REALISM AND ADAPTATION IN DESIGNING HYPOTHETICAL TRAVEL CHOICE CONCEPTS

By Mark Bradley*

1. INTRODUCTION

Stated preference data are collected in the form of preferences or choices between hypothetical alternatives. The data are usually collected by means of an experimental design, which represents each alternative in terms of pre-specified levels across a set of attributes (for example, price, speed, comfort). The preferences or choices from among the alternatives are then used to infer the utility associated with each level of each attribute. This modelling approach is analogous to inferring utility functions from choices among actual alternatives, or revealed preferences. Both types of data are used to estimate models to evaluate new combinations of attributes which may result from a change in product or service design, or from a policy decision or a shift in an external factor (for example, the price of petrol).

An advantage of using hypothetical alternatives is that one can directly study and evaluate products, services or situations which are qualitatively different from those that are commonly encountered. A second major advantage is that each individual can respond to several different hypothetical choice situations; this increases the efficiency of data collection, and often provides enough data to estimate utility functions for each individual. It is therefore not surprising that many of the developments in conjoint techniques have taken place in the field of market research. Green and Srinivasan (1978) provide a comprehensive review of these developments — many of which are still relevant today.

Stated preference methods have also been gaining acceptance in the transport field, largely for the reasons already mentioned. As the focus in transport is often on forecasting demand levels rather than on assessing travellers' preferences or satisfaction, stated preference applications in transport have tended towards the discrete choice methods now commonly used in revealed preference modelling of travel demand (Ben Akiva and Lerman, 1985). An example is the choice simulation technique described by Louviere and Woodworth (1983).

* Hague Consulting Group, The Netherlands. The two projects described in section 4 of this paper were carried out in conjunction with the Transport Studies Unit, Oxford University. The author acknowledges the collaboration of Dr. Peter Jones, of the TSU, and Elizabeth Ampt, of Ampt Applied Research.
Another important distinction between analysis of travel demand and most other areas of consumer research is the high importance of factors such as time constraints, constraints on the availability of alternatives (for example, no car or no bus service), and linkages between individuals (for example, having to get a lift from a certain person at a certain time). These factors tend to limit or influence the choices which people make in actual choice contexts. As the emphasis in transport research is often on predicting choices, there is serious concern that if data from hypothetical situations are used these constraints or influences may not be adequately represented, and that, as a result, the stated choices will not match the choices that would be made in actual situations. This issue — the external validity of stated preference data as a function of the realism and relevance of hypothetical travel choices — is the main focus of this paper.

Internal and external validity

The steps in a typical stated preference study are as follows:

1. Determine what is to be predicted or evaluated.
2. Specify a hypothetical travel context from which to collect data.
3. Determine the target sample for the survey and the method of recruitment.
4. Design the choice or preference alternatives to be presented.
5. Decide on a method for measuring preferences or choices in the survey.
6. Develop the method for presenting the experiment to the respondents.
7. Administer the survey experiment.
8. Estimate random utility models from the stated preferences or choices.
9. Assess the validity of the model predictions, as nearly as possible.
10. Apply the models for evaluation or forecasting.

Most writers who discuss the validity of conjoint analysis methods deal with internal validity. This encompasses steps 4 to 8 above — the statistical design of the survey experiment, the measurement of responses during the survey, and the estimation of preference or choice models from those responses.

Tests of internal validity have included comparisons of different methods of experimental design, different scales for the measurement of responses, different methods of presenting the choice alternatives, different model estimation algorithms, and different levels of aggregation in model estimation. One criterion for judging internal validity is statistical efficiency — the level of significance and model fit achieved. Another is predictive reliability — the success of the estimated model in predicting independent responses to the same hypothetical choice task, either from the same individuals or from a different survey sample. A third, less common, criterion is replicability — the probability that individuals who face the same conjoint tasks twice will give identical responses on both occasions.

