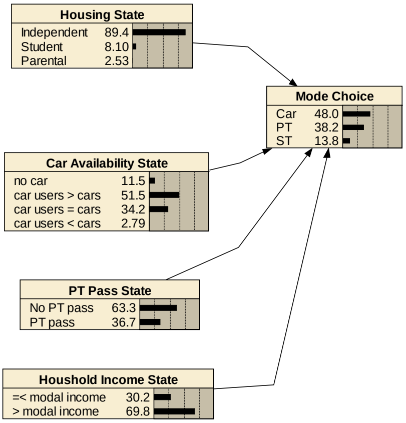


Figure 8 | 3: No evidence entered into the mode choice network

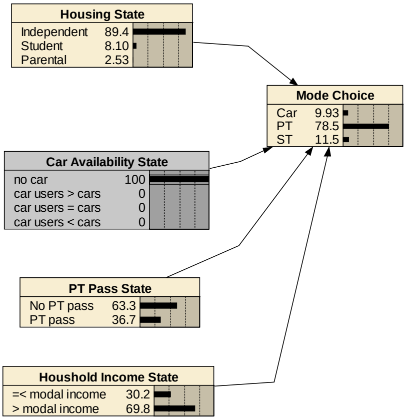


Figure 8 | 4: Evidence "no car" entered into the network

Table 8 | 2: Updated probabilities hard evidence car availability state

Mode Choice	No evidence	Evidence= no car	Evidence= car users> cars	Evidence= car users= cars	Evidence= car users < cars
Car	0.480	0.099	0.444	0.631	0.851
Slow Transport	0.382	0.785	0.397	0.248	0.086
Public Transport	0.138	0.115	0.159	0.121	0.063

Table 8 | 2 shows the probabilities for mode choice without and with evidence entered into the network. In the second column, the probabilities without evidence are listed. The next four columns illustrate the updated probabilities when hard evidence for different “car availability” states are entered into the network.

The initial probabilities and the updated probabilities are compared to analyse the influence of the car availability state. A positive influence of the hard evidence means that the updated probabilities are higher than the initial probabilities and a negative influence means a lower probability.

The initial probability of the mode choice “slow transport” increases from 0.382 to 0.785 when the hard evidence “no car” is entered into the network at the car availability state node. At the same moment the probability for the mode choice “car” decreases substantially, whereas the probability for “public transport” as mode choice option decreases a little. These influences seem logical.

In the following situation, when there are more cars in the household than car users (last column), the probabilities for the “car” increase from 0.480 (initial probability) to 0.851 (updated probability). In this situation, the probabilities for the other two mode choice options (slow transport and public transport) decrease. These influences also seem logical. Influences from other (state) nodes on mode choice are not discussed in this section, but they can be analysed the same way.

4 | Conditional mode choice

As mentioned before, mode choice can only be predicted for a certain year given the life trajectory of previous years with the mode choice network. Obviously there is also an influence of the chosen mode in the previous year(s) on the mode choice in the current year. Since there is no data of previous years in terms of mode choice, this conditional dependency cannot be represented in the BBN estimated. In this section, it is illustrated how the influence of mode choice in the last year can be taken into account when predicting mode choice in the current year, by representing the degree of conditional dependency as a parameter.

Figure 8 | 5 illustrates that given the available data mode choice can be predicted using the mode choice network in the year for which mode choice data is available. However, as illustrated in Figure 8 | 6 mode choice in any given year may also be influenced by mode choice in a previous year. In order to incorporate the influence of the mode choice in the previous year the following method is suggested.

Figure 8 | 5: Mode choice without influence mode choice previous year

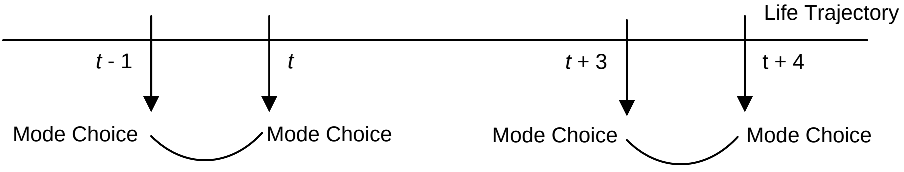


Figure 8 | 6: Mode choice with influence mode choice previous year

The method is based on a distinction of two hypothetical, extreme cases of behaviour. In the first extreme, the individual is maximally consistent in the sense that he/she will always choose the same transport mode as in the previous year when there are no changes in conditions. In the second extreme, the individual is minimally consistent in the sense that mode choice in the previous year has no influence at all on choice probabilities in the current year. In reality, the behaviour will be somewhere in between these two extremes. The method uses a parameter (*alpha*) that indicates on a zero-one scale the degree of history dependence. *Alpha* = 1 represents absence of influence and a*lpha* = 0 means a maximum influence of the previous year choice on the current choice. Posterior probabilities for mode choice can be calculated for each extreme case (as explained below) and actual probabilities can then be determined by calculating a weighted average of the probabilities across the two extremes using *alpha* as a weight.

Given this approach, what remains to be defined is a method to determine choice probabilities for the two extreme cases. The method proposed is based on the following reasoning. The absence of influence case (*alpha* = 1) is straightforward: since there is no influence of history the choice probabilities correspond to the choice probabilities generated by the Bayesian Belief Network (which assumes the same). The full influence case (*alpha* = 0) is more complicated. That is to say, if the BBN does not predict a change in choice probabilities, because no state change occurred, then full influence simply means that the individual chooses the same mode as in the year before, implying that probabilities reduce to a zero-one distribution. However, if there is a state change then choice probabilities change and even if the influence of last year is maximal there is still some proportion of individuals (facing the same circumstances) that do change. In other words, the posterior probabilities are constrained by the requirement that across all individuals under same conditions the probabilities should summate to the probabilities predicted (for that year) by the BBN. The following equations define how posterior probabilities (conditional upon last year choices) are defined based on this logic. Note that under both *alpha* = 0 and *alpha* = 1 the resulting probabilities are consistent with those generated by the BBN. The BBN probabilities are realised only in a different way: all individuals stick maximally to their last choice versus all individuals act as if they are unaware of previous choices.

Since the probabilities for both extremes are consistent with the BBN prediction, the weighted averages determined by *alpha* are so too. The posterior probabilities can be calculated given the following equations:

$$P^0{}_{j|i} = P_j \qquad \forall i$$

$$P^1{}_{j|i} = \min(1; P_j / P_i) \qquad \text{If mode } i \text{ is mode } j$$

$$P^1{}_{j=car|i=PT} = \min(1 - P^1{}_{j=PT|i=PT} - P^1{}_{j=ST|i=PT}; (P^1{}_{j=car} - P^1{}_{j=car|i=car} \cdot P_{i=car} - P^1{}_{j=car|i=ST} \cdot P_{i=ST}) / P_{i=PT}) \qquad \text{If mode } i \text{ is not mode } j$$

$$P_{j|i} = \alpha \cdot P^0{}_{j|i} + (1-\alpha) \cdot P^1{}_{j|i} \qquad \forall i$$

where:

mode has three options: car, PT, and ST

P_i = a-priori probability of mode *i* in year *t*

P_j = a-priori probability of mode *j* in year *t+1*

$P^0{}_{j|i}$ = conditional probability of mode *j* given mode *i* in the *alpha* = 0 situation

$P^1{}_{j|i}$ = conditional probability of mode *j* given mode *i* in the *alpha* = 1 situation

α = *alpha* value

In this way, the following conditions are met:

$$\sum_i P^0{}_{j|i} \cdot P_i = P_j \qquad \forall j$$

$$\sum_i P^1{}_{j|i} \cdot P_i = P_j \qquad \forall j$$

$$\sum_i P_{j|i} \cdot P_i = P_j \qquad \forall j$$

The process of calculating the conditional mode choice is illustrated given Figure 8 | 7. The simulation results for four successive years provide probabilities for mode choice in year 1, 2, 3 and 4. In the first year, the mode choice is determined using Monte Carlo simulation for the probabilities results of the simulation in the corresponding year (illustrated in the first row of the Figure 8 | 7). Mode choice, in the other years, is calculated as described. Thus, the posterior probabilities are calculated and Monte Carlo simulation is used to predict the mode choice.

In the next section the conditional mode choice is calculated given three different *alpha* values.

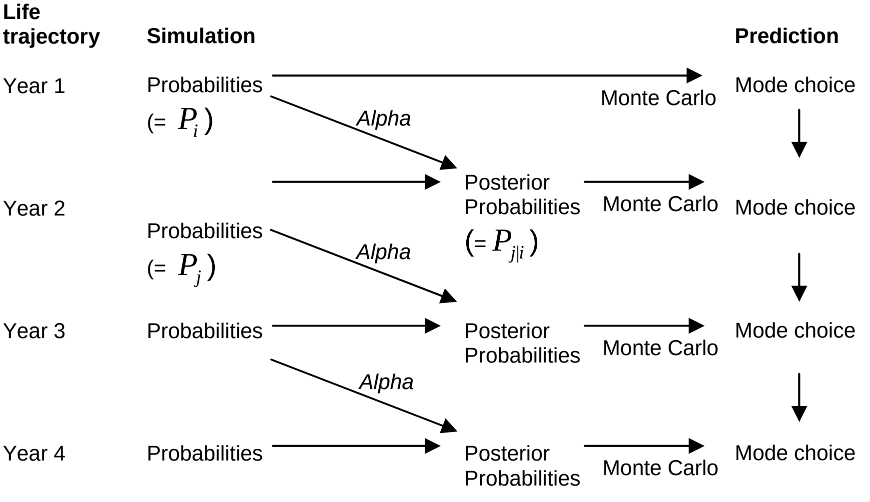


Figure 8 | 7: Calculation of mode choice for several years

5 | Scenario simulation

To illustrate the application of both models a scenario is developed. One simulation run is used to illustrate this scenario. New houses are built every year in the Netherlands. Sometimes there is no space left in the city which results in housing developments outside the city (centre). Those developments are referred to as VINEX in the Netherlands, which is an abbreviation of Vierde Nota Extra. Houses are built for different target groups. It is interesting to study the different life trajectories of the inhabitants of those new housing developments and their mode choice. For example, these dynamics are at the moment not taken into account in the calculation of the traffic performance of a certain neighbourhood.

In this illustration, it is assumed that ten inhabitants move into their houses in the new neighbourhood in 2009. Life trajectories are predicted for these ten inhabitants for eleven years, until the year 2019. The input for the life trajectory network consists of values for all variables/nodes in the life trajectory network, except the occurrence event nodes. This means that hard evidence is entered for the current situation (state nodes), the time ago an event occurred (time ago variables), for how many years ago information is available about that event (history nodes), and for age and gender (personal characteristics). The occurrence of the seven events nodes is predicted for all years. Remember only one simulation run is used in this example. This scenario illustrates the application of the models and is not meant to illustrate differences in the simulations.

Although the simulation involves all years, only three moments in time are illustrated here: the states of the ten inhabitants in the new neighbourhood in the first year (2009), in the year 2013 and finally in the last year (2019). Table 8 | 3 lists the states in the first year at the start of the simulation. In Table 8 | 4 the states in the year 2013 are listed and Table 8 | 5 illustrates the year 2019.

In 2009 40% of the inhabitants are females and 60% males. There are no persons younger than 25, 30% of the persons are between 25 and 31 years of age, 10% is older than 46 and the remaining inhabitants (60%) are between 32 and 46 years of age. All persons live independently, one person has a one person household, 60% of the inhabitants have a two-person household, one person has a three-person household and the rest (20%) belongs to a household of four or more household members. One person is unemployed and nobody is studying. 40% of the inhabitants has more car users in the household than cars and 60% has an equal number of car users and cars. Three persons have a PT pass, and all persons have an above modal household income.

Table 8 | 3: States in year 2009

Age	Gender	Housing	HH	Work	Study	Car	PT pass	Income
30	Female	independent	2	employed	Not studying	Car users < cars	PT pass	> modal income
33	Female	independent	3	employed	Not studying	Car users < cars	No PT pass	> modal income
41	Female	independent	4	employed	Not studying	Car users = cars	No PT pass	> modal income
56	Male	independent	2	unemployed	Not studying	Car users = cars	No PT pass	> modal income
36	Male	independent	2	employed	Not studying	Car users < cars	PT pass	> modal income
30	Female	independent	2	employed	Not studying	Car users = cars	No PT pass	> modal income
37	Male	independent	5	employed	Not studying	Car users = cars	No PT pass	> modal income
32	Male	independent	2	employed	Not studying	Car users < cars	No PT pass	> modal income
31	Male	independent	2	employed	Not studying	Car users = cars	No PT pass	> modal income
42	Male	independent	1	employed	Not studying	Car users = cars	PT pass	> modal income

Table 8 | 4: States in year 2013

Age	Gender	Housing	HH	Work	Study	Car	PT pass	Income
34	Female	independent	2	employed	Not studying	Car users = cars	No PT pass	> modal income
37	Female	independent	4	employed	Not studying	Car users < cars	No PT pass	> modal income
45	Female	independent	4	employed	Not studying	Car users = cars	No PT pass	> modal income
60	Male	independent	2	unemployed	Not studying	Car users = cars	No PT pass	> modal income
40	Male	independent	2	unemployed	Not studying	Car users < cars	PT pass	< modal income
34	Female	independent	3	employed	Not studying	Car users = cars	No PT pass	> modal income
41	Male	independent	3	employed	Not studying	Car users = cars	No PT pass	< modal income
36	Male	independent	2	employed	Not studying	Car users < cars	No PT pass	> modal income
35	Male	independent	2	employed	Not studying	Car users = cars	No PT pass	> modal income
46	Male	independent	2	employed	Not studying	Car users = cars	PT pass	> modal income

Table 8 | 5: States in year 2019

Age	Gender	Housing	HH	Work	Study	Car	PT pass	Income
40	Female	independent	1	employed	Not studying	Car users = cars	No PT pass	> modal income
42	Female	independent	4	employed	Not studying	Car users < cars	No PT pass	> modal income
51	Female	independent	3	employed	Not studying	Car users = cars	PT pass	> modal income
66	Male	independent	2	unemployed	Not studying	Car users = cars	No PT pass	> modal income
46	Male	independent	2	unemployed	Not studying	Car users < cars	PT pass	< modal income
40	Female	independent	4	employed	Not studying	Car users = cars	No PT pass	> modal income
47	Male	independent	4	employed	Not studying	Car users = cars	No PT pass	> modal income
42	Male	independent	1	employed	Not studying	Car users < cars	No PT pass	> modal income
41	Male	independent	2	employed	Not studying	Car users = cars	No PT pass	> modal income
52	Male	independent	1	employed	Not studying	Car users = cars	PT pass	> modal income

In the year 2013, every inhabitant will be four years older. This will result in no persons younger than 31; 10% will be older than 46 and the rest (90%) will be between 32 and 46 years of age. In the simulation, the housing situation will not change for these inhabitants. The one-person household is predicted to change into a two-person household. In total, four household are simulated to change in these five years. In one household, the number of household members is predicted to decrease, while in the other three households the number of members is predicted to increase. In the simulation, someone will lose his job before the year 2013. This will result in two persons who will be unemployed in 2013. There are no changes predicted related to study. Car availability is predicted to increase for one person, while for the other inhabitants car availability will remain stable. Only for one person a change in PT pass is simulated. This person will no longer posses a PT pass; her car availability will increase. This means that the PT pass will be exchanged for an extra car. The modal household income of the person who will be unemployed in 2013 is predicted to decrease.

