Design of stated adaptation experiments: discussion of some issues and experiences

Document status and date:
Published: 01/01/2008

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

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Download date: 17. Nov. 2023

DESIGN OF STATED ADAPTATION EXPERIMENTS: DISCUSSION OF SOME ISSUES AND EXPERIENCES[1]

Kelly van Bladel, Tom Bellemans, Davy Janssens, Geert Wets
Transportation Research Institute, Faculty of Applied Economics, Hasselt University,
Wetenschapspark 5 bus 6, B-3590 Diepenbeek, Belgium, phone +32 11 269158, fax
+32 11 269199, email geert.wets@uhasselt.be

Linda Nijland, Theo A. Arentze, Harry J.P. Timmermans
Eindhoven University of Technology, Urban Planning Group, PO Box 153, 5600 MB
Eindhoven, The Netherlands, phone +31 40 247 3315, fax +31 40 243 8488, email
{e.w.l.nijland; h.j.p.timmermans}@bwk.tue.nl / eirass@bwk.tue.nl

ABSTRACT

Stated preference experiments have become commonly used methods of data collection in transportation research and the increasing importance of individual choice processes in travel behaviour research instigates the use of stated adaptation experiments. Two complementary stated adaptation approaches can be distinguished. The first approach follows the traditional stated preference methods and focuses on the statistical analysis of the variables that affect the individual choice processes; the second approach is not based on strict rules of experimental design and is mainly descriptive in nature. Based on a detailed description of two experiments of the first type, the design and implementation of such experiments are discussed. The most important lesson from our experiences with these stated adaptation experiments is the design of the hypothetical situations: the situations should be realistic for the respondents as well as useful for statistical analyses. Although this is not an easy task, the implementation of the experiments by means of an interactive Internet-based survey is found to be very helpful: such surveys are able to dynamically collect and process personalized data that can be used to design realistic and statistically sound hypothetical situations.

KEYWORDS stated adaptation, time-use, activity-based models

INTRODUCTION

Stated preference and choice experiments have become widely accepted means of data collection in transportation research. Especially, conjoint methods have been widely used. Conjoint preference methods were originally developed for measuring preference or utilities for multi-attribute profiles. Before they were introduced, stated preference methods were typically based on explicitly and separately measuring respondents' evaluations of the various attributes and the relative importance of the weights they attach to these attributes. However, the validity of this so-called compositional measurement approach was increasingly questioned. The measurement problem then becomes "if we would measure the overall evaluation of the multi-attribute profile, can we simultaneously scale the part-worth evaluations, given some assumed integration function". This became known as the conjoint measurement problem and many methodological developments and studies have been conducted ever since. In the early years, linking preferences to choices was based on ad hoc assumptions or questionable approaches such as the exploitation method, but later strategies of creating choice experiments which simultaneously allowed one to test the assumed choice model and derive utility scales were developed.

While the experimental nature of stated preference and choice studies involves expressing an overall evaluation of a series of attribute profiles, respectively choosing between two or more attribute profiles, the focus of stated adaptation experiments shifts towards expressing the likelihood and nature of possible behavioural change. The information processing burden for respondents is usually considerable higher in the sense that the experiments articulate different contexts under which respondents need to indicate if and how they would change their behaviour. It implies that respondents need to interpret the information, build a mental representation of the situation and then simulate their choice process imagining how they would react in reality. Moreover, in many cases, they need to provide information about behavioural change in terms of multiple choice facets. Compared to stated preference and choice experiments, much less is known about stated adaptation experiments. In some case, similar rigorous principles of experimental design are applied to this approach, implying that only the nature of the dependent variable differs: not ratings, rankings or choices, but transition probabilities from current behaviour to new behaviour or the probability of some response to external change. In other studies, the goal is to identify how people will change their behaviour in reaction to some new policy, and more elaborate experiments or games are used.

Thus, researchers have to make a trade-off between task realism and respondent burden. Simplification means that respondents will be better able to provide answers. On the other hand, increased task realism will potentially increase the validity of the responses, unless respondents do not capture the full meaning of the decision context, because of the information processing challenge. If this is the case, the responses may be flawed.

In this paper, we will first briefly discuss some previous applications of stated adaptation experiments reflecting this line of development. Then, against this background, we will describe in some detail our experiences with the design and implementation of specific experiments, and how they were used in the development and estimation of a model which aims at simulating adaptation decisions in activity rescheduling behaviour. More specifically, issues in the design of the experiment, its implementation, respondent burden and qualitative evidence of reliability will be discussed.