As a result of such tests (reported mainly in the marketing science literature), many of the internal properties of stated preference techniques have been documented, and their application generally gives little cause for concern as long as the researcher is aware of the statistical assumptions underlying each technique (see, for instance, the papers by Bates and Louviere in this issue).
REALISM AND ADAPTATION IN TRAVEL CHOICE CONCEPTS

M. Bradley

As mentioned earlier, a more common cause for concern is external validity — the relevance of predictions from stated preference models to choices made in actual situations. Many of the study aspects which influence external validity are in the first and last stages (steps 1 to 3 and 9 to 10 listed above) — the identification of the survey context and sample, and the method by which the results are applied and, if possible, tested. One may notice that only the word “hypothetical” distinguishes these steps from a similar list which might be made for a revealed preference study. Thus, when these issues are discussed in section 2, it is often useful to compare the validation of stated and revealed preference models.

There are also a number of aspects unique to stated preference methods — steps 4 to 8 above — which can be expected to have a direct influence on the external validity of the results. Many recent applications in transport have aimed at making the techniques more realistic in particular choice contexts, and more adaptable to the circumstances of individual respondents. These aspects and developments are discussed in section 3. The discussion refers to several ways of making stated preference experiments more “intelligent” — adaptable to the perceptions and constraints of each respondent.

Section 4 discusses a potentially important development in stated preference methods — the use of microcomputers to design and administer the hypothetical choices in the survey setting. First, comes a description of their use in studies familiar to the author. This is followed by a more general discussion of the advantages, disadvantages and further potential of the computer-based approach.

Section 5 concludes the paper with a brief summary.

2. THE DIMENSIONS OF EXTERNAL VALIDITY

2.1 The objective of the study

The type of external validation which is or should be done for models of travel demand depends a great deal on the subsequent use of the results. The uses to which the models might be put include:

Assessment of preferences for a range of options (that is, to select a policy);

Assessment of the strength of preferences over a range of options (for example, for a social cost-benefit study);

Prediction of shifts in demand from the current situation (for example, to study the profitability of service changes); and

Prediction of absolute demand levels over a range of options (for example, for use in a regional planning model).

For each of these uses, the models may either be applied only to the individuals represented by the survey sample, or may be required to be transferable to a wider geographic or demographic population. Another important aspect of the sample is current behaviour with regard to the travel choice being modelled. This issue is discussed below; it is important in the design of the experimental context and the selection of the sample.
2.2 Choice of a travel context

Since one knows the travel contexts to which eventual models will be applied, specifying the hypothetical context from which to measure choices or preferences may seem a straightforward decision. For example, in order to predict the market penetration of a new mode of travel, one would survey the users of the existing modes, ask them which mode they would choose if the new mode was available, and then repeat the question for various permutations of the mode characteristics. In many cases, however, the design of a hypothetical context is not so simple, and it may be a major influence on the validity of the research.

One issue is whether the situation described is realistic and relevant to the respondents. Consider, for example, the “real time” experiment to measure the value of time losses, in which travellers were offered varying sums of money to miss the train they were waiting for and to take the next one instead (Hauer and Greenough, 1982). This approach made the consequences of the choice less hypothetical, but it was not a choice with which people were familiar. Thus it is not certain whether the responses to this unique experiment would reflect the respondents’ reactions to a sustained policy of reduced fare for increased waiting times, or even whether it would predict their reactions to the same experiment done a second time.

For more hypothetical experiments, the same questions apply. People can rarely respond completely from experience, particularly when presented with new types of travel. The problems can be partially overcome by defining and presenting the alternatives in terms that are familiar to the respondents and relate to their past experience. These issues are discussed further in sections 3 and 4.

Sometimes the objectives of a study are to measure the importance of various service attributes, such as travel time or cost, rather than to predict how people will behave in a certain situation. In such a case the hypothetical context is less obvious, and the issues of experience and familiarity become more important.

An example arises in the question whether to use within-mode or across-mode choices. Offering only variations on a single mode simplifies the context, so that many additional factors regarding the availability and attractiveness of other modes do not need to be considered. For example, if one is only interested in measuring relative preferences for improvements in public transport, other modes can be left outside the experiment. Many regular users of a mode may be more used to selecting from among routes or services within that mode than between modes. For others, however, the opposite may be true, and some flexibility in the experiment would be desirable to adapt to such differences in experience.

