The impact of choice inconsistencies on the valuation of travel time in stated choice studies

Kjartan Sælensminde
Institute of Transport Economics,
Box 6110 Etterstad, N-0602 Oslo, Norway
Tel: +47 2257 3800, Fax: +47 2257 0290, E-mail: ks@toi.no

Abstract:
A new test procedure revealing choices that are jointly inconsistent has been developed and applied to the respondents’ choices in the Norwegian value of time study (Ramjerdi et al. 1997) and a study providing the Norwegian road authorities with valuation of environmental benefits for use in Cost-Benefit Analysis (Sælensminde and Hammer 1993). Our analysis show that lexicographic and inconsistent choices commonly occur in several stated choice tasks and have a significant impact on the valuation of reduced travel time. Both people’s difficulties handling the choice tasks and real preferences cause lexicographic choices. Different abilities to choose are also an important explanation of inconsistent choices. The occurrence of inconsistent choices is shown to be largest first in the choice sequence and is reduced for later choices. These results suggest that the respondents may need more training and help to choose consistent in stated choice studies.

Acknowledgements: This paper was written when the author was visiting researcher at Institute of Transport Studies, University of Sydney, from October 97 to April 98. I am grateful to Ronny Klæboe, Ståle Navrud and Farideh Ramjerdi for valuable comments on previous versions of this paper. Thanks also to Jan-Erik Lindjord for his excellent programming and comments, and to David Hensher for his hospitality. The study is financed by the Research Council of Norway and the Institute of Transport Economics in Oslo.
I. INTRODUCTION

Conjoint analysis, including stated choice methods, have been used for many years in transport economics and marketing research (see e.g. Hensher 1994 for an overview). They are now becoming increasingly popular for the valuation of environmental goods (Widlert et al. 1993, Sælensminde and Hammer 1994, Hanley and Wright 1997, Hanley et al. 1997, Boxall et al. 1996).

Traditionally, however, environmental goods have been valued in studies using the contingent valuation method (CVM). The CVM has been used in many studies the last 30 years and has been subject to much critique and debate (see for example Diamond and Hausman 1994, Hanemann 1994 and Portney 1994 for an overview). One important question in the CVM debate is how the questions should be formulated to make them more familiar to the respondent and in that way eases his/her choice task. As an answer the NOAA panel (Arrow et al. 1993) suggested that the discrete choice (DC) CVM format was preferable to the open-ended (OE) CVM format. But issues concerning statistical uncertainty, stability and bias are not addressed by the panel, and the debate concerning the preferability of the DC-CVM to the OE-CVM is not closed (Neill 1995, Boyle et al. 1996, Frykblom 1997 and Halvorsen and Sælensminde 1998). A move away from valuation of environmental goods by OE- or DC-CVM to use of conjoint analysis/stated choice, with two or more attributes in each choice, will again increase the complexity of the respondent’s choice task. Such a move should therefore not be done without consideration of the type of impact a more complex choice situation would have on the respondent’s ability to state their correct preferences. It is also important to investigate how possible “erronous choices”, due to a too complex choice situation, influence the analysis and therefore the valuation of the goods.

The aim of this study is to investigate if the complex choice situation in stated choice studies result in lexicographic and inconsistent choices, and whether such choices have a significant impact on the valuation of the goods in the study. By lexicographic choices is meant choices were the respondent always chooses that alternative which is best in respect of one particular attribute, e.g. lowest price, and neglect the other attributes. Lexicographic choices do not necessarily mean that the respondent has lexicographic preferences, e.g. that he/she has a willingness to preserve the environment at any cost. Lexicographic choices may be a result of simplification because the respondent finds the choice task too difficult or a result of actual preferences because the presented choice alternatives are too different. It is difficult to decide the cause of lexicographic choices solely

1 According to Foster and Mourato (1997) earlier analyses of lexicographic preferences in the context of environmental valuation studies (e.g. Stevens et al. 1991; and Spash and Hanley 1995) have been based on responses to attitudinal questions indicating a willingness to preserve the environment at any cost.
on the basis of stated choice data, but the inclusion of OE-CVM questions in the current study provide an opportunity for further investigation of this issue. Such an investigation of causes of lexicographic choices has probably not done been carried out previously.

The complex choice task in stated choice studies may, in addition to lexicographic choices, result in inconsistent choices also from respondents who really intended to choose according to their preferences. By inconsistent choices is meant choices that are not consistent with choices the respondent made earlier or later in the presented choice sequence. The reason that inconsistent choices are revealed is that each choice gives an interval of the valuation of each attribute included in the choice task. If the valuation intervals for two choices do not overlap, we find these two choices to be mutually inconsistent. The occurrence of inconsistent choices in stated choice studies has, to my knowledge, not previously been investigated in this manner.

The current empirical study of lexicographic and inconsistent choices uses data from two Norwegian surveys. These were respectively designed to revise the value of travel time and the value of environmental goods used in benefit-cost analysis (BCA). Two experimental designs were implemented in these surveys: a conjoint analysis (stated choice) and an OE-CVM format. Both data sources from the value of time survey are used in the current study. The stated choices in these surveys used in the current study include only three attributes and the context is related to a real journey. Occurrence of “problematic choices” in such simple choice situations may therefore be an important contribution to the discussion of how to design such valuation studies in the future.

The paper commences with an overview of existing literature of lexicographic and inconsistent choices in choice studies. Section three describes how tests of inconsistency and lexicography are carried out. Section four describes the surveys applied here. The fifth section presents the results from the analysis. Finally, section six provides some concluding remarks and present which implications the results of this study may have on the design and analysis of future stated choice studies.

---

2 Foster and Mourato (1997) conclude that it is not possible to discriminate between the hypothesis of genuine lexicographic behaviour and conventional, but strongly skewed, preferences on the basis of non-parametric tests alone. Furthermore, even if it was possible to prove that lexicographic ranking was genuinely taking place, it could always be argued that it was merely being used by respondents as a heuristic device for simplifying the ranking task and did not necessarily reflect the nature of their underlying preferences.

3 In the current study data from the environmental survey is only used to investigate whether demographic variables can explain the occurrence of lexicographic and inconsistent choices.
II. PREVIOUS LITERATURE

In previous literature both simple inspections and parametric approaches have been used in the investigation of problems occurring in choice studies because of the choice task’s complexity. One approach is to inspect each respondent’s choices. In such an approach one can discover the occurrence of lexicographic choices which may constitute a violation against the continuity axiom (see e.g. Widlert 1994), the occurrences of choices that contravene the IIA property and choices that are of incompliance with the axioms of non-satiation and transitivity.\(^4\)

The design of a study by Foster and Mourato (1997) enable them to use this inspection approach to examine all these violations of consumer theory. Another approach, parametric, is to use a statistical estimation method that allows for different error variances within a single model. Scaling, one such parametric approach (Ben-Akiva and Lerman 1985), has mainly been used in studies that use a mixture of revealed preference (RP) and stated preference (SP) data\(^5\). For the topic of this paper, a more interesting use of scaling is as a tool in the investigation of the sources of variance (i.e. “noisy” choices) within a single SP data set (Widlert 1994, Bradley and Daly 1994, Mazzotta and Opaluch 1995, Swait and Adamowicz 1996 and Hensher et al. 1997). In another parametric approach applied by Johnson and Desvousges (1997), each respondent’s multiple observations are used to estimate a model for each individual. A third parametric approach, also applied by Johnson and Desvousges (op.cit.), is to estimate a panel model with respondent-specific scale parameters for the latent random-utility distribution.\(^6\) A fourth parametric approach is to include decision strategy selection as an explicit factor in the choice model (Swait and Adamowicz 1997).

Both inspections of the data and parametric approaches provide the opportunity to investigate how “problematic choices” influence the results from a statistical model. One important difference is between approaches that can identify each respondent’s problematic choices and approaches that only are able to detect that (some) respondents have had more or less problems in the choice task. To see

---

\(^4\) The continuity axiom of consumer theory states that two commodity bundles which are similar to each other will be ranked close together in the consumer’s preference ordering. This axiom precludes lexicographic ordering of bundles by requiring respondents to trade-off gains in one commodity against losses in another. The non-satiation axiom states that a consumer must prefer a bundle, which is in all respect superior to another. The transitivity axiom requires that if a consumer prefers option A over option B and option B over option C, then he must necessarily prefer option A over option C. The IIA (Independence from Irrelevant Alternatives) property, which is central to the logit framework of econometric analysis, requires that the ranking between two bundles in a choice set is not affected by the identity of the remaining bundles in the set.