In the year 2019, four persons will be older than 46 and the rest (60%) will be between 32 and 46 years of age. The housing situation is predicted to be independent for all inhabitants. Compared to the situation in 2013, the number of household members is predicted to decrease in four households and increase in two households. The work and study states will the same as in the year 2013. The car availability state, PT pass state and household income will be the same as six years before.

The simulated life trajectory of the one inhabitant is illustrated in Figure 8 | 8. The states of all events are on the x-axis, the names are listed on the right side of the figure. Each line in the figure refers to an event (i.e., career). For example the top line illustrates the household career for this person. The blocks, squares etc. correspond with an occurrence. Sometimes an event immediately results in a different state and sometimes not. The number of household members decreases from two to one household member. The other events in this life trajectory do not affect the corresponding states. The occurrence in the PT pass career in 2011 does not change the possession of the PT pass. The mode choice was also predicted in a simulation for these ten inhabitants for every year of the life trajectory. Figure 8 | 9 – 8 | 11 illustrate mode choice for this one person.

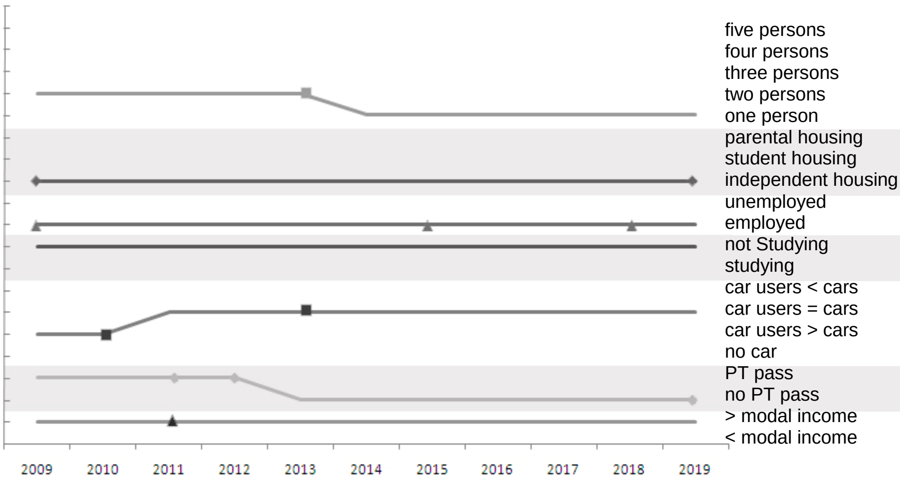


Figure 8 | 8: Simulated life trajectory of person one

Figure 8 | 9 shows that this person is predicted to switch a few times between different transport mode options. First, a switch is predicted from car in 2009 to Public Transport in 2010. Next, the person is predicted to switch back to the car in 2011 and in the following year the person is simulated to switch to Slow Transport. After one year there is a predicted switch back to the car and this mode is predicted to be used until 2018. In 2019, the person will use Public Transport as the main transport mode.

Figure 8 | 10 shows the results for a lower *alpha* value of 0.5. In this case, only two switches are predicted. This person is simulated to switch from car in 2009 to public transport in 2010 and in 2011, a switch back to car as main transport mode is predicted. For the rest of the simulated time period, transport mode remains stable.

Figure 8 | 11, which is based on an *alpha* value of 0.2 shows no switch between the different transport modes. The person uses the car as the main transport mode during these ten years.

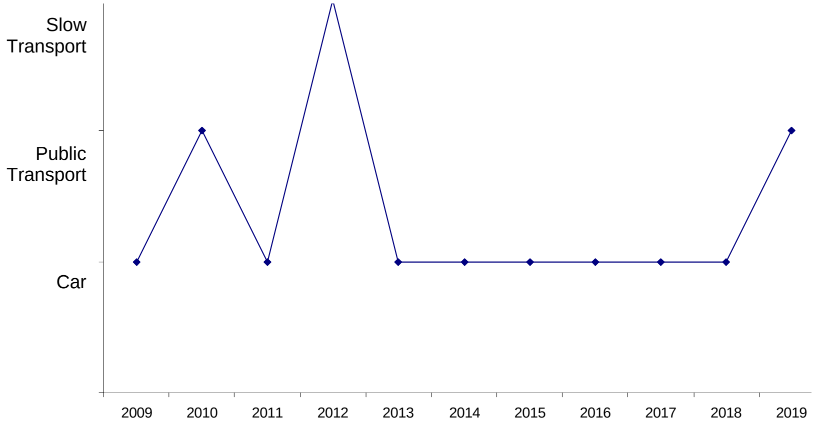


Figure 8 | 9: Mode choice person one with *alpha* value 0.8

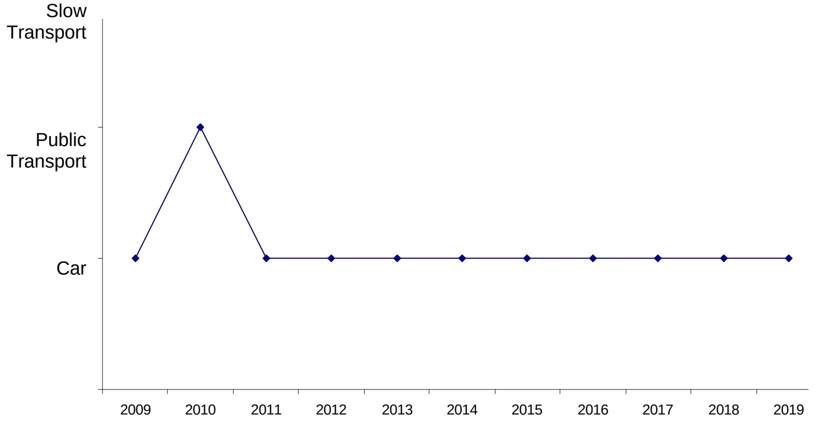


Figure 8 | 10: Mode choice person one with *alpha* value 0.5

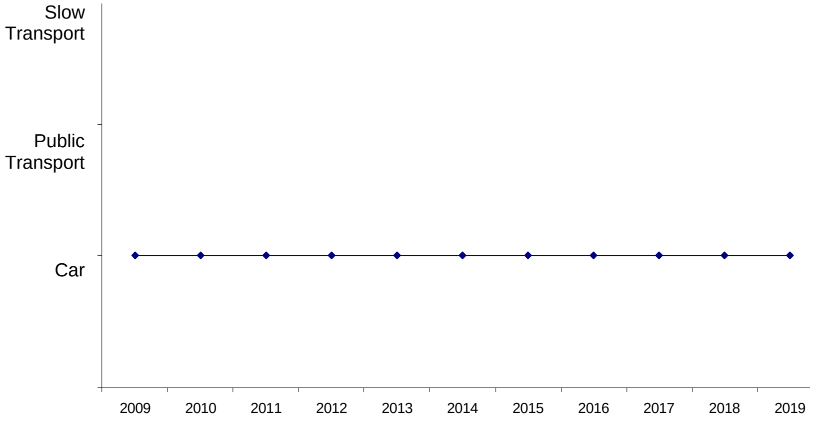


Figure 8 | 11: Mode choice person one with *alpha* value 0.2

These three figures illustrate the effect of the *alpha* value. The lower the *alpha* value, the less switching behaviour is predicted. In this example (*alpha* = 0.2), the transport mode in the first year determines the mode choice for all other years. The higher the *alpha* value, the more switching behaviour is predicted.

The transport mode choice for all ten inhabitants during the eleven years is reported in three tables (Table 8 | 8 - Table 8 | 10). The three tables illustrate the same effect as indicated in the three figures of person's one mode choice.

Table 8 | 6: Mode choice with *alpha* value 0.8

2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
car	PT	car	ST	car	car	car	car	car	car	PT
car	car	car	PT	car	car	car	PT	car	PT	PT
car	car	car	car	car	car	car	PT	car	ST	ST
car	car	car	car	car	car	car	car	car	car	PT
car	PT	PT	PT	car	car	PT	PT	car	ST	ST
car	car	car	car	car	car	car	car	car	car	PT
car	car	car	PT	car	car	car	PT	car	PT	PT
car	car	car	PT	car	car	car	PT	car	PT	PT
car	car	car	car	car	car	car	car	car	car	PT
car	PT	car	ST	car	car	car	PT	car	ST	ST

Table 8 | 7: Mode choice with *alpha* value 0.5

2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
car	PT	car	car	car	car	car	car	car	car	car
car	car	car	car	car	car	car	car	car	car	PT
car	car	car	car	car	car	car	car	car	PT	ST
car	car	car	car	car	car	car	car	car	car	car
car	PT	PT	PT	car	car	car	PT	car	PT	ST
car	car	car	car	car	car	car	car	car	car	car
car	car	car	car	car	car	car	car	car	car	car
car	car	car	car	car	car	car	car	car	car	PT
car	car	car	car	car	car	car	car	car	car	car
car	car	car	car	car	car	car	car	car	PT	ST

Table 8 | 8: Mode choice with *alpha* value 0.2

2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
car	car	car	car	car	car	car	car	car	car	car
car	car	car	car	car	car	car	car	car	car	car
car	car	car	car	car	car	car	car	car	car	Car
car	car	car	car	car	car	car	car	car	car	Car
car	car	car	PT	PT	PT	PT	PT	PT	PT	PT
car	car	car	car	car	car	car	car	car	car	Car
car	car	car	car	car	car	car	car	car	car	Car
car	car	car	car	car	car	car	car	car	car	Car
car	car	car	car	car	car	car	car	car	car	Car
car	car	car	car	car	car	car	car	car	car	Car

Less switching between transport mode choice options are predicted for a lower *alpha* values (Table 8 | 8), while a higher *alpha* value results in more switching behaviour (Table 8 | 6). Modal split for the different years in the simulation is not compared, it is a small simulation sample.

6 | Conclusion

This chapter illustrated the use of the life trajectory model and the mode choice model to analyse direct and indirect effects of variables on other variables in the network. Two examples of entering hard evidence into the network were used for illustration purposes. In the life trajectory network, the influence of the housing event node on the other occurrence event nodes is simulated and in the mode choice network the effect of car availability state on mode choice is explained. Changed probabilities (e.g. updated probabilities) can be higher or lower than the initial probabilities, indicating respectively a positive or negative impact of the hard evidence.

In the life trajectory network, there is an influence of the housing occurrence event on the household and work events. The housing occurrence event showed no major influence on the other occurrence event nodes. Entering hard evidence for other nodes in the network can indicate different direct and indirect influences. The influence from the car availability state on mode choice in the mode choice network is explainable. The influence of other nodes in the network was not included in this example.

In the real-world, people often make a decision about mode choice for a certain trip (for example the work trip) once and do not reconsider all mode options every day. This is habitual behaviour, which can be represented with the influence of mode choice in the previous year on mode choice in the current year. In the mode choice network, there is no link or influence of mode choice in the previous year on mode choice in the current year. However, the posterior mode choice can be calculated given a set of equations. The impact of the influence is controlled with a parameter (*alpha* value). The higher the *alpha* value, the lower the influence of mode choice in the previous year.

Both networks are used for simulation purposes. The life trajectory and mode choice dynamics of ten new inhabitants in a neighbourhood were simulated for the formulated scenario. The life trajectory and the yearly mode choice of these inhabitants are predicted for eleven years, from 2009 until 2019. The current states in every year of the life trajectory give insight in the housing situation, the household composition, the work and education situation and the resources like car availability, PT pass possession and the household income of these ten inhabitants. The life trajectory gives insight in the dynamics of the seven careers of these inhabitants. A person’s life trajectory indicated that an occurrence of an event does not necessarily result in a change in the corresponding state.

As mentioned before, there is no information available about the *alpha* value. Research is necessary to estimate *alpha*. Mode choice data has to be collected for at least two successive years. The context is of course also relevant. Information about the life course events and states in those years needs to be collected as well.

9 | Discussion and conclusions

1 | Introduction

This thesis contributes to the literature on activity-based modelling. More specifically, the focus is on the dynamics underlying activity-travel patterns. The specific focus in this thesis is concerned with life course or life trajectory events that may cause changing needs or preferences and/or influence the constraints that impact activity-travel decisions. This thesis is based on the assumption that life trajectory events may cause individuals and households to change their activity-travel patterns. A modelling approach that allows representing and simulating such dynamics was developed and tested in this thesis. Our special focus was on changes in transport mode choice. In this last chapter the most important results are summarised first. A short discussion follows and possible directions for future research are described.

2 | Short summary (conclusions)

Traditionally, transport mode choice was primarily examined as a stand alone problem. Given a purpose and destination, the choice of transport mode was modelled as a function of the various attributes of the transport mode alternatives. Later, transport mode choice decisions were modelled as part of more comprehensive models (activity-based approach). There is a need in the transport research community to explore and model dynamics in activity-travel patterns along various time horizons. This will lead to dynamic models of behavioural change. In this thesis it is argued that a life course perspective offers some potential advantages in understanding and modelling activity-travel decisions, including transport mode choice. Central concepts in the life course approach are life trajectories, transitions and events. An individual life course is composed of multiple, interdependent careers (i.e. housing, household, education, occupational career) which develop over time in parallel. Earlier life transitions may have a cumulative effect on later life. The concepts of timing, sequencing, duration and spacing are used to describe life events, transitions and trajectories.

The assumed effect of events on activity-travel decisions is captured in terms of a theory of learning and adaptation. Individuals develop and continuously adapt choice rules while interacting with their environment. The context is non-stationary, uncertain and highly dynamic and therefore it is assumed that individuals adapt their behaviour. Under stationary conditions, individuals will show habitual behaviour after some period of time. A life course event is seen as a trigger that may induce individuals and households to reorganise their activities in time and space. A particular event may also lead to other life course events. Thus, life course events may have direct and indirect effects on activity-travel patterns. An event does not necessarily lead to immediate changes in particular facets of activity-travel patterns. Behavioural change may also occur in anticipation of life course events.

