Lexicographic answering in travel choice: Insufficient scale extensions and steep indifference curves?

Marit Killi[*], Åse Nossum and Knut Veisten
Institute of Transport Economics
Department of Economics and Logistics
Gaustadalleen 21, NO-0349
Oslo
Norway
tel: +47 22 57 38 41
fax: +47 22 60 92 00
e-mail: mki@toi.no

EJTIR, **7**, no. 1 (2007), pp. 39-62

Received: June 2005
Accepted: January 2006

This study assesses lexicographic answering in stated choice surveys of travel alternatives. Respondents answering lexicographically in three different data sets are analysed in relation to how important they found the attributes that dominated their choices. Lexicographic answering is also regressed against covariates indicating commuting situation and socio-economic status.
A larger share of those answering lexicographically in relation to one specific attribute stated that this attribute was decisive in their choice compared to the share trading-off attribute levels in choices. Furthermore, a large majority of those answering lexicographically stated that the difference between the highest and lowest values of the attribute, according to which they chose lexicographically, was "very important". Relevant variables explained lexicographic answering in a logistic regression analysis, e.g. that the probability of lexicographic answering with respect to travel time increases with income and travel distance.
Response strategies other than neo-classical trade-off, e.g. simplification with a focus on one attribute alone, cannot be ruled out. However, the results indicate that lexicographic answering is due primarily to steep indifference curves in combination with insufficient attribute scale extension. These findings have implications for choice design and for the treatment of respondents answering lexicographically.

Keywords: design, indifference curves, lexicographic answering, scale extension, stated choice

---

* Corresponding author

1. Introduction

In stated preference surveys on mode of transport, respondents typically face options that differ in travel time, comfort, transfer, fare and so on (Hensher, 1994; Hensher et al., 1988). With repeated choices and change of attribute levels it is anticipated that respondents will trade-off between attributes, enabling an indirect marginal valuation of time, comfort and transfer. However, some respondents answer lexicographically, e.g. they always choose the option with minimum transfer notwithstanding alterations in the level of the other attributes (Sælensminde, 1999; 2001). In this article, we use "lexicographic answers" or "lexicographic answering" as generic terms for a lexicographic choice we observe independently of the preferences, strategies or coincidences that may underlie such lexicographic answers.

The fact that respondents do not trade-off between attributes/options may have implications for the analysis of a specific survey as well as for the general design features of stated preference surveys. According to the standard economic definition of preferences, individuals *should* trade-off (Varian, 1992). They should trade-off even though a tiny reduction in one good must be compensated for by a huge increase in another, i.e. having so-called "steep indifference curves" (Kriström, 1997). If respondents for some reason or another do not trade-off, the corresponding estimation of the attributes' marginal value may be systematically biased (Rouwendal and Blaeij, 2004; Sælensminde, 2001). If one chooses to drop those identified with lexicographic answering from analysis, then data information and statistical power are simply lost. If those dropped in reality have a strong preference for the attribute that governs their lexicographic answering, the parameter estimates may be biased (Johnson and Desvousges, 1997; Lanscar and Louviere, 2006).

In the literature on economics and psychology, lexicographic answering has been assessed mostly as a decision rule that respondents apply when choices are complex (Heiner, 1983; Mazzotta and Opaluch, 1995; Payne et al., 1993; Tversky, 1972; Tversky et al., 1988). Nevertheless, within economics, evaluations have been made as to whether lexicographic answering may indicate real lexicographic preferences (Spash, 2000; Stern, 2000; Sælensminde, 2000; 2006; Veisten et al., 2006). The recent literature indicates that lexicographic answering may have design deficiencies. Following-up on those answering lexicographically with increased compensation for the attribute/good driving the choice may increase the share trading-off (Lockwood, 1999; Rosenberger et al., 2003; Cairns and Van der Pol, 2004; Hojman et al., 2004).

Based on a common analysis of three different stated choice (SC) surveys, we look closely at the issue of "steep indifference curves" and unbalanced design as potential sources of lexicographic answering. The application of this basic term from economics assessing lexicographic answering in SC is novel. One survey comprised car drivers' choices among trip options, with a specific interest in new forms of information about reducing travel time and uncertainty due to delays. Another survey treated public transport passengers' choices among options involving travel time and various alternative quality improvements. The final survey involved public transport alternatives considered by (existing) bus passengers, in addition to a survey for cyclists and car users. All three surveys were carried out using the Internet. Although different approaches were taken and different groups addressed, the studies provide a fairly consistent common assessment of lexicographic answering.

The rest of the article is arranged as follows: In the next section we describe the various reasons behind lexicographic answering and the hypotheses for discriminating between them.

In the third section we describe the three data sets and in the fourth present the results of the data analysis, which are discussed and concluded in the final two sections.

2. Identifying lexicographic answering in stated choice

2.1 Lexicographic answering

Lexicographic answering in stated preference studies implies that the respondent consistently chooses the alternative that is best with respect to one particular attribute (e.g. not trading-off the alternative with lowest travel time against any other alternative). According to Carlsson and Martinsson (2001), lexicographic answering may principally indicate lexicographic *preferences*, steep indifference curves, or a simplification strategy.

In principle, lexicographic answering may follow from lexicographic preference for (at least) one of the attributes/goods in the choice set. For example, the respondent would always prefer the option with lowest price even if the alternative option was just a little more expensive and far better in any other attribute. True lexicographic preferences have been related to basic needs or environmental amenities (Stern, 2000; Spash, 2000; Stevens et al., 1991). A more general and realistic model of lexicographic preference than the typical textbook model (Varian, 1992) is a threshold model that encompasses both lexicographic choices and trade-offs, depending on the fulfilment of the thresholds (Georgescu-Roegen, 1954). A recent application to SC data concludes that a modelling approach adapted to such possible thresholds will yield more correct estimations than multinomial logit or mixed logit models (Cantillo and Ortúzar, 2005). However, safety (accidents) was the main attribute at the non-compensatory threshold stage of the model, while travel time and toll (price) were more relevant at the compensatory trade-off stage. In other cases, if the attribute set contains the only one that matters for the respondent, this could be described as a case of lexicographic *preference*, but it could also be regarded as a design problem of omitted variables (other relevant attributes).

Lexicographic answering may follow from a relatively strong preference for (at least) one of the attributes/goods in the choice set, or "steep indifference curves", i.e. the given interval for this strongly preferred attribute may be too limited for trade-offs with other attributes to be relevant. The adaptation to bundles of attributes/goods may differ greatly from one individual to the next. When presenting choice tasks, the individual's perception of balance between attribute levels will also differ. Although two individuals have a similar current travel pattern (between home and work), they may want very different compensation if one trip attribute deteriorates. Some of the attributes in the trip choice are public goods or, in other terms, are perceived as fixed, e.g. safety, bus service frequency, and even the individual's domicile. The provision of public goods for each individual, be it transport facilities, public health or other goods, is not optimised for each individual in the same (theoretical) sense as private divisible goods are (Kriström, 1997). An individual can only optimise her own use of public transport, not its provision. If an individual finds that the public provision of bus service frequency is too low (steep indifference curve between bus frequency and any other attribute), it may be difficult to adapt choices in a stated preference survey to this particular individual.