\(^5\) See e.g. Ben-Akiva and Morikawa (1990), Bradley and Daly (1992), Hensher and Bradley (1993) and Adamowicz et al. (1994).

\(^6\) This is actual the extreme option of the scaling approach treating each person in the sample as a separate data set with it’s own scale factor.
what is really going on, it may be important to identify both each problematic choice and each respondent’s number of problematic choices. The rest of this section summarises the findings in the studies mentioned above that are specific into the issue about how the respondents choose in different choice tasks.

Widlert (1994) investigated how different aspects of the design influence the results of tasks that include choice, rating or ranking. This study was conducted in order to value different quality attributes connected to train services in Sweden. Widlert’s conclusion is that the most important bias in these data is caused by simplifications the respondents make when they carry out the tasks. This conclusion about simplification is drawn solely on inspection of the occurrence of lexicographic answers in the different selection tasks. Investigation into learning and fatigue by use of the scaling approach enabled Widlert to conclude that there is a clear fatigue effect in the ranking task (ranking of up to 16 cards). However, in the pairwise choice task (12 discrete choices) he does not find any marked trends in the scale factors.

Bradley and Daly (1994) use the scaling approach to investigate whether rank order effects exist in a ranking study, and fatigue effects in a pairwise choice study. The ranking study value attributes related to bus service levels, bus stop facilities and bus vehicle factors in Sweden. The choice study is conducted to study train/car mode choice for intercity travel in the Netherlands. These studies showed that the amount of unexplained variance increase as rankings become lower, and as the number of pairwise choices increases. By use of simulation tests they conclude that the scale effects are real-world phenomena due to differences in the way in which real respondents make decisions at different points in the ranking or choice process.

Bradley and Daly further conclude that for the pairwise choice data, the addition of the scale factors to account for respondents fatigue did not change the relative magnitude of the model coefficients. For the rank-order data, on the other hand, some coefficients became more or less important relative to the others. They explain this by the fact that the pairwise choice data was collected using a computer-based approach, which presented the pairs of alternatives in a different random order for each respondent. As a result, any adverse influence of the order-related fatigue effect might have been “randomised out” of the data. Another possible explanation is that addition of scale factors in the models is not sufficient to correct for the effects of problematic choices. This view is supported by the fact that exclusion of problematic choices, as in Foster and Mourato (1997) and the current study, show significant effect on the valuations.

Johnson and Desvousges (1997) investigated the behavioural and statistical properties of a rated pair data set generated by adaptive conjoint analysis procedures. Included in the environmental related attributes valued in this study were the number of lakes with fish consumption bans, the number of cancer cases
and the number of respiratory cases among the population. By use of each respondent’s multiple observations to estimate individual models they test for consistency and transitivity in each respondent. If this estimation procedure did not converge or otherwise failed, which happened for 52 of the 205 respondents, the respondents were referred to as “nonconvergent”. Johnson and Desvousges conclude that these nonconvergent respondents’ pairwise ratings are generally inconsistent, are insensitive to large differences in attribute levels, or employ very few response categories. The primary cause of failure in the estimation procedure was insufficient variation in individual responses (i.e. lexicographic choices).

Johnson and Desvousges suggest that explanations for such patterns may include protesting some aspect of the elicitation, inattention to or failure to understand the task, ill-defined preferences and fatigue, but they do not distinguish further between these possible explanations. Unfortunately, they are unable to identify any attitudinal or socioeconomic differences (e.g. age and education) that distinguish between the two groups. This result may be driven by the fact that the failure test used by Johnson and Desvousges may not be a very precise procedure in identifying which choices are inconsistent and which choices that are lexicographic.

By use of a panel model with respondent-specific scale parameters for the latent random utility distribution, Johnson and Desvousges (op.cit.) show that the distribution of the scale parameters from “nonconvergent” respondents are more noisy than from “convergent” respondents and that they represent a relatively uninformative ratings pattern. However, a re-estimation of the aggregate model after trimming observations with small scale estimates shows that uninformative ratings have little effect on the WTP estimates and that the reduction in sample size offsets the expected gain in efficiency in this case. Unfortunately, many “nonconvergent” choices are in the trimmed sample and the results may depend on this fact. They also conclude that aggregate models (including both “convergent” and “nonconvergent” respondents) tend to suppress individual inconsistencies. Bradley and Daly (1994) arrived at the same conclusion, but again, maybe the panel model is not sufficient to correct for the effect of problematic choices either.

Mazzotta and Opaluch (1995) use the scaling approach and data from a landfill siting survey to explore Heiner’s (1983) hypothesis concerning a gap between the cognitive ability of decision-makers and the difficulty of decisions (the C-D gap). They assume that the heteroscedasticity arises from variations in the complexity of choices, defined in terms of the number of attributes that differ across alternatives. Their results suggest that the C-D gap becomes significant for comparisons of alternatives with more than three attributes (complexity level 3), and that simplified decision strategies were employed beyond complexity level 4. Unfortunately, due to deficiencies in their data set they could not test for lexicographic choices as a specific simplifying choice strategy. In addition, their
tests for other simplifying strategies were inconclusive. They suggest that this may be a result of mixed strategies used by individuals, or that different choice strategies are used by different respondents, but Mazzotta and Opaluch do not explicitly test for inconsistent choices. Mazzotta and Opaluch also examined the implications of complexity for calculating welfare effects in the form of monetary values of attributes. Their results show that one of the attributes, wildlife habitat, is valued about 2.5 times higher at complexity level 4 as compared to complexity level 3. If it is assumed that actual experience with a good increases the ability to choose, regardless of the complexity level, this result is supported by Cameron and Englin (1997) who conclude that estimated WTP tends to fall with increased experience, i.e. use of the good. These results contrast the results of Bradley and Daly (1994) and Johnson and Desvousges (1997).

Swait and Adamowicz (1996) characterise task demands in a more general manner than Mazzotta and Opaluch (1995) and incorporate them into random utility models of choice. In the context of Swait and Adamowicz (1996) it is suggested that the complexity of the decision problem will affect the ability to choose, and thus for any given individual, ability to choose will differ depending on the task demand. Both entropy and cumulative entropy are included in their model. Employing six case studies which examine choice within very different product classes, they find that complexity affects variance in a non-linear fashion with both very difficult and very easy choices resulting in near random choice behaviour. They conclude that “choice complexity” is an important factor to consider when modelling choice behaviour, both at the task design stage and during econometric model estimation.

In Swait and Adamowicz (1997) measures of entropy and cumulative entropy (see Swait and Adamowicz 1996 above) are used in a model that allows for changes in decision making strategies over ranges of task complexity. Their empirical analysis illustrates that a distinct processing strategy arises in cases with high levels of task complexity. This processing strategy appears to focus on only two characteristics of the task, greatly simplifying the choice task for respondents. They suggest the use of such models to test whether changes in complexity result

---

7 In addition to the effect of complexity this result in Mazzotta and Opaluch (1995) might also be affected by changed context and interaction effects (i.e. substitution and/or complementary effects) between the attributes (Sælensminde 1998). But the fact that the context is similar for the different complexity levels and that interaction effects in most cases will reduce the valuation when more attributes are included and probably will affect more than one attribute, supports the hypothesis that the complexity level is the reason for these results.

8 Like Mazzotta and Opaluch (1995), Swait and Adamowicz (1997a) present an analysis of context effects on choice in the same tradition as Heiner (1983). De Palma et al. (1994) present a more formal examination of the processing limitation argument. In their model an individual with lower ability to choose will make more errors in comparisons of marginal utilities. The implications of limited processing capability in their model are that individual heterogeneity in the ability to choose produces widely different choices, even if in all other aspects the individuals are identical.
in changes in processing strategy and drive consumers toward non-compensatory choice heuristics.

Hensher et al. (1997) is not explicit regarding the questions about how respondents make choices in discrete choice studies, but they investigate the effect of relaxing strong assumptions on the variance structure (specifically heteroscedasticity) in the random utility model. This is essentially the same procedure as is used by Bradley and Daly (1994) and Widlert (1994). In one of the case studies by Hensher et al., regarding preferences for shopping outlets for apparel purchases, they conclude that the responses of individuals who dislike shopping are less consistent (i.e. have a lower scale/ higher variance) than respondents who like to shop. They further conclude that these subjects exhibit homogeneity of tastes but different variances in utility, and that this conclusion is a contradiction to what one find if one assumed a homoscedastic variance structure in the model. Hensher et al. suggest that future research should be directed towards models that are more able to capture differences in variances, as well as empirical work directed towards understanding sources of differences in variances and their role in choice.