To complicate matters, suppose it is also required to predict changes in demand. In the example of an improvement in public transport, it might be possible to include the best alternative mode as a choice option. If, however, the sample consists mainly of current users, and the within-mode-options involve only improvements or small changes, the alternative mode will very rarely be chosen! Any changes in demand are more likely to arise from changes in the frequency of travel than from mode switching. Again, some flexibility would be useful to adapt the experimental context to different existing situations.

Another question is whether the context is made to represent a single hypo-
theoretical day or journey, or to represent more general behaviour. In many studies, one would like to strike a balance between providing a great deal of contextual detail to familiarise the choice context to the specific respondent, and being able to apply the responses to a wider population and more general circumstances.

For example, the type of detail often useful might include time constraints, interactions with other travellers or household members, and external circumstances such as weather or the amount of traffic. To mention and record these items, it is usually necessary to refer to a specific day or journey, with the risk of selecting a very atypical case. A good compromise seems to be to make sure that all important factors are recorded, and that the day or journey which is reflected most heavily in the choice context is a fairly typical one (if such a thing exists). Later, when stating their preferences, respondents can be asked whether this would also reflect their behaviour on a more regular basis. Also note that, if one has a large and representative sample, basing the experiment on an atypical instance may be valid — provided that the hypothetical choice context remains relevant.

2.3 Sampling strategy

In the preceding discussion distinctions were raised between “current users” and “non-users”, and between “typical” and “atypical” journeys. The issues of sample composition and sampling strategy are thus important in the design of an appropriate stated preference context. As in revealed preference studies, the basic requirement is to have a sufficiently large and representative sample. Because conjoint experiments are statistically efficient, samples are typically smaller than for comparable revealed preference surveys. There are other important differences as well.

In stated preference studies it is possible to base a model only on individuals of a certain type — for example, frequent users of a public transport service, who are asked to make within-mode comparisons of service improvements. Use of such a self-selective or choice-based sample can ensure that the survey context is relevant to all respondents, but of course can provide little evidence about the behaviour of other groups — for example, new users who might be attracted by the same service improvements.

Where forecasts of demand are needed, it is often necessary to survey many types of individuals in order to obtain results which are transferable to the entire population. In revealed preference models, if the proportion of the sample currently choosing a certain option is over-represented, or “enriched” in comparison with the general population, a simple correction can usually be made in model estimation to avoid statistical bias. In stated preference data, however, there may be additional biases arising from the way different types of travellers perceive or interpret the choice context.

Returning to the example of an across-mode experiment, there is a chance that respondents will presume that their current mode is superior, and not consider the full information given for the other alternatives. On the other hand, some respondents might indicate that they would switch from their current mode, whereas, in reality, certain constraints or habits would prevent them from doing
so. Neither of these possible sources of bias is well understood, but it is at least encouraging that they work in opposite directions, and that making the choice contexts as explicit, realistic and interesting as possible should reduce the bias in any experiment. In some cases, it may be possible to design different experiments for different groups of users, or to make one experiment adaptable to their varying perspectives.

2.4 Model application and validation

In any discussion of the external validity of stated preference data, a key issue is the way in which the results are applied. The case when one is modelling preferences across a number of alternatives, or the relative importance of various attributes, is quite different from the situation where one is attempting to produce forecasts of choice behaviour.

In the former case, there are no objective data from observable behaviour to validate against, so one is left to rely mainly on the internal validity of the technique. Here, it is important to see that the experimental design and sampling procedures are carried out properly in the statistical sense, and that the choice context and alternatives are clear to the respondents.

In other words, the validity of the application requires that the preference functions have been estimated reliably for the chosen sample. This has been tested in practice by also collecting data on actual choices for the sample and estimating analogous revealed preference models (Louviere et al., 1981; Bates, 1986; Wardman, 1986 and in this issue). Such comparisons have generally yielded similar models, but with more precise estimates from the stated preference data. This precision may be due to the efficiency of using experimental designs, or to the relative absence of external unobserved factors — such as choice constraints or incomplete information — which are a more likely source of error in choice data from actual situations.