OVERVIEW OF STATED ADAPTATION STUDIES

The analysis of travel behaviour is an important aspect of transportation research and stated preference experiments have become commonly used methods of data collection in this domain, because the direct observation of real travel did not provide the data necessary to realistically predict and simulate traffic flows and how they react to external changes. Whereas the focus of transportation policy used to be on expanding and improving the supply side (infrastructure, transport systems, etc.), nowadays it is on managing the demand for transportation. This implies that individual choice processes become increasingly important and exactly these processes are ideally examined by means of stated preference, and more specifically stated adaptation experiments.

Although stated preference experiments are frequently applied in transportation research (see Hensher, 1994, and Louviere *et al.*, 2000, for an overview) and in spite of the methodological analyses and improvements to this technique that are developed within the domain of transportation (e.g. Ben-Akiva and Lerman, 1985; Caussade *et al.*, 2005; Hess *et al.*, 2006; Van der Waerden, 2006), relatively little is still known about stated adaptation techniques and at the moment there is no universal definition, although it is generally accepted that stated adaptation experiments are developed with a focus on individual choice processes. Two opposite approaches exist.

The first approach follows the strict principles of experimental design that are used in stated preference techniques: the individual must choose between a series of predefined behavioural responses when presented with one or several hypothetical situations that are precisely described based on a variety of influential attributes that can take different values. Only the nature of the dependent variable differs from traditional stated preference experiments, since the responses are transition probabilities from current behaviour to new behaviour instead of ratings, rankings or choices of alternatives. This approach has the advantage that it satisfies a number of statistical assumptions and requirements, so that fairly standard statistical techniques and models can be used to analyse the data and estimate the effect of the influential variables included in the experiment. Sometimes, change is related only to the attributes of the choice alternatives, but often, anticipated change (also) relates to the decision context, and this case will be more demanding for respondents. If change only applies to the attributes of current choice alternatives, respondents only need to process such changes and map them to their current preference structures and decision strategies. The experimental task then involves changing current behaviour or not and respondents have to assess whether the indicated change in the attributes will induce them to change their behaviour. It means that respondents only need to check whether the change in the current non-chosen alternative generates an increase in utility that is higher than the utility difference between the current chosen alternative and the changed alternatives. In more complex experiments, however, the decision context will (also) change. This is more demanding in the sense that the changed context may be less or even unfamiliar, implying that respondents first need to interpret the data, cognitively construct the appropriate decision context, apply their current preference and decision strategies to the constructed decision context, mentally simulate the consequences of choice options, judge the consequences and report choices or changes in behaviour, which may involve multiple choice facets. Hence, respondents may have difficulties to realistically evaluate the decision context. Respondent burden may also become high, if choice situations differ from the subject's experiences. The unfamiliar choice context and the respondent

burden can bias the data. Researchers have tried to develop hypothetical situations that are as close as possible to the subject's experiences. To check how well subjects can relate to the presented hypothetical situations, information about (parts of) the subject's real activity-travel schedules can be collected. This information can also be used to construct realistic stated choice situations, but then interactive interview or survey techniques are needed to collect and process these data (Faivre D'Arcier, 2000). Another way to enhance the realism of stated preference or stated adaptation choices is taking revealed preference situations as the starting point for the stated choice experiments, although more complex models are then sometimes required, since certain statistical assumptions required by traditional logit models are no longer satisfied (Train and Wilson, 2008). In transportation research the interest in stated adaptation experiments of this type has recently grown, resulting in experiments reported by authors such as Arentze et al. (2004), Nijland et al. (2006), and Wang et al. (2003).

In addition to studies that are based on experiments, especially in the early years of activity-based research in transportation, games have been used to obtain a better insight into adaptive behaviour. In such games, subjects are free to indicate every behavioural response he/she considers feasible: the responses are not predefined by the researcher (Axhausen and Sammer, 2001; Faivre D'Arcier, 2000; Faivre D'Arcier et al., 1998). Since these experiments are not based on rigorous rules of experimental design and subsequent statistical analysis, the researcher has more flexibility in designing choice situations that are not distant from the subject's experiences. This stated adaptation approach is mainly used to analyse how individuals will change their behaviour in reaction to some new transport policy or to assess the social acceptability of transport policy measures. These studies can also be used to examine which variables affect certain choice processes and should be incorporated in the utility function. Furthermore, the reasons why the same situational change causes a completely different reaction for different socio-economic groups can be revealed by these stated adaptation games. Mostly it is difficult to derive statistically significant conclusions from these experiments, but they often have only descriptive purposes. Sometimes, the different individual responses are grouped into categories and statistically analysed, but this often involves loss of data or requires the imposition of simplifying assumptions. For these studies, personal interviews are usually the preferred survey method. Examples of this approach can be found in Andrey et al. (2004), Faivre D'Arcier et al. (1998) and Roorda and Andre (2007). The problem of cognitive representation and mental simulation however also applies to this type of studies. However, because the researcher is less trying to create the same decision context for all respondents or to even systematically vary the decision context, the problem may be less of an issue in the contexts of games compared to systematic experiments.