Foster and Mourato (1997) use a contingent ranking survey to value health effects on the general public and effects on farmland bird species in the UK. Their survey was designed to permit empirical tests of compliance with the IIA property and the axioms of non-satiation, transitivity and continuity. The results show that 58% of the respondents passed the IIA test on every occasion and the remaining 42% only failed, on average, one out of four times. These findings indicate that respondents do not systematically contravene the IIA property. Each respondent was given two opportunities to fail a non-satiation test. 83% of the sample consistently passed this non-satiation test, 12% failed in one occasion and 5% failed in both occasions. 9 Foster an Mourato presented about 43% of the respondents to a transitivity test, each of them being given on average two opportunities to pass or fail the test. 82% of these respondents consistently passed the transitivity test, while 12% failed on every occasion. This is more than double the failure rate for the non-satiation test, but Foster and Mourato comment that this is not surprising given that the transitivity test is considerably more demanding than the test for non-satiation. 10 To test for violations of the continuity

9 By giving the respondents only two opportunities to fail the test, respondents with “choice problems” and which therefore chose randomly in “doubt cases” had a 25% chance of passing the test in both occasions. Likewise, they had a 50% chance of failure in one occasion and a 25% chance of failure in both occasions. Consequently, with the occurrence of 5% failure in both occasions and 12% failure in one occasion one can assume that 5% or 6% of the respondents with “choice problems” passed both tests by chance and therefore are included in the 83% that always passed this test. This may be troublesome when Foster and Mourato (1997) compare the model results for non-failure respondents with failure respondents.

10 In addition to the comments in the previous footnote, it may be important that if a large (?) part of the respondents only got one opportunity to pass or fail the test, these respondents will pass (and
axiom Foster and Mourato test whether the respondents devise various “lexicographic algorithms” for answering the questionnaire. One such algorithm is “price-health”; i.e. rank options first on the basis of the price attribute and settle any ties with reference to the health attribute. Not surprisingly, only 18% of the respondents gave rankings that always match up with such a strong form of lexicographic algorithms. “Reassuringly”, they write in their summary remarks, “respondents whose rankings were consistent with particular lexicographic algorithms, were found to be significantly more likely to have expressed strong attitudinal preferences in favour of their highest ranking criterion at an earlier stage in the questionnaire.”

A number of statistical tests were performed by Foster and Mourato (1997) to establish whether failure of rationality tests was random or systematically related either to observable characteristics or to design features of the questionnaire. Thus, a systematic relationship between test failures and educational achievement would suggest that these might have largely been a reflection of cognitive difficulties with the questionnaire as opposed to genuine irrationality. Similarly, a systematic relationship between test failure rates and question order would suggest that these may have been largely attributable to fatigue effects as people got towards the end of the questionnaire. Foster and Mourato report that in neither case was there any substantial variation in failure rates against either of these variables. These results are probably affected by the fact that the failure tests conducted only expose less than half of the respondents that have problems carrying out the task. This may also be a reason why Foster and Mourato have to conclude, “the test failures are diluted by the rational majority” when they compare the results from models with and without the inclusion of these test failures. However, a Hausman test strongly rejects the null hypothesis of equality between the two sets of coefficients and they therefore further conclude, “these findings indicate that the parametric analysis is highly sensitive to the presence of irrational responses in the dataset. It is therefore important to design contingent ranking surveys in such a way as to permit non-parametric tests of the fundamental axioms of consumer theory so that respondents who violate these assumptions can be identified and may be removed from the sample prior to estimation if deemed necessary. The fact that this has not been standard in the

fail) with 50% chance by choosing randomly. This may also be an explanation of the result that more respondents failed the transitivity test than the non-satiation test.

Note that this represents a strong form of lexicographic ordering in that the algorithms do not simply refer to a single lexicographic criterion but assume that options which tie on the first criterion are sorted with reference to a second. Such a strong test is probably only able to detect respondents that very consciously sort the alternatives and will not detect respondents that only use a sorting algorithm for one attribute and choose randomly with regard to the other attributes. Therefore, such a strong test will not detect all respondents that use a lexicographic choice algorithm to simplify the task.
literature to date raises questions about the accuracy of valuation estimates extracted from existing studies.”

Bates (1994) points out that the danger of presenting respondents with tasks that are too complicated is well known to practitioners. What is much less clear, however, are where the boundaries lie, and, in particular, how they might vary between respondents of different intellectual ability. According to Bates, there is a need for much more research into how respondents actually carry out the tasks, using de-briefing techniques and alternative decision rules. As a minimum, the data should be checked for possible lexicographic effects.

Inspired by Bates (1994), the current study investigate the effect of lexicographic choices, the effect of inconsistent choices (by use of a procedure illustrated by a “ray diagram”, also described by Bates) and how such problems vary between respondents. In addition to an inspection of the occurrence of lexicographic choices, the current study investigates the cause of lexicographic answers by use of the results from additional OE-CVM questions and socioeconomic variables in statistical models. An explicit investigation of how each respondent has performed the choice task is also undertaken and the conclusions about learning and fatigue from this procedure are compared to results from the parametric scale approach.

III. TESTS OF CONSISTENCY AND LEXICOGRAPHY

The tests undertaken in the current study include all the choices that each respondent has made. This is a strong and comprehensive test procedure that will detect all present inconsistencies. Test procedures that only make use of a few of the respondents’ choices (see e.g. Foster and Mourato 1997) may detect some of the violations against the axioms of consumer theory, but will not detect inconsistencies between all the choices made by each respondent.

This chapter introduces and explains those tests used to investigate whether the respondents choose consistently, or if they choose in a lexicographic manner in the discrete choice task. The connection between the consistency test and the lexicographic test procedure is explained.

Test of consistency

The consistency test is based on the assumption that the respondents have a given preference structure and that their choices are based on rationality in the way usually done in the consumer theory. The data used for these tests are choices between two different journeys with the same mode. Three attributes; price, travel
time and headway (or chance of delay) describes the journeys.\textsuperscript{12} The preference structure is based on the assumption that the respondents prefer to pay less, use less travel time and get a better headway (or less chance of delay). Bates (1994) shows how a “ray diagram” can be used in the design of choice studies. In the current study such a ray diagram is used to test if the choices made by each respondent are mutually consistent.

Figure 1 presents four different discrete choices. Each choice is between two long distance train journeys described by the three attributes price, travel time and headway. The data used in the current study include nine such choices. The respondents’ task is to choose, in each choice, if they prefer the journey described on the left hand side (LHS) or the journey described on the right hand side (RHS). Each of the four choices in Figure 1 is illustrated graphically in the “ray diagram” in Figure 2.\textsuperscript{13} Ray numbered (1) in Figure 2 is a graphically representation of choice number (1) in Figure 1 and represents the following. If the respondent’s WTP for reduced headway is zero, he/she will choose LHS or RHS if his/her WTP for reduced travel time is more or less, respectively, than 100 NOK/h. 100 NOK/h is therefore a point on the axis denoted “valuation of travel time”. Similarly, if the respondent’s WTP for reduced travel time is zero, we achieve the point 50 NOK/h on the axis denoted “valuation of headway”. The line between the two points represents linear combinations of the valuation of travel time and headway and, according to the continuity axiom it is assumed that the respondents trade-off gains in travel time against losses in headway, and vice versa. In this case it is also assumed that there are no interactions between travel time and headway that could make the shape of line other than linear.

<table>
<thead>
<tr>
<th>Choice no. 1</th>
<th>Choice no. 2</th>
<th>Choice no. 3</th>
<th>Choice no. 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LHS</td>
<td>RHS</td>
<td>LHS</td>
</tr>
<tr>
<td>Price (NOK)</td>
<td>300</td>
<td>200</td>
<td>300</td>
</tr>
<tr>
<td>Travel time (hour)</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Headway (hour)</td>
<td>4</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 1: Four different choices between journeys described by the three attributes price, travel time and headway. In each choice the respondent is to choose if he/she prefer the journey presented on the left hand side (LHS) or the journey presented on the right hand side (RHS).