Errors arise often from incomplete information. The burden falls on the stated preference researcher to take account of the full range of items which matter in real-world situations — including choice constraints, lack of information, and interactions with other travellers. Of course, many excluded factors can be “absorbed” in estimation if we use alternative-specific constants, as is often done in revealed preference modelling. Stated preference surveys, however, often present a good opportunity to use such information — both in presenting a more realistic choice experiment and in developing a more complete forecasting model. We will return to these issues in sections 3 and 4.

In forecasting studies, models based on stated preference data are further calibrated against revealed preference samples from similar contexts (see Kocur et al., 1982). Usually only two parameters are estimated — a scale coefficient on the utility calculated from the SP model, and an additional constant; this makes it possible to use a much smaller revealed preference sample than would be needed to estimate a full model. Tests of this nature have given quite encouraging results — often with scale coefficients not significantly different from unity.

As forecasting models are eventually relied upon to predict changes which will result from new services or policies, the strongest test of external validity is the
ability to forecast market shifts correctly. Here, stated preferences would seem to have an advantage over revealed preferences. Instead of inferring the effects of new policies or travel options from cross-sectional evidence, one can study them more directly in a choice experiment. Nevertheless, for this kind of validation the survey sample must be followed and their choices and alternatives must be recorded both before and after a service or policy is actually introduced.

It is perhaps not surprising that before-and-after validation has not been common in practice. Finding an appropriate study context and time scale can be difficult; it can be difficult also to procure funding for continued research after the policy has already been implemented. Also, unless the sample size or the predicted changes are quite substantial, other random influences over the intervening period could cause quite a large error in the validation. Still, as stated preference applications become more common, and as data collection and follow-up methods become more efficient, there appears to be a great potential for research of this sort.

In practice, a less formal type of validation is also common. Models based on stated preference data are used to make forecasts, and the results are compared with “reference” values such as elasticities of fare or petrol price. Such checks are useful, but it is not always clear how much importance should be placed on them — particularly in view of the wide range of choice contexts considered. Thus, a larger body of evidence from formal validation studies would be very useful in assessing the results of each new stated preference study.

There are, in addition, a number of issues in designing a choice simulation experiment which, one feels intuitively, should have an influence on how closely the stated choices reflect what might occur in reality. They are evident more from the comments and reactions of interviewers and respondents than from any formal validation. These issues, discussed in the following section, generally relate to how far the specific circumstances and beliefs of the respondent are taken into account in creating the choice situations, and how these situations are in turn presented to the respondent.

3. ISSUES IN SURVEY DESIGN

3.1 Selection of choice attributes and levels

The first task in “filling in” the choice context is to select the set of variables which are likely to be most important. The set normally includes those which are commonly found to be important (cost, walk time, in-vehicle time, etc.), plus any instrumental factors in the policies or scenarios to be studied. Sometimes, however, the set of factors important to the respondent is not the same as might be deduced from most existing models.

One such example is waiting time. It may be difficult for a respondent to imagine a scenario where the usual waiting time for a bus is increased to 10 minutes. The change is more believable if put in terms of a decrease in the frequency or reliability of the service — and these measures are more relevant for policy analysis. In general, any variable which might be expected to vary
considerably from day to day might better be broken down to isolate the reasons for the variability. In the United Kingdom Value of Time Study (Bradley et al., 1986), the public transport experiments included both service frequency and delay/reliability variables, while the car driver experiments distinguished time spent in free flow and in congested conditions.

Another point is that different factors may be important to different people, so that, for some, important attributes are missing from the experiment. An example might be the lack of space on buses to ride with baby carriages. Often it is sufficient just to record the point and continue with the experiment. If, on the other hand, something is missing which is very important to the respondent and which affects the credibility of the other variables — for example, reliability of service — the results may be less valid. When demand changes are to be predicted, rather than just trade-offs, information about other important variables is even more vital.