Bayesian Belief Networks is the approach adopted in this thesis to model the direct and indirect effects of life course effects on transport mode choice. More complex causation patterns can be included and results can be directly interpreted in terms of the classified events. Such networks need as input empirical data to learn the structure of the network and the conditional probability tables of the variables that are identified to be relevant.

Data was collected using a retrospective Internet-based survey. Retrospective surveys, especially when administered through the Internet, are a good alternative for (quasi-)longitudinal data collection methods, like panel surveys, repeated cross sectional surveys, and cohort pseudo-panel surveys. One would expect that the quality of data coming from a retrospective survey depends on the nature of the event about which information is collected and on the time elapsed between the occurring of the events and the time of the retrospective survey. The quality of the data was tested and the results were positive. In case of life course events memory lapses are less of a problem. Life course events can be better recollected than other events. In this study, the results of the reliability and validity tests of the collected data showed that item non-response in general was relatively low, especially for those life course events that serve as markers unfolding one's life. A statistical analysis suggested that memory / cohort effects were not found for the more salient life course events, such as housing, work and study related events. Memory may have an effect in reporting of events in case of income and transport mode related events (car availability and PT pass). The study illustrated that certain details of events, such as housing type and housing state are more difficult to recall.

The time effect of an influence of life course events on mode choice was tested with a simple multinomial logit model. The results support the conclusion that a certain time influence exists in the response to events. The data of the retrospective Internet-based survey was used as input for two Bayesian Belief Networks, a life trajectory and a mode choice network. A year is chosen as the unit of analysis for these networks. Both networks were successfully learned from the data. The first network can be used to simulate a person's life trajectory and the second network can be used to predict mode choice for an individual at a certain time given the individual life trajectory.

The goodness-of-fit of the learned Bayesian Belief Networks was assessed on the basis of the log likelihood statistic. The values indicated that both networks perform relatively well. It was also investigated whether the life trajectory network was capable of reproducing observed characteristics of complete life trajectories. The observed and predicted life trajectories were compared in terms of the following criteria: the number of occurrences, interval times between occurrences of events, simultaneous occurrences of events and sequence of occurrences of events. The life trajectory network reproduced the number of occurrences in the life trajectories quite well. In general, the network predicted more or less the same means of interval times for the events, except for the PT pass event. The network was less successful in predicting correctly the observed incidence of synchronic events. The results of the sequence alignment analysis indicate that the network predicts the sequence of the occurrences in the life trajectories relatively good. The modal split (car, public transport and slow transport) of the predicted mode choice was compared with the observed mode choice. Results indicated a relatively small over prediction of public transport and under prediction of car and slow transport. This suggests that the mode choice network is able to simulate more or less the same mode choice as registered in the data.

The learned networks were used to study direct and indirect effects of one variable on other variables in the network. The described effects seem logical. A simulation illustrated the dynamics of the lives of ten inhabitants of a newly build neighbourhood. It showed that, insight in dynamics of life trajectory events and mode choice can lead to a better understanding which can support the development of better or different policy measures.

3 | Discussion and future research

This study has indicated that Bayesian Belief Networks (BBNs) are a potentially powerful approach for modelling direct and indirect influences between variables. The potential advantage of BBNs over other techniques, like hazard models and multiple-spell duration models, is that more complex causation patterns can be included and that the results can be directly interpreted in terms of the classified events. On the other hand, whereas duration model have been explicitly developed for addressing time and duration, Bayesian Belief Networks have not. The Bayesian Belief Networks developed in this study rely on a translation of continuous time variables into discrete categories. The reason is that many of the algorithms work best when a finite number of parent states is considered. However, continuous variables may provide a behaviourally more realistic representation of human behaviour. In duration modelling, for example hazard models, time can be handled as continuous variable. The choice of one year as the unit of observation was an operational decision. This choice will impact the conditional probabilities and will also impact the number of zero cells. It is important therefore to systematically investigate the sensitivity of predictions for the choice of time resolution.

In any case, a difficulty of life course analysis may be the lack of longitudinal data. In the context of this study for example, in the Netherlands data about the occurrences of life course events, such as housing, household, work, study, car availability, public transport pass and household income were not available. Retrospective surveys administered through Internet, as used in this study, may be the best alternative, but are not without limitations. Crucial in this context is the way in which memory is triggered. In our survey, respondents answered questions about seven life course events separately, in seven matrices. Sometimes an occurrence in one career is related to occurrences in another career. The memory of one event can therefore also be triggered by the recollection of another event. Placing occurrences of different careers on an interactive time line may offer some advantages compared to the technique used in this study. Every event (e.g. career) can have a separate time line, but all time lines are in one computer screen. This technique can also be applied to collect information about other activity-travel behaviour aspects in retrospect. Respondents can link their behaviour to life course events, which can be recollected very well. Of course, it is important to use an appropriate time period for recollection. This means that the time elapsed between the behaviour of interest (or phenomena) and the time of the survey should not be too long. Future research should systematically compare the performance of alternative recall triggers.

In this study there was no information available about the underlying reasons of occurrences of events. Therefore, it was not possible to investigate if occurrences were in anticipated direction or not. People can react or anticipate to the occurrence of events. If questions about the underlying reasons would be asked, analyses on pro-active versus reactive behaviour can be conducted.

The results of our pilot studies indicated that people had substantial difficulty in recalling their use of transport mode. For that reason, in the retrospective Internet-based survey respondents had to answer questions about their mode choice for different purposes at the time of data collection (and not in retrospect). This decision implies however that the direct and indirect influence of other events on changes in transport mode choice was not part of the model. To alleviate this problem, a modelling strategy was developed to incorporate the influence of mode choice in the previous year on the current year. The parameter representing this influence could not be estimated. This limitation means that either it is necessary to further explore other options for collecting information about the transport mode careers (and other facets of activity travel patterns for that matter), or to collect data to estimate the temporal influence parameter. If it can be assumed that this influence does not change significantly over time, data of mode choice for two consecutive years would suffice. Alternatively, assuming that recall of car purchasing behaviour has a higher reliability, car purchases can be treated and modelled as events, while transport mode choice can be modelled as a function of available resources, and the usual attributes.

Although this thesis was positioned in the context of dynamic activity-based models of transport demand, only transport mode choice was deliberately considered to explore the modelling approach. Future research should however also consider other facets used in activity-travel models, like location choice, duration, etc.. Provided data about these facets are available, these facets can be added to the mode choice network in the future.

The validation tests reported in this thesis are based on the same sample (700 respondents) used for the learning of both networks. This means that the validation results are limited and therefore preliminary. If a larger sample would be available, it would be better to divide the data collection into two samples (training and testing data). One sample can then be used for the modelling (training) and the other sample can be used for validation purposes (testing). Over fitting may have occurred, since splitting the sample was assumed not to be realistic.

Further extension of the networks could also include the reaction of individuals to changes in the transportation system, spatial or economical context. It is interesting for example to study whether people adapt their behaviour when new mode options become available or when public transport options for certain trips are more frequently available.

Finally, the current networks are based on individual life trajectories and not on trajectories of households. Households are a combination of one or more individuals and can be formed, dissolved, split in two or combined. Therefore, it is difficult to model life trajectories for households. Decisions made during the life course are often household decisions. In the current model, there is no guarantee of consistency between events of multiple individuals from the same household. The model can be extended with other household characteristics and the past experiences of household occurrences. Another interesting extension is the link between social networks and life trajectories of people. People exchange information with each other through social interaction. Social networks are also dynamic and change over time. In future research it is important to extend the network in this way. It could lead to better simulation results for dynamic behaviour.

Bibliography

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl and J. Beckman (Eds.), *Action-control: From cognition to behavior* (p. 11-39). Heidelberg, Germany: Springer.

Ajzen, I. (1991). The Theory of Planned Behaviour, *Organizational behaviour and human decision processes*, 50, p. 179-211.

Ajzen, I., and Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Ajzen, I. and Madden, T.J. (1986) Prediction of goal-directed behavior: Attitudes, intentions, and perceived behavioral control. *Journal of Experimental Social Psychology* 22: 453–474.

Arentze, T.A. and Timmermans, H.J.P., eds. (2000). *ALBATROSS: A Learning Based Transportation Oriented Simulation System*, Eindhoven: European Institute of Retailing and Services Studies.

Arentze, T.A., and Timmermans, H.J.P. (2003). Modeling learning and adaptation processes in activity-travel choice A framework and numerical experiment. *Transportation*, Volume 30, p. 37 – 62.

Arentze, T.A. and Timmermans, H.J.P. (2004). A learning-based transportation oriented simulation system. *Transportation Research. Part B: Methodological*. vol. 38. no. 7. p. 613-633.

Arentze, T.A. and Timmermans, H.J.P. (2007). Modelling dynamics of activity-travel behaviour. In Cheng, C.H. and Ho, S.C(eds.): *Transportation Systems: Engineering & Management*. Proceedings of the 12th international Conference of Hong Kong Society for Transportation Studies. Hong Kong: HKSTS.

Arentze, T.A., Timmermans, H.J.P., Jorritsma, P. and Olde Kalter, M.J. (2008). More Gray Hair – But for Whom? Scenario-Based Simulations of Elderly Activity Travel Patterns in 2020. *Transportation*, 35, p. 613-627.

Baddeley, A. (1979). The Limitations of Human Memory: Implications for the Design of Retrospective Surveys. In: *the Recall Method in Social Surveys*, L. Moss and H. Goldstein, University of London Institute of Education.

Balmer, M., Meister, K., Rieser, M., Nagel, K. and Axhausen, K.W. (2008). Agent-based simulation of travel demand: Structure and computational performance of MATSim-T, paper presented at the *2nd TRB Conference on Innovations in Travel Modeling,* Portland, June 2008.

Behrens, R and Del Mistro, R. (2006). Shocking Habits, Methodological issues in analysing changing personal travel behaviour over time. *Proceedings 11th IATBR Conference* (CD-ROM), Kyoto, Japan.

Beige, S. and Axhausen, K.W. (2008). The Ownership of Mobility Tools During the Life Course. *Proceedings 87th TRB* Conference (CD-ROM), 08-1094, Washington D.C..

Beige, S. (2008). *Long-term and mid-term mobility decisions during the life course*, Dissertation, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

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

Ben-Akiva, M.E., Bowman, J.L. and Gopinath, D. (1996). Travel demand model system for the information area. *Transportation*, **25**, p. 241-266.

Best, S. and Krueger, B. (2004). *Internet Data Collection*, SAGE Publications.

Borgers, A.W.J., Timmermans, H.J.P. and van der Waerden, P.J.H.J., (2002). Patricia: predicting activity-travel interdependencies with a suite of choice-based, interlinked analyses. *Transportation Research Record* 1807, p. 145–153.

Bowman, J.L. (1995): *Activity based travel demand model system with daily activity schedules*. MSc thesis in transportation, Massachusetts Institute of Technology.

Bowman, J.L., Bradley, M., Shiftan, Y., Lawton, T.K., and Ben-Akiva, M.E. (1998). Demonstration of an activity-based model system for Portland. Paper presented at the 8th World Conference on Transport Research, Antwerp.

Bowman, J.L., and Ben-Akiva, M.E. (1999). The day activity schedule approach to travel demand analysis. Paper presented at the 78th Annual Meeting of the Transportation Research Board, Washington, D.C., January 10-14.

Centraal Bureau voor Statistiek (2004), Statistisch jaarboek 2004, website CBS: www.cbs.nl (pagina 339, 348)

Cheng, J. (1998). Learning Bayesian networks from data: An information theory based approach, *Doctoral Dissertation*, Faculty of Informatics, University of Ulster, U.K.

Christian, L.M., Dillman, D.A. and Smyth, J.D. (2005). Instructing Web and Telephone Respondents to Report Date Answers in Format Desired by the Surveyor, *Social and Economic Science Research Center*, Report 05-067, Washington.

Cooper, G.F., Herskovits, E. (1992). *A* Bayesian Method for the induction of probabilistic networks from data, *Machine Learning*, 9, 309-347.

Cowell, R.G. and Dawid, A.P. (1992). Fast retraction of evidence in a probabilistic expert system, *Statistics and Computing*, 2: 37 – 40.

Dargay, J.M. and Vythoulkas, P.C. (1999). Estimation of a Dynamic Car Ownership Model, *Journal of Transport Economics and Policy*, Volume 33, Part 3, p. 287-302.

de Leeuw, E.D. (2001). Reducing Missing Data in Surveys: An Overview of Methods? *Quality & Quantity* 35, Kluwer, pp. 147–160.

Dempster, A.P., Laird, N., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, volume: 39, issue: 1, pp. 1-38.

Dijst, M. (1995). *Het Elliptisch Leven*, Ph.D. dissertation, KNAG, Utrecht.

Dykstra, P.A. and Van Wissen, L.J.G. (1999). Introduction: the life course approach as an interdisciplinary framework for population studies, in: P.A. Dykstra and L.J.G. Van Wissen (eds.) *Population Issues; An Interdisciplinary Focus*, New York, Kluwer Academic, pp. 1–22.

Econometric Software (2003), LIMDEP, Econometric Software Inc., New York and Sydney.

Elder, G.H., Jr (1985). Perspective on the Life Course, in: G.H. Elder Jr., *Life course dynamics: trajectories and transitions, 1968 – 1980*. Ithaca, Cornell University Press, p. 23 – 49.

Elder, G.H., Jr. (2000). The life course, in: E. F. Borgatta (ed.) *Encyclopedia of Sociology*, 1614-1622, Macmillan Reference, New York.

Ettema, D.F., Borgers, A.W.J., and Timmermans, H.J.P. (1995). SMASH (Simulation Model of Activity Scheduling Heuristics): empirical test and simulation issues. Paper presented at the Conference on Activity Based Approaches: Activity Scheduling and the Analysis of Activity Patterns, Eindhoven, May.

Ettema, D.F., Daly, A., de Jong, G. and Kroes, E. (1997). Towards an applied activity-based travel demand model. Paper presented at the IATBR Conference, Austin, Texas.

Ettema, D.F., Tamminga, G.F., Timmermans, H.J.P. and Arentze, T.A. (2005). A micro-simulation model system of departure time using a perception updating model under travel time uncertainty. *Transportation Research Part A: Policy and Practice*, **39**, p. 325 – 344.

Ettema, D.F. and Timmermans, H.J.P. (2006). Costs of travel time uncertainty and benefits of travel time information: Conceptual model and numerical examples. *Transportation Research Part C 14*, p. 335–350

Feijten, P. (2005). Life events and the housing career: a retrospective analysis of timed effects. Dissertation. Delft, the Netherlands, Eburon Publishers.