\textsuperscript{12} In the choice between two longer car journeys the third attribute was automatic speed control; i.e. how many such speed control units that were to be used. No preference structure was assumed in the design of the study for that attribute.

\textsuperscript{13} All the choices in Figure 1 are such that by choosing the most expensive journey the respondent gets an improvement in both travel time and headway. This is done to simplify the graphical illustration.
If the respondent in choice (1) chooses RHS, i.e. he/she is not willing to pay 100 NOK more for a journey where both travel time and headway are improved. Such a choice indicates that his/her valuation of reduced travel time is less than 100 NOK/h and that his/her valuation of reduced headway is less than 50 NOK/hour. In this case, the valuation of travel time and headway therefore lies in the area restricted by the two axes and ray numbered (1). This conclusion is based on the *a priori* assumptions about the preference structure and that an improvement in both travel-time and headway is not valued higher than the aggregate of an isolated valuation of these attributes.\(^{14}\)

![Figure 2: “Ray-diagram” based on the four choices in figure 1.](image)

In a similar manner, as for choice (1), it is possible to decide which side of the rays numbered (2), (3) and (4) the respondent’s valuation is positioned given the choices he/she does in choice numbered (2), (3) and (4). If the valuation area determined by one choice is part (not part) of the area determined by another choice that means that these two choices are mutually consistent (inconsistent).

---

\(^{14}\)This is an assumption that may be wrong for some respondents. Sælensminde (1995) show that a valuation of travel time and headway in a package may give a higher valuation than the sum of the valuation of travel time and headway valued separately. For most of the respondents these package effects are negative. *A priori*, it is difficult to decide the sign of such package effects for the individual respondent.
Two degrees of consistency

The least restrictive test of consistency is when each choice made by one respondent is tested only against each of the respondent’s previous choices separately. In the examples presented in Figure 1 and 2 this means that choice numbered (2) is tested against choice numbered (1). According to Figure 2, choice number (1) and (2) are mutually consistent irrespective of the choice made in these tasks. In similar manner, we see that choice number (1), (2) and (3) are all mutually consistent when tested against each other separately. A failure of this least restrictive test can be illustrated by the choice of RHS in choice numbered (1) and LHS in choice numbered (4).

The most restrictive test of consistency is when each choice made by one respondent is tested against the “aggregate” of the previous choices. This means that choice (2) is first tested against choice (1). If he/she, for example, chooses RHS in both these choices, his/her valuation of travel time and headway lies in the area restricted by the two axes and the rays numbered (1) and (2) in Figure 2. When the respondent makes the third choice, this choice will not be consistent with the “sum” of the two previous choices unless he/she chooses RHS. To use the aggregate of the previous choices is therefore a more restrictive consistency test than if the choices are tested against each other separately.

The most restrictive consistency test is probably in best agreement with the idea of rationality in economic theory. Despite this, the tests of consistency in this study are based on the least restrictive test. The reason for this choice lies in the fact that the respondents are engaged in a cognitive process during the course of the experiment (Johnson and Desvousges 1997, and Swait and Adamowicz 1996). The first choices may therefore be more troublesome for the respondents than later choices in the choice sequence. This is also the reason why Bradley and Daly (1994) and Johnson and Desvousges (1997) advise against use of an adaptive design in choice studies.

Testing of lexicographic choices in the choice study

Lexicographic choices in choice studies occur as a result of simplification because the respondent finds the choice task to difficult to handle or as a result of too large

---

15 If one choice (e.g. no. 3) is inconsistent with a previous choice (e.g. no. 2) then the previous choice (no. 2) will be inconsistent with the subsequent choice (no. 3). Therefore it is sufficient to test each choice for consistency against previous choices to know “what’s going on”.

16 Inconsistent choices may also influence tests of the effects of repeated observations from the same respondent in stated choice studies. See, for example, Ouwersloot and Rietveld (1996) who, without considering inconsistencies, conclude that repeated observations do not cause significant autocorrelation problems.
differences in attribute levels; i.e. as a result of real preferences. It is difficult to decide if a respondent has chosen lexicographically because he/she wanted to simplify the choice task or because the differences in attribute levels were too large.

In the test of lexicographic choices performed in the current study it is only considered whether the respondent has consistently chosen the alternative with the best level for one of the attributes included in the task. For example, if he/she has chosen the alternative with lowest price in all the choices, we will say that he/she has chosen lexicographic regarding to the price attribute. In similar manner we will investigate if he/she has “sorted” the choice task according to one of the other attributes. It is not investigated in this study how the lexicographic respondents made their choice if the level was equal on the attribute they used for sorting (see for example Foster and Mourato 1997 for the results of such a strong test). This simple test is used because, according to the discussion above, a stronger test will probably not detect all respondents that use a lexicographic choice algorithm to simplify the task.

More choices in the task means less “apparent” lexicographic choices

In tests of lexicographic choices it is important to be aware of the share of “natural” of “apparent” lexicography. The share of “apparent” lexicographic choices are determent by the probability that a random choice gives as a result that that particular choice is judged as lexicographic according to the procedure described above. For each discrete random choice this probability is 3/4 if the attribute has two levels and 2/3 if the attribute has three levels. Table 1 shows how the share of “apparent” lexicographic choices is reduced when the respondents are given more choices.

Table 1: The share of “apparent” lexicographic choices for an attribute in a choice task is reduced by the increased number of choices and increased number of levels of the attribute.

<table>
<thead>
<tr>
<th>No. Of discrete choices</th>
<th>Attribute, 2 levels</th>
<th>Attribute, 3 levels</th>
<th>Attribute, 4 levels</th>
</tr>
</thead>
</table>

17 By random choice we mean a choice made without considering the levels of the attributes, or just by chance selection (or lot).

18 The probability that the level of the attribute, of the randomly chosen alternative, is better or equal to the level of that same attribute of the alternative not chosen. This is the same algorithm that is used to investigate the share of lexicographic choices in the current study.

19 An attribute with two levels (high and low) gives four different combinations in the two choice alternatives: 1) low/low, 2) low/high, 3) high/low and 4) high/high. If we assume that these four combinations have equal probability as a choice in the choice task, a random choice will be classified as a lexicographic choice in 3 of 4 cases; i.e. better or equal. Correspondingly, an attribute with three levels gives nine different combinations and a random choice will be classified as lexicographic in 6 of 9 cases.
The share of “apparent” lexicographic choices in a real choice task will be reduced if dominant choices (i.e. choices where one of the alternatives are better or equal for all the attributes’ levels) are not part of the task. Dominant choices may be used as an introduction to the task (Bradley and Daly 1994) and/or for test purposes (Foster and Mourato 1997).

**Lexicographic choices are consistent choices**

By use of the least restrictive consistency test all the respondents who are making lexicographic choices according to the price attribute will be classified as consistent respondents. The reason for this is that the use of the price attribute as *numéraire* in the consistency test causes that choices with equal levels of the price attribute are inconclusive and therefore excluded from the test. A respondent who always chooses that alternative with the lowest price, will (in Figure 2) always lie to the left of the rays, i.e. all his/her choices are mutually consistent.

This unambiguous relation between lexicographic choices and consistency is not present for the attributes not chosen as *numéraire*. The reason for this is that choices where the levels for these factors are equal are included in the consistency test. Equal levels will appear either as vertical or horizontal rays in the raydiagram. Nevertheless, this may be a problem because the respondents who choose lexicographically may choose randomly if the levels of his sorting attribute are equal. Such choices therefore may, or may not, be detected as inconsistent in the consistency test.