If the survey format is interactive, as in home or on-board interviews, it may be possible to include one or more “free” attributes, to be mentioned by the respondent and added into the experiment by the interviewer. As such attributes could be different for each individual, this is most feasible for qualitative variables which can be treated on two levels — the current one and a more acceptable one. A potential problem with this approach, besides the complexity for the interviewer, is that the set of attributes is no longer the same across the sample. In practice, variables not considered by individual respondents could be assigned an importance ranging from none at all to that of the least important of the included variables, and one could test the sensitivity of forecasts to this assumption.

Customisation can also be important in selecting the levels of the attributes included in the experiment. As a start, an option with all attributes at their actual levels can be included in the choice set in order to put the other options into perspective. Going further, the levels in the experiment can be based on the actual reported values — for example, 10 per cent higher than at present and 10 per cent lower than at present. Both these approaches can be followed in virtually any survey format, as long as the researcher has information on the actual values in due time to prepare the choice options for the respondents.

Ranges and degrees of variation

The realistic ranges of the attributes may also vary across the sample. In an interactive survey format, these ranges can be identified at both the high and the low ends, and the levels offered to the respondent can be restricted accordingly. It is most important to detect the limit of realism for improvements. For deterioration, it is likely that the levels will become unacceptable before they become unrealistic. Unacceptable levels could be included in the experiment without affecting the validity of the results, but this is probably not true of unrealistic levels. Still, it is most efficient to focus on levels at which respondents are willing to trade for improvements in other factors.

A final issue here is the units in which the levels are presented. Discrete qualitative variables are usually described in two or three levels — the current situation and some policy change of interest, in one or both directions. Even quantitative variables, however, have some discreteness as perceived by the
REALISM AND ADAPTATION IN TRAVEL CHOICE CONCEPTS

M. Bradley

respondent. Variables of a mainly spatial character, such as walk time to bus stop, may signify specific locations to the respondent — with differences of at least a few minutes between them. Also, the fact that most people report actual times and costs in units of 5 suggests that they do not readily perceive changes much smaller than this, particularly in the face of day-to-day variability. It would not be valid to offer hypothetical changes smaller than those that can be perceived in specific choice situations (even though people may actually respond to smaller shifts which persist over a long period).

In dealing with the problems mentioned above, initial qualitative research is a very important first step to determine which level, attribute, and choice context descriptions will be relevant and understandable to the majority of the target sample. Survey piloting provides another important check. Additional steps to account for variations between individuals can also be incorporated, as described in section 4.

**Complexity and competitiveness**

The concepts of experimental design are generally used to arrange the attribute levels into a series of choice situations for choosing, ranking, or rating. This is usually done by means of a factorial design which optimises the statistical independence of the alternatives (Kocur et al., 1982). One to three sets of choice tasks, each requiring 8 to 16 responses, is a standard size.

A key issue in creating a design is *complexity*. Studies have shown that people give the most reliable responses when assessing changes in only two or three factors simultaneously (Huber and Hansen, 1986). More complex choice tasks may actually lead people to use simpler, non-trading choice rules, such as considering only one attribute at a time lexicographically (Johnson and Meyer, 1984). This implies a strategy of starting with the simplest tasks, varying only two factors, and then slowly graduating to more complex tasks with a maximum of three or four varying attributes. Often the first few “warm up” choices are not used in estimation.

The complexity of the response task can be expected to influence the amount of error present in the measured responses. Survey piloting and subsequent debriefing can uncover general difficulties, and checks can be incorporated in the survey to identify individuals with poor understanding (for example, a choice between two alternatives where one is better than the other along every dimension). A truly interactive survey procedure could identify such difficulties and immediately probe further or provide further instruction.

A second key issue is *competitiveness*. How close in attractiveness should the choice alternatives be? Very dominant options are usually to be avoided (except as a check), because the strength of dominance cannot be identified with non-metric scales such as ranks, and they tend to “push” metric responses (ratings) to the other end of the scale, and thus decrease the response variation relative to the measurement error. Also, the optimal design for statistical efficiency is one which is concentrated where the slope of the utility function is most sensitive to changes in the attribute levels (Gunn, 1981), and this is most likely in competitive situations.