RECENT EXAMPLES AND EXPERIENCES

This section presents the experiences of the authors with stated adaptation experiments in the context of estimation of the Aurora model (Timmermans et al., 2001; Joh et al., 2006). We will describe the design and implementation of the experiments and discuss issues related to data quality and respondent burden. More details about the experiments are provided in Bladel et al. (2006) and Nijland et al. (2006, 2007).

Purpose

The experiments were conducted to test the basic assumptions of Aurora (Timmermans *et al.*, 2001; Joh *et al.*, 2006), a model which was developed to predict and simulate activity rescheduling behaviour. The model assumes that individuals experience a certain level of utility if they participate in activities and that activity generation and the subsequent (re)scheduling decisions will be mainly driven by this utility, which is a complex and continuous function of the time available for the activity, D_a, the time passed since the last time the activity has been performed, T_a, and the time pressure on the daily activity-travel schedule measured as V_a, the discretionary time during the day. In the Aurora model, the effect of D_a on the utility is represented by an asymmetrical S-shaped function and a similar utility function describes how the need or urgency of an activity grows over time. Following functional form is suggested in the model (Joh *et al.*, 2003):

$$U_a = \frac{U_a^{\max}}{\left(1 + \gamma_a \exp\left[\beta_a(\alpha_a - D_a)\right]\right)^{1/\gamma_a}}$$

$$with \quad U_a^{\max} = \frac{U_a^x}{1 + \exp\left[\beta_a^x(\alpha_a^x - T_a)\right]}$$

where
D_a is the available time;
T_a is the activity history;
U_a^{max} is the asymptotical upper limit of the U_a-function;
U_x is the asymptotical upper limit of the U_a^{max}-function;
α_a, β_a and γ_a are activity-specific parameters for the U_a-function;
α_a^x, β_a^x, γ_a^x and U_a^x are activity-specific parameters for the U_a^{max}-function.

Empirical validation of the Aurora model has long been based on cross-sectional data and complex estimation techniques like purpose-built genetic algorithms (Joh *et al.*, 2003), due to the complex functional form of the assumed utility functions and to lack of appropriate data. A more rigorous test of the basic assumptions underlying the model and the estimation of adaptive behaviour requires controlled experiments. These experiments serve two purposes. First, experimental design data allow testing the validity of the assumed S-shaped utility functions. This is not a stated adaptation experiment in these sense that the focus is not on adaptation but on testing and estimating the utility function that is assumed to trigger behavioural adaptation. A second experiment involved asking respondents how they would change behaviour under changing conditions. The task can be expressed as follows:

Assume you intend to conduct activity A today. For the activity including travel time you have M minutes. You want to conduct the activity at location L and you arrive there by transport mode T. Unfortunately, today you have encountered a delay with as a consequence that the available time (for activity and *travel) has been reduced to R minutes. After this, you should be back for another activity. What would you do in this situation? Indicate for each of the following options the probability that you would choose this.*

Thus, rather than an all-or-nothing response, a respondent could indicate a probability. In this way, the respondent could take into account variation in circumstances on factors that are not controlled in the experiment. The following choice options were included:

1. I change location
2. I change transport mode
3. I shorten the duration of the activity
4. I change the location and transport mode
5. I change the location and shorten the duration of the activity
6. I change the transport mode and shorten the duration of the activity
7. I change the location, the transport mode and the duration
8. I cancel the activity

This is a true stated adaptation experiment. Both types of experiments share the problem of task realism and respondent cognitive burden. In light of page limitations and the possibility of communicating shared experiences, in this paper we will focus on the experiments testing the utility function. Two similar experiments were conducted: one in The Netherlands and one in Flanders. In both experiments, subjects were asked to express their activity-scheduling choice in various hypothetical situations that were constructed by varying the levels of the three key influential variables D_a, T_a and V_a and a time-of-day variable, I_a. In case of the Flemish experiment, the purpose was to evaluate the effect of D_a, T_a, V_a and I_a, as well as the influence of location, L_a, and accompanying persons, C_a, on the activity utility. The different situations that were presented to the subjects in this experiment thus included more variables. The experiment was also used to assess how I_a, L_a and C_a enter the S-shaped utility function, since the functional specification of these variables is not presupposed by the theory. A final research aim of the experiment was to examine whether different groups of the population have different utility functions. The assumption that women and men experience different levels of utility for the same activity was tested by examining the influence of gender on the supposed utility functions.

