Journal of Choice Modelling, 2(2), pp 216-236
www.jocm.org.uk

Factors affecting the amount of effort expended in responding to questions in behavioural choice experiments

Peter Bonsall[1,*] and Bill Lythgoe[1,†]

[1] Institute for Transport Studies, University of Leeds, Leeds, LS2 9JT, UK

Received 7 May 2009, revised version received 21 August 2009, accepted 9 November 2009

Abstract

A brief review of theoretical and empirical literature on the effort expended in making choices is followed by the description of an experiment in which participants were asked to respond to a series of road charging scenarios. Participants were asked to estimate the charges payable, to indicate their confidence in that estimate, to say how difficult they had found it to understand the scenario, and to indicate whether they would change their behaviour if the charges were introduced. The time they took to answer each question was recorded and background data was collected on their personal characteristics and attitudes as well as on their attitude to decision-making. Analysis of the resulting data shows that the time taken to estimate the charge and decide on a response varied not only with the characteristics of the scenario and with the order of presentation (indicating a learning, or fatigue, effect), but also with personal characteristics (notably age, educational attainment and self-reported decision-making style). The time taken to estimate a charge was significantly (and positively) related to the degree of difficulty reported and, although additional time devoted to making an estimate generally resulted in more accurate estimates, for the simplest scenarios, increased time taken was associated with reduced accuracy.

These findings, particularly on the existence of groups with very different levels of motivation to make careful assessment of costs and benefits, and on the factors influencing the time taken to respond to questions, have implications for our understanding of real-world decision-making and for the way that we should collect and categorise choice data and model the decision-making process. The use of simple questions designed to elicit participants' attitudes to decision-making is seen to show great promise in this context.

Keywords: Latency, Effort, Complexity, Accuracy, Error, Need for Cognition, Tolerance of Ambiguity

* Corresponding author, T: +44-1133435335, F: +44-1133435334, p.w.bonsall@its.leeds.ac.uk
† T: +44-11333438768, F: + 44-11333435334, w.f.lythgoe@its.leeds.ac.uk

Licensed under a Creative Commons Attribution-Non-Commercial 2.0 UK: England & Wales License
http://creativecommons.org/licenses/by-nc/2.0/uk/

ISSN 1755-5345

1 Introduction

1.1 Objective and Scope

This paper reports results from a controlled experiment during which the time taken by participants to answer each question in an extended questionnaire was logged. It seeks to explore the factors influencing these latency values (and related variables such as accuracy achieved and difficulty reported) and, in so doing, to throw new light on issues which have arisen in previous work on decision latencies, task performance, attitude to decision-making, learning effects, and respondent fatigue.

1.2 Background – Previous Exploration of Latencies

Expenditure of effort is, of course, difficult to measure directly and many authors, have used decision latency (time taken to make a decision) as a proxy. There is a substantial literature on the time and/or effort expended by people when making choices. This literature includes both theoretical considerations and empirical results from a range of experiments.

Theoretical considerations lead us to expect that the time taken to make choices generally increases with the complexity or difficulty of the task being performed. Numerous studies have confirmed this expectation (e.g. Wise and Cain (2000) find that the time taken to discriminate between stimuli varies inversely with the difference between them) and, since decision time latencies are now produced as a matter of course by widely used software, it is becoming commonplace to use decision latencies as a qualitative measure of task complexity.

However, the relationship between task complexity and decision latency is not linear. There is well established theory (e.g. Newall and Simon 1972) and empirical evidence to suggest that, if a task is perceived as too difficult, some individuals will abandon any attempt to achieve a "full" resolution of the task and will instead adopt a simplified decision making procedure which might sometimes result in less time being taken to make the most complicated choices (e.g. Pollay 1970a,b; Onken et al. 1985).

Bettman *et al*. (1990) examined the time taken by subjects to undertake a range of tasks and compared them with subjects' own assessment of effort expended. They concluded that the time taken, and the estimate of effort, reflected the number and type of elementary information processes required to complete the task. They noted that different individuals required different amounts of time, and reported more effort, to complete different types of elementary information processes and that the individuals who found a given process (e.g. multiplication) particularly difficult were more likely to resort to heuristic solutions.

The time taken to make a choice can be expected to vary between individuals depending on their personal decision-making styles. It will also reflect circumstances such as the extent of any distraction, the time pressure they are under and their current mental and motivational state. For example, Ratcliff and Rouder (1998) have shown that, in experimental conditions, latencies are reduced when instructions to subjects encourage a speedy response. Espinoza-Varas and Watson (1994) have shown that decision latencies vary with the decision criteria being employed.

For any given individual, the time taken to make a choice will crucially depend on their familiarity with the task. Generally, the time taken by a given individual will reduce as they become more familiar with it. This effect is noted in the context of

repeated choice experiments by authors such as Haaijer *et al*. (2000) and Rose and Black (2006) who note that decision latencies are much greater for the first few questions than for subsequent presentations of similar questions. This phenomenon is generally interpreted as a learning effect reflecting the fact that, with experience, people learn how to process the information more rapidly – for example by ignoring information which they perceive as irrelevant.

Haaijer *et al*. (2000) use a filtering process to normalise the latencies for individual responses to remove differences between participants and order-related differences. They demonstrate that models which use these normalised latencies are better able to predict participants' choices. Interestingly, they note that the latency parameter was negative in one experiment but positive in another (indicating that increased latency is associated with more systematic decisions in one experiment but with less systematic decisions in the other) – a fact whose significance we will revisit later in the paper.

Rose and Black (2006) extended the work of Haaijer *et al*. (2000) and noted that models could also be improved by including the effect that decision latencies have on parameter variances. They concluded that the high variances associated with the first few decisions were an indication that respondents had not yet learned how to make the decisions and so could be used as a basis for excluding data from responses to the first few questions asked.

2 The Experiment

Our experiment was not designed simply to explore influences on decision latencies. Its broader aim was to explore the relationship between participants' attitudes to a controversial policy intervention (road charging) and their engagement with the task of assessing the implications of complex examples of such policies and how they might respond to them[1]. This broader objective determined the structure and overall content of the questionnaire but left room to explore the factors affecting latencies for different types of question in different circumstances. The experimental protocol, including a list of questions asked, is provided in the Appendix 1 but, for convenience, is briefly summarised here.

There were three groups of questions. In the first group, participants were presented with descriptions of road charging schemes in an entirely hypothetical network (five such schemes were presented in randomised order, each with a different level of complexity – the most complex having charges which were different in each of three zones and two time periods). After each description participants were asked to estimate what the charge would be for a specified journey in that network, how certain they were about that estimate, and how easy it had been to understand the price structure.

In the second group of questions, participants were given descriptions of charging schemes which might be introduced in their home city (five such schemes were

[1] The work, described in Rößger et al. (2008), was conducted within a European Union funded project (DIFFERENT) which was exploring the optimal degree of complexity in infrastructure charges. It grew out of previous work by Bonsall *et al*. (2007a, b) which had shown that many people have a limited ability or motivation to respond complex price signals and a limited ability to assess the implications of complex pricing regimes, and work by Schade and Schlag (2003) which suggested that antipathy towards a policy instrument might reduce an individual's willingness to engage with the assessment of that instrument.

presented in randomised order, each accompanied by an appropriate map and each having a different level of complexity). After each description they were asked to estimate what such a scheme might cost them per month (assuming no change in behaviour), how confident they were of that estimate, how effective they thought such a scheme might be, whether they expected to be better or worse off if such a scheme were introduced, what response they thought they would be most likely to make, whether they thought the charge fair or unfair, whether they would approve its introduction, and how complicated they thought it had been to understand.