Figure 1 depicts a situation with steep indifference curves between attribute *x* and attribute *y*, and the attribute range for the attributes in the choice set. This assumes that the individual has preferences that can be represented as neo-classical trade-off preferences. The levels x_m and y_m refer to *medium* levels, possibly given from the current level of consumption, for example a typical trip to work where *x* and *y* could represent, say, travel time and travel comfort. From the situation depicted the respondent will avoid combinations where one of the attributes has the low level, x_l or y_l, since the high level of the other attribute, y_h or x_h, would not compensate for that decrease. The required compensation level is beyond the scale extension of the attributes. Although such a simple figure does not comprise all elements of an actual choice task, it illustrates how insufficient scale extensions for some "high valuers" may yield lexicographic answering that indicates neither lexicographic preferences nor choice task simplification (Rosenberger et al., 2003).

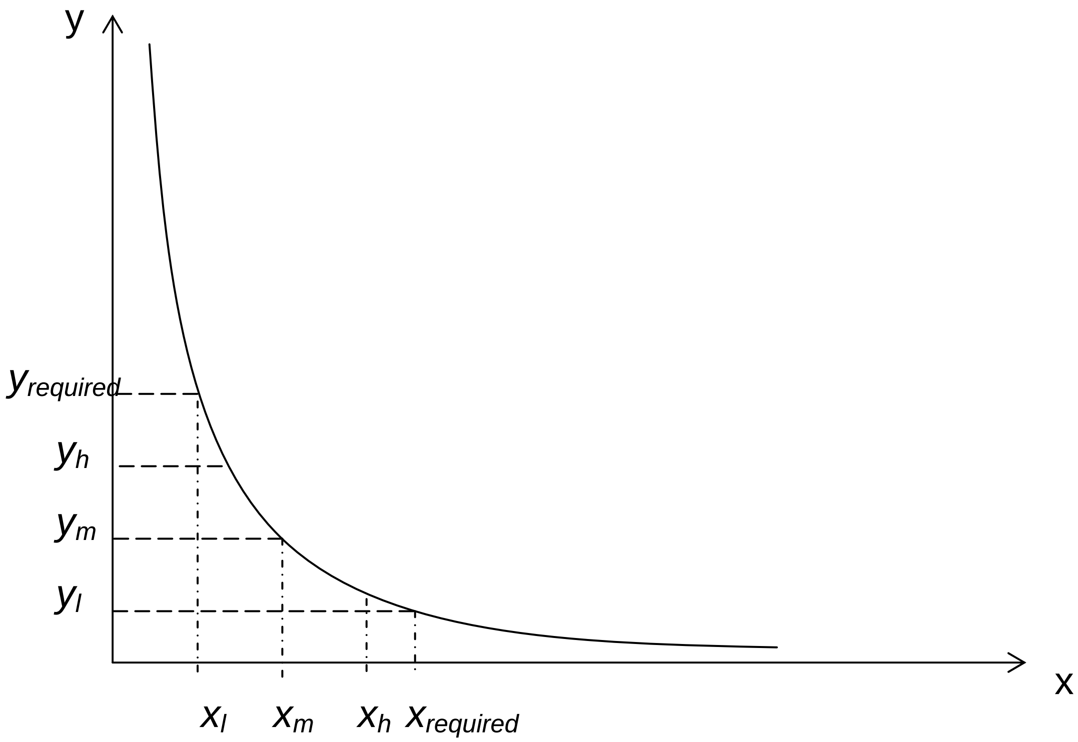

Figure 1. Steep indifference curve for attribute x vs. attribute y – within the attribute range (scale extension) from lowest levels to highest levels (l = low, m = medium, h = high).

Lexicographic answering may follow from an answering strategy of simplifying choice by focusing on one or only a few attributes. Lexicographic answering as a rational approach for handling complex decisions has been put forward as a general explanatory model of behaviour (Encarnación, 1990; Heiner, 1983; Tversky, 1972). The individual may value all attributes, but because of a choice task too complex to optimise or because of lack of time (high valuation of time) the individual may pick just one or a few attributes as the overall most important and disregard the rest. This may misrepresent preferences according to the strict trade-off model (where time-use in the survey is disregarded), although still represent individuals' actual choices.[1]

2.2 Hypotheses to be tested

There is no one question or variable in the three surveys that provides a single test of the cause of lexicographic answering. Individuals identified as answering lexicographically can be evaluated with respect to a series of responses to other survey questions. This should shed some light on the causes of lexicographic answering.

Main hypothesis: Steep indifference curves constitute one main cause of lexicographic answering. This is driven by unbalanced design, in terms of irrelevant or insufficient attribute ranges for allowing trade-offs, or possibly, omission of relevant attributes.

Premise A: If answers to questions other than the choice questions indicate preference for the lexicographically chosen attribute, this is an indication of steep indifference curves rather than use of a pure simplification strategy.

Sub-hypothesis 1: *Here the null hypothesis is that there are equal shares in the group answering lexicographically and in the group not answering lexicographically stating that the range of the interval given for that attribute is "very important".*

Sub-hypothesis 2: *Under this null hypothesis there are equal shares stating that only one attribute is of importance for choice in the group answering lexicographically and the group not answering lexicographically.*

If sub-hypotheses 1 and 2 are rejected it will be taken as an indication that the lexicographical answering is governed by true preferences for increasing the level of the specific attribute. However, the question about the attribute's importance for choice was given after the choice task, and may be plagued with rationalisation in ex-post judgments (Festinger, 1957; Montgomery, 1987).

Premise B: Lexicographic answering caused by a strategy of simplification should be higher in the last part (sequence) of the choices, as a result of fatigue (Caussade et al., 2005). Alternatively, one could propose that lexicographic answering due to simplification should be higher in the first part of the choices if the respondent needed time to learn the choice task.[2]

Sub-hypothesis 3: *Under this null hypothesis the share of lexicographic answering is equal between choice sequences.*

---

[1] One may distinguish between *strong* lexicographic answering, that would comprise the first two explanations above, and *weak* lexicographic answering (Foerster, 1979). Weak lexicographic answering comprises part (not all) of the response/simplification strategies that do not follow from neo-classical trade-offs. It is not obvious, for example, that strategies such as lexicographic semi-ordering (Tversky, 1969) and some types of elimination of aspects (Tversky, 1972) or threshold preferences (Encarnación, 1964) are observed as lexicographic answering. Only simplification strategies with respect to best/worst level of one attribute would generally be observed as lexicographic answering.