Design and Application

Experiment 1. The aim of this experiment was to estimate the influence of the factors, available time (D_a), the time elapsed since last performance of the activity (T_a), the amount of available discretionary time (V_a) and the time-of-day (I_a), on the propensity of individuals to schedule a given activity on a given day. An experiment was designed in which different levels of the four variables were shown in hypothetical situations. For both T_a and D_a the number of levels was set to four and for V_a and I_a three levels were chosen. Because we assumed that there is no interaction between the factors, an experimental design with 16 situations was enough to estimate the supposed effects. To reduce respondent-burden, the questionnaire covered a single activity for each respondent. Thus, in this experiment each respondent received 16 scenarios concerning one activity. The hypothetical situations had the following general form:

Assume today is a day on which you have V_a hours of discretionary time. The last time you conducted activity A is T_a days ago. Assume furthermore that it is now I_a and that at the

current moment you have D_a minutes available to conduct the activity including travel time. What would be the probability that you choose the option of conducting the activity now (as opposed to delaying it to a later moment of the day or a later day)?

Respondents were asked to assign a percentage to the option of conducting the activity now. By indicating a probability, the respondent could take circumstances and factors that were not included in the experiment into account. To avoid order effects, the 16 hypothetical scenarios were shown either in the original or in the reversed order and a respondent was randomly assigned to an order.

The survey was developed to collect data for a representative set of flexible, frequently scheduled out-of-home activities. Only the activities that meet those conditions are considered relevant for the S-shaped functions of the Aurora model. Flexible means that the start time and the duration are not completely fixed on the day of performance of the activity. It is still possible to conduct the activity earlier or later or to shorten or prolong the activity. The following activities were included in the survey:

- Daily shopping
- Non-daily shopping: buying clothes
- Visiting relatives/friends
- Visiting a café, bar, discothèque
- Sports free (not in union/club context)
- Walking in a park or nature
- Touring by bike
- Touring by car/motor bike

During the test period it became clear that it was easier for the respondents to answer the questions for more specific activities. Therefore, the key activity "buying clothes" was chosen to represent the quite diverse group of non-daily shopping activities. Furthermore, in case of the social activity "visiting relatives/friends", respondents were asked to consider only a specific relative or friend, they frequently visit, when answering the questions.

For all eight activities included in the survey, the same number of levels was chosen. In this way, the same experimental design could be used for each activity. The levels of V_a were operationally defined as the time left after subtracting the time for sleeping and mandatory activities. The same set of three levels was used for each respondent, viz. 15, 10 and 6 hours. Roughly, these values represent a non-work day, a part-time workday and a full workday (including travel time). For time-of-day (I_a) the levels 'morning', 'afternoon' and 'evening' were chosen. Due to the fact that individuals may be quite diverse with regard to the normal frequency and time needed to implement an activity of a certain type, the levels of T_a and D_a had to be based on the usual frequency and duration of the subject. Therefore, five duration classes (including travel) and five frequency classes were defined. For every class, four levels were chosen in such a way that the range between the extremes within the class was subdivided into approximately equal intervals. In a part of the questionnaire preceding the stated adaptation experiment, respondents had to indicate their normal frequency, duration and travel time of the considered activity. Based on this information, a duration group and a frequency group were allocated to the respondents and the levels of the classes were applied to the choice experiment. Table 1 shows the T_a and D_a groups and their group-specific levels.