In the final group of questions, participants were asked to categorise their attitude to decision-making and to provide socio-economic data including gender, age, employment status, educational background and household income. Previous research into attitudes to decision-making has employed extensive batteries of questions to explore people's need for cognition, their need to evaluate and their tolerance of ambiguity (see, respectively: Cacioppo and Petty 1982; Blair *et al*. 1996; and Budner 1962 or Lamberton *et.al*. 2005). However, as is generally the case with field questionnaires, it would not have been practical to extend an already lengthy question by requiring participants to answer almost one hundred further questions. We sought instead to see whether the use of a single question drawn from each of the full batteries would provide a useful and practical approach by which to classify participants' attitudes to decision-making[2].

The original experiment was conducted in January 2008 and the sample was extended with further data collection in Spring 2009. Three methods of recruitment were used; (i) emails to staff at Leeds University and to the UOLDS[3] panel, (ii) posters at strategic locations around the university campus and (iii) direct contact with people at locations within the university.

The questionnaire was programmed for CASI (Computer Assisted Self-Interviewing) and could have been administered remotely on participants' own computers. However, in order to minimize external influences on the latencies, all data was collected in a controlled laboratory environment in which all the computers were all of the same speed and were displaying the briefing material in an identical way. The controlled environment also made it possible to ensure that participants were not interrupted during the task and that each participant had the same amount of preparation time.

Data was collected from 199 participants. Data from 6 participants was rejected as incomplete. The profile of participants is shown in Appendix 2. Although the sample was not designed to be representative of any specific population, it turns out to be fairly representative of drivers within the Leeds area – all be it with a bias towards students, people with degrees, females and people on higher incomes.

[2] We recognize, of course, that we cannot claim that our single questions are fully representative of the batteries from which they were drawn and we therefore do not claim that our categorization of attitudes to decision-making is the same as that used in previous literature in that field.

[3] The UOLDS (University of Leeds Driving Simulator) panel comprises drivers who have participated in driver behaviour studies in the recent past or who have registered their interest in doing so in forthcoming studies. The panellists were recruited from a combination of the social, personal and research networks of university staff and students and therefore are predominantly, though not exclusively, from the university community.

3 The Results

This section of our paper sets out the results in a logical order – moving from general observations, through tabulations and graphical displays, to modelling. The implications of these results are considered in the Discussion section.

An initial investigation of the data revealed some outlying latencies. At the low end, latencies recorded for one participant were such as to suggest that that individual had completed the experiment with barely enough time to read the questions; data for this participant were excluded from further analyses. At the high end, overall latencies for some 15 participants were substantially higher than for the rest; closer investigation showed that these high overall latencies were caused by extra-ordinarily high latencies for one or two questions but that the questions associated with these outlying latencies varied from case to case. We cannot know whether the long latencies were due to difficulties experienced with specific questions or to extraneous causes (e.g. needing to sneeze) but the laboratory supervisor's report indicated that some participants had asked for assistance during the experiment. Latencies in such cases will depend on how quickly the supervisor attended to them and are therefore unreliable. For some of the following analyses the original latency values have been truncated at the population mean plus three Standard Deviations for that question, in others they have been omitted.

Figures 1, 2 and 3 are based on data truncated as described. Figure 1 shows the latency and participant-assessed degree of difficulty for estimating the charge of each of the five hypothetical charging regimes. As explained in Appendix 1, the schemes differed in terms of their inherent complexity and were labelled A-E (A being the simplest and E the most complex).

As expected, the graph reveals a positive relationship between the complexity of a hypothetical charging regime and the time taken to estimate the cost of the specified journey (i.e. question *H1*) under that regime; participants required more than twice as long to calculate the charge for the most complex pricing scheme than for the simplest one. The graph also shows that the reported degree of difficulty in understanding the scheme (question *H3*) also rises with the objective degree of complexity. The relationship between the two items (latency and reported difficulty) was confirmed by a regression model which showed that 10% of the variation in reported difficulty was explained simply by the latency.

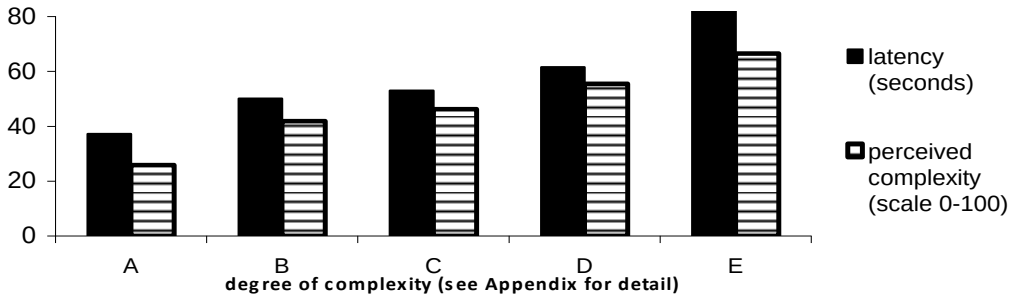

Figure 1: Mean time taken to estimate charge for, and reported difficulty understanding, different hypothetical schemes

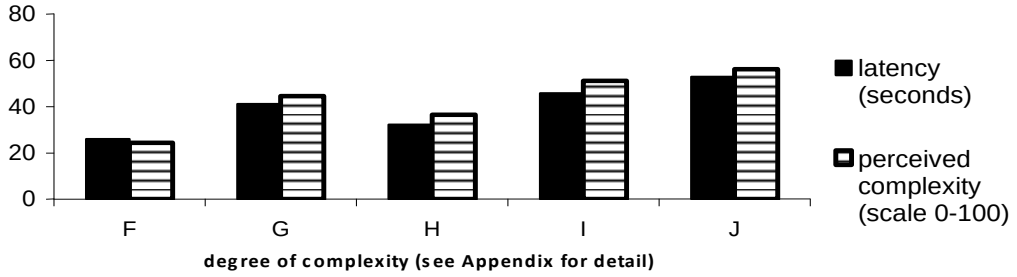

Figure 2: Mean time taken to estimate charge for, and reported difficulty understanding, different Leeds schemes

Figure 2 shows the equivalent relationships for the Leeds schemes; the striped bars again follow the solid bars indicating that reported difficulty is again related to latency. An interesting difference between Figures 1 and 2 is that, for the Leeds schemes (Figure 2), regime G has much higher latency, and is reported to be markedly more difficult to understand, than scheme H (Figure 1 showed no significant difference between schemes B and C). It seems that, for "real" schemes, spatial differentiation (having multiple zones) had much more effect on latencies than temporal differentiation (having multiple time periods).

Figure 3 shows, for the five hypothetical schemes, how latency reduces with order of presentation. All five schemes show the expected downward trend (indicative of a learning effect) but it is interesting to note that whereas, for the most complex scheme (E), the learning effect continues undiminished up to the fifth presentation, for the simpler schemes (A, B and C) the effect is much diminished after the second presentation.