[2] However, there could also be reason for the assumption that the respondent needed to "learn" a consistent simplification strategy (Payne et al., 1993; Gigerenzer and Todd, 1999).

If sub-hypothesis 3 is rejected, it is indicated that the lexicographical answering is governed to some extent by a simplification strategy. This test can be described as *weak* in the sense that falsification (of the simplification as explanation of lexicographic answering) will be the result of retaining, not rejecting, the null hypothesis (type I error).

Premise C: Multivariate logistic regression can be applied to indicate if lexicographic answering with respect to one attribute is relevantly explained from covariates, e.g. that lexicographic answering with respect to time is explained by income and travel distance.

Sub-hypothesis 4: *Under this null hypothesis, relevant covariates show no correlation with lexicographic answering.*

If sub-hypothesis 4 is rejected, it is indicated that at least some of those answering lexicographically do so as a reflection of true preferences for increasing the specific attribute's level, i.e. steep indifference curves.

3. Data

3.1 Three surveys

Three different Internet-based SC surveys were carried out in 2002 and 2003 using the Sawtooth software (Sawtooth Software, 2002). The object of all three studies was to assess travellers' preferences for commuter attributes. This section gives a brief description of the three surveys.

The first survey dealt with travellers' valuations of traffic information when commuting by car (Killi and Samstad, 2002). It involved the following attributes: cost (price), travel time, travel time variability, congestion time and new forms of information to reduce travel time and uncertainty due to delays. The survey was carried out focusing on travel information connected with the trip to work in morning rush hours. Car users experiencing congestion problems comprised the target population. Participants were recruited along the road during morning rush hours; of 1733 cards handed out with information about the survey and giving the website address to the questionnaire, 278 persons fully completed the questionnaire on the Internet, yielding a response rate of 17.5%. The recruiting method provided no opportunity to follow-up with reminders. This is referred to as the "Traffic-info survey".

The second survey dealt with travellers' valuations of time in public transport in Oslo (Nossum, 2003). Finding indicators of passengers' preferences for alternative quality improvements was the motivation behind the survey. The sample was chosen at random from the Norwegian Population Register, and everyone selected received a letter by post with an Internet address, user name and password to log on to the survey. The overall response rate was 30%, including 879 respondents. This is referred to as the "Oslo survey".

The third survey was about a transport investment programme in Tønsberg, a town southeast of Oslo with a population of approximately 35,000 inhabitants (Vibe et al., 2004). Among those sampled for SC were bus passengers and car drivers. Samples of these two travel mode groups were chosen randomly from telephone numbers in Tønsberg. Each respondent was offered a home interview if they could not, or would not, reply via the Internet. Those who preferred a self-administered Internet interview received an Internet address, a user name and a password by e-mail. The home interviews were carried out using a laptop connected to the Internet via a mobile telephone and the questionnaire was a bit shorter than the self-administered Internet questionnaire. The response rate was 31.5%, including 1105 respondents. This is referred to as the "Tønsberg survey".

The three surveys were similar in design. Five parts of the questionnaire can be distinguished for our study:

- Questions about the usual/last trip, establishing the medium attribute levels (for travel time, fare/price/cost, etc.).
- Questions to establish the importance of the difference between the calculated maximum and minimum levels of the attributes in the SC.
- SC sequences (3, 4 or 5, each with 6 or 9 pairwise choices).
- Question to clarify what attributes were important for the choices
- Demographic/socioeconomic variables.

At the start of all three surveys, respondents were asked about characteristics of the last journey or a representative journey, including travel time, travel cost and time spent waiting for public transport or driving in congested traffic. The responses produced medium attribute levels for many of those included. Then, before the SC sequences, the respondents were asked about the importance of the implied attribute range, e.g. ±25% of the medium level from the usual/last trip. The SC choices included three to five choice sequences, separated with new introductions and (some) new attributes. After the SC sequences, respondents were asked about attribute importance for choice. Finally, they were asked about background variables such as age and household income.

3.2 Adaptive stated choice designs

The adaptive choice-based module (ACA) in the Sawtooth software was used to design the SC sequences in all three surveys. After establishing medium attribute levels based on individual-specific journeys, and denoting the importance of the calculated individual attribute ranges, a crude set of estimates for the individual's utilities is produced. The design of the pairwise choices is then chosen by the computer to provide the most incremental information, taking into account what is already known about the individual's utilities (Sawtooth Software, 2002).[3]

Before facing the SC sequences the respondent had to answer the following question: "If two journeys are equal in all other ways, how important would the difference between highest and lowest price be for you?", and similarly for other attributes. The respondents ticked off on a scale from 1 ("not important") to 5 ("very important") in the "Traffic-info survey" and the

[3] Johnson et al. (2003) relate adaptive choice-based designs to D-optimality. D-optimality is principally based on maximisation of the Fisher information matrix (or minimisation of the determinant of the inverse of the variance-covariance matrix – minimising the D-error). Adaptive choice-based designs address a main challenge of implementing D-optimality – "that with nonlinear models the researcher must know the parameter values before deriving his or her design" (Kanninen, 2002). Individuals' value parameters are estimated from the initial questions about the typical (or last) journey and about the importance of the calculated attribute ranges. Johnson et al. (2003) also stress "four requirements of a good design": orthogonality, level balance, minimal overlap and utility balance. They find that "D-error is one way of trading-off these goals". However, Kanninen (2002) clarifies that orthogonality, level balance, minimal overlap and utility balance represent constraints that may actually reduce efficiency, such that adaptive choice-based designs may not be D-optimal designs.

"Oslo survey", while a 4-point scale was applied in the "Tønsberg survey". As indicated, the answers to these questions were applied in the Sawtooth software to set the initial values in the choice design. These answers afford the possibility to assess connections between lexicographic answering and the stated importance of the calculated attribute range (sub-hypothesis 1) – normally ±25% of the medium level given from the respondent's representative journey.

In each choice situation, in all three surveys, two trips, A and B, with three or four attributes were presented on the screen. The respondent could indicate which of the two trips he preferred by ticking either "Definitely A", "Probably A", "Don't know", "Definitely B" or "Probably B".[4] From a sequence of choices it is then possible to estimate willingness to pay for the included attributes. An example of a pairwise choice is given in figure 2.

If these trips were identical in all other respects, which one would you choose?

Car trip A		*Car trip B*
Cost: €1.80		*Cost: €2.26*
Travel time: 33 min		*Travel time: 27 min*
Information of type A, B and C		*Information of type A*

☐	☐	☐	☐	☐
Definitely A	***Probably A***	***Don't know***	***Probably B***	***Definitely B***

Figure 2. Example of a pairwise choice from the "Travel-info" survey (third sequence).