Table 1 Choice of levels for T_a and D_a by group

	Frequency (T)			Required time (D)	
Group	Mean frequency condition	T levels 1-4	Group	Mean time condition	D levels $1-4$ (min.)
1	≥ 2 x a week OR	1 day	1	≤ 45 min.	15
	≥ 7 x per month OR	2 days			45
	≥ 74 x per year	4 days			30
		7 days			90
2	1 x per week OR	3 days	2	46–90 min.	30
	3–6 x per month OR	7 days			60
	31–73 x per year	10 days			90
		14 days			150
3	2 x per month OR	7 days	3	91–150 min.	60
	18–30 x per year	14 days			120
		21days			150
		1 month			210
4	1 x per month OR	14 days	4	151–210 min.	120
	8–17 x per year	21 days			180
		1.5 months			210
		2 months			270
5	< 1 x per month OR	1 month	5	> 210 min.	150
	< 8 x per year	2 months			240
		4 months			300
		6 months			360

Because a paper-and-pencil version of the questionnaire would be impracticable, the survey was internet-based. This enabled us to adapt several parameters of the experiment to the responses of the respondents, for example, the assignment of the activity, routing through the questionnaire and the levels of T_a and D_a.

Approximately 350 individuals, who had indicated in an earlier survey to be willing to participate again in another survey, were approached by email. Furthermore, they were asked to forward the invitation to their friends and acquaintances. Incentives were awarded in two ways. First, five vouchers of 50 Euro were allocated through a lottery, to respondents who successfully "recruited" three other respondents. Second, in an additional lottery, twenty vouchers of 25 Euro were allocated to respondents. In addition, the following procedure was used to increase the number of observations. After finishing the questionnaire, each respondent was invited to fill in the questionnaire a second time for a different activity. Finally, 328 questionnaires were completed by 188 respondents.

Experiment 2. Analogous with the Dutch experiment, the Flemish experiment focused on the utility of particular data points of the S-shaped utility functions, namely those data points that contain relevant information for the validation of the theory and the model estimation. Each data point on the utility function represents a particular situation that is characterised by specific levels of D_a, T_a, V_a, I_a, L_a and C_a and the utility of an activity in that particular situation is assumed to be a linear function of the part-worth utilities that are associated with the specific levels of D_a, T_a, V_a, I_a, L_a and C_a. As is the case for the first experiment, different choice situations were designed by varying the levels of the influential variables and subjects were asked whether or not they would schedule a specific activity in the presented hypothetical situation. For the estimation procedure, the interpretation of the part-worth utilities and the choice of the base levels for D_a, T_a and V_a, the approach of the first

experiment was copied. Both experiments followed an experimental design to choice situations/decision contexts and the assumption that there was no interaction between the different influential variables was also adopted for the Flemish experiment. A simple design could thus be used: for each hypothetical situation all the variables except for one are kept at a fixed level, so following exemplary choice situation is repeatedly presented to the respondents:

Assume it is time of day I_a and today you have V_a hours of discretionary time. You have D_a minutes available at present to conduct activity a, including travel time. What would be the probability that you choose to execute this activity [alone / with others] immediately (instead of performing it later (later today, tomorrow,…)) on location L_a, if you last performed activity a T_a days ago?

Since the theory of the S-shaped functions is only relevant for flexible non-routine activities that are frequently scheduled, not all activities are suitable to test the validity of the model. Two specific activity types, namely "daily shopping" and "social visit", were chosen for this experiment, because both of them are generally flexible and a large number of individuals conduct these activities on a frequent and regular basis. Furthermore, the activity types Daily Shopping and Social Visit are supposed to be wide enough to provide useful information and yet limited enough to be homogeneous.

The choice of the different levels of D_a, T_a, V_a, I_a, L_a and C_a was the next step in the design of the experiment. For both D_a and T_a the different levels were carefully chosen over the whole range of the S-functions in order to obtain as much information on the part-worth utilities as possible. The classification of subjects in D_a and T_a groups for each activity type, based on the average D_a and T_a values, happened parallel to the classification presented in Table 1. An important difference with the first experiment was the extra level of D_a and T_a that was incorporated in the design: this fifth level was defined at a very low D_a or T_a value, because findings of the first experiment revealed that four levels for D_a and T_a were not enough to unambiguously confirm the theory of the S-shaped utility functions. The levels of V_a were defined to approach an average working day (6 hours), half-time working day (9 hours) and day off (15 hours). The small discrepancy between the levels of V_a for the two experiments are due to different time use patterns in the Netherlands and Flanders. The levels of I_a were assigned the values 'morning', 'afternoon' and 'evening'. Subjects were invited in the beginning of the survey to answer activity-specific questions about activity locations and to order them according to preference. This information is subsequently used in the choice situations to personalize the levels of L_a: the first level of L_a is assigned the name of the most preferred location of the subject, the second level the name of the second preferred location. Finally, C_a was constructed as a binary variable that took the values 'alone' or 'with others'.