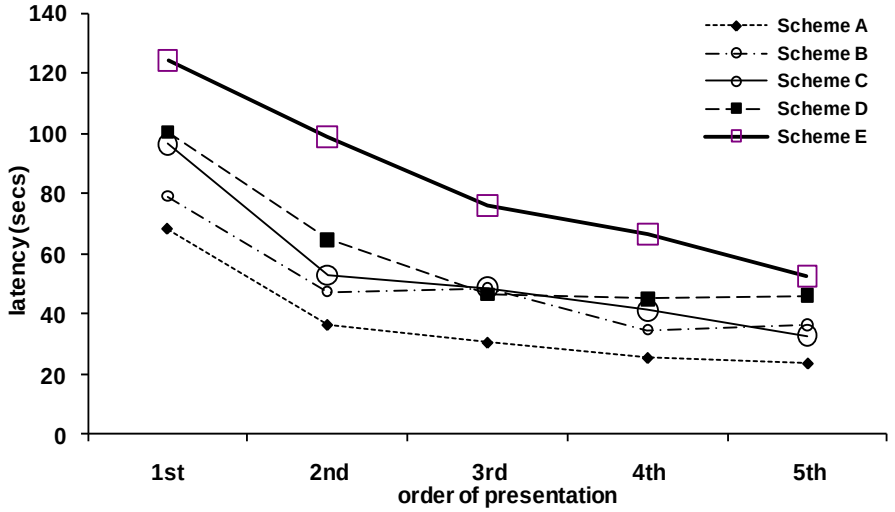

Figure 3: Relationship between latency and order of presentation (for the five hypothetical schemes)

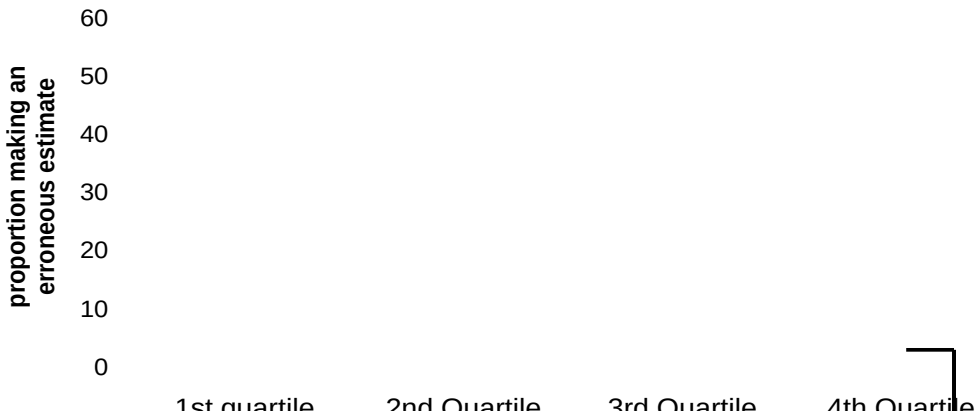

Figure 4: Relationship between latency and probability of error

Figure 4 shows, for hypothetical schemes A and E, the relationship between the time taken to calculate the charge (*H1*) and the probability of making an inaccurate estimate (the x-axis is a series latency bins – the first being those participants whose latency for H1 was in the lowest quartile, while the y-axis shows the proportion of people in each latency quartile who provide an inaccurate estimate of the charge).

For the simplest scheme (A), the probability of making an error increases with the time taken - suggesting perhaps that the task is so simple that extra time brings no increase in accuracy but is indicative of people finding the task difficult. For the most complex scheme (E), the probaility of making an error tends, if anything, to decease with time taken - suggesting perhaps that, for complex tasks, extra time is generally rewarded by greater accuracy (we note that, for scheme E, the third and fourth quartiles are likely to produce more error than the second quartile and suggest that this may indicate that a tail of people find the task so difficult that no amount of extra time will result in greater accuracy).

An analysis was conducted to determine how long participants were taking to answer each question. Although the average time taken to answer each question was 53.8 seconds, most took fewer than 10 seconds. Unsurprisingly, the time taken to answer the questions which required participants to estimate a charge (*H1* and *L1*) took the longest to complete (56.9 and 39.5 seconds respectively) and the first such question (i.e. for the first-presented hypothetical scheme and for the first presented Leeds scheme) took the longest of all (92.9 and 70.8 seconds respectively).

Tabulation of the time taken to complete the whole exercise, and to answer different types of question, showed that the values differed for different subgroups of participants. Although the times taken by different subgroups were not statistically different from those for the whole population, it appears that people over 35 years of age took noticeably longer to complete the experiment. More detailed investigation showed that females and people without a degree tended to take longer to estimate hypothetical charges (*H1*) and/or less time to assess their own decision style (*E2*, *E3* and *E4*). The strength of the learning effect (measured as the time taken to answer the first presented *H1* question relative to that taken to answer the other four *H1* questions) was much lower for people who had no degree[4] but was relatively strong for people who claimed (via *E2*) to gain satisfaction from completing tasks that have

[4] This difference was the only one to achieve statistical significance.

required a lot of mental effort. Ratios were computed to indicate the amount of care taken by different types of participant when estimating the hypothetical charges. Compared to other participants, those claiming to gain satisfaction from mentally demanding tasks showed low ratios for error/latency and for uncertainty/latency (uncertainty being measured via *H2*). The highest error/latency ratio was for people with low incomes and the highest uncertainty/latency ratio was for people without a degree.

Graphs and tabulations are useful but any serious attempt to determine significant influences on latencies and accuracies achieved requires the construction of models. Ideally, the models would permit detailed exploration of joint impacts, and allow for correlation between explanatory variables and the theoretical lower bound on latencies[5]. Resources did not permit such an approach in the current study and our exploration of the data has therefore been limited to the construction of regression models. The implication of this limitation will be discussed later.

Table 1 contains nine of the models which were tested and shows values for the variables which offered significant explanation (at 5%) of the dependent variable. Except where indicated otherwise, the models were run on a dataset which excluded observations for which latencies were unusually high (> mean + 3 SD) and were run using SAS stepwise regression (with inclusion criterion set to 5%).

Models 1 to 7 explore the extent to which the time taken by participants to respond to questions was significantly related to any aspect of the question, the scheme being considered or of their own characteristics.

From model 1 we deduce that the time taken to complete the experiment is not well explained by any of our explanatory variables but that there is a tendency for people with a low need to evaluate (not feeling the need to consider all pros and cons before making a decision) to take less time and for people over 35 to take more time.

From model 2 we deduce that the time taken to estimate the hypothetical charges is strongly affected by an order effect[6]. We also note that more time is taken to estimate the charge for schemes which are inherently complex (those whose description necessarily involved more numbers) or which were assessed by the participant as being complicated. Over and above this, we note that people who took a long time to answer the other questions were likely to take more time to estimate the charges. The fact that people with a degree took less time to estimate the hypothetical charges is interesting (and gratifying for those us involved in teaching!).

From model 3 we deduce that the time taken to estimate the "real" charges (those in Leeds) is again influenced by an order effect[7]. We also again see a positive relationship with the perceived complexity and with the time that participants took to answer other questions. There are, however, some interesting differences between models 3 (which relates to "real" charges) and 2 (which relates to hypothetical charges): when dealing with "real" charges, the latency is related to the number of

[5] Strictly speaking, the fact that latencies cannot be less than zero should be allowed for in the specification of explanatory models.