To avoid this source of confusion, the respondents that have chosen lexicographically, thus violating the axiom of continuity, are detected and described first in the current study. These lexicographic respondents are removed before the consistency test. This procedure tries to ensure that the respondents included in the consistency test are those who have actually considered the levels of more than one attribute. The share of respondents in this group that have

<table>
<thead>
<tr>
<th></th>
<th>75.0 %</th>
<th>66.7 %</th>
<th>62.5 %</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>56.3 %</td>
<td>44.4 %</td>
<td>39.1 %</td>
</tr>
<tr>
<td>3</td>
<td>42.2 %</td>
<td>29.6 %</td>
<td>24.4 %</td>
</tr>
<tr>
<td>4</td>
<td>31.6 %</td>
<td>19.7 %</td>
<td>15.3 %</td>
</tr>
<tr>
<td>5</td>
<td>23.7 %</td>
<td>13.2 %</td>
<td>9.5 %</td>
</tr>
<tr>
<td>6</td>
<td>17.8 %</td>
<td>8.8 %</td>
<td>6.0 %</td>
</tr>
<tr>
<td>7</td>
<td>13.3 %</td>
<td>5.9 %</td>
<td>3.7 %</td>
</tr>
<tr>
<td>8</td>
<td>10.0 %</td>
<td>3.9 %</td>
<td>2.2 %</td>
</tr>
<tr>
<td>9</td>
<td>7.5 %</td>
<td>2.6 %</td>
<td>1.5 %</td>
</tr>
</tbody>
</table>

20 Unfortunately, we can not be absolutely certain that respondents who have not chosen lexicographically have considered the levels of more than one attribute. The reason for this is that some respondents may have simplified the decision by choosing for example, the journey on the
chosen inconsistently may give an indication of how cognitively demanding the choice task has been. There were no tests of violation of the non-satiation axiom in the data used in the current study; all inconsistencies detected are therefore due to violations of the axiom of transitivity.

IV. DESCRIPTION OF THE SURVEYS

Data from two different surveys are used in this paper to study lexicography and inconsistency in stated choice studies. The data from The Norwegian value of time survey is used as the main data source and the data from an environmental survey is used as a supplementary source. This section mainly describes how the choice tasks and the OE-CVM questions were formulated in the value of time survey and only briefly presents the first choice task in the environmental study.  

*The Norwegian value of time survey*

The Norwegian value of time survey consisted of personal interviews undertaken in 1994-1996 of car drivers and travellers by public transport. The respondents were randomly drawn from the population in the largest cities in Norway. 2568 of the interviews made in the value of time survey are used in the current study. Respondents who recently had undertaken a private journey first answered questions in the stated choice experiment and were then asked more directly about their willingness to pay for a reduction of 25% in the travel time on that particular journey. Four typical choices from the stated choice sequence are shown in Figure 1. The direct questions used to find the respondents' willingness to pay for reduced travel time on the particular car journey were formulated as so-called transfer price questions. What has been called transfer price questions in the context of value of time estimation are very similar to OE-CVM questions (see, for example, Layard and Glaister 1994 p.258 for an outline). The respondents should be quite familiar with their preferences for reduced travel time on a left hand side on all choices or by choosing alternately “left, right, left,…”. By the use of such choice strategies the respondents are not judging the levels of the attributes. Tests for these types of simplification strategies have not been carried out in this study.

---

21 Only the first choice task in the environmental study is used in the current study. This task is similar (the same attributes and the same context) to the choice tasks in the value of time study. The complete questionnaires from both studies are available from the author on request.

22 The value of time survey also includes data for business travellers and travellers with air and ferry. These data are not included in the current study, but valuations from these travellers are reported in Ramjerdi et al. (1997).

23 They were also asked about their willingness to pay for a reduction of 10% in the travel time on that particular journey and about what compensation that was needed if the travel time were increased with 25%. The results from these questions are reported in Ramjerdi et al. (1997).
particular journey when they were presented the OE-CVM questions, since it followed the choice experiment.

In the stated choice task the respondents were to choose between two journeys using the travel mode they had used on the particular journey. The journeys were described by three attributes: price, in-vehicle travel time and a third attribute. The third attribute was for long (more than 50 km) car journeys “automatic speed control” or so-called “photo-boxes” and for short car journeys the third attribute was “chance of delay”. In the choice tasks presented to travellers by public transport the third attribute was “headway” or “chance of delay”. In addition to the actual mode choice task, the respondents were also given a choice task between two journeys with their most preferred alternative transport mode. In this task all the travellers with car had to choose between two journeys with public transport.

The survey was administered using the MINT computer-assisted personal interview (CAPI) software. Three design variables were used, each with three or four levels. By use of SPEED an orthogonal fractional factorial design of sixteen alternatives was randomly selected for each respondent from the full factorial design of all possible combinations. Respondents were presented with a series of nine pairwise choices from among these sixteen options. Pairs of alternatives were presented in random order. Binary logit models were estimated by ALOGIT on the pairwise choice data, using linear functions of the three design variables.24

An environmental survey

The environmental survey consisted of 1680 personal interviews undertaken in 1993 of car drivers and travellers by public transport randomly drawn from the population in Oslo and the neighbouring county of Akershus. Respondents who recently had undertaken a private local journey first answered the questions on a simple stated choice task and then they were given more complicated choice tasks also including environmental attributes. Only the first choice task, completed by 1473 respondents, is used in the current study.

In the stated choice task the respondents were to choose between two journeys with the travel mode they had used on the particular journey. The journeys were described by three attributes: price, in-vehicle travel time and “walking time to parking place” for car travellers, and “chance of seating” for travellers by public transport. As in the travel time study, the environmental survey was administered using MINT. A similar design procedure was used and similar binary logit models were used for estimation. One important difference between the two surveys is

24 MINT, SPEED and ALOGIT are all software from Hague Consulting Group.
that respondents were presented with a series of only four pairwise choices in the first choice task of the environmental survey.

IV. THE RESULTS

This chapter first gives an overview of the occurrence of lexicographic choices and the impact of such choices of the valuation of reduced travel time based on the discrete choice data in a logit model. Thereafter, the reason for such choices is discussed by use of the valuation of travel time based on OE-CVM data from the same respondents. In the next part of the chapter the same procedure is used in the investigation of inconsistent choices. The last part of the chapter investigates whether the valuations from OE-CVM show systematic differences between respondents with consistent and inconsistent choices in the stated choice data, and if the occurrence of inconsistent choices differs during the choice sequence.

Occurrence and the effects of lexicographic choices

Table 2 shows that there is a larger share of lexicographic choices for long journeys than for short journeys. The most used “sorting attributes” for short car journeys and long car journeys are price and travel time, respectively. Price is mostly used as a “sorting attribute” both for short and long journeys by public transport.

Table 2: The share of lexicographic choices for the different attributes (used as a “sorting attribute” by the respondents) and modes.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Long journeys (more than 50 km)</th>
<th>Short journeys (less than 50 km)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Car (n=685)</td>
<td>Coach (n=440)</td>
</tr>
<tr>
<td>Price</td>
<td>7.6</td>
<td>11.4</td>
</tr>
<tr>
<td>Travel time</td>
<td>18.1</td>
<td>5.0</td>
</tr>
<tr>
<td>3. attribute*</td>
<td>2.6</td>
<td>8.2</td>
</tr>
<tr>
<td>Total share</td>
<td>28.3</td>
<td>24.5</td>
</tr>
</tbody>
</table>

*For the car mode of travel the third attribute is automatic speed control for long journeys and chance of delay for short journeys. For public transport the third attribute is headway for long journeys and headway or chance of delay for short journeys.

According to the results in Table 3, the lexicographic choices in the discrete choice data increase the valuation of travel time for long car journeys. This is a natural consequence of the fact that travel time is the main “sorting attribute” for long car journeys. For long journeys with public transport and for short journeys the lexicographic choices decrease the valuation of travel time. This result is
explained by the fact that price is the main “sorting attribute” for these journeys. For long journeys with coach and for short journeys it is not possible to estimate a valuation of travel time for the respondents that have chosen lexicographically. The reason for this is that relatively few of these respondents have used travel time as the “sorting attribute”, and consequently too few respondents in these groups show willingness to pay for reduced travel time in the stated choice data.
Table 3: Valuation of travel time from the discrete choice data. Unit: NOK/h. Standard error in parenthesis.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Long journeys (more than 50 km)</th>
<th>Short journeys (less than 50 km)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Car</td>
<td>Coach</td>
</tr>
<tr>
<td>All respondents</td>
<td>85 (3.1)</td>
<td>35 (4.1)</td>
</tr>
<tr>
<td>Respondents that chose lex.</td>
<td>120 (12.4)</td>
<td>Wrong sign</td>
</tr>
<tr>
<td>Respondents that don’t chose lex.</td>
<td>76 (2.9)</td>
<td>44 (4.2)</td>
</tr>
</tbody>
</table>

Lexicographic choices are partly a result of real preferences

As described, lexicographic choices may be a result of simplification of the choice task or a result of real preferences. Respondents that use the price attribute as a “sorting attribute” (i.e. always choose the alternative with lowest price) because they want to simplify the task will, in the analysis of the stated choice data, get a lower valuation (if it is possible to get a valuation) than their correct valuation. Respondents that use the price attribute as a “sorting attribute” because they are not willing to pay the amount that is the difference between the two alternatives will, in the analysis of the stated choice data, get a higher valuation than their correct valuation. To decide if a specific respondent has chosen lexicographically because of simplification or real preferences one must know his/her correct valuation.