At a more subjective level, Huber and Hansen (1986) found that choices
between closely competitive options were more interesting for the respondents and more reliably answered, even though somewhat more difficult.

A difficulty in design is to determine before the study takes place which choice options will be competitive for most people. Knowledge of the order of preference for the levels within each attribute will usually enable one to avoid extremely dominant options, and survey piloting can provide additional guidance.\(^1\) Even more useful is an "intelligent" design, as described in section 4.

### 3.2 Design of response measurement scales

The type of response asked for is important, not only on a statistical level (see the paper by Louviere in this issue), but also in terms of realism and complexity. The most common types are choices between options, rankings across options, and scoring of each option on a metric scale. Simple choices have been advocated as being more analogous to decisions made and data collected in actual situations, and more familiar tasks for the respondents (Louviere and Woodworth, 1983). Ranking and scaling data tend to be more efficient for estimating individual model parameters, but are usually more difficult for the respondent, and require more assumptions to be made before choices can be predicted.

In practice, it is common to use ratings or rankings as the main data source for estimation, and to include a number of simple choice tasks at the end, to be used for further calibration or internal validation of the estimated models. Another approach to increase the information from choice data is to ask first for the choice among a set of options, and then elicit a scaled response for the strength, or certainty, of that preference. This second response could be measured on a simple scale such as definitely/probably/unsure, or could be a more demanding task such as identifying the point where the preferred option would change — as in transfer price questions (see, for example, Gunn, 1981). In combination with a ranking or choice task, it is possible to ask for an additional set of responses giving the frequency or likelihood of use for each option. This approach gives metric measures of demand which can be used to provide forecasts or elasticities.

### 3.3 Presentation of choice tasks

The final issues in this section are more speculative than those mentioned earlier, as presentation is more in the realm of art than of science. Certain types of attributes can often be perceived more clearly if presented in forms other than verbal descriptions. For example, qualitative factors can often be depicted by drawings or photographs; spatially-based attributes can be displayed in maps; time constraints or event timings can be laid out on timelines or timetables; and factors showing variability or reliability could be shown graphically as distributions. The presentation format can also be used to clarify or familiarise the whole context of the choice task. An example here is laying out options for changes in public transport fares and services in the form of brochures, timetables, and/or tickets which passengers can select (Sheldon and Steer, 1982).

\(^1\) See also, in this context, the paper by Fowkes and Wardman in this issue.
REALISM AND ADAPTATION IN TRAVEL CHOICE CONCEPTS

M. Bradley

It is possible that the form of presentation will influence the results, and thus the validity, of the experiment. A simple example is that much effort may have been put into depicting certain variables, such as qualitative aspects, while relatively little emphasis has been given to the more straightforward attributes of times and costs. The balance in the presentation could in turn affect the relative attention paid by the respondent to evaluating the attributes.

A more complex situation arises when two or more of the attributes may be somehow related to each other. This often occurs when a change in one qualitative attribute also implies other qualitative changes — which may or may not be attributes included in the experiment. In an experiment investigating route choice for bicyclists (Bradley and Bovy, 1986), there was evidence that a certain type of cycle path — one of the attributes — alone implied certain levels of traffic congestion and surface quality, even though these two attributes were also included in the experiment. This type of interacting perception of attributes can clearly be influenced by how they are defined and presented (though the selection of the variables and the experimental design are also important). It would seem especially difficult to avoid such interactions when using drawings or photographs — which are often easier to understand than words but cannot be as easily focussed on particular qualities or concepts.

In combination with all the factors mentioned earlier, presentation may determine how well the experiment reflects real choice situations, and thus may influence the external validity. Because the number of possibilities is virtually infinite, it is not possible to assess specific approaches, but only to say that the presentation should be as concise, realistic, and balanced across the choice attributes as possible. The next section discusses a particular form of presentation — the portable micro-computer.