Fellendorf, M., Haupt, T., Heidl, U. and Scherr, W. (1997). PTV vision: Activity-based demand forecasting in daily practice. In D.F. Ettema and H.J.P. Timmermans (eds.): *Activity-based Approaches to Activity Analysis*, Pergamon Press, Oxford, p. 55 – 72.

Fishbein, M., and Ajzen, I. (1975). Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Addison-Wesley, Reading, MA.

Fosgerau, M. (1998). PETRA: an activity based approach to travel demand analysis. Paper presented at the 8th World Conference on Transport Research, Antwerp.

Fujii, S.R., Kitamura, R. and Monma, T., (1998). A utility-based Micro-simulation Model System of Individual's Activity Travel Patterns, Washington, DC, USA.

Fung, R.M. and Crawford, S.L. (1990). Constructor: a system for the induction of probabilistic models, Proceedings of the Seventh National Conference of Artificial Intelegence.

Gärling, T., Brännäs, K., Garvill, J., Golledge, R.G., Gopal, S., Holm, E. and Lindberg, E. (1989). Household activity scheduling. In *Transport policy, management and technology towards 2001: selected proceedings of the Fifth World Conference on Transport Research* vol. 4, Ventura, CA: Western Periodicals, 235–48.

Gärling, T. (1992) Determinants of everyday time allocation. *Scandinavian Journal of Psychology* 33: p. 160–169.

Giele, J.Z. and Elder, G.H., Jr. (1998). Life Course Research: Development of a Field, in: J.Z. Giele, and G.H. Elder Jr., *Methods of Life Course Research: Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage, p. 5 – 27.

Golob, T.F. (1980). Transit Riders' Attitudes to Bus and Train. unpublished research documents.

Hagestad, G.O., and Neugarten, B.L. (1985). Age and the life course, in: R.H. Binstock and E. Shanas (eds.), *Handbook of aging and the social sciences*, New York: Van Nostrand Reinhold, p. 35 - 61.

Han, Q. and Timmermans, H.J.P. (2006). Interactive learning in transportation networks with uncertainty, bounded rationality, and strategic choice behaviour: quantal response model. *Transportation Research Record*, *1964* (1), 27-34.

Hareven, T.K. (1977). Family Time and Historical Time. *Daedalus*, 106 (2), p. 57 – 70.

Heckerman, D., Geiger, D. and Chickering, D.M. (1995). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data*, Machine* Learning, volume 20, issue 3, pp. 197-243.

Herskovits, E. and Cooper, G.F. (1990). Kutato: An entropy-driven system for the construction of probabilistic expert systems from databases. Proceedings of the sixth International conference on Uncertainty in Artificial Intelligence (pp. 54-62). Cambridge, MA.

Hollingworth, B.J. and Miller, E.J. (1996), Retrospective interviewing and its application in study of residential mobility, *Transportation Research Record*, 1551, 74-81.

Hugin Expert A/S (1995-2005), Software program, Aalborg, Denmark

Huigen, P.P.P. (1986). Binnen of buiten bereik?: een sociaal-geografisch onderzoek in Zuidwest-Friesland. Amsterdam/Utrecht: KNAG/Universiteit van Utrecht, Nederlandse Geografische Studies 7.

Joh, C.H., Arentze, T.A. and Timmermans, H.J.P. (2001). Pattern recognition in complex activity-travel patterns: A comparison of Euclidean distance, signal processing theoretical, and multidimensional sequence alignment methods, *Transportation Research Record*, 1752, 16-22

Jones, P.M., Dix, M.C., Clarke, M.I. and Heggie, I.G. (1983). *Understanding travel behaviour*. Aldershot: Gower.

Kitamura, R. and Fujii, S.R. (1998). Two computational process models of activity-travel choice. In: T. Gärling, T. Laitila and K. Westin (eds.): *Theoretical Foundations of Travel Choice Modeling*, Elsevier, Oxford, p. 251 – 279.

Klöckner, C. (2004). How single events change travel mode choice a life span perspective. Paper presented at the *3rd International Conference of Traffic and Transport Psychology*, 5-9 September, Nottingham, UK.

Kruskal, J.B. (1983). An overview of sequence comparison. In: Sankoff, D., Kruskal, J.B. (Eds.), Time Warps, StringEdits, and Macromolecules: The Theory and Practice of Sequence Comparison. Addison-Wesley, London, pp. 1–44.

Krygsman, S., Arentze, T.A. and Timmermans, H.J.P. (2007). Capturing tour mode and activity choice interdependencies: A co-evolutionary logit modelling approach. *Transportation Research. Part A: Policy and Practice*. vol. 41. no. 10. p. 913-933.

Lam, W. and Bacchus, F. (1994). Learning Bayesian belief networks: An approach based on the MDL principle, Computational Intelligence, Volume 10 issue 4.

Lanzendorf, M. (2003). Mobility biographies. A new perspective for understanding travel behaviour. Paper presented at *the 10th International Conference on Travel Behaviour Research*, Luzern, Augustus.

Lanzendorf, M. (2006). Key events and their effect on mobility biographies. The case of child birth. *Proceedings 11th IATBR Conference* (CD-ROM), Kyoto, Japan.

Lauritzen, S.L. (1995). The EM algorithm for graphical association models with missing data, Computational Statistics & Data Analysis, 19: 191 – 201.

Lenntorp, B. (1976). Paths in space-time environment: A time geographic study of possibilities of individuals, the Royal University of Lund, Department of Geography, Lund Studies in Geography, Series B, *Human Geography* no. 44.

Louviere, J.J. (1988), *Analyzing decision making: Metric conjoint analysis*. Newbury Park, CA: Sage.

Madsen, A.L., Lang, M., Kjærulff, U.B. and Jensen, F. (2004). The Hugin Tool for Learning Bayesian Networks, *Computer Science*, Volume 2711, p. 594 - 605.

Maslow, A.H. (1954): *Motivation and personality*. New York: Harper & Row.

Mayer, K.U. (1986). Structural constraints on the life course, *Human Development*, 29(3), p. 163–170.

Mayer, K.U. and Tuma, N.B. (1990). *Event history analysis in life course research*. Madison, W.I.: University of Wisconsin Press.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, **672**, p. 72-77.

McNally, M.G. and Rindt, C.R. (2008). The Activity-based Approach, in: D.A. Hensher and Button, K.J. (eds.) *Handbook of Transport Modelling*, 2nd edition, Elsevier, Amsterdam, p. 55 – 73.

Mohammadian, A. and Miller, E.J. (2003). Dynamic modelling of Household automobile transactions. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1831, p. 98 – 105.

Miller, E.J., and Roorda, M.J. (2003). Prototype Model of Household Activity-Travel Scheduling. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 1831: p. 114 – 121.

Miller, K. (2005). *Communications theories: perspectives, processes, and contexts*. New York: McGraw-Hill.

Mulder, C.H. (1993). *Migration dynamics: A life course approach*. Dissertation. Amsterdam, Thesis Publishers.

Neapolitan, R.E. (1990). *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*, John Wiley & Sons, New York.

Netemeyer, R.G., Burton, S. and Johnston, M. (1991). A comparison of two models for the prediction of volitional and goal-directed behaviors: A conformatory analysis approach. *Social Psychology Quarterly* 54: p. 87–100.

Norsys Software Corp. (1997) Netica Application User's Guide Version 1.05, Norsys Software Corp., Vancouver, Canada.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*: network of plausible inference, Morgan Kaufmann.

Pendyala, R.M., Kitamura, R., and Kikuchi, A. (2004). FAMOS: the Florida Activity Mobility Simulator. Presented at the Conference on "Progress in Activity-Based Analysis", Maastricht, The Netherlands, May 28-31, 2004

Pendyala, R.M., Kitamura, R., Kikuchi, A., Yamamoto, T., and Fujii, S. (2005). Florida Activity Mobility Simulator: Overview and Preliminary Validation Results. *Transportation Research Record: Journal of the Transportation Research Board*, Volume 1921: p. 123-130.

Prillwitz, J. and Lanzendorf, M. (2006a). Impact of life course events on car ownership. *Proceedings 85th Transportation Research Board Annual Meeting* (CD-ROM), January, Washington, D.C.

Prillwitz, J. and Lanzendorf, M. (2006b). The importance of life course events for the daily travel behaviour: A panel analysis. Paper presented at the 11th *International Conference on Travel Behaviour* Research, Kyoto, August 2006.

Ramoni, M. and Sebastiani, P. (1997). Discovering Bayesian networks in incomplete databases, *Technical report KMI-TR-46*, Knowledge Media Institute, The Open Universite.

Salomon, I. (1983). Life styles – a broader perspective on travel behaviour. In: Carpenter, S., P. Jones (eds.), *Recent Advances in Travel Demand Analysis*. Gower, Aldershot, Hants, pp. 290 – 310.

Schifter, D.E. and Ajzen, I. (1985). Intention, perceived control, and weight loss: An application of the theory of planned behavior. *Journal of Personality and Social Psychology* 49: 843–851.

Schwarz, N. (1997). Questionnaire design: the rocky road from concepts to answers. In: *Survey Measurement and Process Quality*, L. Lyberg et al. (cds), Wiley, New York, pp. 29-45.

Shiftan, Y., Kaplan, S., and Hakkert, S. (2003). Scenarion building as a tool for planning a sustainable transportation system. Transportation Research Part D 8, p. 323-342

Silverfrost, Fortran (FTN95), © 1999-2009 Silverfrost Limited

Simma, A. and Axhausen, K.W. (2003). Commitments and modal usage: Analysis of German and Dutch panels, *Transportation Research Record*, 1854, p. 22 - 31.

Sparmann, U. (1980). ORIENT: Ein Verhaltensorientiertes Simulstionsmodell zur Verkehrsprognose. Schriftenreihe des Institutes für Verkehrswesen der Universität Karlsruhe, Karlsruhe.

Spirtes, P., Glymour, C. and Scheines, R. (1990). Causality from probability, *proceedings of Advanced Computing for the Social Sciences*, Williamsburgh, VA.

Spirtes, P., Glymour, C. and Scheines, R. (1991). An algorithm for fast recovery of sparse causal graphs, *Social Science Computer Review*, volume 9, pp. 62-72.

Sprites, P., Glymour, C., Scheines, R., Heckerman, D., Meek, C. and Cooper, G.F. (2000). *Causation, Prediction and Search*, MIT Press, 2nd edition.

Srinivas, S., Russell, S. and Agogino, A. (1990). Automated construction of sparse Bayesian networks from unstructured probabilistic models and domain information, in Henrion, M., Shachter, R.D., Kanal, L.N. and Lemmer, J.F. (Eds.), *Uncertainty in artificial intelligence 5*, Amsterdam: North-Holland.

States, D.J., and Boguski, M.S. (1991). Similarity and homology. In: Gribskov, M., Devereux, J. (Eds.), Sequence Analysis Primer. Stockton Press, New York, pp. 89–157.

Stanbridge, K., Lyons, G. and Farthing, S. (2004). Travel behaviour change and residential relocation. Paper presented at *the 3rd International Conference of Traffic and Transport Psychology*, 5 9 September, Nottingham, UK.

Steck, H. and Tresp, V. (1999). Bayesian belief networks for data mining, Proceedings of The 2nd Workshop on Data Mining und Data Warehousing als Grundlage moderner enscheidungsunterstuetzender Systeme, DWDW99, Sammelband, Universität Magdeburg.

Strack, F. and Martin, L. (1987). Thinking judging and communicating: a process account of context effects in attitude survey. In: *Social Information Processing and Survey Methodology*, H. Hippler et al. (cds), Springer verslag, New York, pp. 123-148.

Suzuki, J. (1996). Learning Baysesian belief networks based on the MDL principle: An efficient algorithm using branch and bound technique, *Proceedings of the international conference on machine learning*, Bari, Italy.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.

Tourangeau, R. (1984). Cognitive science and survey methods: a cognitive perspective. In: *Cognitive Aspects of Survey Methodology: Building a Bridge between Disciplines*, T. Jabine et al. (cds), National Academic Press, Washington DC, pp. 73-100.

Van der Vaart, W. (1996). Inquiring into the past: Data Quality of Response tot Retrospective Questions. Dissertation. Vrije Universiteit, Amsterdam.

Veldhuisen, K.K., Kapoen, L.L. and Timmermans, H.J.P. (2000). RAMBLAS: A regional planning model based on the microsimulation of daily activity travel patterns. Environment and Planning A, volume **32**, p. 427 - 443

Verhoeven, M., Arentze, T.A., Timmermans, H.J.P. and van der Waerden, P.J.H.J. (2005). Modelling the Impact of Key Events on Long-Term Transport Mode Choice Decisions: A Decision Network Approach Using Event History Data. In *Transportation Research Record*: Journal of the Transportation Research Board, No. 1926, TRB, National Research Council, Washington, D.C., 2005, pp. 106-114.

Verhoeven, M., Arentze, T.A., Timmermans, H.J.P. and van der Waerden, P.J.H.J. (2006). Modelling the Influence of Structural Lifecycle Events on Transport Mode Choice Decisions Using a Structure Learning Algorithm. *Proceedings 11th IATBR Conference* (CD-ROM), Kyoto, Japan.

Verhoeven, M., Arentze, T.A., Timmermans, H.J.P. and van der Waerden, P.J.H.J. (2007). Examining temporal effects of lifecycle events on transport mode choice decision. The International Journal of Urban Sciences. vol. 11. 1-14.

Vovsha, P., Bradley, M., Bowman, J.L. (2005). Activity-based travel forecasting models in the United States: progress since 1995 and prospects for the future. In: Timmermans, H.J.P. (eds.) *Progress in Activity-Based Analysis*. Elsevier, Oxford England.

Van der Waerden, P.J.H.J., and Timmermans, H.J.P. (2003). Key events and critical incidents influencing transport mode choice switching behaviour: an exploratory study. *Proceedings 82nd Annual Meeting of the Transportation Research Board)*, January 12-16, Washington, D.C., (CD-ROM: 13 pp.

Van der Waerden, P.J.H.J., Borgers, A.W.J. and Timmermans, H.J.P. (2003). The influence of key events and critical incidents on transport mode choice switching behaviour: a descriptive analysis. *Proceedings of the IATBR Conference*, Lucerne, August 10-16 (CD-ROM: 24 pp.).

Wall, R., Devine-Wright, P. and Mill, G.A. (2007). Comparing and Combining Theories to Explain Proenvironmental Intentions: The Case of Commuting-Mode Choice, *Environment and Behavior*, 39: p. 731 - 753.

Wallace, C., Korb, K.B. and Dai, H. (1996). Causal discovery via MML, *Proceedings of the thirteenth international conference on machine learning (ICML'96)*, Morgan Kaufmann Publishers, San Francisco CA USA, pp. 516-524.