[6] Model 2 used the sequence (*order*), another model, not shown here for reasons of space, using the difference between the first and subsequent presentations (*firstD*), offered a similar degree of explanation and had very similar parameter values (the value of *firstD* was 41.98 with a *t*-value of 19.26).

[7] Another model using sequence (*order*) rather than *firstD*, offered a somewhat lower degree of explanation (adj. R^2= 0.36) but had similar parameter values (the value of *order* was -7.79 with a *t*-value of -15.02).

Table 1: Regression models

Model		1†	2	3	4	5	6	7	8†	9
Dependent variable		Total time	H1 Time	L1 time	L4 time	L7 time	L8 time	L9 time	E3 time	Error in H1
Constant		**883.11 (21.44)**	**21.83 (4.74)**	**-36.86 (-8.32)**	**5.28 (7.33)**	**3.19 (3.45)**	**1.42 (2.29)**	**2.17 (5.09)**	1.48 (1.68)	11.80 (0.93)
Explanatory variables:	*nzones*		*	**11.33 (7.13)**	*	*	*	*		*
	nperiods		*	**4.82 (3.12)**	*	*	*	*		*
	nnos		**7.53 (14.37)**	*	*	*	*	*		*
	complic		**0.179 (5.50)**	**0.08 (3.30)**	*	*	*	*		**0.45 (2.61)**
	order		**-12.02 (19.99)**		**-1.09 (-8.56)**	**-0.77 (-5.07)**	**-0.38 (-3.66)**	**-0.18 (-2.58)**		*
	firstD			**34.03 (19.20)**	**3.67 (7.95)**	**4.02 (7.78)**	**3.50 (9.56)**	**1.05 (4.19)**		**50.64 (4.08)**
	LschemeID			**24.28 (11.33)**	*	*	*	*		
	othertime		**0.0499 (12.70)**	**0.0423 (13.19)**	**0.0062 (10.69)**	**0.0050 (7.24)**	**0.0070 (14.92)**	**0.0034 (10.55)**	**0.0067 (7.22)**	*
	forrevD	*	*	*	*	*	*	*	*	*
	mentalSatD	*	*	*	**0.56 (2.16)**	**0.64 (2.07)**	*	**0.35 (2.43)**	*	**-34.95 (-3.54)**
	lowNTED	**-141.68 (-2.31)**	*	*	*	*	*	*	**1.77 (2.19)**	*
	clearcutD	*	*	*	*	*	*	**0.36 (2.46)**	*	*
	femD	*	*	*	*	*	*	*	*	*
	oldD	**99.32 (3.09)**	*	*	*	*	*	**-0.30 (-2.03)**	*	*
	employedD	*	*	*	*	*	**-0.57 (-2.20)**	*	*	*
	degreeD	*	**-5.35 (-2.70)**	*	*	**-1.00 (-2.81)**	**0.57 (2.30)**	**0.34 (1.98)**	*	*
	richD	*	*	*	*	*	*	*	*	*
	ExtraLat									**-0.36 (-2.04)**
	Uncert									**6.63 (2.75)**
Adjusted R^2		0.07	0.48	0.43	0.38	0.42	0..36	0.18	0.22	0.06
Root MSE		221.07	25.89	20.94	4.00	3.30	3.17	2.17	2.88	149.25
Number of obs.		193	950	942	947	477	946	951	193	965

Notes for Table 1:
* indicates explanatory variables considered stepwise, but which were not significant at 5% (blank cells indicate variables not considered).
† models 1 and 8 were run using truncated values for unusually high latencies (versions of these models run excluding data with such latencies had failed to provide any significant explanation).

For definition of variables see continuation of Table 1 on next page /continued

Table 1: Regression models (cont'd)

Definition of variables:	
Total time	total time (secs) to complete experiment (after completion of tutorial)
H1 time	time (secs) to estimate journey cost under hypothetical regime
L1 time	time (secs) to estimate monthly cost under Leeds regime
L4 time	time (secs) to decide whether would be better or worse off due to Leeds regime
L7 time	time (secs) to decide on likely behavioural response to Leeds regime
L8 time	time (secs) to decide on fairness of Leeds regime
L9 time	time (secs) to decide whether they approve of Leeds regime
E3 time	time (secs) to respond to the question on their need to evaluate
Error in H1	Difference (absolute) between the participant's estimate of H1 and true value of H1
nzones	number of zones in described scheme
nperiods	number of time periods in described scheme
nnos	number of numbers employed to describe the scheme
complic	participant's assessment of how complicated the scheme had been for them to understand (from H3 for hypothetical regime and from L10 for Leeds regime) (from 0 = "easy" to 100 = "difficult")
order	order in which the regime appeared to the participant
firstD	dummy (=1 if scheme was first to be shown to participant)
LschemeID	dummy (=1 if scheme was Leeds scheme I – see description in Appendix)
othertime	total time (secs) to complete all other questions (excluding current one)
forrevD	dummy (=1 if participant thinks council's main purpose of scheme was to raise revenue for non transport purposes (E1 = 3)
mentalSatD	dummy (=1 if participant claimed to gain satisfaction from completing a mentally demanding task (E2 = 1))
lowNTED	dummy (=1 if participant claimed not to need to evaluate all pros and cons before making decisions (E3 = 3 or 4))
clearcutD	dummy (=1 if participant claimed only to like dealing with problems which have a clear cut solution (E4 = 1 or 2))
femD	dummy (=1 if participant was female)
oldD	dummy (=1 if participant was over 35 years old)
employedD	dummy (=1 if participant was employed)
degreeD	dummy (=1 if participant had a degree)
richD	dummy (=1 if participant's annual household income exceeds £30,000)
ExtraLat	H1Time – (H1 time as predicted using model 2)
Uncert	participant's degree of uncertainty as to the accuracy of their cost estimate (H2)

zones and time periods (with an additional effect associated with the most complicated scheme) rather than simply to the number of numbers required to describe the scheme; the effect of having a degree is no longer apparent; and the overall level of explanation is lower.

From model 4 we deduce that the time taken by participants to decide whether they would be better or worse off due to a specified Leeds regime was again influenced by an order effect (over and above the effect of being first-presented), and that more time was taken by people who took longer to answer the other questions and by those claiming (via *E2*) to derive satisfaction from completing mentally demanding tasks.

Model 5 suggests that the time taken by a participant to decide on their likely behavioural response to a specified Leeds regime was, again, influenced by an order effect (over and above the effect of being first-presented) and by the time they took to

answer the other questions, and that people claiming to gain satisfaction from mentally demanding tasks were again taking more time than others. Interestingly, after allowing for these effects, people with a degree took less time than others to answer this question (another model, not shown here, showed that the time taken to decide on the effectiveness of a specified Leeds regime was affected by the same variables and to similar extents, but that people with degrees actually took longer than other people – perhaps they were considering the issues more thouroughly?).

From model 6 we deduce that the time taken to decide on the fairness of a specified Leeds regime is influenced by whether it was the first presented scheme (with a relatively weak order effect beyond that) and by the time taken by that participant to answer other questions. After allowing for these effects, less time was taken by people in employment and more by those with a degree.