The "Traffic-info survey" contained three choice sequences with nine pairwise choice situations in each sequence (27 pairwise choices per respondent in total). The cost (price) and travel time of the two attributes were included in all sequences, and a third attribute was specific to each of the three sequences: arrival time variation, congestion time and type of traffic information. The public transport studies in Oslo and Tønsberg contained three to five choice sequences with six choice situations in each sequence (18 - 30 pairwise choices per respondent in total). The attributes in the "Oslo survey" were price, travel time, comfort, headway, walking time, delays and transfer. Price was included in all the sequences. Bus trip attributes in the "Tønsberg survey" were price, travel time, headway, walking time, delays and transfer. The "Tønsberg survey" for bus passengers was similar in design to the "Oslo survey". For car drivers, the attributes in the "Tønsberg survey" were parking fee (price), travel time and distance from parking place. In all the sequences there were profiles with three attributes, except for the third sequence in the "Oslo survey", which had profiles with four different attributes.

The total number of pairwise choices per respondent was relatively high in all three surveys. However, each choice sequence was separated by new introductions, and the number of pairwise choices in each sequence was close to optimal in terms of minimising error variance (Caussade et al., 2005). Furthermore, the data enable effects from possible fatigue to be tested (sub-hypothesis 3).

[4] In a pilot to the "Traffic-info survey" both "equivalent" and "don't know" were included. Even if "equivalent" may be deemed closer to "indifferent", "don't know" is rather close to the wording that Ortúzar and Garrido (1994) applied instead of "indifferent" (arguing, as we would also do, that people are not clear about the meaning of "indifferent"): "I cannot decide".

The attributes in all three surveys had three levels – a best level, a medium level and a worst level. The medium level of each attribute derived from the responses to the introductory question about "a representative journey". The worst and best levels were set as ±25% of the medium levels, as a default, based on results from a pilot survey and experiences from previous studies (Killi et al., 2001; Norheim and Stangeby, 1993; Norheim et al., 1994; Kjørstad, 1995). Yet, as will be indicated, ±25% could yield a fairly limited attribute range for those having, for example, short travel time in their usual/last journey. There were some exceptions to the ±25% default: (*i*) In the "Traffic info" survey, the attribute ranges were set to ±10% of the medium level in the third choice sequence, because the pilot study showed that otherwise travel time and price would be preferred to information in almost all choice situations; (*ii*) In the "Tønsberg survey", the walking time attribute ranges were set to ±50% of the medium, because it was acknowledged that ±25% was not sufficient (Ortúzar and Rodríguez, 2002); and (*iii*) some other attributes were not customized in this manner at all (e.g. delays), but fixed to four (instead of three) levels. In general, all three surveys confronted the respondents with a relatively low information load in terms of attributes and levels (Caussade et al., 2005; Mazzotta and Opaluch, 1995).

After the third choice sequence, the respondents were asked how many of the attributes (e.g. among price, travelling time, delay, etc.) had been decisive in their choices. If only one, they were asked which one and why it was decisive. The answers help us clarify whether the respondents were aware of their lexicographic answers – if conscious, it was thus more likely preference-related (sub-hypothesis 2).

To some extent a lexicographic answering from strong preference may be correlated with individual characteristics. For example, valuation of travel time savings has been found to increase with income and travel distance (Fosgerau, 2006). The same relationship may also hold for headway (waiting time) reduction and minimising transfers. Other travel attributes, too, such as comfort, may correlate with individual characteristics like age. It is therefore assumed that lexicographic answering with respect to some attribute will correlate with demographic variables (sub-hypothesis 4).

4. Results

4.1 Identified lexicographic answering

An answer was determined as lexicographic if the respondent in all six/nine pairwise choices had chosen the trip option with the best level for one particular attribute (Norheim, 2003; 2004). Choice of option *A* was defined as including both "*Definitely A*" and "*Probably A*". However, the classification of lexicographic answering to only "*Definitely A*" may be restricted. At the end of this section we present some results based on this stricter definition of lexicographic answering.

In the "Traffic-info survey", 21-24% of the respondents answered lexicographically, primarily with respect to travel time. In the "Oslo survey", 26-33% answered lexicographically, mostly with respect to the price attribute; 22-30% of bus passengers in the "Tønsberg survey" answered lexicographically in that survey too often due to price but also to transfer and delay. Among car users in the "Tønsberg survey", 41% answered lexicographically, almost all according to price (tables 1-3).

Table 1. Lexicographic answering in the "Traffic-info survey"

Sequence	N	Price	Travel time	Arrival time variation	Total
1	313	2.2%	21.7%	0.3%	24.3%
Sequence	N	Price	Travel time	Congestion time	Total
2	297	5.1%	19.2%	0.0%	24.2%
Sequence	N	Price	Travel time	More information	Total
3	277	4.0%	17.0%	0.4%	21.3%

Table 2. Lexicographic answering in the "Oslo survey"

Sequence	N	Price	Walking time	Headway		Total
1	778	24.3%	1.4%	7.7%		33.4%
Sequence	N	Price	Travel time	Transfer		Total
2	770	14.0%	5.6%	7.5%		27.1%
Sequence	N	Price	Travel time	Comfort	Delays	Total
3	762	13.6%	2.9%	5.9%	3.5%	26.0%

Table 3. Lexicographic answering in the "Tønsberg survey"

Sequence	N	Price	Walking time	Headway	Total
1	307	14.3%	1.6%	6.1%	22.1%
Sequence	N	Price	Travel time	Transfer	Total
2	303	6.9%	0.3%	23.1%	30.4%
Sequence	N	Price	Travel time	Delays	Total
3	120	11.6%	0.8%	10.8%	23.3%
Sequence	N	Price (parking fee)	Travel time	Distance from parking place	Total
4 (car users)	423	38%	0.7%	2.6%	40.9%

In the "Traffic-info survey", 8.3% of the respondents answered lexicographically due to travel time in all three sequences. In the "Oslo survey", 5.8% of the respondents answered lexicographically due to price in all three sequences. Not more than 1% answered lexicographically with respect to travel time in either sequence.

4.2 Importance of the attribute intervals

The results from the "Traffic-info survey" and the "Tønsberg survey" are given in tables 4, 5 and 6.[5]

[5] As a result of a programming error in the "Oslo survey", the quality of data from this particular question was not acceptable (Nossum, 2003).