4. CHOICE EXPERIMENTS BY COMPUTER

The preceding discussion has shown the advantages of adapting choice experiments to individual circumstances. This can be done in defining the choice context, selecting the choice attributes and their levels, creating the choice tasks, and presenting them to the respondent. This degree of flexibility, however, is more than even the most experienced interviewer could manage while still providing statistically relevant data. An ever more attractive solution is to use portable computers to help the interviewer, in the course of the experiment, and perhaps even to present the choice scenarios. The flexibility of this technology must of course be weighed along with its other possible advantages and disadvantages. This section helps to assess the balance by looking at stated preference surveys in which personal computers have been used with an emphasis on two specific studies of travel demand in which the author has recently been involved.

4.1 Urban bus passengers

The first example was a study of preferences of urban bus passengers for the Transport and Road Research Laboratory in the UK. In the initial phase of this

131
study (Ampt et al., 1986), the interviewers were equipped with a small portable computer as an aid in customising the levels of fare, frequency, and walk, wait and travel times, to encompass those reported by the respondents for their actual bus journeys. A written questionnaire form was used to prompt for these actual levels; the information was immediately entered into the computer, which then generated a set of choice options, using a fairly simple algorithm. The interviewer then copied these choice options on to cards for the respondent, highlighting the attributes for which the options differed. Maps were used to clarify variables such as walk and in-vehicle times. Attributes such as variability in wait time and in-vehicle time, and timetable options for a given service frequency, were also included, but no special presentation methods were used.

By the second phase of this study (Bradley et al., 1987), the technology and prices of portable computers had improved a great deal. Now, it was feasible to use the computer to administer the entire survey, retaining only a few “human touches” such as a bus route map and a sheet for recording additional comments. This made the task somewhat easier for the interviewer, allowing automatic branching of questions and automatic changes to the question prompts, using answers to previous questions. The data were also automatically coded and stored; this saved work after the interviews.

In addition, the improving graphical capabilities of portable computers gave better presentation of difficult variables. For example, colour shadings were used to portray service timings, time constraints, and possible variations in wait and journey times, along a “timeline” showing the relevant times of the day. This made it clear when certain options would entail a risk of violating time constraints, or when the timetable for a certain option fitted best with the person’s other activities. Differences between the choice options were also automatically highlighted by colours. Most respondents reported that they understood these displays after some explanation by the interviewer, and that they used the information on timings and variability in making their choices. An example of the choice screen is shown in Figure 1.

A more powerful portable computer also made possible a more adaptable method for creating the choice attributes and alternatives. Respondents were able to nominate attributes not included in the original set — usually qualitative aspects of the service, with which they were not satisfied. The various attributes were then ranked in importance, and this information was used to create the initial preference weights for an iterative experimental design algorithm.

The algorithm used followed the example of adaptive conjoint analysis (Johnson, 1985). The computer is used in the course of the interview to create optimal choice tasks — those that will be most competitive and interesting to the respondent. After each new response (or set of responses) the estimates of the preference weights for each attribute are updated, using least-squares regression. The new weights are input to another algorithm, which attempts to create new choice options which will be competitive and will provide efficient data for estimation. This iterative process continues either till the preference function for the individual has been estimated to within a specified confidence interval, or till a specified maximum number of choice tasks has been completed.

Use of a least-squares technique requires that the preferences between the
choice alternatives be assessed by the respondent on a metric scale. The data can then be re-analysed after the interview by non-metric estimation methods, such as logit, which generally require too many observations and computations to be used inside the survey program.

This particular application was designed to be as adaptable as possible in creating realistic choice contexts for urban bus service and in presenting them in a clear and interesting manner to the respondent. Respondents were also asked about the mode which they would use if not travelling by bus, and this option was included in the experiment to provide choice information for modelling mode or frequency changes.
4.2 Peak-spreading policies

The second example is a technique developed for the South Australia Department of Transport in Adelaide, in order to predict responses to peak-spreading policies such as increasing peak-period parking charges or changing school hours (Jones et al., 1986, 1987). Such policies may involve retiming various travel activities, and may affect several members of the respondent’s household. Conventional experimental design approaches, even in their most flexible form, are not capable of dealing with such complex household travel patterns and time constraints. In this study the approach used was that each relevant household member provided the times of all travel and non-travel activities for a selected day. These were then portrayed in full-colour graphics, to lay out the activity patterns in both timetable and timeline formats, including details of times when people travel together. An example of the screen is shown in Figure 2.