The Flemish experiment was also implemented as a dynamic Internet-based survey: a series of Internet pages were connected to a database in real-time, so that the subject-specific information about activities, average D_a and T_a, locations, etc. could be easily retrieved and used in the choice situations. The sample size of this experiment, however, was considerably smaller than the sample size of the first experiment, since the survey was initially implemented as a pilot study. Subjects were contacted by email from the authors' circle of acquaintances. They were invited to fill out the survey and they were asked to forward the received e-mail to other people who might be willing to participate in the experiment. After

three weeks, the pilot study of the experiment was closed and it turned out that 61 subjects successfully filled out the survey.

Experiences

Overall, the experiments succeeded in collecting the intended data: the effects of the different influential variables on the activity utility could be examined in detail and are shown to satisfy most of the postulated expectations. Based on the Dutch experiment, little conclusive evidence was found for the hypothesis of the S-shaped utility function, perhaps due to the missing level of D_a and T_a, but by including the fifth level of D_a and T_a in the Flemish experiment, the S-shape of the utility function could be confirmed, especially for T_a. The large drawback of the Flemish experiment is the limited number of subjects: no statistically significant conclusions could be drawn because the different D_a and T_a groups contained too few subjects. Yet, in spite of the small sample size, at least the shape of the function as visualized appeared to support the theory.

Given the recruiting methods for both surveys, verifying how many respondents were initially contacted, was not possible. This means that no regular non-response rate could be computed. The drop-out rate of the first survey was only 7%, calculated as the number of subjects that started the questionnaire compared with the number of respondents that successfully answered all questions. Most of the subjects that dropped out, did not complete the experimental task of the survey (12 out of the 23 drop-outs). This may suggest that the majority of the respondents did not decide to discontinue the experiment because it was too difficult or too demanding. Having said that, comments of some respondents which were invited to assess the reliability of the experiments, revealed that the decision contexts were not varied enough and were found unrealistic for some of the subjects. This suggests that the problem of cognitive burden may apply to a subset of respondents. Whether this would have any implications on the outcomes of the study, depends on the question whether activity scheduling of this group differs from those who completed the experiment. Because the experiment was placed at the end survey, non-completion may also be caused by the length of the questionnaire.

The Flemish experiment was completed by 88 subjects during the three weeks it was online and 61 of them successfully completed the survey. For this experiment, follow-up interviews revealed that respondents found the survey rather lengthy, in spite of the automated omission of non-relevant questions, e.g. questions about activity types that the respondent rarely scheduled. The experiment could be improved by eliminating some of the socio-demographic questions that proved to be unused in the subsequent analyses or by presenting only one activity type to each respondent, as was the case in the Dutch experiment. The choices to be made in the experiments were found to be most difficult. Subjects indicated they had to think consciously about decisions usually taken subconsciously because of reinforcement and known conditions. They also had to imagine themselves repeatedly in situations that were not always completely familiar to the respondent. This lack of familiarity increased cognitive burden.

These findings amplify the importance of the decision made in both experiments to increase the realism of the choice situations by using personalized values for the explanatory variables. The majority of the respondents did indicate they found the variable's levels were well-chosen, and presented values realistic. This suggests that the classification of the subjects into different D_a and T_a groups was successful. The only exception to this result occurred for

Social Visit: in both experiments some subjects indicated an average D_a of 12 hours and more for this activity type, which biased the realism of the values associated with the highest D_a group, as the levels of this group were based on an average D_a of 4.5 hours, with a maximum value of 6 hours. Nevertheless, for most subjects the hypothetical situations were reasonably realistic and accordingly they procured more reliable answers and more precise research results than would have been obtained by presenting all subjects with the same standardized values.

A vast majority of subjects had no comments on the explanatory variables used in the experiments, although the results indicated that the influence of time pressure was rather limited. Yet, a small number of respondents indicated that – especially for the activity types "going out" and "visiting relatives/friends" - some of the explanatory variables used, were normally not considered to evaluate activity scheduling or that other variables than D_a, T_a, V_a and I_a (e.g. the day of the week, weather conditions) also affect the utility of the activity. In both experiments the effect of these additional variables was set off by inviting the subjects to give a probability instead of a yes/no answer for each of the scenarios. It is not completely clear whether respondent burden was reduced by this increased flexibility, but it should be noted that not all statistical software packages are able to cope with non-binary response data on an individual level.