Wagner, P. and Nagel, K. (1999). Microscopic modelling of travel demand: The home-to-work-problem. Proceedings 78th Annual Conference of the Transportation Research Board, Washington, D.C., January 10 – 14.

Wang, D. and Timmermans, H.J.P. (2000). A conjoint-based model of activity engagement, timing, scheduling and stop pattern formation. Paper presented at the 79th Annual Meeting of the Transportation Board, Washington, DC, January.

Wen, C. and Koppelman, F.S. (2000). A conceptual and methodological framework for the generation of activity-travel patterns. *Transportation* 27, p. 5–23.

Wermuth, N. and.Lauritzen, S.L. (1983). Graphical and recursive models for contingency tables, *Biometrika Trust*, Vol. 70, No. 3, pp. 537-552

Willekens, F. (1987). Migration and development: a micro-perspective. *Journal of Institute of Economic Research*, 22, 2.

Willekens, F. (1991). Understanding the interdependence between parallel careers, in: J.J. Siegers, de J. De Jong-Gierveld and E. van Imhoff (eds.), *Female labour market behaviour and fertility: A rational-choice approach*. Berlin, Springer, p. 2 – 31.

Yagi, S. and Mohammedian, A. (2007). Policy Simulation for New BRT and Area Pricing Alternatives Using an Opinion Survey in Jakarta. *Proceedings 86th Transportation Research Board Annual Meeting*, January, Washington, D.C.

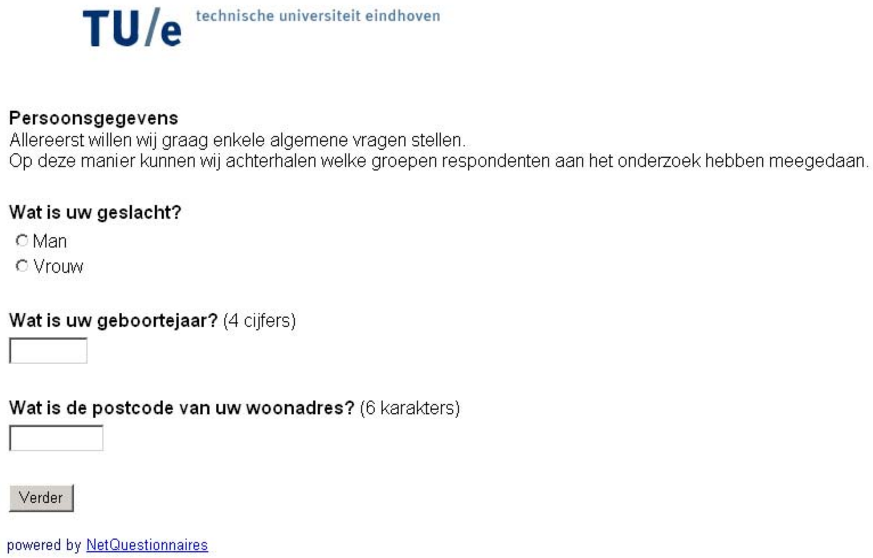
Zimmerman, C.A. (1982). The life cycle concept as a tool for travel research. *Transportation* 11, p.51-69

Appendix

This appendix consists of four parts. The first part illustrates the Internet-based survey which took place in 2004. All questions were in Dutch. First, the print screens of the survey are given. They give insight in the structure of the survey and questions asked (Appendix 1A). Next, a list with variables and classes is given (in English) for a more detailed insight into the collected data (Appendix 1B). The second part illustrates the classification of the data collection into life course events. This means that the type of occurrence defined in the Internet-based survey are combined into new classes. For example, in the Internet-based survey there was a distinction between moving to a student room for the first time or for a second (or more) time. Both answer categories are combined into the occurrence student housing. All these translations are listed in Appendix 2. The routing of the Internet-based survey, as explained in chapter six, is illustrated in Appendix 3. In total there were eigth possible routings in the Internet-based survey. One routing is explained in chapter six, all routing . are illustrated in Appendix 4.

Appendix 1A | Questions Internet-based survey

<u>Household and personal characteristics:</u> gender, year of birth, zipcode, education, living conditions, marital state, number of children, household members, hours of paid work, household income.


TU/e technische universiteit eindhoven

Persoonsgegevens
Allereerst willen wij graag enkele algemene vragen stellen.
Op deze manier kunnen wij achterhalen welke groepen respondenten aan het onderzoek hebben meegedaan.

Wat is uw geslacht?
- Man
- Vrouw

Wat is uw geboortejaar? (4 cijfers)

Wat is de postcode van uw woonadres? (6 karakters)

Verder

powered by NetQuestionnaires


TU/e technische universiteit eindhoven

Wat is uw hoogst genoten afgeronde opleiding?
- wetenschappelijk onderwijs (doctoraal, universiteit)
- hoger beroepsonderwijs en wetenschappelijk onderwijs-kandidaats (HBO, PABO, HTS, HEAO)
- hoger algemeen en voorbereidend wetenschappelijk onderwijs (HAVO, VWO, MMS, HBS)
- middelbaar beroepsonderwijs (MBO, MTS, MDS)
- middelbaar algemeen voortgezet onderwijs (MAVO, (M)ULO)
- lager beroepsonderwijs (LBO, LTS, Huishoudschool)
- lager onderwijs
- anders, namelijk

Hoe zou U uw woonsituatie het beste kunnen omschrijven?
- zelfstandig wonen
- op kamers wonen
- bij ouders wonen

Terug Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

Wat is op dit moment uw burgerlijke staat?

- ○ alleenstaand
- ○ samenwonend
- ○ gehuwd
- ○ gescheiden
- ○ weduwe / weduwnaar
- ○ anders, namelijk ______

Hoeveel kinderen heeft U?

- ○ geen
- ○ één kind
- ○ twee kinderen
- ○ meer dan twee kinderen, namelijk: ______

Uit hoeveel personen bestaat het huishouden waarvan U deel uitmaakt?
(Uzelf, eventueel partner en INWONENDE kinderen OF ouders en broers/zussen)

______

Terug | Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

Hoeveel uren betaald werk verricht U?

- ○ geen betaald werk
- ○ < 11 uur per week
- ○ 12 - 19 uur per week
- ○ 20 - 34 uur per week
- ○ > 35 uur per week

Wat is ongeveer het totale inkomen van uw huishouden?

- ○ beneden 1 keer modaal
- ○ rond 1 keer modaal (± 25.000 euro bruto/jaar)
- ○ tussen 1 en 2 keer modaal
- ○ rond 2 keer modaal (± 50.000 euro bruto/jaar)
- ○ boven 2 keer modaal
- ○ weet ik niet

Terug | Verstuur

powered by NetQuestionnaires

Availability of transport mode: possession of driver's licence (and year), numbers of cars in the household, availability car, sharing car with others, possession of a bike, possession of a public transport pass;

technische universiteit eindhoven

Gegevens vervoermiddel bezit
Hierna volgen enkele vagen over uw huidige vervoermiddel bezit.

Bent U in het bezit van een autorijbewijs?

- ja
- nee

Verder

powered by NetQuestionnaires

technische universiteit eindhoven

In welk jaar heeft U uw autorijbewijs behaald?

Beschikt uw huishouden over minimaal één auto?

- ja
- nee

Terug Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

Beschikt U regelmatig over een auto?

- ja, altijd
- ja, in overleg
- nee

Met hoeveel andere personen in uw huishouden deelt U deze auto (exclusief uzelf)?

Hoeveel auto's heeft uw huishouden in het totaal?

Terug Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

Bent U in het bezit van een fiets?

- ja
- nee

Welke Openbaar Vervoerkaart bezit U?

- GEEN
- OV studentenkaart
- OV jaarkaart
- NS jaarkaart
- Jaartrajectkaart
- Maandtrajectkaart
- Maandnetkaart
- Voordeeluren kaart
- Stad- of Streekabonnement
- anders, namelijk

Terug Verstuur

powered by NetQuestionnaires

<u>Occurrence of lifecycle events:</u> (a) change in residential location, (b) change in household composition, (c) change in work location, (d) change in study location, (e) change in car possession and availability, (f) change in availability of public transport pass, and (g) change in household income.

Respondents indicated whether they experienced these events, and, if so, they indicated in a matrix the timing of the event (month and year), the cause of the change (i.e., the specific type of event) that took place and the before and after situation for every change to a maximum of ten changes.

TU/e technische universiteit eindhoven

Gegevens over gebeurtenissen
Hierna volgen enkele vragen over 7 verschillende gebeurtenissen.

Deze gebeurtenissen zijn:
1) verandering van **woonadres**;
2) verandering van **huishoudsamenstelling** *(alleen als U zelfstandig woont)*;
3) verandering van **werklocatie**;
4) verandering van **studielocatie** *(indien U geboren bent na 1969)*;
5) verandering van **autobeschikbaarheid** *(indien U in het bezit bent van een autorijbewijs)*;
6) verandering van het bezit van **openbaar vervoerkaart**;
7) verandering van **huishoud-inkomen** *(alleen als U zelfstandig/op kamers woont)*.

Per gebeurtenis dient U een tabel in te vullen. In deze tabel worden voor elke verandering naar een aantal details gevraagd.
Bij elke gebeurtenis staat uiteraard nog de benodigde uitleg.

Deze gegevens zijn van groot belang voor het onderzoek.
Met behulp van deze gegevens proberen wij de invloed van gebeurtenissen op de vervoermiddelkeuze te achterhalen.

Suggestie: gebruik eventueel pen en papier als geheugensteuntje.

Verder

powered by NetQuestionnaires

technische universiteit eindhoven

Gegevens over gebeurtenissen
Hierna volgen enkele vragen over 7 verschillende gebeurtenissen.

Deze gebeurtenissen zijn:
1) verandering van **woonadres**;
2) verandering van **huishoudsamenstelling** *(alleen als U zelfstandig woont)*;
3) verandering van **werklocatie**;
4) verandering van **studielocatie** *(indien U geboren bent na 1969)*;
5) verandering van **autobeschikbaarheid** *(indien U in het bezit bent van een autorijbewijs)*;
6) verandering van het bezit van **openbaar vervoerkaart**;
7) verandering van **huishoud-inkomen** *(alleen als U zelfstandig/op kamers woont)*.

Per gebeurtenis dient U een tabel in te vullen. In deze tabel worden voor elke verandering naar een aantal details gevraagd. Bij elke gebeurtenis staat uiteraard nog de benodigde uitleg.

Deze gegevens zijn van groot belang voor het onderzoek.
Met behulp van deze gegevens proberen wij de invloed van gebeurtenissen op de vervoermiddelkeuze te achterhalen.

Suggestie: gebruik eventueel pen en papier als geheugensteuntje.

Verder

powered by NetQuestionnaires

technische universiteit eindhoven

VERANDERING WOONADRES

Hoe vaak bent U van woonadres veranderd sinds U het ouderlijk huis heeft verlaten?
Onder verandering van woonadres worden de volgende veranderingen verstaan:
- op kamers gaan wonen (een onzelfstandige woning betrekken in de stad waar U studeert)
- verhuizen naar een andere studentenkamer (niet binnen 1 studentenhuis)
- zelfstandig gaan wonen (een zelfstandige woning betrekken)
- gaan samenwonen (bij uw partner intrekken of samen op een nieuw adres gaan wonen)
- een (andere) woning huren
- een (andere) woning kopen
- weer bij ouders intrekken

Geef nu aan of U dit nooit of een (aantal) keer heeft meegemaakt.

○ nooit
○ een (aantal) keer

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

VERANDERING WOONADRES

Hoe vaak bent U van woonadres veranderd sinds uw 16e jaar?

Onder verandering van woonadres worden de volgende veranderingen verstaan:

- op kamers gaan wonen (een onzelfstandige woning betrekken in de stad waar U studeert)
- verhuizen naar een andere studentenkamer (niet binnen 1 studentenhuis)
- zelfstandig gaan wonen (een zelfstandige woning betrekken)
- gaan samenwonen (bij uw partner intrekken of samen op een nieuw adres gaan wonen)
- een (andere) woning huren
- een (andere) woning kopen
- weer bij ouders intrekken

Geef nu aan of U dit nooit of een (aantal) keer heeft meegemaakt.

- nooit
- een (aantal) keer

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

VERANDERING WOONADRES

- op kamers gaan wonen
- verhuizen naar een andere studentenkamer (niet binnen 1 studentenhuis)
- zelfstandig gaan wonen
- gaan samenwonen
- een (andere) woning huren
- een (andere) woning kopen
- weer bij ouders intrekken

Maand geef hier aan wanneer dit heeft plaatsgevonden, eventueel kunt U 'geen idee' kiezen.
Jaar geef hier aan wanneer dit heeft plaatsgevonden.
Van vul hier de straat & woonplaats in van de oude locatie.
Naar vul hier de straat & woonplaats in van de nieuwe locatie.
Oorzaak maak hier een keuze uit de mogelijkheden.
Woningtype geef hier aan om wat voor type woning het gaat op de nieuwe locatie.
Huur/koop geef hier aan of uw nieuwe woning een huur- of koopwoning is.

Vul hieronder voor elke verandering van woonadres de gegevens in.

	maand	jaar	van	naar	oorzaak	woningtype	huur/koop
meeste recent							
keer daarvoor							
keer daarvoor							
keer daarvoor							
keer daarvoor							
keer daarvoor							
keer daarvoor							
keer daarvoor							
keer daarvoor							
keer daarvoor							

op kamers gaan
andere studentenkamer
zelfstandig gaan wonen
samenwonen
woning huren
woning kopen
bij ouders intrekken
anders

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

VERANDERING HUISHOUDSAMENSTELLING

Hoe vaak is uw huishoudsamenstelling veranderd sinds U zelfstandig woont?
(toe/af name van het aantal gezinsleden)
Onder verandering van huishoudsamenstelling worden de volgende veranderingen verstaan:

- geboorte/adoptie kind
- verlaten van het ouderlijk huis (door uzelf of één van uw kinderen)
- samenwonen
- uit elkaar gaan
- trouwen
- scheiden
- overlijden gezinslid
- er komt een persoon inwonen
- die persoon verlaat het huis weer

Geef nu aan of U dit nooit of een (aantal) keer heeft meegemaakt.

- ○ nooit
- ○ een (aantal) keer

Terug Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

VERANDERING HUISHOUDSAMENSTELLING

- geboorte/adoptie kind
- verlaten van het ouderlijk huis (door uzelf of één van uw kinderen)
- samenwonen
- uit elkaar gaan
- trouwen
- scheiden
- overlijden gezinslid
- er komt een persoon inwonen
- die persoon verlaat het huis weer

Maand geef hier aan wanneer dit heeft plaatsgevonden, eventueel kunt U 'geen idee' kiezen.
Jaar geef hier aan wanneer dit heeft plaatsgevonden.
Van vul hier het oude aantal personen in.
Naar vul hier het nieuwe aantal in.
Oorzaak maak hier een keuze uit de mogelijkheden.