Table 4. Importance of the attribute intervals connected with the lexicographic answers – "Traffic-info survey"

Degree of importance	Sequence 1		Sequence 2		Sequence 3		Total
	Price	Travel time	Price	Travel time	Price	Travel time	
1 Not important	0	1	0	1	0	2	4
2	1	0	0	0	0	1	2
3 Of some importance	1	5	2	5	3	3	19
4	0	3	5	2	3	10	23
5 Very important	5	59	8	49	5	31	157
Total	7	68	15	57	11	47	205

In the "Traffic-info survey", nearly 77% (157/205) of the lexicographic answers came from respondents who stated that the difference between the highest and lowest value of that particularly attribute was "very important" to them. This is considerably higher than the share of those not answering lexicographically (42%), and the difference is significant at the 1% level (Pearson χ^2 equal to 87.6, p<0.001 with *d.f.*=1).
For travel time alone, the share is even higher. Price had few observations in the "Traffic-info survey" and is disregarded from the further discussion. Regarding travel time, 86-87% stated that the difference between the highest and lowest value of that particularly attribute was "very important" in the first two sequences. When reducing the best/worst values to ±10% of the medium value in the third sequence, compared to ± 25% in the first two sequences, the share of "very important" dropped to 66%. This sensitivity with respect to attribute range is also significant at the 1% level (Pearson χ^2 equal to 22.5, p<0.001 with *d.f.*= 1).[6]

Table 5. Importance of the attribute intervals connected with the lexicographic answers – public transport passengers in the "Tønsberg survey"

Degree of importance	Sequence 1			Sequence 2			Sequence 3			Total
	Price	Walking time	Headway	Price	Travel time	Transfer	Price	Travel time	Delays	
1 Not important	2	0	0	0	0	1	0	0	0	3
2	6	1	1	0	0	4	0	0	0	12
3	13	0	3	2	0	15	1	1	2	37
4 Very important	23	4	15	19	1	50	13	0	11	136
Total	44	5	19	21	1	70	14	1	13	188

In the public transport part of the "Tønsberg survey", slightly less than 75% (136/188 yields 72%) of the lexicographic answers came from respondents who stated that the difference between the highest and lowest value of that particular attribute was "very important" to them. This is considerably higher than the share for those not answering lexicographically (39%), and the difference is significant at the 1% level (Pearson χ^2 equal to 80.8, p<0.001 with *d.f.*=1).

[6] In some very few cases the respondent answered lexicographically with respect to one attribute, but stated that the difference in the highest and lowest levels of that attribute was "not important". Looking more closely at those four responding like this in the "Travel-info study", one finds that all had very short travel distance (to work). Thus "travel distance" would remain virtually unchanged even with ±25%.

Table 6. Importance of the attribute intervals connected with the lexicographic answers – car drivers in the "Tønsberg survey"

Degree of importance	Car users			Total
	Price (Parking fee)	Travel time	Distance from parking place	
1 Not important	6	0	0	6
2	9	0	0	9
3	21	1	3	25
4 Very important	123	2	8	133
Total	159	3	11	173

In the car drivers' part of the "Tønsberg survey", 77% (133/173) of the lexicographic answers came from respondents who stated that the difference between the highest and lowest value of a particular attribute was "very important" to them. This is considerably higher than the share for those not answering lexicographically (23%), and the difference is significant (Pearson χ^2 equal to 42.8, p<0.001 with $d.f.$=1).

Looking at price (parking fee) separately, we find that 77% of the lexicographic answers due to price came from respondents who stated that the difference between the highest and lowest value was "very important" to them. This is significantly higher than the share (60%) for trade-off answers (Pearson χ^2 equal to 13.6, p<0.001 with $d.f.$=1).

There was therefore accordance between weighing the importance of the attribute interval and answering lexicographically with respect to that attribute. Sub-hypothesis 1 is rejected based on these results.

4.3 Only one attribute of importance for choices

Generally speaking, there was no full agreement on whether a respondent answered lexicographically stating that one attribute only was decisive in his/her choice. In the "Traffic-info survey", no more than 36% of the respondents who answered lexicographically stated that one attribute only (travel time) had been decisive in their choices. However, this is three times the share in the group trading-off between attributes (12%), and the difference is significant at the 1% level (Pearson χ^2 equal to 17.5, p<0.001 with $d.f.$=1). In the "Oslo survey", 10% of the respondents who answered lexicographically said that one attribute only had been decisive in their choices, versus 6.5% in the group trading off – a difference significant at the 10%-level (Pearson χ^2 equal to 3.32 with p=0.068, $d.f.$=1). The corresponding figure for those answering lexicographically in the public transport part of the "Tønsberg survey" was 32%, which was about twice the share of the group trading off. This difference is also significant at the 10%-level (Pearson χ^2 equal to 3.37 with p=0.067, $d.f.$=1). Sub-hypothesis 2 is rejected on the basis of these results.

Table 7a. Lexicographic answering versus number of attributes of significance, the "Traffic-info survey"

	Only one attribute decisive for the respondent	Several attributes decisive
Lexicographic answering	21	38
Travel time	21	
Price	0	
Information	0	
Trading-off	27	191

Table 7b. Lexicographic answering versus number of attributes of significance, the "Oslo survey"

	Only one attribute decisive for the respondent	Several attributes decisive
Lexicographic answering	20	178
Price	17	
Travel time	1	
Comfort	1	
Delays	1	
Trading-off	35	529

Table 7c. Lexicographic answering versus number of attributes of significance, the "Tønsberg survey"

	Only one attribute decisive for the respondent	Several attributes decisive
Lexicographic answering	9	19
Price	5	
Travel time	0	
Delays	4	
Trading-off	15	77

In the "Traffic-info survey", all those stating that one attribute only was decisive in their choice pointed to travel time, and none to price. In the other two studies, price and, to some extent, delays were the attributes most respondents stated as being decisive in their choice.

4.4 Fatigue or learning?

Respondents took about 25-35 minutes to accomplish the surveys, and approximately 10% did not complete them. One possible reason for simplifying the heuristics in SC studies could be tiredness, and if lexicographic answering reflects simplification an increasing trend in lexicographic answers should be registered towards the end of the survey. It has to be noted that some of the attributes changed from one sequence to the next, thus obstructing clear-cut comparison between sequences; for example, the degree of perceived response exigency/difficulty varied between sequences (Stopher and Hensher, 2000). Still, there is no indication of an increasing number of respondents simplifying due to tiredness in the course of the choice sequences (table 8).

Table 8. Lexicographic answers according to price and travel time, per cent of total samples

	Price			Travel time		
	Traffic-info (%)	**Oslo (%)**	**Tønsberg (%)**	**Traffic-info (%)**	**Oslo (%)**	**Tønsberg (%)**
Sequence 1	2.2	24.3	14.3	21.7	*	*
Sequence 2	5.1	14.0	7.0	19.2	5.5	0.3
Sequence 3	3.5	13.7	11.6	17.0	4.0	0.8

* Not an attribute in this sequence.