If one assumes that the respondents’ correct valuations can be found by use of the OE-CVM questions, these data can be used as an indication of the reason for lexicographic choices. If the respondents that have chosen lexicographically by use of the price as the “sorting attribute” have done this because this is according to their real preferences, one can expect that the valuation from the OE-CVM data is lower for this group than for the group that have not chosen lexicographically. If real preferences have caused that other attributes are the “sorting attribute” one can expect that the valuation from the OE-CVM data is higher for this group than for the group that have not chosen lexicographically. If the lexicographic choices are caused by simplification, one can expect that the valuation from the OE-CVM data to be about the same for the lexicographic group and the non-lexicographic group. These expectations rely on the assumption that other factors are not influencing the valuation.

As can be seen from Table 4, respondents that have used price as the “sorting attribute” (denoted: lex. price) have a significantly lower valuation of travel time than respondents that made non-lexicographic choices (denoted: non-lex.). (Short journeys by public transport is an exception.) It can also be seen that respondents who have used travel time as the “sorting attribute” (denoted: lex. travel time) for most of the modes have a significantly higher valuation of travel time than
respondents who made non-lexicographic choices. (The exception this time is short journeys with car where only 2 respondents have used travel time as a “sorting attribute” and both of them have a WTP of 0 in the OE-CVM questions.)

A conclusion from this investigation is that lexicographic choices in this study are at least partly a result of the respondents’ real preferences. This conclusion is somewhat weak because of the fact that more than half of the respondents who have used price as a sorting attribute have a WTP larger than zero in the OE-CVM questions, but did not expressed their WTP in the choice task by not making lexicographic choices. One explanation may be that some of the respondents used the choice task to determine their preferences. Another explanation may be that stated choice tasks capture respondents’ relative valuation rather than their absolute valuation (Roe et al. 1997 and Sælensminde 1998.)

Table 4: Valuation of travel time*,** from the OE-CVM data. Unit: NOK/h. Standard error presented in parenthesis.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Long journeys (more than 50 km)</th>
<th></th>
<th>Short journeys (less than 50 km)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Car (2.3)</td>
<td></td>
<td>Coach (1.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coach (3.1)</td>
<td></td>
<td>Train (3.2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Train (3.6)</td>
<td></td>
<td>Car (2.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Public transport (1.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>62 (2.3)</td>
<td></td>
<td>38 (1.7)</td>
<td></td>
</tr>
<tr>
<td>Lex. total</td>
<td>75 (5.4)</td>
<td></td>
<td>33 (3.1)</td>
<td></td>
</tr>
<tr>
<td>Lex. price</td>
<td>25 (4.3)</td>
<td></td>
<td>39 (3.2)</td>
<td></td>
</tr>
<tr>
<td>Lex. travel time</td>
<td>98 (6.9)</td>
<td></td>
<td>47 (8.3)</td>
<td></td>
</tr>
<tr>
<td>Non-lex.</td>
<td>57 (2.3)</td>
<td></td>
<td>39 (2.0)</td>
<td></td>
</tr>
</tbody>
</table>

* The valuation of travel time is an average of WTP to obtain a 25% reduction in travel time and WTP to avoid a 25% increase in travel time. This makes the results from the OE-CVM data comparable to the results from analysis of the discrete choice data.

** Respondents that used headway as a sorting factor have a higher valuation of headway than respondents that did not make lexicographic choices. This is not showed in the table.

Again, this shows that real preferences that cause a lexicographic choice procedure in a choice task are not the same as lexicographic preferences. It is also important to stress that even if these data show that the lexicographic choices are partly a result of real preferences, this result can hardly be generalised. The reason for this is that the share of lexicographic choices that is caused by simplification probably will increase with the number of attributes in the task and if the respondents have less a priori knowledge of the attributes. In the current study the choice task involves only three attributes and those three are well known to the respondents. It is therefore expected that these choice tasks are considered as relatively easy by most of the respondents and that they don’t find it necessary to

---

25 In addition, this result reveals an inconsistency between the discrete choice data and the OE-CVM data. Such inconsistencies are not discussed in the current paper.
simplify the choice task by choosing lexicographically. This view is supported by findings in Mazzotta and Opaluch (1995) concerning the C-D gap referred above.

*The occurrence of inconsistent choices*

As explained, in the tests of consistency undertaken in the current study all respondents that have made lexicographic choices are removed before the tests. A respondent has either chosen such that all his choices are mutually consistent (i.e. zero inconsistent choices) or such that one or more choices are inconsistent with other choices. For example, if choices 7 and 9 are not consistent with the previous choices 1, 2, 3, 4, 5 and 6, the test will show that two choices are inconsistent with previous choices (i.e. IPC=2). But the test will also show that six choices are inconsistent with subsequent choices (i.e. ISC=6). In Table 5 the number of inconsistent choices for each respondent is defined as the minimum of IPC and ISC.

Table 5: The share of respondents with a different number of inconsistent choices and the average number of inconsistent choices for different modes. (Respondents that have chosen lexicographically are removed before the consistency test.)

<table>
<thead>
<tr>
<th>Number of inconsistent choices</th>
<th>Long journeys (more than 50 km)</th>
<th>Short journeys (less than 50 km)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Car (n=490)</td>
<td>Coach (n=330)</td>
</tr>
<tr>
<td>0</td>
<td>33.1</td>
<td>25.2</td>
</tr>
<tr>
<td>1</td>
<td>39.0</td>
<td>31.8</td>
</tr>
<tr>
<td>2</td>
<td>16.9</td>
<td>24.2</td>
</tr>
<tr>
<td>3</td>
<td>7.6</td>
<td>11.8</td>
</tr>
<tr>
<td>4</td>
<td>2.4</td>
<td>5.9</td>
</tr>
<tr>
<td>5</td>
<td>1.0</td>
<td>0.9</td>
</tr>
<tr>
<td>6</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average number of inconsistent choices</td>
<td>1.10</td>
<td>1.45</td>
</tr>
</tbody>
</table>

For long journeys, Table 5 shows that car travellers chose more consistently than travellers by public transport. The reason for this may be that longer car journeys are undertaken more frequently than similar journeys by public transport and that car travellers therefore have more knowledge about their preferences than travellers by public transport.
For short journeys it is shown that car travellers chose less consistently than travellers by public transport. The reason for this is probably that the attribute “price per journey” is a common attribute in public transport, but less familiarly associated with local car journeys.

Table 5 shows a large difference in consistency between long and short journeys by public transport. The main explanation for this result is probably differences in travel frequency between short and long journeys and therefore a much clearer preference structure (i.e. knowledge about the goods in the choice task) for short journeys than for long journeys.

**Cognitive ability affects the valuation from the discrete choice data**

Table 6: Valuation of travel time from the discrete choice data for respondents with a different number of inconsistent choices. Unit: NOK/h. Standard error presented in parenthesis. (Respondents that have chosen lexicographically are removed before the consistency test.)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Long journeys (more than 50 km)</th>
<th>Short journeys (less than 50 km)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Car</td>
<td>Coach</td>
</tr>
<tr>
<td>All non-lex. respondents</td>
<td>76 (2.9)</td>
<td>44 (4.2)</td>
</tr>
<tr>
<td>Non-lex. and 0 incons. choices</td>
<td>71 (3.4)</td>
<td>37 (4.4)</td>
</tr>
<tr>
<td>Non-lex. and 1 incons. choices</td>
<td>69 (3.9)</td>
<td>16 (9.4)</td>
</tr>
<tr>
<td>Non-lex. and 2 incons. choices</td>
<td>89 (9.0)</td>
<td>43 (13.5)</td>
</tr>
<tr>
<td>Non-lex. and 3+ incons. choices</td>
<td>95 (21.4)</td>
<td>101 (14.8)</td>
</tr>
</tbody>
</table>

* All data are included in the analysis.
* Respondents with I) headway as third attribute in the exercise and 2 inconsistent choices and II) delay as third attribute in the exercise and 0 inconsistent choices are excluded from the analysis. This is done because these groups cause problems (wrong signs and not significant parameters) for the analysis where all data are included. The result is a “pattern” more similar to the other modes.