Vul hieronder voor elke verandering van huishoudsamenstelling de gegevens in.

	maand	jaar	van	naar	oorzaak
meest recent					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					

geboorte/adoptie
verlaten van het ouderlijk huis
samenwonen
uit elkaar gaan
trouwen
scheiden
overlijden gezinslid
inwonen persoon
persoon verlaat het huis
anders

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

VERANDERING WERKLOCATIE

Hoe vaak bent U van werklocatie veranderd nadat U school of studie heeft verlaten?
Bijbaantjes dienen buiten beschouwing gelaten te worden.

Onder verandering van werklocatie worden de volgende veranderingen verstaan:
- starten met eerste baan
- veranderen van werkgever (andere baan)
- overplaatsing naar andere locatie
- tijdelijk zonder baan zitten
- met vut of pensioen gaan

Geef nu aan of U dit nooit of een (aantal) keer heeft meegemaakt.

- nooit
- een (aantal) keer

Terug Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

VERANDERING WERKLOCATIE

- starten met eerste baan
- veranderen van werkgever (andere baan)
- overplaatsing naar andere locatie
- tijdelijk zonder baan zitten
- met vut of pensioen gaan

Maand geef hier aan wanneer dit heeft plaatsgevonden, eventueel kunt U 'geen idee' kiezen.
Jaar geef hier aan wanneer dit heeft plaatsgevonden.
Van vul hier de straat & woonplaats in van de oude werklocatie.
En bij **naar** van de nieuwe werklocatie.
Oorzaak maak hier een keuze uit de mogelijkheden.

Vul hieronder voor elke verandering van werklocatie de gegevens in.

	maand	jaar	van	naar	oorzaak
meest recent					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					

eerste baan
andere baan
overplaatsing
tijdelijk geen baan
vut of pensioen
anders

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

VERANDERING OPLEIDINGSLOCATIE

Hoe vaak bent U van opleidingslocatie veranderd sinds uw 16e jaar?
Onder verandering van opleidingslocatie worden de volgende veranderingen verstaan:
- verandering van opleidingsinstelling (bijvoorbeeld vanwege verhuizing)
- een nieuwe opleiding beginnen
- een opleiding beeindigen (zonder diploma)
- een opleiding afronden (met diploma)
- cursus of certificaat beginnen (minimale duur 6 maanden)
- cursus of certificaat afronden

Geef nu aan of U dit nooit of een (aantal) keer heeft meegemaakt.

- nooit
- een (aantal) keer

Terug | Verder

powered by NetQuestionnaires

technische universiteit eindhoven

VERANDERING OPLEIDINGSLOCATIE

- verandering van opleidingsinstelling (bijvoorbeeld vanwege verhuizing)
- een nieuwe opleiding beginnen
- een opleiding beeindigen (zonder diploma)
- een opleiding afronden (met diploma)
- cursus of certificaat beginnen (minimale duur 6 maanden)
- cursus of certificaat afronden

Maand geef hier aan wanneer dit heeft plaatsgevonden, eventueel kunt U 'geen idee' kiezen.
Jaar geef hier aan wanneer dit heeft plaatsgevonden.
Van vul hier de naam opleidingsinstituut & woonplaats in van de oude opleiding in.
En bij **naar** van de nieuwe opleiding.
Oorzaak maak hier een keuze uit de mogelijkheden.

Vul hieronder voor elke verandering van opleidingslocatie de gegevens in:

	maand	jaar	van	naar	oorzaak
meest recent					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					

veranderen van opleidingsinstelling
nieuwe opleiding beginnen
opleiding beeindigen
opleiding afronden
cursus of certificaat beginnen
cursus of certificaat afronden
anders

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

VERANDERING AUTO BESCHIKBAARHEID

Hoe vaak is de beschikking over een auto veranderd sinds U zelfstandig woont?
Hieronder worden de volgende veranderingen verstaan:
- meer autogebruikers per huishouden
- minder autogebruiker per huishouden
- meer auto's per huishouden
- minder auto's per huishouden

Geef nu aan of U dit nooit of een (aantal) keer heeft meegemaakt.

- nooit
- een (aantal) keer

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

VERANDERING AUTO BESCHIKBAARHEID

Hoe vaak is de beschikking over een auto veranderd sinds U het ouderlijk huis heeft verlaten?
Hieronder worden de volgende veranderingen verstaan:
- meer autogebruikers per huishouden
- minder autogebruiker per huishouden
- meer auto's per huishouden
- minder auto's per huishouden

Geef nu aan of U dit nooit of een (aantal) keer heeft meegemaakt.

- nooit
- een (aantal) keer

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

VERANDERING AUTO BESCHIKBAARHEID

Hoe vaak is de beschikking over een auto veranderd sinds U uw rijbewijs heeft behaald?
Hieronder worden de volgende veranderingen verstaan:

- meer autogebruikers per huishouden
- minder autogebruiker per huishouden
- meer auto's per huishouden
- minder auto's per huishouden

Geef nu aan of U dit nooit of een (aantal) keer heeft meegemaakt.

- ○ nooit
- ○ een (aantal) keer

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

VERANDERING BESCHIKBAARHEID AUTO

- meer autogebruikers per huishouden
- minder autogebruiker per huishouden
- meer auto's per huishouden
- minder auto's per huishouden

Maand geef hier aan wanneer dit heeft plaatsgevonden, eventueel kunt U 'geen idee' kiezen.
Jaar geef hier aan wanneer dit heeft plaatsgevonden.
Oorzaak maak een keuze uit de mogelijkheden.
Van maak hier een keuze uit de mogelijkheden voor de oude situatie.
En bij **naar** voor de nieuwe situatie.

Vul hieronder voor elke verandering van de beschikbaarheid van een auto de gegevens in:

	maand	jaar	oorzaak	van	naar
meest recent					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					

meer gebruikers
minder gebruikers
meer auto's
minder auto's

geen auto
1 auto, 1 gebruiker
1 auto, > 1 gebruiker
2 auto's, 1 gebruiker
2 auto's, 2 gebruikers
2 auto's, > 2 gebruikers

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

VERANDERING OV KAART BEZIT

Hoe vaak is uw openbaar vervoerkaart bezit veranderd sinds uw 16e jaar?
Onder verandering in openbaar vervoerkaart bezit worden de volgende veranderingen verstaan:

- starten met aanschaf OV kaart
- stoppen met aanschaf OV kaart

Geef nu aan of U dit nooit of een (aantal) keer heeft meegemaakt.

- nooit
- een (aantal) keer

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

VERANDERING OV-KAART BEZIT

- starten met aanschaf OV kaart
- stoppen met aanschaf OV kaart

Maand geef hier aan wanneer dit heeft plaatsgevonden, eventueel kunt U 'geen idee' kiezen.
Jaar geef hier aan wanneer dit heeft plaatsgevonden.
Oorzaak maak een keuze uit de mogelijkheden.
Van geef hier aan om welke kaart het ging.
Naar geef hier aan om welke kaart het ging.

Vul hieronder voor elke verandering van openbaar vervoerkaart bezit de gegevens in:

	maand	jaar	oorzaak	van	naar
meest recent					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					

oorzaak:
starten met aanschaf OV kaart
stoppen met aanschaf OV kaart

van:
GEEN
OV studentenkaart
OV jaarkaart
NS jaarkaart
Jaartrajectkaart
Maandtrajectkaart
Maandnetkaart
Voordeeluren kaart
Stad- of Streekabonnement
anders

Terug Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

VERANDERING HUISHOUD-INKOMEN

Hoe vaak is het huishoud-inkomen noemenswaardig veranderd sinds U zelfstandig woont?
Onder verandering in huishoud-inkomen worden de volgende veranderingen verstaan:
- verdiener erbij
- verdiener eraf
- salaris sprong
- anders

Geef nu aan of U dit nooit of een (aantal) keer heeft meegemaakt.

○ nooit
○ een (aantal) keer

Terug Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

VERANDERING HUISHOUD-INKOMEN

Hoe vaak is het huishoud-inkomen noemenswaardig veranderd sinds U het ouderlijk huis heeft verlaten?
Onder verandering in huishoud-inkomen worden de volgende veranderingen verstaan:
- verdiener erbij
- verdiener eraf
- salaris sprong
- anders

Geef nu aan of U dit nooit of een (aantal) keer heeft meegemaakt.

○ nooit
○ een (aantal) keer

Terug Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

VERANDERING HUISHOUD-INKOMEN

- verdiener erbij
- verdiener eraf
- salaris sprong
- anders

Maand geef hier aan wanneer dit heeft plaatsgevonden, eventueel kunt U 'geen idee' kiezen.
Jaar geef hier aan wanneer dit heeft plaatsgevonden.
Van maak hier een keuze voor de voor-situatie.
En bij **naar** voor de na-situatie in.
Oorzaak maak hier een keuze uit de mogelijkheden.

Let op: het kan zijn dat de voor en na situatie gelijk zijn.

Vul hieronder voor elke verandering van huishoud-inkomen de gegevens in:

	maand	jaar	van	naar	oorzaak
meest recent					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					
keer daarvoor					

oorzaak:
verdiener erbij
verdiener eraf
salaris sprong
anders

van:
onder 1 keer modaal
rond 1 keer modaal
(25.000 euro bruto/jaar)
tussen 1 en 2 keer modaal
rond 2 keer modaal
(50.000 euro bruto/jaar)
boven 2 keer modaal

Terug Verstuur

powered by NetQuestionnaires

Current travel behaviour per trip purpose: (a) work, (b) study, (c) grocery, (d) shopping and (e) sport.

The respondent answered for each trip purpose questions in a matrix about trip frequency, mode choice, alternative mode choice, destination, travel distance, travel time and they also indicated the start time for the trip from home to their destination and for the trip back to home.


TU/e technische universiteit eindhoven

Gegevens over uw huidige vervoermiddel gebruik
Nu volgen enkele vragen over uw huidige gebruik van vervoermiddelen.
De vragen worden weergegeven in de vorm van een tabel.
U kunt het beste de tabel rij voor rij (dus per doel) invullen.

Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

Wilt U per doel (werk, studie, boodschappen doen, winkelen, sport) voor de meeste voorkomende verplaatsing de volgende gegevens willen invullen:

Frequentie geef hier aan hoe vaak U deze verplaatsing gemiddeld maakt, dus aantal keer per dag/week/maand /jaar (0 keer kan ook).
(hoofd)Vervoermiddel kies hier het vervoermiddel dat U meestal gebruikt voor deze verplaatsing.
(alternatief)Vervoermiddel kies hier het vervoermiddel dat U ook wel eens gebruikt voor deze verplaatsing.
(eind)Bestemming vul hier de straat & woonplaats in van uw meest voorkomende doelbestemming.

Let op: Alle velden dienen ingevuld te worden!

	frequentie	per	(hoofd)vervoermiddel	(alternatief)vervoermiddel	(eind)bestemming
werk					
studie					
boodschappen					
winkelen					
sport					

per: dag, week, maand, jaar

(alternatief)vervoermiddel: n.v.t., geen alternatief, auto als bestuurder, auto als passagier, fiets, lopend, trein, bus, anders

Wilt U per doel (werk, studie, boodschappen doen, winkelen, sport) voor de meeste voorkomende verplaatsing de volgende gegevens willen invullen:

Heenreis geef hier aan wanneer U deze verplaatsing meestal maakt.
Terugreis geef hier aan wanneer U deze verplaatsing meestal maakt.
Afstand vul hier het aantal km in van uw woonadres tot de (eind)bestemming.
Reistijd vul hier het aantal minuten in dat U erover doet om met het (hoofd)vervoermiddel van uw woonadres op uw (eind)bestemming te komen.

Let op: Alle velden dienen ingevuld te worden!

	heenreis	terugreis	afstand	reistijd
werk				
studie				
boodschappen				
winkelen				
sport				

heenreis: n.v.t., voor 9 uur, tussen 9-12 uur, tussen 12-16 uur, tussen 16-19 uur, na 19 uur

Terug Verstuur

powered by NetQuestionnaires

Perception of trip conditions: (a) comfort, (b) safety, (c) privacy, (d) environmental damage, (e) expenses, and (f) time.

The respondents scored the different trip conditions with a score of 0 – 100 (unfavourable – favourable) depending on different situations, like mode choice or a combination of mode choice and weather conditions.

TU/e technische universiteit eindhoven

Hoe beoordeelt U de mate van veiligheid in elk van de volgende situaties?
Onder veiligheid wordt verstaan risico's op ongelukken, openbare veiligheid enzovoorts.

0 wil zeggen zeer ongunstig (dus zeer onveilig).
100 wil zeggen zeer gunstig (dus zeer veilig).

U mag per situatie elk getal tussen de 0 en de 100 invullen.

	beoordeling
situatie 1: auto als bestuurder	
situatie 2: auto als passagier	
situatie 3: fiets	
situatie 4: bus	
situatie 5: trein	

Terug Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

Hoe beoordeelt U de mate van privacy in elk van de volgende situaties?
Onder privacy wordt verstaan privé of openbaar gebruik van het vervoermiddel.

0 wil zeggen zeer ongunstig (dus zeer weinig privacy).
100 wil zeggen zeer gunstig (dus zeer veel privacy).

U mag per situatie elk getal tussen de 0 en de 100 invullen.

	beoordeling
situatie 1: auto als bestuurder	
situatie 2: auto als passagier	
situatie 3: fiets	
situatie 4: bus	
situatie 5: trein	

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

Hoe beoordeelt U de mate van milieuschade in elk van de volgende situaties?
Onder milieuschade wordt verstaan uitlaatgassen, geluidsoverlast enzovoorts.

0 wil zeggen zeer ongunstig (dus zeer veel milieuschade).
100 wil zeggen zeer gunstig (dus zeer weinig milieuschade).

U mag per situatie elk getal tussen de 0 en de 100 invullen.

	beoordeling
situatie 1: auto als bestuurder	
situatie 2: auto als passagier	
situatie 3: fiets	
situatie 4: bus	
situatie 5: trein	

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

Hoe beoordeelt U de mate van comfort in elk van de volgende situaties?
Onder comfort wordt verstaan beschutting tegen weersinvloeden, zitcomfort enzovoorts.

0 wil zeggen zeer ongunstig (dus zeer oncomfortabel).
100 wil zeggen zeer gunstig (dus zeer comfortabel).