From model 7 we deduce that the time taken to decide whether they approve of a specified Leeds regime is affected by the order in which it was presented (with being presented first having a relatively small effect beyond that) and by the time taken by that person to answer other questions. We note that more time was taken by people claiming to derive satisfaction from completing a complex mental task or to feel comfortable only when dealing with problems which have a clear cut solution. We also note that, over and above the "other questions" effect, more time was taken by participants with a degree and less by those over 35.

From model 8 we deduce that the time taken by participants to respond to the question on their need to evaluate is correlated with the time they took to answer other questions and, perhaps surprisingly, that it was longer for those claiming a no desire to consider all pros and cons before making a decision.

Model 9 seeks to explain the error that participants make in estimating the charge for the hypothetical schemes (calculable this because we know what the correct charge was). The results are interesting although the model does not provide a good explanation of the data. It is clear that errors are greater for the first presented scheme (with no significant further reduction beyond that), for schemes perceived as complicated and when the participants express uncertainty as to the accuracy of their estimate. We also note that less error is made by participants who claim to derive satisfaction from completing a complex mental task and when participants took longer than normal (as represented by the variable *Extralat*) to provide the estimate[8].

Another model, not shown in Table 2, indicated that, as anticipated in Figure 1, the participant's assessment of the degree of difficulty experienced in understanding a charging regime is positively associated with the time taken to estimate the charge.

4 Discussion and Conclusions

Using fairly simple analyses we have shown that the time taken to estimate the charge and decide on a response varied not only with the characteristics of the scenario (with more time taken to assess scenarios which were objectively more complicated) and with the order of presentation (indicating a learning, or fatigue, effect), but also with personal characteristics (notably age, educational attainment and attitude to decision-

8 Another model, from which *Extralat* was excluded, showed no significant relationship between error *H1time* – thus confirming that error is related to latency only when we consider latency over and above that which would be predicted for this participant and scheme.

making). We now discuss the significance and utility of these findings – but begin the discussion with a question….

4.1 Are Latencies Indicative of Effort?

Many previous researchers have assumed that latencies are an indication of effort expended. When dealing with differences in the time that one person takes to answer different questions this assumption is probably fairly safe (and our analysis suggests that when someone devotes extra time to making an estimate, the resulting estimate is more accurate). However, when comparing latencies of different people it would be dangerous to assume that this necessarily reflects different expenditures of effort – high latencies might simply be an indicator of a more relaxed, unhurried approach to the task.

Our analyses have shown that different types of people took different lengths of time to complete the questionnaire and that they took different amounts of time on different questions. For example, although they took longer overall, people over 35 took less time to indicate the extent to which they approved of each scheme (*L9*). Similarly, although people with a degree took less time to estimate charges (*H1*) and to indicate what their considered behavioural response might be (*L7*), they took longer to consider whether a given scheme was likely to be effective (*L3*) or fair (*L8*) and to indicate whether they approved of it (*L9*).

Answers to our questions on attitudes to decision-making (*E2*, *E3* and *E4*) helped to explain why different people took different lengths of time to answer different types of question. It appears that participants who reported (via *E2*) gaining satisfaction from completing mentally demanding tasks took longer to assess scheme effectiveness, to decide whether it would leave them better or worse off, to report their considered behavioural response, and to indicate the extent of their approval of the scheme. Those who claimed (via *E3*) not to feel a need to evaluate all pros and cons before making decisions took less time to complete the experiment, and those who reported (via *E4*) only liking to deal with problems which have a clear cut solution took longer to decide whether they would approve or disapprove of the scheme. The ratios of error to latency and of uncertainty to latency may be taken as indicators of "lack of care" and we note that these ratios were particularly low for people who gain satisfaction from mentally demanding tasks. It was also apparent that these people, and those who like to evaluate all the pros and cons, displayed stronger learning effects than other participants. Questions *E2*, *E3* and *E4* seem to be revealing not only different attitudes to decision making but also different propensities to engage with mental tasks.

Comparison of efforts expended by different people is clearly a complex issue whose resolution requires neural or physiological data but, meanwhile it seems that an indication of effort can be found by calculating the additional latency (beyond that which is expected for a given participant and task), and that questions on attitudes to decision-making can provide pointers to their likely degree of engagement with a task.

4.2 Factors Affecting Latencies

There is a clear and unsurprising link between latencies and the difficulty which participants report as having had in understanding the schemes. As expected, and in line with previous research, we find that latencies for the estimation of charges are related to task complexity. However, our new evidence allows us to extend this general finding to suggest that, in hypothetical contexts, complexity can be measured by the number of numbers used to describe the problem and that, in our experiment, after allowing for "the number of numbers", no additional effect was found for the number of zones or the number of time periods. However, where the context is "real", we find that the number of numbers has less impact than the more tangible indicators (the number of zones and the number of time periods). We also note that, in our "real" context, spatial complexity (number of zones) appears to have more influence than temporal complexity (number of time periods).

As expected, different types of question were associated with different latencies. In our experiment, questions requiring the participant to estimate charges were much more time consuming than those which asked them to assess the fairness or effectiveness of the scheme or indicate their likely behavioural response to it. We note that the inherent complexity of the scheme (objective or perceived) had no significant effect on the time participants took to decide on its effectiveness, whether it would leave them better or worse off, how to respond, whether it was fair, or whether they approved of it. Taken together with the fact that there was little evidence of any learning effect (reduction of latencies after the first presentation) for these questions, this might lead to the rather alarming conclusion that participants did not feel the need to consider the scheme details when answering such questions. The fact that participants took relatively little time to indicate their likely behavioural response to the charging regimes might itself be of concern to analysts who deal with stated choice data but it should be noted that, since the response questions (*L6* and *L7*) were asked after participants had estimated the charges considered whether they would be better or worse off if they continued driving, they were perhaps only the final stage of a considered decision on an appropriate behavioural response.

4.3 The Relationship between Latency and Accuracy

Although, taken as a whole, the results do not show any clear relationship between time taken to estimate a charge and the accuracy of the resulting estimate, Closer examination reveals some very interesting relationships.

Figure 4 showed that the relationship between latency and likelihood of error differs according to the complexity of the task. For very simple tasks extra time does not seem to result in reduced errors, rather it seems to reflect the degree of difficulty being experienced by the participant. However, for tasks requiring some effort, extra time generally results in fewer errors. This interpretation might explain the finding by Haaijer *et el*. (2000) that their normalised latencies were associated with more systematic decisions in one experiment but with less systematic decisions another; we note that they found an association with more systematic decisions in an experiment dealing with a relatively complex choice context whereas they found an association with less systematic decisions came from an experiment where, if the respondent recognised the attributes, the task was very simple. This again suggests that, in potentially simple choice contexts, extra time may indicate that the decision-maker

was experiencing a problem whereas, when the choice context is complex, extra time is an indication that more effort has been expended.

Model 9 showed that accuracies were significantly increased when a participant spent more time than would be expected (given the nature of the task and their personal characteristics). Other analyses showed that the time taken to estimate a charge was significantly (and positively) related to participant's assessments of the degree of difficulty they had experienced in understanding that scenario, and that accuracy is much increased for second and subsequent presentations.