The valuations of travel time savings from the discrete choice data in table 6 show significant differences between respondents that have chosen consistently (denoted: Non-lex. and 0 incons. choices) and respondents that have one or more inconsistent choices. This is a general result for short journeys, but also longer journeys by public transport show significant differences between the group with 3 or more inconsistent choices and those groups with no inconsistencies. For longer car journeys the valuations show the same pattern, but the differences between the groups are not so large and not significant.

The differences in valuations between the groups in Table 6 may be explained by differences in real preferences or by differences in the cognitive ability to express...
real preferences in the discrete choice task the respondents were presented. The valuations from OE-CVM in Table 7 show no systematic differences between respondents with consistent and inconsistent choices in the stated choice data. This result indicates that there is no difference in real preferences between these groups, and that differences in the cognitive ability to express the real preferences in the discrete choice task seems like the most reasonable explanation to the differences shown in Table 6.²⁶

Table 7: Valuation* of travel time from the OE-CVM data. Unit: NOK/h. Standard error presented in parenthesis. (Respondents that have chosen lexicographic are removed before the consistency test.)

<table>
<thead>
<tr>
<th>Sample</th>
<th>Long journeys (more than 50 km)</th>
<th>Short journeys (less than 50 km)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Car</td>
<td>Coach</td>
</tr>
<tr>
<td>All non-lex. respondents</td>
<td>57 (2.3)</td>
<td>39 (2.0)</td>
</tr>
<tr>
<td>Non-lex. and 0 incons. choices</td>
<td>62 (4.3)</td>
<td>35 (2.6)</td>
</tr>
<tr>
<td>Non-lex. and 1 incons. Choices</td>
<td>54 (3.4)</td>
<td>38 (4.4)</td>
</tr>
<tr>
<td>Non-lex. and 2 incons. choices</td>
<td>53 (5.0)</td>
<td>42 (4.4)</td>
</tr>
<tr>
<td>Non-lex. and 3+ incons. choices</td>
<td>59 (7.6)</td>
<td>43 (4.1)</td>
</tr>
</tbody>
</table>

* The valuation of travel time is an average of WTP to obtain a 25% reduction in travel time and WTP to avoid a 25% increase in travel time. This makes the results from the OE-CVM data comparable to the results from analysis of the discrete choice data.

Relationships between observable respondent characteristics and the tendency to choose lexicographically and inconsistently

²⁶ Table 6 shows that inconsistent choices disturb model results and increase the valuation of travel time from the stated choice data. Together with the result that the occurrence of inconsistent choices is largest first in the choice sequence (Table 9), one possible solution to the problem with inconsistent choices might be to exclude the first (one or two) choices from the analysis. After analysis based on each choice separately it was decided to exclude the two first choices from the analysis because these choices cause insignificant parameters. The valuation of travel time (NOK/hour) for “all non-lex. respondents” when the two first choices are excluded are 81, 47, 58, 37, 19 and 34 for the modes from left to right in Table 6, respectively. These results are still above the valuations from the group with 0 inconsistent choices (and the valuation from the OE-CVM) and the reason is probably that there are enough inconsistencies left in the remaining choices to affect the valuations. These analyses suggest that the exclusion of the first choices in the choice sequence seems to be no solution to the problem caused by inconsistent choices in stated choice data.
If inconsistent choices in general cause significant effects on the valuation from discrete choice analysis, it is important to discuss whether such choices should be deleted from the analysis. It might be objected that the removal of “noisy” responses from the survey sample potentially creates a different kind of problem – one of self-selection bias – if test failures are systematically related to observable respondent characteristics. If this were so, it would be a case of trading-off the bias created by the inclusion of “noisy” responses against the bias created by the loss of population representation in the “cleaned” sample. Unfortunately, the data from the value of time study do not include education, which is an important variable for testing whether a relationship exists between observable respondent characteristics and tendency to choose lexicographically or inconsistently. However, by use of the first choice task from an environmental study that is similar to the choice tasks in the value of time study, this relationship can be investigated. The results from logistic regressions to explain lexicographic and inconsistent choices by socioeconomic variables in a stated choice task are reported in Table 8.

Table 8: Logistic regression to explain lexicographic and inconsistent choices by socioeconomic variables in a stated choice task.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Logit model for lexicographic choices</th>
<th>Logit model for inconsistent choices</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parameter estimate</td>
<td>Standard error</td>
</tr>
<tr>
<td>Gender (0=female, 1=male)</td>
<td>-0.1539</td>
<td>0.1188</td>
</tr>
<tr>
<td>Age</td>
<td>-0.0053</td>
<td>0.0051</td>
</tr>
<tr>
<td>Income (in 1000 NOK)</td>
<td>-0.0860</td>
<td>0.0001</td>
</tr>
<tr>
<td>Education (no. of years)</td>
<td>-0.0534</td>
<td>0.0193</td>
</tr>
<tr>
<td>Pensioner (0=no, 1=yes)</td>
<td>0.2097</td>
<td>0.2414</td>
</tr>
<tr>
<td>Difficult to choose? (0=no, 1=yes) *</td>
<td>-0.1427</td>
<td>0.1575</td>
</tr>
<tr>
<td>Difficult to concentrate? (0=no, 1=yes) *</td>
<td>0.3399</td>
<td>0.1614</td>
</tr>
<tr>
<td>Constant</td>
<td>1.7823</td>
<td>0.4297</td>
</tr>
<tr>
<td></td>
<td>n=1473</td>
<td></td>
</tr>
</tbody>
</table>

* Answers to control questions asked after the respondents had completed the choice task.
Results from OE-CVM questions may also be used as an explanatory variable in the logit model for lexicographic choices, but OE-CVM questions, valuing reduced travel time, were not included in the environmental survey. However, by use of the data from the value of time survey, logistic regression models including answers to the OE-CVM questions are used to explain lexicographic choices. These models confirm the above results that real preferences are also an important explanation why respondents make lexicographic choices in this particular case. Given space limitations, details of these models are not reported here.

Table 8 shows strong statistically significant relationships between the variables “education” and “difficult to concentrate” and the tendency to choose lexicographic in this survey. In addition, the variable education was the only significant explanatory variable in the model to explain inconsistent choices. This result is in contrast to the results from Foster and Mourato (1997) and Johnson and Desvousges (1997), and is probably achieved as a result of more comprehensive tests for “noisy” choices in the current study. Therefore, in the context of the present sample, the potential problems created by removal of “noisy” responses from the survey may be an issue for further research.

The occurrence of inconsistent choices is large first in the choice sequence

If the choice task is difficult to understand at first glance one can imagine that the respondents need some kind of education or training in how to choose such that they are able to express their real preferences in the framework presented to them. If this is the case, that the respondents need more training in the choice process, one will expect that the occurrence of inconsistencies is larger among the first choices in the choice sequence than among later choices. If the choice task is cognitive demanding and/or the choice sequence is long, one might suspect that this could result in fatigue effects. In the presence of fatigue effects one would expect that the occurrence of inconsistent choices becomes larger later in the choice sequence. One might see a combination of both learning and fatigue effects in stated choice data.

In the current study consistency tests are made both for choice task 1, nine choices between two journeys with the respondents’ actual mode, and choice task 2, nine choices between two journeys with an alternative mode. The results of these tests are presented in Table 9. (As in Table 5, the share of inconsistent choices in Table 9 is defined as the minimum of IPC and ISC.)

27 These results are achieved despite the fact that only four choices in the environmental survey data cause more “apparent” lexicographic choices. This strengthens the hypotheses that simplification also is an important reason why respondents make lexicographic choices.
For public transport (both long and short journeys) and for short car journeys Table 9 shows that the share of inconsistent choices is less for choice task 2 (alternative mode) than for choice task 1 (actual mode). Such results suggest that it is the choice task itself that is problematic for the respondents and that it is not particularly more difficult to make such choices between modes they use more seldom. For long car journeys there are more inconsistent choices in task 2 than task 1. A possible explanation is that respondents who have used a car on a long journey very seldom, or never, use public transport for such journeys. The unfamiliar context in choice task 2 presented to car users on long journeys seems therefore to increase the choice problems more than these are reduced by the learning process.