If respondents find the choice task difficult to understand at first glance, one might imagine they needed training in the first pairwise choices before learning to express their preferences in the following choices. If this is the case, and using the same reasoning of lexicographic simplification, one would expect most lexicographic answering in the first choice sequence. This seems to be the case for the price attribute in the Oslo survey and for travel time in the Traffic-info survey. However, evidence in the previous sections suggests that the high degree in lexicographic answering in the first sequences of these surveys might just as well be a consequence of deliberate trade-off behaviour. Anyhow, the sub-hypothesis 3, that the share of lexicographic answering is equal between choice sequences, has to be rejected in all studies.

4.5 Multivariate logistic regression of lexicographic answering

4.5.1 Regression modelling per attribute

Multivariate logistic regression was run to analyse what characteristics contributed to a higher probability of lexicographic answering. The parameters of these variables show incremental effects for the lexicographic group (for the probability of a lexicographic answering outcome) compared to the trading-off reference group (Greene, 1993).

For the "Traffic-info survey", the regression model was limited to those answering lexicographically according to travel time. For the "Oslo survey", regression models were run for those answering lexicographically according to price, travel time, headway, transfer and comfort. For the "Tønsberg survey", regression models were run for price and transfer. The number of respondents answering lexicographically with respect to other attributes was too small to be used in logistic regression analysis. After several runs and taking correlations between the right-hand side variables into account, a set of model estimates was chosen.

The explanatory variables in all three surveys were age, household income and gender. In addition, the "Traffic-info survey" used "Kmdist", indicating distance between home and work, and "Diff", which is a dummy variable equal to 1 if the respondent has answered that it is difficult to choose between the two journeys. In both the "Oslo survey" and the "Tønsberg survey" we included the explanatory variables travel time, price and employment. The last explanatory variable, "Scale-imp" is a dummy that equals 1 if the respondent has answered that the difference between highest and lowest value is "very important".[7]

[7] In the "Tønsberg-survey" a four-point scale was applied. In the "Oslo study" no such variable was included.

4.5.2 Lexicographic answering with respect to minimum travel time

Lexicographic answering is modelled as a logistic regression model of travel time in all three sequences in the "Traffic-info survey" and in the second sequence in the "Oslo survey". The results are presented in table 9a and 9b and show that income is significant in all four sequences and travel distance in two sequences. Stating that the difference between highest and lowest travel time is "very important" is significant in all three possible sequences.
The probability of answering lexicographically with respect to travel time *increases* with:

- Higher household income
- Stating that the difference in the levels of travel time interval is "very important"
- Longer travel distance/travel time (in two of four sequences)

Table 9a. Logistic regression – lex travel time, "Traffic-info survey"

	Constant	Kmdist	Diff	Age	Income	Gender	Scale-imp
Estimate seq1	-2.82***	0.02***	-0.18	-0.02	0.11***	0.34	1.18***
T value	-3.48	2.86	-0.34	-1.20	2.97	0.97	2.73
Estimate seq2	-2.94***	0.01	-0.38	-0.01	0.09***	0.22	1.30***
T value	-3.63	1.14	-0.64	-0.44	2.58	0.61	3.03
Estimate seq3	-2.25***	0.01	-0.11	-0.03	0.13***	-0.47	1.30***
T value	-2.71	0.75	-0.17	-1.39	3.38	-1.06	3.49

*** 0.01 Significance level, ** 0.05 Significance level, * 0.10 Significance level

Table 9b. Logistic regression – lex travel time, "Oslo survey"

	Constant	Employment	Price	Kmdist	Age	Income	Gender
Estimate	-5.53***	0.05	-0.01	0.05***	0.01	0.65*	0.56
T value	-4.72	0.12	-1.13	5.63	0.77	1.78	1.61

*** 0.01 Significance level, ** 0.05 Significance level, * 0.10 Significance level

Higher valuation of possible travel time reduction should increase with travel/commuting distance (Kmdist) and with income. Shorter travel time may then be far more important to these respondents than lower price (steep indifference curve for travel time with respect to price).
Stating that the difference between highest and lowest travel distance is "very important" (Scale-imp) is also a reasonable contributor to increasing the probability of lexicographic answering. On the other hand, how difficult the respondents found choosing between two journeys (Diff) did not have any measurable effect on whether the respondents answered lexicographically or not.
In the analysis of the second and third sequences, the travel distance variable (Kmdist) was not significant. As indicated, in sequence 3 it was applied ±10% of the medium level as highest and lowest value of the attributes (not ±25%). This might partly explain this observation.

4.5.3 Lexicographic answering with respect to minimum price

Lexicographic answering with respect to price is modelled for all three sequences in the "Oslo survey", for the first sequence for public transport in the "Tønsberg survey", and for the first sequence for car-use in the “Tønsberg survey”. Household income turns out to be significant in all four public transport sequences, travel time in two and age in only one. The probability of answering lexicographically with respect to low price in public transport increases with lower income. The probability of answering lexicographically according to low parking fee for car users is higher if travel time is low (table 9c).

Table 9c. Logistic regression – lex price, “Oslo survey” and “Tønsberg survey”

	Constant	Employment	Price	Travel time	Age	Income	Gender	Scale-imp
“Oslo survey”								
Estimate Seq1	-1.99***	0.11	0.002	-0.003	0.01**	-0.44**	0.13	-
T value	-3.68	0.51	1.00	-0.60	2.00	2.40	0.75	-
Estimate Seq2	-2.20***	-0.22	0.01	-0.01	0.01	-0.59***	0.25	-
T value	-3.35	-0.89	1.40	-1.14	1.25	-2.53	1.15	-
Estimate Seq3	-1.37**	0.51*	0.00	-0.03***	0.002	-0.85***	0.21	-
T value	-2.06	1.83	0.00	4.00	0.25	-3.61	0.97	-
“Tønsberg survey”								
Estimate Seq1	35.98	0.45	0.04	-0.03	-0.02	-1.11*	-0.09	0.29
T value	1.57	1.05	0.67	-1.29	-1.58	-1.89	-0.25	0.64
Estimate car	-0.71	-0.17	-	-0.01**	0.00	-0.20	0.16	0.43
T value	-0.80	-0.71	-	2.00	0.00	-0.89	0.80	1.31

*** 0.01 Significance level, ** 0.05 Significance level, * 0.10 Significance level

Note that parking fee (price) was not customised for car users, in contrast to price for public transport passengers.

4.5.4 Lexicographic answering with respect to “no transfer”

Transfer was an attribute in the second sequence in both the public transport studies. The model estimation shows that there is a higher probability of answering lexicographically with respect to “no transfer” if the travel time by bus is relatively short. Higher age (in the “Oslo survey”) and lower price (in the “Tønsberg survey”) corresponded with higher probability of answering lexicographically due to no transfer. Stating that it is important to have no transfer instead of transfer with ten minutes waiting time in the “Tønsberg survey” increases the probability of answering lexicographically (table 9d).