4.4 Use of Latencies to Identify "Good" Data – the Interplay of Learning and Fatigue Effects

Initial examination of latencies revealed some outliers; those at the low end were such as to suggest that participants could not have given serious attention to the task in hand. Several of those at the high end were subsequently found to be associated with interruptions to the experimental process. These findings would support the use of a lower bound, and perhaps also an upper bound, on latencies as the basis for exclusion of some observations from a dataset.

In line with previous research, and consistent with the existence of a learning effect, we found a strong order effect in the response latencies. However, our results show that the strength and duration of the order effect is very different for different questions – being very strong for the calculation questions (*H1* and *L1*) but imperceptible for those which sought an assessment of the scheme or an indication of behavioural response. Moreover, the strength and duration of the order effect also varied with the complexity of the scheme (being stronger, and lasting longer, for the most complex scheme); reductions in latency following the first presentation were apparent for all schemes but further reductions were apparent up until the fifth presentation only for the most complex scheme.

Rose and Black (2006) proposed examining decision latencies to determine the point in a repeated sequence of questions when participants have learned how to process the information. They suggested that, having thus identified data from the period during which learning was still occurring, better fitting models could be built if such data were ignored. We certainly see the value of such an approach. However, given that we found the strength of the learning effect to vary by type of question and to be different for different types of people[9], we suggest that further improvements to models might be obtained by being more discriminating in the selection of which data to exclude (e.g. using the trend in normalised latencies to identify the appropriate cut off for each question and participant, rather than devising a general rule such as "exclude the first five observations").

A note of caution is appropriate here; although it is possible that high decision latencies are a signal that the participant has not yet learned how to process the experimental data, the reduction in the time taken to make choices during a choice experiment might also be evidence of an experimentally-induced fatigue effect whereby participants simply ignore information which they find difficult to deal with – in which case one might wish to exclude the data from questions with low latencies

[9] The learning effect was weakest for people who have no degree and strongest for people who report gaining satisfaction from mentally demanding tasks or a desire to consider all pros and cons before making a decision (via *E2* and *E3* respectively).

rather than from those with high latencies. One might perhaps be particularly suspicious of low latencies in data from experiments in which participants were motivated by a reward simply for completing a task (see Bonsall 2002).

The difference between a legitimate learning effect and an artificial experimental effect is, of course, difficult to determine. The development of an effort-minimisation strategy whereby people concentrate their attention on information which is easiest to process may reflect the widespread use of heuristics for real world decisions. However, low latencies become suspect if they are likely to have been stimulated by the experimental context. It is clearly a worrying possibility that long sequences of similar questions in a stated choice experiment might encourage development of an effort minimisation strategy.

One might perhaps distinguish between a legitimate learning effect and an artificial (experimentally-induced) effort-reduction effect by noting whether the latencies reduce after the first question (in which case one might deduce that it is a learning effect) or after several (in which case one might begin to suspect an effort-minimisation strategy). Reassuringly, we, like Rose and Black, find that response time latencies reduce most significantly between the first and second choices and even though they continue to fall throughout the sequence, we find no evidence of any acceleration in the rate of decline after several repetitions.

4.5 The Significance of an Individual's Attitude to Decision-making

The results from our simple questions on attitudes to decision-making are very exciting. One of our most important findings is perhaps that the accuracy of estimates was not related to latency, or even very strongly to expressed confidence in the estimate, but that it reflected the level of satisfaction which the participant claimed to derive from completing mentally demanding tasks.

As has been noted above, participants claiming to gain such satisfaction exhibited a stronger learning effect, made more accurate estimates of the charges, and took more care in making their estimates. They also took longer to assess scheme effectiveness, to decide whether it would leave them better or worse off, to report their considered behavioural response, and to indicate the extent of their approval of the scheme. Those who reported feeling no need to evaluate all pros and cons took less time to complete the experiment. Those who reported wanting to work on problems only if they had a clear cut solution took longer to decide whether they would approve or disapprove of the scheme.

It seems that our question *E2*, in particular, provides a simple, and thus very practical, indicator of a participant's attitude to mental tasks and that this is strongly related to the likelihood of them being able or willing to make accurate estimates. We are not aware of any previous example of this relationship having been revealed and suggest that it may have important implications for the understanding of decision making and for the segmentation of models.

4.6 Recommendations for Further Work

Our results confirm the value of latency data as a source of insight into the duration of learning processes and effort expended and as a basis for exploration of differences between the decision making processes of different types of participant. We suggest that there is considerable scope for more work in this area.

At a practical level we strongly recommend collection of latencies (appropriately normalised to allow for person-specific effects) to help identify data which ought to be excluded or subject to more detailed investigation. High latencies may indicate an incomplete learning effect while low latencies may indicate fatigue or experimentally-induced lack of engagement – or the use of heuristic decision rules.

The relationships revealed in Figure 4, and in Haaijer *et al*. (2000), suggest that high latencies might be indicative of problems in simple choice contexts but of effort in more complex choices. If so, this further reinforces the need to be circumspect in the use of high (or low) latency values to exclude data from choice experiments – one might quite unintentionally exclude the most important data! Thought should perhaps be given to supplementing data on latencies with data on participants' assessment of the difficulty they had experienced in making the choice (we suggest this may be more reliable than asking them to indicate how much confidence they have in their choice). It might be very interesting to explore the extent to which these variables are providing different types of insight or may be considered as alternatives.

Our experiment was conducted in controlled ("laboratory") conditions in order to reduce the risk that variance in latencies might be caused by external distractions. Latencies collected in uncontrolled conditions are likely to contain more "noise" and hence to be less useful. Work is underway to compare the variance in latencies collected in laboratory conditions with those collected from an online version of the same experiment.

We see particular scope for using decision latencies in repeated choice experiments to study the duration of learning processes, the onset of fatigue effects and the relationship between the two. Segmentation of sequential choices into initial choices with high latencies, subsequent choices with lower, but stable, latencies, and later choices with decreased latencies might be a particularly fruitful first step. Separate analysis of data from these three groups might be very rewarding.

Previous work on attitudes to decision-making has used batteries of questions which are quite time consuming for the participant. Incorporation of these batteries into an already long questionnaire results in an excessively long questionnaire and so risks low response rates and consequential bias. Although we would not claim that our questions *E2*, *E3* and *E4* have the strength of the full batteries from which they were drawn, they can easily be included at the end of a questionnaire which has been mainly concerned with other issues and they do appear to be capturing important attributes of participants which help to explain not only their decision latencies but also the accuracy of their estimates. This finding has great potential significance for data collection practice. It would, for example, be very interesting to see whether, having included such questions in a stated preference questionnaire, the resulting variables could further improve the robustness of the choice models or provide the basis for appropriate segmentation of participants. We are pursuing this avenue of research and would recommend others to do likewise.

Our work in this field is ongoing. We will shortly report on a comparison of results from our Leeds experiment with a matched experiment conducted in Dresden, on latencies collected via the web, and on a more detailed analysis of the relationship between decision latencies, accuracies, approval and other factors affecting engagement with the topic. Our further analysis will explore the extent to which the use of more sophisticated modelling methods (allowing for the lower bound on latencies and explicitly recognising the dependencies and joint influences of different factors) can provide any further insight. To this end we intend to explore the use of MANOVA, SURE and Tobit approaches.