U mag per situatie elk getal tussen de 0 en de 100 invullen.

	beoordeling
situatie 1: auto	
situatie 2: fiets bij neerslag	
situatie 3: fiets bij droog weer en 10 graden	
situatie 4: fiets bij droog weer en 20 graden	
situatie 5: OV bij neerslag zonder overstap	
situatie 6: OV bij neerslag met één overstap	
situatie 7: OV bij neerslag met twee overstappen	
situatie 8: OV bij droog weer zonder overstap	
situatie 9: OV bij droog weer met één overstap	
situatie 10: OV bij droog weer met twee overstappen	

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

Hoe zeker bent U van uzelf als het gaat om het inschatten van de kosten voor de verschillende vervoermiddelen?

0 wil zeggen zeer ongunstig (dus zeer onzeker).
100 wil zeggen zeer gunstig (dus zeer zeker).

U mag per situatie elk getal tussen de 0 en de 100 invullen.

	beoordeling
situatie 1: auto	
situatie 2: bus	
situatie 3: trein	

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

Hoe zeker bent U van uzelf als het gaat om het inschatten van de reistijd voor de verschillende vervoermiddelen?

0 wil zeggen zeer ongunstig (dus zeer onzeker).
100 wil zeggen zeer gunstig (dus zeer zeker).

U mag per situatie elk getal tussen de 0 en de 100 invullen.

	beoordeling
situatie 1: auto	
situatie 2: fiets	
situatie 3: bus	
situatie 4: trein	

Terug Verstuur

powered by NetQuestionnaires

Stated preference part was used for evaluation of the selected trip conditions. The stated preference part of the survey used an orthogonal fraction of a full factorial design to generate profiles in terms of the benefit variables. Respondents were asked to indicate their preference for each profile on a 0-100 scale.

TU/e technische universiteit eindhoven

Beoordeling van scenario's

U krijgt nu 9 scenario's van mogelijke verplaatsingen voorgelegd.
Elk scenario is beschreven op basis van de eerder beschreven aspecten
(comfort, veiligheid, privacy, milieuschade, kosten en tijd).
Elk aspect kan nu alleen de **waarde 20, 50 of 80** aannemen op
de schaal van **zeer ongunstig** naar **zeer gunstig**.

Wij vragen U om deze scenario's te beoordelen met een **score** tussen de **0 en 100**.

Als eerste volgt een voorbeeld, dat U ook moet invullen.
Daarna volgen nog 9 andere scenario's van mogelijke verplaatsingen.

Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

Bij onderstaande verplaatsing is
- het comfort (80), **gunstig** (veel comfort)
- de veiligheid (20), **ongunstig** (weinig veiligheid)
- de milieuschade (80), **gunstig** (weinig milieuschade)
- de privacy (20), **ongunstig** (weinig privacy)
- de kosten (20), **ongunstig** (veel euro's)
- de tijd (20), **ongunstig** (veel minuten)

voorbeeld:

comfort	80
veiligheid	20
milieuschade	80
privacy	20
kosten	20
tijd	20

Geef met een score tussen 0 en 100 aan hoe U deze verplaatsing zou waarderen.
Waarbij 0 zeer ongunstig is, 100 zeer gunstig is.

SCORE:

Terug Verder

powered by NetQuestionnaires

Bij onderstaande verplaatsing is
- comfort (80), **gunstig** (veel comfort)
- veiligheid (80), **gunstig** (veel veiligheid)
- milieuschade (80), **gunstig** (weinig milieuschade)
- privacy (80), **gunstig** (veel privacy)
- kosten (80), **gunstig** (weinig euro's)
- tijd (80), **gunstig** (weinig minuten)

scenario 1:

comfort	80
veiligheid	80
milieuschade	80
privacy	80
kosten	80
tijd	80

Geef met een score tussen 0 en 100 aan hoe U deze verplaatsing zou waarderen.
Waarbij 0 zeer ongunstig is, 100 zeer gunstig is.

SCORE:

Terug Verder

powered by NetQuestionnaires

scenario 2:

comfort	80
veiligheid	80
milieuschade	80
privacy	80
kosten	20
tijd	80

Geef met een score tussen 0 en 100 aan hoe U deze verplaatsing zou waarderen.
Waarbij 0 zeer ongunstig is, 100 zeer gunstig is.

SCORE:

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

scenario 3:

comfort	20
veiligheid	20
milieuschade	20
privacy	80
kosten	80
tijd	80

Geef met een score tussen 0 en 100 aan hoe U deze verplaatsing zou waarderen.
Waarbij 0 zeer ongunstig is, 100 zeer gunstig is.

SCORE:

Terug Verder

powered by NetQuestionnaires

technische universiteit eindhoven

scenario 4:

comfort	20
veiligheid	20
milieuschade	80
privacy	80
kosten	20
tijd	20

Geef met een score tussen 0 en 100 aan hoe U deze verplaatsing zou waarderen.
Waarbij 0 zeer ongunstig is, 100 zeer gunstig is.

SCORE:

Terug Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

scenario 5:

comfort	20
veiligheid	80
milieuschade	20
privacy	20
kosten	80
tijd	20

Geef met een score tussen 0 en 100 aan hoe U deze verplaatsing zou waarderen.
Waarbij 0 zeer ongunstig is, 100 zeer gunstig is.

SCORE:

Terug Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

scenario 6:

comfort	20
veiligheid	80
milieuschade	80
privacy	20
kosten	20
tijd	80

Geef met een score tussen 0 en 100 aan hoe U deze verplaatsing zou waarderen.
Waarbij 0 zeer ongunstig is, 100 zeer gunstig is.

SCORE:

Terug Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

scenario 7:

comfort	80
veiligheid	20
milieuschade	20
privacy	20
kosten	20
tijd	80

Geef met een score tussen 0 en 100 aan hoe U deze verplaatsing zou waarderen.
Waarbij 0 zeer ongunstig is, 100 zeer gunstig is.

SCORE:

Terug Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

scenario 8:

comfort	80
veiligheid	20
milieuschade	80
privacy	20
kosten	80
tijd	20

Geef met een score tussen 0 en 100 aan hoe U deze verplaatsing zou waarderen.
Waarbij 0 zeer ongunstig is, 100 zeer gunstig is.

SCORE:

Terug Verder

powered by NetQuestionnaires

TU/e technische universiteit eindhoven

scenario 9:

comfort	80
veiligheid	80
milieuschade	20
privacy	80
kosten	20
tijd	20

Geef met een score tussen 0 en 100 aan hoe U deze verplaatsing zou waarderen.
Waarbij 0 zeer ongunstig is, 100 zeer gunstig is.

SCORE:

Terug Verstuur

powered by NetQuestionnaires

Appendix 1B | Variables Internet-based survey

For every part of the survey the variables and the classes are listed here. The predefined answers are between brackets. If there was no list with answers it is noted with [open question].

Part 1: Personal and household characteristics

Gender [Male; Female]

Year of Birth [open question]

Zip code [open question]

Education [in Dutch: wetenschappelijk onderwijs; hoger beroepsonderwijs; hoger algemeen en voorbereidend wetenschappelijk onderwijs; middelbaar beroepsonderwijs; middelbaar algemeen voortgezet onderwijs; lager beroepsonderwijs; basis onderwijs; other, like…]

Living situation [independent living ; student living ; parental living]

Marital state [alone; living together; married; divorced; widow/widower; different, like…]

Number of **children** [none; one child; two children; more than two children, …]

Number of persons in the **household** [open question]

Work hours [no paid job; < 11 hours a week; 12 – 19 hours a week; 20 – 34 hours a week; > 35 hours a week]

Household income [below 1 x modal income; around 1 x modal income; between 1 and 2 x modal income; around 2 x modal income; above 2 x modal income; I don’t know] Modal income was defined as +/- 25.000 euro bruto/year

Part 2: Possession and Availability Transport Modes

Driver’s licence [yes; no]

Number of years **driver’s licence** [open question]

One **car** in possession of the household [yes; no]

Car available [yes, always; yes, after some consideration; no]

Car sharing with .. others [open question]

Total **number of cars** in the household [open question]

Possession of a **bike** [yes; no]

Possession of **PT pass** [None; OV student pass; OV year pass; NS year pass; Yearly route pass, Monthly route pass; benefit hours pass; City or region pass; different, like…]

Part 3: Occurrence Lifecycle Events

1. Housing event

Experienced an occurrence [never; a few times]

Timing: since leaving parental house OR since age 16

Month [No idea; list with all months]

Year [choice of years]

Before situation [open question]

After situation [open question]

Type of change [first student room; another student room; living on your own; living together; rent a house; buy a house; live with parents; different]

Housing type [studentroom; flat; appartement; terrace house; semi-detached house; detached house; place for elderly people; different]

Payment [rented; bought]

2. Household event

Experienced an occurrence [never; a few times]

Timing: since living independent

Month [No idea; list with all months]

Year [choice of years]

Before situation [open question]

After situation [open question]

Type of change [birth/adoption; leaving parental house; living together; breaking up; getting married; getting divorced; passing of family member; person moving in; person leaving; different]

3. Work event

Experienced an occurrence [never; a few times]

Timing: since leaving school

Month [No idea; list with all months]

Year [choice of years]

Before situation [open question]

After situation [open question]

Type of change [first job; another job; job transfer; temporarily no job; retirement; different]

4. Study event

Experienced an occurrence [never; a few times]

Timing: since age 16

Month [No idea; list with all months]

Year [choice of years]

Before situation [open question]

After situation [open question]

Type of change [change school location; start new study; quit study; graduation; start course/certificate; start course/certificate; different]

5. Car availability event

Experienced an occurrence [never; a few times]

Timing: since living independent OR leaving parental house OR driver's licence

Month [No idea; list with all months]

Year [choice of years]

Before situation [no car; one car, one car user; one car, > one car user; two cars, one car user; two cars, two car users; two cars, > two car users]

After situation [no car; one car, one car user; one car, > one car user; two cars, one car user; two cars, two car users; two cars, > two car users]

Type of change [more car users; less car users; more cars; less cars]

6. PT pass event

Experienced an occurrence [never; a few times]

Timing: since age 16

Month [No idea; list with all months]

Year [choice of years]

Before situation [None; OV student pass; OV year pass; NS year pass; Yearly route pass, Monthly route pass; benefit hours pass; City or region pass; different]

After situation [None; OV student pass; OV year pass; NS year pass; Yearly route pass, Monthly route pass; benefit hours pass; City or region pass; different]

Type of change [start PT pass, stop PT pass]

7. Household Income

Experienced an occurrence [never; a few times]

Timing: since living independent OR leaving parental house

Month [No idea; list with all months]

Year [choice of years]

Before situation [below 1 x modal income; around 1 x modal income; between 1 and 2 x modal income; around 2 x modal income; above 2 x modal income]

After situation [below 1 x modal income; around 1 x modal income; between 1 and 2 x modal income; around 2 x modal income; above 2 x modal income]

Type of change [extra person in household with income; fewer persons with income; salary raise; different]

Part 4: Transport Mode Choice

Five different trip purposes: work, study, groceries, shopping and sport

Frequency [open question] **every** … [day; week; month; year]

Main transport mode [not applicable; car as driver; car as passenger; bike; walking; train; bus; different]

Alternative transport mode [not applicable; no alternative; car as driver; car as passenger; bike; walking; train; bus; different]

Destination [open question]

Departure time [not applicable; before 9h00; between 9h00 - 12h00; between 12h00 – 16h00; between 16h00 – 19h00; after 19h00]

Departure time return [not applicable; before 9h00; between 9h00 - 12h00; between 12h00 – 16h00; between 16h00 – 19h00; after 19h00]

Distance [open question]

Travel time [open question]

Part 5: Judgement (CPT)

Safety: car as driver; car as passenger; bike; bus; train

Privacy: car as driver; car as passenger; bike; bus; train

Environment: car as driver; car as passenger; bike; bus; train

Comfort: car; bike with rain; bike with dry weather and 10 degree; bike with dry weather and 20 degree; PT with rain without transfer; PT with rain and one transfer; PT with rain and two transfers; PT with dry weather without transfer; PT with dry weather and one transfer; PT with dry weather and two transfers

Costs: car; bus; train

Travel time: car; bike; bus; train

Part 6: Judgement (stated preference)

Nine designs with six aspects (safety; comfort; privacy; environment; costs and travel time) of two levels (80=favourable, 20=not favourable)

Appendix 2 | Classification life course events

Classes of the Internet-based survey are on the left side and the new classes are on the right side. The question type of change from the matrix question is recoded into 2 or 3 classes.

Event: Housing (type of change)

1 = first student room	2 = student housing
2 = move to other student room	2 = student housing
3 = independent living	1 = independent housing
4 = living together	1 = independent housing
5 = rent a house	1 = independent housing
6 = buy a house	1 = independent housing
7 = moving in with parents	3 = parental housing
8 = other	1 = independent housing

Event: Work (type of change)

1 = first job	1 = employed
2 = different job	1 = employed
3 = job transfer	1 = employed
4 = temporarily no job	2 = unemployed
5 = retirement	2 = unemployed
6 = other	1 = employed

Event: Study (type of change)

1 = change education center	1 = studying
2 = start a new education	1 = studying
3 = stop with education	2 = not studying
4 = graduate	2 = not studying
5 = start with certificate/course	1 = studying
6 = stop with certificate/course	2 = not studying
6 = other	2 = not studying

Classes of the Internet-based survey are on the left side and the new classes are on the right side. Before and After situations are recoded into new classes.

Event: Household (Before/After)

# Household members	1 = one person
	2 = two persons
	3 = three persons
	4 = four+ persons

Classification is the last step. This means that a transfer from for example 4 to 5 persons is listed as occurrence “change, increase of number of household members” (instead of “change, same number of household members” from 4+ persons to 4+ persons).

Event: Car availability (Before/After)

1 = no car	0 = no car
2 = 1 car, 1 car user	2 = # cars = # car users
3 = 1 car, > 1 car user	1 = # cars < # car users
4 = 2 cars, 1 car user	3 = # cars > # car users
5 = 2 cars, 2 car users	2 = # cars = # car users
6 = 2 cars, > 2 car users	1 = # cars < # car users

Event: Public transport pass (Before/After)

1 = no pass	0 = no PT pass
2 = OV student pass	1 = PT pass
3 = OV year pass (all PT)	1 = PT pass
4 = NS year pass (only train)	1 = PT pass
5 = year route pass	1 = PT pass
6 = month route pass	1 = PT pass
7 = month pass	1 = PT pass
8 = discount pass	1 = PT pass
9 = City or Region pass (bus)	1 = PT pass
10 = other	1 = PT pass

Event: Household Income (Before/After)

1 = below modal income	1 = class one (<= modal income)
2 = 1x modal income	1 = class one (<= modal income)
3 = between 1x and 2x modal income	2 = class two (> modal income)
4 = 2x modal income	2 = class two (> modal income)
5 = above 2x modal income	2 = class two (> modal income)

Appendix 3 | List of occurrences and states of all life course events

In the first colum the occurrences are listed and in the second column the different event states are listes. In the thesis there is often a referrence to subevent. The numbers used for these subevents are listed in the left column.