Table 9d. Logistic regression – lex “no transfer”, “Oslo survey” and “Tønsberg survey”

	Constant	Employ-ment	Price	Travel time	Age	Income	Gender	Scale-imp
“Oslo survey” Sequence 2								
Estimate	0.15	0.01	0.001	-0.05***	-0.02**	0.01	-0.28	-
T value	0.20	0.01	0.33	-4.08	-2.2	0.03	-1.00	-
“Tønsberg survey” Sequence 2								
Estimate	8.05	0.01	-0.10**	-0.06**	-0.004	0.21	-0.34	1.37**
T value	0.47	0.02	-2.28	-2.20	-0.45	0.53	-1.11	2.43

*** 0.01 Significance level, ** 0.05 Significance level, * 0.10 Significance level

4.5.5 Lexicographic answering with respect to minimum headway

Lexicographic answering for minimum headway was modelled for the second sequence in the "Oslo survey". The probability of answering lexicographically with respect to headway was greater when travel time was lower and age was higher (table 9e).

Table 9e. Logistic regression – lex headway, "Oslo survey"

	Constant	Employment	Price	Travel time	Age	Income	Gender
Estimate	-0.78	-0.21	-0.01	0.02**	-0.02**	0.43	-0.44
T value	-0.99	-0.60	-1.30	2.00	-2.00	1.46	-1.55

*** 0.01 Significance level, ** 0.05 Significance level, * 0.10 Significance level

4.5.6 Lexicographic answering with respect to comfort

The third SC sequence in the "Oslo survey" included a comfort attribute with three levels: Being seated for the entire journey, being seated for some part of the journey and standing for the entire journey. With increasing age, a higher probability of answering lexicographically was found with respect to comfort (a seat all the way).

Table 9f. Logistic regression – lex comfort, "Oslo survey"

	Constant	Employment	Price	Travel time	Age	Income	Gender
Estimate	-0.60	-0.20	0.001	0.01	-0.03***	0.13	-0.44
T value	-0.75	-0.51	0.33	0.88	-3.09	0.39	-1.40

*** 0.01 Significance level, ** 0.05 Significance level, * 0.10 Significance level

4.5.7 Indications from the regression modelling

In all regression models, some relevant demographic/psychographic variable or travel aspect is found to increase the probability of lexicographic answering with respect to one specific attribute. As expected, lexicographic answering with respect to minimum travel time corresponds with higher income, expressed importance of this attribute, and initial long travel distance. In contrast, lexicographic answering with respect to minimum price corresponds with lower income. Furthermore, and also as expected, lexicographic answering with respect to seating comfort and avoiding transfer corresponds with higher age.[8] This indicates that at least some lexicographic answering is driven by preferences for improving the attribute chosen lexicographically. Thus, sub-hypothesis 4 can be rejected.

[8] The models presented in the Tables 9a-9f include parameters with low t-ratios. Generally the final models are not chosen only with respect to some goodness-of-fit criterion, but rather with respect to theoretical considerations (including income and, to some extent, employment) and consistency in terms of included background variables (including age and gender in all models). The models show test results rather than best descriptions of what explains some particular lexicographic answer.

4.6 Multivariate logistic regression with a stricter definition of lexicographic answering

A definition of lexicographic answering could very well be restricted only to those respondents answering "*Definitely A*" or "*Definitely B*" (omitting "*Probably A*" and "*Probably B*"). Such strict lexicographic answering was found for only a small fraction of the respondents in the "Traffic-info survey". In the two public transport studies the strict definition resulted in approximately half as many lexicographic answers as the wider (less strict) definition. The reduction was quite evenly distributed among the attributes. In the "Oslo survey", 9-18% answered lexicographically due to the strict definition (compared to 26-33% for the wider definition). In the "Tønsberg survey", between 12% and 16% of bus passengers and 23% of car users answered lexicographically in the strict sense (compared to 22-30% and 41%, respectively, when applying the wider definition). It could be added that 17% of those who answered strictly lexicographically due to price in the "Oslo survey" did so in all three sequences.
An alternative multiple regression analysis for strict lexicographic answering was performed only for the two public transport studies.[9] For the "Oslo survey", the regression modelling involved price from three sequences and headway from one. For the "Tønsberg survey", it involved transfer and parking fee (price). The numbers of respondents answering strictly lexicographically with respect to other attributes were too small to enable analysis. In general, the results show that fewer coefficients are significant using the strict definition compared to the wider definition. Some of the key results are:

- The probability of answering lexicographically according to price is higher if household income is lower ("Oslo survey") and travel time by car is lower ("Tønsberg survey"), which was also found in the analysis applying the wider definition of lexicographic answering.
- The probability of answering lexicographically due to parking fee is significantly higher if the respondent finds the difference between higher and lower parking fee important. This was not significant applying the wider definition.
- For the other modelling, no significant impact was found from travel time and age on the probability of answering strictly lexicographically according to headway, and no significant impact from price, travel time, or the difference between no transfer and transfer (Scale-imp) on the probability of answering strictly lexicographically according to no transfer.

5. Discussion and conclusions

Lexicographic answers potentially constitute a serious problem in SC analysis, especially if it is not known whether they contribute to increasing or decreasing the valuation of the attribute. Respondents simplifying the decision-making process by focusing on one attribute only will lead to an overestimation of the relative valuation of this attribute. If, on the other hand, lexicographic answers are a result of poor balancing of the attribute levels in the design, there is a risk that the valuation of the attribute they choose will be underestimated (Rouwendal and Blaeij, 2004; Sælensminde, 2001).

[9] In the "Travel-info study" the number of respondents answering strictly lexicographic was too small to run a multivariate logistic regression.

Previous studies have shown that respondents' tendency to simplify the decision-making process by stating lexicographic answers will vary with different design. One fundamental feature is the complexity of the design (Heiner, 1983), which also drives the degree of learning and the effects of fatigue (Swait and Adamowicz, 1996). The design of the three studies explored in this article was relatively simple, with paired comparisons and only three attributes, the risk of lexicographic simplification thus being minimum. Despite this, lexicographic answering was observed, although one-fifth to one-third is at the lower end of what has been observed in SC (Sælensminde, 2000; Widlert, 1998). It could not be excluded that a limited part of the lexicographic answering was due to simplification in the learning part (first sequence) of the choice task (sub-hypothesis 3 rejected).

Although the three surveys reported here have limitations in terms of sampling and response rates, their results point in the same direction. Most of the respondents who answered lexicographically with respect to an attribute also stated that the attribute, and its range, was important for them. A larger share among those answering lexicographically stated that the range of the interval given for that attribute was important (sub-hypothesis 1 rejected). Furthermore, the share stating that one attribute only was of importance for choice was higher among those answering lexicographically than among those trading-off (sub-hypothesis 2 rejected).