With this approach, instead of creating alternatives from a pre-specified set of attributes, the interviewers first asked the respondents to state how their existing travel patterns would have to change to accommodate a variety of travel options. These options might include changing modes, travelling earlier, or travelling later. The survey program detected automatically when these changes might violate time constraints or affect other household members; this was intended to ensure that the stated options were well thought out and were feasible, given the activity patterns of the household. Then, when the alternatives had been considered by all household participants, they were ranked against each other under various policy scenarios.

4.3 Advantages of computers in interviewing

Both survey approaches described above have been administered by market research staff with no specialist knowledge of transport research. From a practical standpoint, the advantages of computer-based interviewing are:

(1) an interesting and flexible presentation format;
(2) a format which is consistent across interviewers and respondents;
(3) automatic question branching and prompting;
(4) automatic data coding and storage;
(5) the ability to incorporate checks to avoid inconsistent or wrongly entered answers.

There appear to be very few disadvantages, apart from the initial capital cost and training involved in using the computers. For these examples, one day of interviewer training was required, and this appeared adequate. Automatic interviewer “help” commands were available in both computer programs to handle any on-site difficulties. For further interviews with the same staff, it is likely that even less training would be required than for conventional surveys, as the computer is programmed to do much of the work automatically. As it also avoids many of the costs of survey production, preparation, and coding, it seems likely that over a number of surveys the overall field costs would not increase substantially.
From a methodological standpoint, computer-based interviewing opens up many new possibilities for creating realistic stated preference experiments. In the two examples given, we have tested the limits of the approach by incorporating adaptable attribute sets, attribute levels, and experimental designs — as well as graphical presentation. The price to be paid for this flexibility is the increase in programming and survey piloting required for each new study, and the need to use more complex disaggregate model estimation techniques to handle choice situations which differ by respondent (these are the same techniques that are used with disaggregate data in revealed preference).

Computer-based interviewing may also be useful as a method of presenting
more conventional stated preference experiments. The potential improvement in respondents’ comprehension and response accuracy would seem to warrant a shift towards automated data collection, though comparative studies would be useful to confirm this. One would expect the relative advantage of the computer to increase as the experiment becomes more complex and adaptive. As software and computer-based survey techniques develop, it is conceivable that the conventional stated preference methodology of the future will no longer be within the scope of manual survey techniques.

As we depart progressively further from conventional conjoint techniques, we can find additional roles for computer-based surveys in studying travel demand. One such role is to study the dynamics of the choice process itself. Smith et al., (1984) got respondents to search for a residence, with the computer to simulate the information available in the housing market. The resulting data were used to estimate a heuristic rule-based model of the housing search process.

In another example, Mahmassani and Chang (1985) studied day-to-day adjustments in departure time for the work trip, simulating the resulting congestion and delays. If the respondents in such experiments could receive information and give responses remotely by computer, it would become more feasible to contact them repeatedly over a number of days or months. As research into dynamic choice processes gains momentum, such possibilities may become important.

5. SUMMARY AND CONCLUSIONS

Stated preference methods are now commonly used for analysing transport policy, but some remain sceptical about their external validity in relation to actual choice situations. This paper has discussed some aspects of designing a stated preference experiment which would be expected to influence the validity of the choice responses. A recurrent theme of the discussion is that it is important to be able to adapt the hypothetical choice contexts so that they are realistic and relevant to individual respondents; it is also important to be able to present the choice tasks in a way which reflects the context being studied.

The portable computer provides a means of adapting experiments during each interview, and of presenting choice options clearly. We have discussed some of its possible advantages and disadvantages, as shown in the still limited number of examples of its use. It appears that computer-based interviewing is particularly suitable for complex experiments which take into account differences between individual respondents, but that it may also improve the quality of more conventional stated preference data.

REFERENCES


136
REALISM AND ADAPTATION IN TRAVEL CHOICE CONCEPTS

M. Bradley