Acknowledgements

We are happy to acknowledge the financial support of the European Union via the DIFFERENT project and are pleased to be able to thank our colleague Phil Wickham for his help in the design and implementation of the experiment. We are also pleased to be able to acknowledge the contributions of our collaborators in the DIFFERENT project, notably Jens Schade and Lars Rößger, for their input to the experimental design and interpretation of results. This paper is a revised version of one presented at the International Conference on Choice Modelling held in Harrogate in March 2009. We are also grateful to two anonymous referees for their helpful comments on the original draft of this paper.

References

Bettman, J., Johnson, E.J. and Payne, J.W. (1990) A component analysis of cognitive effort in choice, *Organizational Behaviour and Decision Processes*, 45 111-139.

Blair, W., Jarvis, G. and Petty, R.E. (1996) The Need to Evaluate, *Journal of Personality and Social Psychology*, 70(1), 172-194.

Bonsall, P.W. (2002) Motivating the respondent: how far should you go? In Mahmassani (ed.) *In perpetual Motion: Travel Behavior Research Opportunities and Applications Challenges*, Pergamon.

Bonsall, P.W., Shires, J., Maule, J., Matthews, B. and Beale J. (2007a) Responses to complex pricing signals: theory, evidence and implications for road pricing, *Transportation Research Part A*, 41(7), 672-683.

Bonsall P.W., Shires, J., Link, H., Becker, A. Papaioannou, P. and Xanthopoulos, P. (2007b) Complex congestion charging – an assessment of motorists' comprehension and the impact on their driving behaviour, Proceedings of the European Transport Conference, Nordwijk, October 2007, PTRC.

Budner, S. (1962) Intolerance of ambiguity as a personality variable, *Journal of Personality*, 30, 29-50.

Cacioppo, J.T. and Petty, R.E. (1982) The need for cognition, *Journal of Personality and Social Psychology*, 42, 116-131.

Espinoza-Varas, B and Watson, C.S. (1994) Effects of decision criterion on response latencies in binary decisions, *Percept Psychopys*, 55(2), 190-203.

Haaijer, R., Kamakura, W. and Wedel, M. (2000) Response latencies in the analysis of conjoint experiments, *Journal of Marketing Research*, 37, 376-382.

Lamberton, B., Fedorowitz, J. and Roohani, S.J. (2005) Tolerance for ambiguity and IT competence among accountants, *Journal of Information Systems*, 19(1), 75-95.

Newall, A. and Simon, H.A. (1972) *Human Problem Solving*, Englewood Cliffs, NJ. Prentice Hall.

Onken, J., Hastie, R. and Revelle, W. (1985) Individual differences in the use of simplification strategies in a complex decision-making task, *Journal of Experimental Psychology: Human Perception and Performance*, 11(1), 14-27.

Pollay, R.W. (1970a) A model of decision times in difficult decision situations, *Psychological Review*, 77(4), 273-281.

Pollay, R.W. (1970b) The structure of executive decisions and decision times. *Administrative Science Quarterly*, 15(4), 249-260.

Ratcliff R and Rouder, J.N. (1998) Modelling response times for two-choice decisions, *Psychological Science*, 9, 347-356.

Rose, J.M. and Black, I.A. (2006) Means matter but variances matter too: decomposing response latency influences on variance heterogeneity in stated preference experiments, *Market Letters*, 17, 295-310.

Schade, J. and Schlag, B. (2003) Acceptability of urban transport pricing strategies, *Transportation Research Part F: Traffic Psychology and Behaviour*, 6(1), 45–61.

Rößger, L., Schade, J., Obst, D., Gehlert, T., Schlag, B., Bonsall, P. and Lythgoe, W. (2008) Psychological constraints of user reactions towards differentiated charging, Deliverable 4.2 of the DIFFERENT project, University of Edinburgh Napier.

Wise, P.M. and Cain, W.S. (2000) Latency and accuracy of discriminations of odor quality *between binary mixtures and their components, Chemical Senses,* 25(3), 247-265.

Appendix 1: Description of the Experimental Protocol

Participants were recruited for "a study of travel costs in Leeds" and were offered a payment of £5. Those who confirmed that they drove on a regular basis in the city of Leeds were allocated an appointment to attend the laboratory. On arrival at the laboratory they were welcomed, their name was checked against the attendance list and they were taken to a computer in a small booth and instructed on how to log on. Once logged on, the software gave them a brief tutorial explaining how to interact with the software (typing in an answer or moving a cursor to a position along a scale) and gave them a test question. The system clock began when they had completed this test question.

The main survey was split into three sections. The first section sought the participants' response to each of five charging scenarios for a hypothetical journey (the journey was always the same; a 20 mile journey from A to B starting at 0740 and travelling at a constant 30 miles per hour - and they were helpfully informed that it would therefore take 40 minutes and that they would arrive at B at 0820). Each scenario included one or more distance-based charges operating in specified parts of the city during specified hours and was described such that the participant had sufficient information to calculate the price that they would have to pay to make that journey under the specified pricing regime. The five schemes (A-E), which were presented in randomized order, differed in terms of their inherent complexity (A had one zone and one time period, B had one zone and one time period, B had two zones and one time period, C had one zone and two time periods, D had two zones and two time periods, E had three zones and two time periods). Figure A1 shows how they were presented.

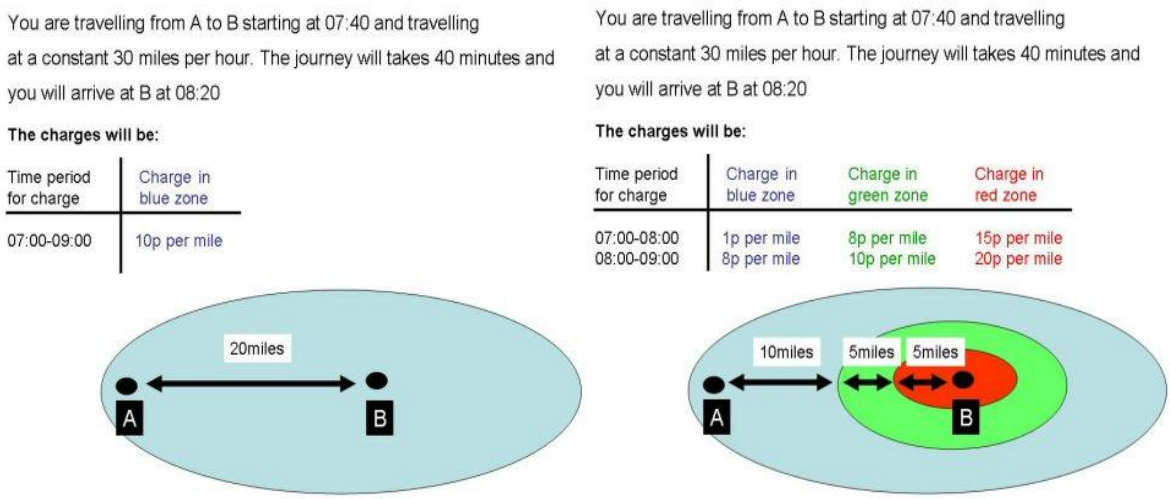

Figure A.1: Screen dumps for Schemes A (left) and E (right)