Software packages created to develop web-based surveys, can also pose challenges. The first experiment was developed with the online software of NetQuestionnaires. Some problems arose when generating the online questionnaire. First, the values of the T_a and D_a levels could not be calculated based upon the answers of questions that preceded the stated adaptation experiment. This meant that for the first experiment the classes of T_a and D_a levels had to be created as shown in Table 1, which resulted in 800 additional questions. Another problem was the allocation of the activities. The questionnaire only covered a single activity. Therefore, one activity, which the respondent conducts on a regular basis and considers to be flexible, had to be assigned to the respondent. This procedure was (at that time) impossible within the software package of NetQuestionnaires. The problem was solved by splitting the survey in a registration part and a main part. In the registration part, questions about the flexibility and frequency of conducting the activity had to be answered for each activity. After this part, one activity was assigned manually. These limitations do not restrict survey implementation when a programming language such as Perl is used, as was the case for the second survey: no extra questions were needed to classify the respondents into T_a and D_a groups and the survey was presented to the respondents as one questionnaire. Although more programming experience is needed to efficiently use such a language, the use of these programming languages enables more complex computations and real-time data processing, which facilitates the survey course for the subjects.

Internet-based applications are know to have several advantages over traditional survey instruments, such as time and cost efficiency, better data quality, etc. A well-known bias of Internet-based experiments is the underrepresentation of older-age and lower-education groups in the sample. This underrepresentation is indeed detected in both stated adaptation experiments, but exactly these socio-economic groups are also underrepresented when using more conventional survey instruments. It is thus more important to assure that a sample is representative in terms of travel behaviour, than in terms of socio-economic characteristics. For lack of sound reasons why Internet users' travel behaviour would be significantly different from the travel behaviour of people that do not use the Internet, a balanced Internet

sample could provide useful results (Arentze *et al.*, 1997; Arentze and Timmermans, 2002). For the two stated adaptation experiments under consideration, it would have been impossible to implement them in a paper-and-pencil format and highly inefficient to collect the data based upon personal interviews. The personalized questions, the complex survey routing and the calculations of the values used in the stated adaptation situations would have been impracticable to include in a survey that was not computer-based.

CONCLUSIONS AND DISCUSSION

Stated preference experiments have become commonly used methods of data collection in transportation research, because the analysis of travel behaviour requires data that cannot be captured by traditional revealed preference surveys. The shift in transportation policy from improving the supply side of transportation to travel demand management, increases the importance of individual choice processes that determine which activities people undertake and the subsequent travel episodes. Stated adaptation experiments are increasingly used to examine these choice processes. If hypothetical situations as well as possible behavioural responses of individuals are defined in advance, as is the case in the experiments of Nijland *et al.* (2006) and van Bladel *et al.* (2006), individual choices can be efficiently analysed. Both experiments were successful in collecting the envisioned data, because respondents were presented with realistic behavioural responses in reply to the different hypothetical situations. These situations could be efficiently described in terms of a limited number of well-known and easily interpretable attributes.

The design and implementation of good stated adaptation experiments can be challenging. As for most stated preference experiments, the selection of the explanatory variables and the choice of their values has to be carefully considered, the size of the sample has to be large enough to allow for statistically significant outcomes and the survey itself has to be designed in a respondent-friendly way, so that respondent burden is minimised. But the most important lesson learnt from the experiments in Nijland *et al.* (2006) and van Bladel *et al.* (2006) is the significance of the hypothetical situations' design: it is fundamental that the hypothetical situations are designed to be as realistic as possible for the respondents, because unfamiliar situations can cause respondent burden and unreliable survey results. Given the restrictions that experimental design can impose, this is not an easy task, but interactive surveys allow the researcher to check how well the respondents can relate to the presented hypothetical situations or to construct realistic stated choice situations. In Nijland *et al.* (2006) and van Bladel *et al.* (2006) Internet-based applications are used to collect and process respondent-specific data that is subsequently used to successfully design the hypothetical situations. Recommended is a software package or programming language without limitations: not every application is capable of creating complex stated adaptation experiments.

Given the increasing interest in individual choice processes within the domain of transportation, further research into various types of stated choice experiments is to be expected. Enhancements are currently pursued with respect to survey design and implementation. Furthermore, advanced statistical techniques that require less restrictive assumptions are developed and software packages provide integrated modules for these more complex models. The continuous improvements in computer efficiency and newly emerging technologies support these developments.

REFERENCES

Andrey, J., K.R. Burns and S.T. Doherty (2004). Toward sustainable transportation: Exploring transportation decision making in teleworking households in a mid-sized Canadian city. *Canadian Journal of Urban Research* **13**, pp.257-277.