<u>Housing event</u>

	No change	independent living
1	Change, independent housing	student living
2	Change, student housing	parental living
3	Change, parental housing	

<u>Household event</u>

	No change	one household member
4	Change, decrease household members	two household members
5	Change, increase household members	three household members
6	Change, same household members	four or more household members

<u>Work event</u>

	No change	employed
7	Change, employment	unemployed
8	Change, unemployment	

<u>Study event</u>

	No change	studying
9	Change, study	not studying
10	Change, no study	

<u>Car availability event</u>	
No change	no car
11 Change, decrease car availability	car users > cars
12 Change, increase car availability	car users = cars
13 Change, same car availability	car users < cars
<u>PT pass event</u>	
No change	No PT pass
14 Change, stop PT pass	PT pass
15 Change, start PT pass	
16 Change, change PT pass	
<u>Household Income event</u>	
No change	< = modal income
17 Change, decrease household income	> modal income
18 Change, increase household income	
19 Change, same household income	

Appendix 4 | Routing Internet-based survey

Routing 1

living situation= independent housing (1)

age => 35 years old (2)

driver's licence= yes (Y)

Event	recall period	routing
Housing	since Tx= leaving parental house	(living situation=1)
Household	since Tx= leaving parental house	(living situation=1)
Work	since Ty= leaving school	
Study	skipped	(age=2)
Car availability	since Tz= living independent	(living situation=1) (driver's licence=Y)
PT pass	since age 16	
Household Income	since Tz= living independent	(living situation=1)

Housing			Tx							T
Household			Tx							T
Work				Ty						T
Study										
Car availability					Tz					T
PT pass	T16									T
Household Income					Tz					T
	past									present

Tx = leave parental house

Ty = leave school

Tz = live on your own

Routing 2

living situation= independent housing (1)

age => 35 years old (2)

driver's licence = no (N)

Event	recall period	routing
Housing	since Tx= leaving parental house	(living situation=1)
Household	since Tx= leaving parental house	(living situation=1)
Work	since Ty= leaving school	
Study	skipped	(age=2)
Car availability	skipped	(living situation=1) (driver's licence=N)
PT pass	since age 16	
Household Income	since Tz= living independent	(living situation=1)

Housing			Tx							T
Household			Tx							T
Work				Ty						T
Study										
Car availability										
PT pass	T16									T
Household Income					Tz					T
	past									present

Tx = leave parental house

Ty = leave school

Tz = live on your own

Routing 3

living situation= independent housing (1)

age < 35 years old (1)

driver's licence = yes (Y)

Event	recall period	routing
Housing	since Tx= leaving parental house	(living situation=1)
Household	since Tx= leaving parental house	(living situation=1)
Work	since Ty= leaving school	
Study	since age 16	(age=1)
Car availability	since Tz= living independent	(living situation=1)
		(driver's licence=Y)
PT pass	since age 16	
Household Income	since Tz= living independent	(living situation=1)

Housing			Tx							T
Household			Tx							T
Work				Ty						T
Study	T16									T
Car availability					Tz					T
PT pass	T16									T
Household Income					Tz					T
	past									present

Tx = leave parental house

Ty = leave school

Tz = live on your own

Routing 4

living situation= independent housing (1)

age < 35 years old (1)

driver's licence = no (N)

Event	recall period	routing
Housing	since Tx= leaving parental house	(living situation=1)
Household	since Tx= leaving parental house	(living situation=1)
Work	since Ty= leaving school	
Study	since age 16	(age=1)
Car availability	skipped	(living situation=1) (driver's licence=N)
PT pass	since age 16	
Household Income	since Tz= living independent	(living situation=1)

Housing			Tx							T
Household			Tx							T
Work				Ty						T
Study	T16									T
Car availability										
PT pass	T16									T
Household Income					Tz					T
	past									present

Tx = leave parental house

Ty = leave school

Tz = live on your own

Routing 5

living situation= student housing (2)

age < 35 years old (1)

driver's licence = yes (Y)

Event	recall period	routing
Housing	since Tx= leaving parental house	(living situation=2)
Household	skipped	(living situation=2)
Work	since Ty= leaving school	
Study	since age 16	(age=1)
Car availability	since Tx= leaving parental house	(living situation=2) (driver's licence=Y)
PT pass	since age 16	
Household Income	since Tx= leaving parental house	(living situation=2)

Housing			Tx							T
Household										
Work				Ty						T
Study	T16									T
Car availability			Tx							T
PT pass	T16									T
Household Income			Tx							T
	past									present

Tx = leave parental house

Ty = leave school

Routing 6

living situation= student housing (2)

age < 35 years old (1)

driver's licence = no (N)

Event	recall period	routing
Housing	since Tx=leaving parental house	(living situation=2)
Household	skipped	(living situation=2)
Work	since Ty=leaving school	
Study	since age 16	(age=1)
Car availability	skipped	(living situation=2) (driver's licence=N)
PT pass	since age 16	
Household Income	since Tx= leaving parental house	(living situation=2)

Housing			Tx							T
Household										
Work				Ty						T
Study	T16									T
Car availability										
PT pass	T16									T
Household Income			Tx							T
	past									present

Tx = leave parental house

Ty = leave school

Routing 7

living situation= parental housing (3)

age < 35 years old (1)

driver's licence = yes (Y)

Event	recall period	routing
Housing	since age 16	(living situation=3)
Household	skipped	(living situation=3)
Work	since Ty= leaving school	
Study	since age 16	(age=1)
Car availability	since Ti= driver's licence	(living situation=3) (driver's licence=Y)
PT pass	since age 16	
Household Income	skipped	(living situation=3)

Housing	T16									T
Household										
Work				Ty						T
Study	T16									T
Car availability			Ti							T
PT pass	T16									T
Household Income										
	past									present

Ty = leave school

Ti = driver's licence

Routing 8

living situation= parental housing (3)

age < 35 years old (1)

driver's licence = no (N)

Event	recall period	routing
Housing	since age 16	(living situation=3)
Household	skipped	(living situation=3)
Work	since Ty= leaving school	
Study	since age 16	(age=1)
Car availability	skipped	(living situation=3) (driver's licence=N)
PT pass	since age 16	
Household Income	skipped	(living situation=3)

Housing	T16									T
Household										
Work				Ty						T
Study	T16									T
Car availability										
PT pass	T16									T
Household Income										
	past									present

Ty = leave school

Author index

Subject index

Summary

Modelling life trajectories and mode choice using Bayesian Belief Networks

Traditionally, transport mode choice was primarily examined as a stand alone problem. Given a purpose and destination, the choice of transport mode was modelled as a function of the various attributes of the transport mode alternatives. Later, transport mode choice decisions were modelled as part of more comprehensive models (activity-based approach). There is a need in the transport research community to explore and model dynamics in activity-travel patterns along various time horizons. This will lead to dynamic models of behavioural change. In this thesis, it is argued that a life course perspective offers some potential advantages in understanding and modelling activity-travel decisions, including transport mode choice. Central concepts in the life course approach are life trajectories, transitions and events. An individual life course is composed of multiple, interdependent careers (i.e. housing, household, education, occupational career) which develop over time in parallel. Earlier life transitions may have a cumulative effect on later life. The concepts of timing, sequencing, duration and spacing are used to describe life events, transitions and trajectories.

The assumed effect of events on activity-travel decisions is captured in terms of a theory of learning and adaptation. Individuals develop and continuously

adapt choice rules while interacting with their environment. The context is non-stationary, uncertain and highly dynamic and therefore it is assumed that individuals adapt their behaviour. Under stationary conditions, individuals will show habitual behaviour after some period of time. A life course event is seen as a trigger that may induce individuals and households to reorganise their activities in time and space. A particular event may also lead to other life course events. Thus, life course events may have direct and indirect effects on activity-travel patterns. An event does not necessarily lead to immediate changes in particular facets of activity-travel patterns. Behavioural change may also occur in anticipation of life course events.

Bayesian Belief Networks is the approach adopted in this thesis to model the direct and indirect effects of life course effects on transport mode choice. More complex causation patterns can be included and results can be directly interpreted in terms of the classified events. Such networks need as input empirical data to learn the structure of the network and the conditional probability tables of the variables that are identified to be relevant.

Data was collected using a retrospective Internet-based survey. Retrospective surveys, especially when administered through the Internet, are a good alternative for (quasi-)longitudinal data collection methods, like panel surveys, repeated cross sectional surveys, and cohort pseudo-panel surveys. One would expect that the quality of data coming from a retrospective survey depends on the nature of the event about which information is collected and on the time elapsed between the occurring of the events and the time of the retrospective survey. The quality of the data was tested and the results were positive. In case of life course events memory lapses are less of a problem. Life course events can be better recollected than other events. In this study, the results of the reliability and validity tests of the collected data showed that item non-response in general was relatively low, especially for those life course events that serve as markers unfolding one's life. A statistical analysis suggested that memory / cohort effects were not found for the more salient life course events, such as housing, work and study related events. Memory may have an effect in reporting of events in case of income and transport mode related events (car availability and PT pass). The study illustrated that certain details of events, such as housing type and housing state are more difficult to recall.

The time effect of an influence of life course events on mode choice was tested with a simple multinomial logit model. The results support the conclusion that a certain time influence exists in the response to events. The data of the retrospective Internet-based survey was used as input for two Bayesian Belief Networks, a life trajectory and a mode choice network. A year is chosen as the unit of analysis for these networks. Both networks were successfully learned from the data. The first network can be used to simulate a person's life trajectory and the second network can be used to predict mode choice for an individual at a certain time given the individual life trajectory.

The goodness-of-fit of the learned Bayesian Belief Networks was assessed on the basis of the log likelihood statistic. The values indicated that both networks perform relatively well. It was also investigated whether the life trajectory network was capable of reproducing observed characteristics of complete life trajectories. The observed and predicted life trajectories were compared in terms of the following criteria: the number of occurrences, interval times between occurrences of events, simultaneous occurrences of events and sequence of occurrences of events. The life trajectory network reproduced the number of occurrences in the life trajectories quite well. In general, the network predicted more or less the same means of interval times for the events, except for the PT pass event. The network was less successful in predicting correctly the observed incidence of synchronic events. The results of the sequence alignment analysis indicate that the network predicts the sequence of the occurrences in the life trajectories relatively good. The modal split (car, public transport and slow transport) of the predicted mode choice was compared with the observed mode choice. Results indicated a relatively small over prediction of public transport and under prediction of car and slow transport. This suggests that the mode choice network is able to simulate more or less the same mode choice as registered in the data.

The learned networks were used to study direct and indirect effects of one variable on other variables in the network. The described effects seem logical. A simulation illustrated the dynamics of the lives of ten inhabitants of a newly build neighbourhood. It showed that, insight in dynamics of life trajectory events and mode choice can lead to a better understanding which can support the development of better or different policy measures.

Curriculum Vitae

Marloes Verhoeven was born on 11th of April 1979 in Oss. In 1997 she started her study at the Department of Architecture, Building and Planning of the Eindhoven University of Technology. During her study, she was a student assistent and between 2001 and 2002 a board member of CHEOPS, the student association. She obtained her Master degree in 2004.

After her graduation, she decided to continue as a PhD candidate in the DDSS research program. Together with several other PhD students she started a PhD network in 2006 within the Department of Architecture, Building and Planning. In 2007 she won a price for the best journal paper of the International Journal of Urban Sciences. During her PhD research, she combined research with activities of the PhD network as well as teaching Bachelor students.

Her research interests are in the areas of human behaviour and decision making processes, urban and transport planning, activity-travel behaviour, and Internet-based survey techniques.

In the beginning of 2009, she already accepted a new job at a transportation consultancy bureau in Eindhoven in parking. Together with Eindhoven University of Technology she started a panel for parking related surveys. This is an excellent opportunity to continue working together with the Urban Planning Group.

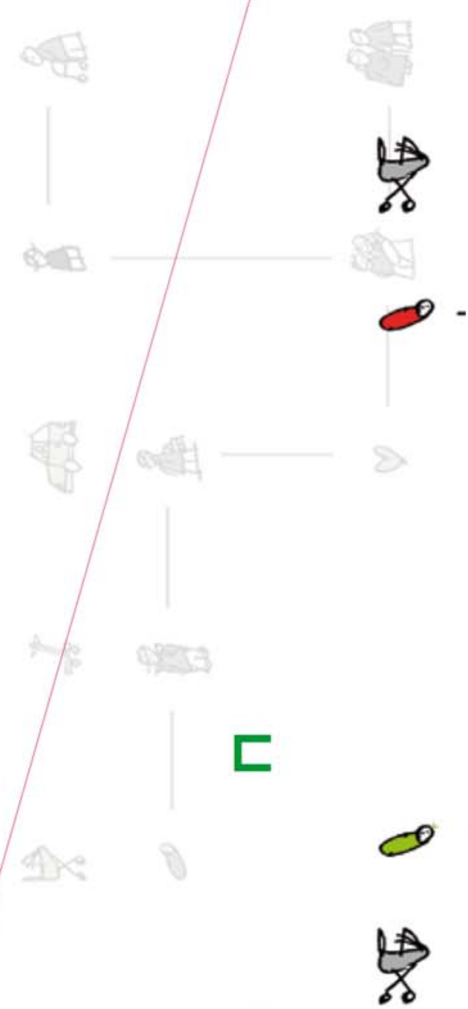
Where innovation starts

Modelling life trajectories and mode choice using Bayesian Belief Networks

An important limitation of operational activity-based models is that these models simulate activity-travel patterns of a population for a single day. Another limitation of these models is the lack of simulating explicitly how individuals and households react to changes. Thus, there is a need to explore and model dynamics in activity-travel patterns along various time horizons.

This PhD thesis contributes to the emerging, but still scarce literature on long-term dynamics in the activity-based approach. The specific focus in this thesis is concerned with life course or life trajectory events that may cause changing needs or preferences and / or influence the constraints that impact activity-travel decisions.

This thesis is based on the assumption that life trajectory events may cause individuals and households to change their activity-travel patterns. A modelling approach, based on Bayesian Belief Networks, that allows representing and simulating such dynamics is developed and tested in this thesis. The focus is especially on changes in transport mode choice.



/ faculty of architecture building and planning

TU/e Technische Universiteit **Eindhoven** University of Technology