Table 9: The share of inconsistent choices for each choice number in the choice sequence. Choice task 1 (=actual mode) / Choice task 2 (=alternative mode). (Respondents that have chosen lexicographically are removed before the consistency test.)

<table>
<thead>
<tr>
<th>Choice number</th>
<th>Long journeys (more than 50 km)</th>
<th>Short journeys (less than 50 km)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Car (n=490/492)</td>
<td>Coach (n=330/328)</td>
</tr>
<tr>
<td>1</td>
<td>25 / 37</td>
<td>35 / 29</td>
</tr>
<tr>
<td>2</td>
<td>18 / 17</td>
<td>23 / 17</td>
</tr>
<tr>
<td>3</td>
<td>14 / 17</td>
<td>16 / 13</td>
</tr>
<tr>
<td>4</td>
<td>9 / 14</td>
<td>16 / 13</td>
</tr>
<tr>
<td>5</td>
<td>10 / 16</td>
<td>14 / 13</td>
</tr>
<tr>
<td>6</td>
<td>10 / 11</td>
<td>9 / 12</td>
</tr>
<tr>
<td>7</td>
<td>9 / 11</td>
<td>12 / 9</td>
</tr>
<tr>
<td>8</td>
<td>8 / 9</td>
<td>10 / 6</td>
</tr>
<tr>
<td>9</td>
<td>8 / 6</td>
<td>11 / 8</td>
</tr>
</tbody>
</table>

Table 9 also shows a declining share of inconsistency for both choice tasks 1 and 2. This indicates that training is necessary in order to achieve less inconsistent choices and that fatigue is no problem in this study. The relatively short duration of the interview (15-20 minutes), combined with a relatively easy choice task possibly explains why we observe little sign of fatigue effects in this study.

The share of inconsistent choices in Table 9 is defined as the minimum of IPC and ISC. This is the same way as the number of inconsistent choices for each respondent is defined in Table 5. If both IPC and ISC are tabled one can see a possible fatigue effect, but the learning problem first in the choice sequence seems to dominates later fatigue problems.
V. CONCLUDING REMARKS

In this paper a test procedure that gives the possibility of identifying each respondent’s mutual inconsistent choices is presented. Together with a simple test for lexicographic choices this gives the opportunity to examine “problematic choices” more thoroughly than by other approaches. The empirical analysis illustrates that lexicographic and inconsistent choices are common in several stated choice tasks and that such choices have a significant impact on the valuation of reduced travel time. Other important findings are that both different abilities to choose and real preferences cause lexicographic choices. Different abilities to choose are also an important explanation of inconsistent choices. The occurrence of inconsistent choices is shown to be largest at the beginning of the choice sequence, and is reduced for later choices. These results suggest that the respondents may need more training and help to choose consistently in stated choice studies.

In a discussion of those implications resulting from this and similar studies, upon the direction of further research, it is probably wise to commence with possible explanations of “noisy choices” in stated choice studies. Unfortunately, we can not assume that genuine irrational individuals and people unwilling to complete the task in a proper manner are not included in stated choice surveys. Such respondents may use the choice task to give protest answers, but their choices will probably be either lexicographic or inconsistent. They can therefore be detected in the test procedure used in the current study. If this was the only reason why “problematic choices” occur these choices could easily be deleted from the sample. Unfortunately, other possible explanations why lexicographic and inconsistent choices occur may be that the respondents a) have difficulties with the task, b) have unstable or ill defined preferences, c) learn about their actual preferences during the task, d) are indifferent to the presented options, e) become fatigued during the task, or f) have preferences not covered by the presented alternatives. Fortunately, the occurrence of “problematic choices” caused by the latter reasons may be reduced by improvements in study design, by giving more help to the respondents during the task and by use of better statistical models.

The results of this, and other studies, show that the first choices may be problematic. It is therefore a good idea to give the respondents a few learning choices to ensure that they understand the task (Bradley and Daly 1994). Another attempt that may ease the task for the respondents is to undertake tests of lexicography and consistency during the choice task similar to that undertaken in multi-attribute utility theory (MAUT) (see for example Baron and Greene 1996, Gregory et al. 1994 and Loomes 1994). This will probably help the respondents to answer consistently according to their real preferences, but may result in a more time consuming data collection procedure. Letting the respondents become more aware of their preferences before the choice task begins may reduce problems caused by unstable or ill-defined preferences and learning about own preferences.
during the task. This may be done by use of simple questions about how they prefer the actual goods compared to other goods, if they are willing to pay for improvements in the actual goods quality at all and possible by use of OE-CVM questions like those used in the current study.

Indifference problems in the choice task may be reduced by giving the respondent the opportunity to choose that he/she is “indifferent” to the presented alternatives and/or to state that he/she “don’t know” which alternative he/she prefers. By giving such choice alternatives one can probably reduce the number of “noisy choices”, but if the respondents use these alternatives as an easy way to complete the task one will probably also reduce the number of informative choices. If “noisy choices” generally cause such large problems as in the current study, the inclusion of “indifferent” and/or “don’t know” as choice alternatives will probably be an improvement of stated choice studies. If the respondents that choose “indifferent” really are indifferent or use this choice alternative as a simplification of the task, this can be detected in the test procedure used in the current study.

Preference uncertainty may also be explicitly modelled in the analysis of the data. By use of a follow-up certainty question or by letting the respondents directly indicate how certain they are of their choice one can incorporate the respondents’ uncertainty in the analysis. Li and Mattson (1995) use such an approach in a discrete choice CVM study. They conclude that ignorance of preference uncertainty would lead to a seriously upward-biased mean estimate. The fact that DC-CVM and stated choice are related methods makes this an interesting result. If those respondents that make inconsistent choices state that they are more uncertain about their choices in the choice task than respondents that choose consistently, then this approach will result in less weight on inconsistent choices (i.e. respondents with task problems) in the analysis of the data. Preference uncertainty may therefore be a possible explanation of why respondents that chose inconsistently had a higher valuation of travel time in the current stated choice study. Champ et al. (1997) state that use of a follow-up certainty question is a promising approach as a means of providing a lower bound to the theoretical construct. They conclude that respondents who answer consistently are more certain of their real preferences.

As illustrated in the current study, “noise” in a stated choice study is more than unexplained variation in a statistical model. If one investigates a data sample as a whole, it will in most cases seem both non-lexicographic and mostly consistent. This fact that one person’s lexicographic or inconsistent choices together with other persons’ choices seems “reasonable” results in very few of the lexicographic and inconsistent choices being detected as model “outliers”; i.e. choices with low probability of the observed choice. Therefore, it is probably not enough to design statistical models without a thorough investigation of each respondent’s choices.
In the data collection we should use designs that collect more signals and less noise, and in the data analysis we should use models that can separate signals from noise. (In order to deal with these issues Swait and Adamowicz 1996 present a promising approach.) But there is obviously a limit regarding how much noise can be separated from the signals in a model and therefore how much noise such models can stand and yet produce a useful valuation of the goods of interest. For further investigation of such limits a direct test procedure which can investigate if each respondent has completed the task in a proper manner, like the one presented in the current study, is valuable. It is valuable because it can be used to investigate how different choice complexity levels influence the share of “problematic choices” and therefore produce input to the discussion of “optimal complexity levels”. And it is valuable because it can be used to investigate how much of the “problematic choices” are detected, and if they are handled satisfactorily by statistical models which claim to “separate signals from noise”.

Almost all empirical choice modelling work in the literature assumes that individuals behave in a compensatory (i.e. non-lexicographic and consistent) fashion. Further, the specification of choice models tends to assume a utility maximising, full information, indefatigable decision-maker who is able to assign values to alternatives, and choose the alternatives with the highest value, independent of context. If the occurrence of “problematic choices” is generally large in stated choice studies, it is questionable whether the data meet the assumptions upon which the analysis relies. It is therefore important to detect whether limits in people’s cognitive abilities make some data collection methods, or complexity levels of such methods, unsuitable as a framework for people to state their preferences. It may, for example, be a problem if only highly educated people are able to choose in a consistent manner in stated choice studies and that their preferences therefore will count more than that of others.

REFERENCES


Baron, J. and J. Greene (1996): Determinants of Insensitivity to Quantity in Valuation of Public Goods: Contribution, Warm Glow, Budget Constraints,


The effect of heteroscedasticity, *Forthcoming in Land Economics*.