Arentze, T.A., H.J.P. Timmermans, F. Hofman and N. Kalfs (1997). Data needs, data collection and data quality requirements of activity-based transportation demand models. Proceedings of an International Conference on Transportation Survey Quality and Innovation, Resource Paper Workshop, Grainau, Germany.

Arentze, T., F. Hofman and H.J.P. Timmermans (2004). Predicting multi-faceted activity-travel adjustment strategies in response to possible congestion pricing scenarios using an Internet-based stated adaptation experiment. *Transport Policy* **11**, pp.31-41.

Arentze, T.A. and H.J.P. Timmermans (2002). Response rates, sampling bias and reliability of Internet-based travel surveys: Experiences with a stated adaptation experiment. Research Report: EIRASS, Eindhoven, The Netherlands.

Axhausen, K.W. and G. Sammer (2001). "Stated responses": Überblick, Grenzen, Möglichkeiten. *Internationales Verkehrswesen* **53**, pp.274-278.

Ben-Akiva, M. and S.R. Lerman (1985). *Discrete choice analysis: Theory and application to travel demand*. The MIT Press, Cambridge (Mass) / London.

Caussade, S., J. de Dios Ortuzar, L.I. Rizzi and D.A. Hensher (2005). Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B* **39**, pp.621-640.

Faivre D'Arcier, B. (2000). Hypothetical situations: The attempt to find new behavioral hypotheses. *Transportation Research Circular* **E-C008**, TRB, Washington, DC.

Faivre D'Arcier, B., O. Andan and C. Raux (1998). Stated adaptation surveys and choice processes: Some methodological issues. *Transportation* **25**, pp.169-185.

Hensher, D.A. (1994). Stated preference analysis of travel choices: The state of practice. *Transportation* **21**, pp.107-133.

Hess, S., K.W. Axhausen and J.W. Polak (2006). Distributional assumptions in mixed logit models. Proceedings of the 85th Annual Meeting of the Transportation Research Board, Washington, DC, United States.

Joh, C.H., T.A. Arentze and H.J.P. Timmermans (2003). Estimating non-linear utility functions of time use in the context of an activity schedule adaptation model. Paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland.

Joh, C.H., T.A. Arentze and H.J.P. Timmermans (2006). Measuring and predicting adaptation behavior in multi-dimensional activity-travel patterns. *Transportmetrika* **2**, pp.153-173.

Louviere, J., D.A. Hensher and J. Swait (2000). *Stated choice methods: Analysis and applications*. Cambridge University Press, New York, United States.

Nijland, L, T.A. Arentze, A. Borgers and H.J.P. Timmermans (2006). Modelling complex activity-travel scheduling decisions: procedure for the simultaneous estimation of activity generation and duration functions. Paper presented at the 11th International Conference on Travel Behaviour Research, Kyoto, Japan.

Nijland, L, T.A. Arentze, A. Borgers and H.J.P. Timmermans (2007). Activity-travel rescheduling decisions: Estimating parameters of response patterns. Proceedings of the

86th Annual Meeting of the Transportation Research Board, Washington, DC, United States.

Roorda, M. and B. Andre (2007). A stated adaptation survey of activity rescheduling: Empirical and preliminary model results. Proceedings of the 86th Annual Meeting of the Transportation Research Board, Washington, DC, United States.

Timmermans, H.J.P., T.A. Arentze and C.H. Joh (2001). Modelling the effects of anticipated time pressure on the execution of activity programs, *Transportation Research Record* **1752**, pp.8-15.

Train, K. and W.W. Wilson (2008). Estimation on stated-preference experiments constructed from revealed-preference choices. *Transportation Research Part B* **42**, pp.191-203.

van Bladel, K., T. Bellemans, G. Wets, T.A. Arentze and H.J.P. Timmermans (2006). Fitting S-shaped activity utility functions based on stated-preference data. Paper presented at the 11th International Conference on Travel Behaviour Research, Kyoto, Japan.

Van der Waerden, P., A. Borgers, H.J.P. Timmermans and M. Bérénos (2006). Order effects in stated choice experiments: A study of transport mode choice decisions. Proceedings of the 85th Annual Meeting of the Transportation Research Board, Washington, DC, United States.

Wang, D., J. Li and H.J.P. Timmermans (2003). Measuring stated bifurcation points in traveler adaptation processes: The example of long distance transport mode choice. Proceedings of the 82th Annual Meeting of the Transportation Research Board, Washington, DC, United States.