For each of the five schemes, three questions were asked while the appropriate map and description was displayed on screen:

H1 ***How much do you think the congestion charge for the specified journey would be (in pounds and pence)?*** *...*

H2 ***How certain are you of that estimate?*** *Potential responses were: 1 "very certain", 2 "to within 5% (i.e. plus or minus x)", 3 "to within 10% (i.e. plus or minus x)", 4 "to within 25% (i.e. plus or minus x)", 5 "to within 50% (i.e. plus or minus x)", 6 "to within 100% (i.e. plus or minus x)", 7 "even less sure than that". In each case, x was calculated from their estimate at H1)*

H3 ***How complicated do you think this congestion charging scheme is to understand?*** *(on a scale from 0="easy" to 100= "difficult").*

The second section of the questionnaire was concerned with five charging schemes for the city of Leeds. The schemes, which were presented in randomized order, differed in terms of their inherent complexity. Scheme details were summarized on an annotated map (Figure A2 shows the presentation of scheme I) and in text. The texts for schemes F-J were:

F "A £5 charge to use any roads in inner or outer zone between 0700 and 0900"

G "A £4 charge to use any roads in outer zone plus £6 charge to use any roads in inner zone (between 0700 and 0900 in each case)"

H "Charges to use any roads in inner or outer zone; £4 between 0700 and 0800, or £6 between 0800 and 0900"

I "Charges to use any roads in outer zone; £3 between 0700 and 0800, or £5 between 0800 and 0900. Charges to use any roads in inner zone; £5 between 0700 and 0800, or £6 between 0800 and 0900"

J "Charge to use Arterial roads in inner or outer zone: £4. Charge to use major arterial roads in inner or outer zone: £6"

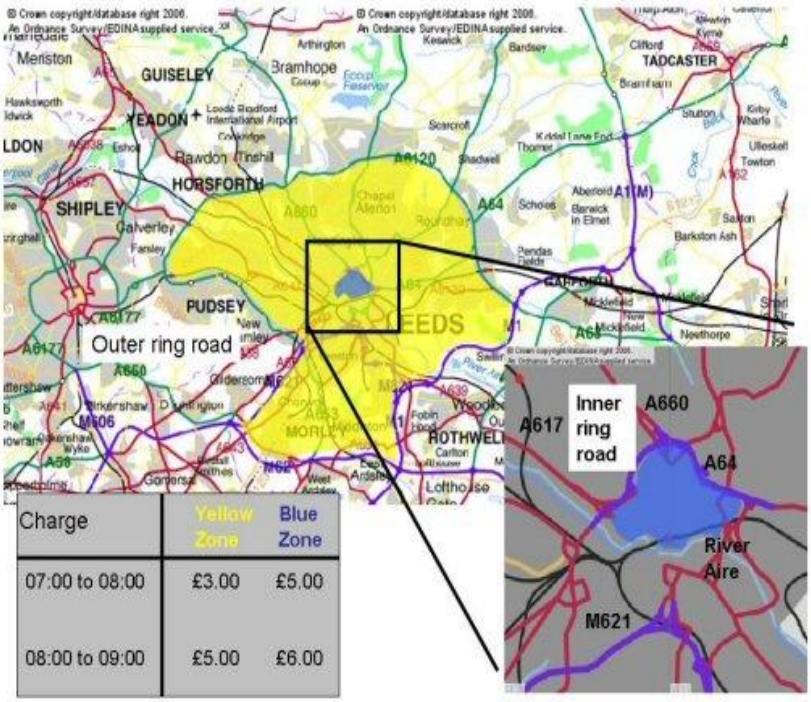

Figure A.2: Screen dump for Leeds Scheme I

For each of the five Schemes ten questions were asked while the map and description were displayed on screen:

L1 ***How much do you think this congestion charge would cost you per month?*** (assume that you carry on making all the journeys that you make in a typical month)…….pounds and pence

L2 ***How certain are you of that estimate?*** (options as per H2)

L3 ***How effective do you think this charge would be in reducing congestion?*** (on a scale from *0 Ineffective to 100 Effective*)

L4 ***On balance, if you carried on driving and had to pay the charges, do you think that you personally would be better off (because of driving conditions) or worse off (because of having to pay the charge)?*** (on a scale from *0 Better off to 100 Worse off*)

L5 ***How sure are you that you would be this much better off or worse off?*** (on a scale from *0 Just a guess, to 100 Completely sure*)

L6 ***If this congestion charge were to be introduced, would you need time to decide what to do?*** (the options being: *1 No, I would simply carry on driving as now, 2 No, I definitely reduce the amount of driving I do in the charge zone, and 3 Yes, I would want to think carefully about the costs and benefits of all the alternatives- then decide what to do*)

L7 ***Do you think that, when you have had time to think more carefully, you would be more likely to …(***the options being: *1 Carry on driving as now, 2 Reduce the amount of driving I do in the charge zone, and 3 Really don't know what you would do*)

L8 ***Different People would be affected by the scheme in different ways. Overall, do you think that a charge is fair or unfair?*** (on a scale from *0 Fair to 100 Unfair*)

L9 ***Would you approve or disapprove of this congestion charging scheme?*** (on a scale from *0 Disapprove to 100 Approve*)

L10 ***How complicated do you think this congestion charging scheme was to understand?*** (on a scale from *0 Easy to 100 Difficult*)

The third section included nine questions designed to obtain background information on the participant and their attitudes:

E1 ***What do you think would be the Council's main reason for introducing road charges like the ones we have described?*** (the options being: *1 To raise revenue for investing in public transport, 2 To raise revenue for investing in roads, 3 To raise revenue for non-transport purposes, 4 To reduce the amount of traffic on the roads, and 5 Don't know*)

E2 ***In general, when you complete a task that has required a lot of mental effort, do you feel satisfied - or relieved that it is over?*** (the options being: *1 Mainly satisfied, 2 Mainly relieved, and 3 Half and half*)

E3 ***In general, would you say that you are the sort of person who likes to work out all the pros and cons before making a decision?*** (the options being: *1Yes certainly, 2 Yes, 3 No, and 4 Certainly not*)

E4 ***In general, would you say that you are the sort of person who likes to work on a problem only if there is a possibility of coming up with a clear-cut and unambiguous answer?*** (the options being: *1 Yes certainly, 2 Yes, 3 No, and 4 Certainly not*)

(Questions E5-E9 respectively sought gender, age, employment status, educational background and household income).

Appendix 2: Profile of Participants

Attribute	Category	N*	%
Gender	Male	91	47.2
	Female	102	52.8
Age	17-25	31	16.1
	26-35	79	40.9
	36-45	50	25.9
	46-55	25	13.0
	56-69	8	4.1
Employment status	Employed	148	76.7
	Self-employed	6	3.1
	Retired/ Unemployed/ Home maker	4	2.1
	Student	35	18.1
Educational background	School level qualifications	28	14.5
	Degree	95	49.2
	Professional qualifications	18	9.3
	Degree and professional qualifications	52	26.9
Annual Household Income	Up to £13,499	19	9.8
	£13,500 - £29,999	65	33.7
	£30,000 - £49,999	64	33.2
	£50,499 - £74,999	35	18.1
	£75,000 or more	10	5.2