The respondents who answered lexicographically also followed-up with reasoning on why only one attribute was important for their choice. Between 55% and 86% stated that the *other* attributes were of little importance. Fewer than 10% (in all three surveys) answered that the reason for their weighing only one attribute was that the difference in high and low value of this attribute was large. Only a small fraction claimed that they weighed only one attribute in terms of facilitating the choice task. The main indication from these answers is that the lexicographic answering was due to the omission of relevant attributes or insufficient attribute scales. Among the majority of respondents stating that several attributes were decisive in their choices, some still answered lexicographically (17% in the "traffic-info survey"). The main indication is that the attributes and profiles have not been sufficiently well balanced, i.e. insufficient attribute scales and steep indifference curves.

The estimated probability of answering lexicographically, given from the multiple regression analysis, was fairly well explained by relevant covariates, e.g. that the probability of answering lexicographically according to travel time increased with income, actual travel time, and stated importance of this attribute. Thus, sub-hypothesis 4 can also be rejected. The coefficients obtained with correct signs and significant values support the alternative hypothesis: "part of the lexicographic answering is due to steep indifference curves". A considerable proportion of those answering lexicographically with respect to an attribute did so because of their preferences for these attributes, i.e. they did not find that the given level on, for example, travel time could be matched by the improved level on the other attributes. The attribute scale extension was insufficient.[10]

[10] We acknowledge that a generalised lexicographic answering structure following refined economic definitions (Encarnación, 1990; Georgescu-Roegen, 1954) or psychological definitions (Mazzotta and Opaluch, 1995; Tversky, 1972), and not omitting the possibility of trade-off, may constitute a more realistic choice model than the neo-classical exchange model. However, applying relatively simple design travel choices can probably be analysed adequately within trade-off-based modelling. Alternatively, the given levels of travel time may have been above and beyond a maximum acceptable threshold (while no threshold is passed for other attributes), for which no improvement in price or other attributes could match (Cantillo and Ortúzar, 2005). If we had had a split design of paired choices with different attribute ranges, we could have applied a test between lexicographic

Steep indifference curves being a primary cause of lexicographic answering in simplistic SC surveys will have clear design implications. An adaptive design was applied in the three surveys, with initial attribute levels set from the respondents' own journeys. Firstly, with heterogeneity in the initial attribute levels, use of percentage differences relative to the initial level may complicate the objective of sufficiently well balanced attribute levels. Most levels were calculated as ±25% of the medium level. If the initial medium level was "small" ("short" travel time, "low" price, etc.), the range of the attribute interval would also be "small". This may have produced situations where the change in level for this attribute would be ignorable in the trade-offs. For example, a respondent's actual journey could have a "relatively low" price and average levels for travel time, and according to this respondent's preference he/she could have steep indifference curves between these two attributes in terms of demanding a "large" reduction in price for an increase in travel time. However, if price is relatively low at the outset, a 25% reduction could be not enough to compensate for any given increase in travel time. One possible solution to this imbalance problem could be to apply the ±X% design only for attributes that are above a certain level at the outset (not "too small") and apply fixed increments/decrements for attributes with initial low levels.

The adaptive choice-based designs such as ACA do not solve all problems in the search for making choices individually relevant. An initial adaptation based on a representative journey may represent an amendement compared to just updating from previous choices. But with adaptive design there is the risk that the answers to the first pairwise choices in a sequence, those that are used to determine the attribute levels in a later choice situation, are not accurate. Respondents may find the choice alternatives difficult to grasp at the beginning and need some training. The first choices may therefore be inconsistent and not represent the true preferences (Bradley and Daly, 1994; Johnson and Desvousges, 1997). Then respondents can change their values during the choice sequences, such that the initial crude estimates of an individual's utilities (part-worths) may not yield the best design of the following choices (Johnson et al., 2003; Hensher and Rose, 2005).

Initial levels of a "representative journey" and statements about the importance of attributes and attribute ranges do not provide enough information for us to be able to assess whether the respondent has steep indifference curves between some of the attributes. Applying a set of trade-off questions between pairs of attributes, i.e. after stating the initial values of the typical journey and before the full choice sequences, could provide more precise information about the true trade-off levels between each attribute (Fearnley and Sælensminde, 2001). Notwithstanding, taking into consideration steep indifference curves, by increasing the attribute range and/or including pairwise trading-off between attributes, may also have adverse effects. Increasing the attribute range could complicate the choice task, resulting in more inconsistency and, possibly, in simplifying lexicographic answering (Cantillo and Ortúzar, 2005). An extra set of attribute trade-off questions, before the main choices, would also augment the response burden. More research is needed to assess the effect of steep indifference curves for SC and lexicographic answering and to evaluate design implications.

---

threshold preferences and steep indifference curves (Lockwood, 1999; Rosenberger et al., 2003). Furthermore, with another split design based on different choice complexity (e.g. number of alternatives) we could have applied this to test between threshold preferences or steep indifference curves, on the one hand, and simplification strategy, on the other (Caussade et al., 2005).

Acknowledgement

This study was funded by the Research Council of Norway through the programmes "Methodological challenges related to valuation of non-market goods" (155683) and "Strategic market analyses" (171388). The surveys reported in the study were obtained from projects financed by the Norwegian Public Roads Administration, the Norwegian National Rail Administration, AS Oslo Sporveier, Greater Oslo Passenger Transport Ltd, and Oslo County Council. The article benefited greatly from the comments of two anonymous referees to this journal. We also thank Jon Martin Denstadli, Kjartan Sælensminde and Ronny Klæboe for valuable input. We bear sole responsibility for remaining errors and omissions.

References

Arentze, T., Borgers, B, Timmermans, H. and DelMistro, R. (2003). Transport stated choice responses: effects of task complexity, presentation format and literacy. *Transportation Research Part E*, vol. 39, pp. 229-244.

Bates, J.J. (1998). Reflections on stated preference: theory and practice. In: Ortúzar, J. de D., Hensher, D.A. and Jara-Diaz, S.R. (eds.) *Travel Behaviour Research: Updating the State of Play*. Pergamon, Oxford.

Bradley, M. and Daly, A. (1993). New analysis issues in stated preference research. *Proceedings 21st PTRC Summer Annual Meeting*, Institute of Science and Technology, University of Manchester, UK.

Bradley, M. and Daly, A. (1994). Use of logit scaling approach to test for rank-order and fatigue effects in stated preference data. *Transportation*, vol. 21, pp. 167-184.

Cairns, J. and Van der Pol, M. (2004). Repeated follow-up as a method for reducing non-trading behaviour in discrete choice experiments. *Social Science and Medicine*, vol. 58, pp. 2211-2218.

Cantillo, V. and Ortúzar, J. De D. (2005). A semi-compensatory discrete choice model with explicit attribute threshold of perception. *Transportation Research Part B*, vol. 39, pp. 641-657.

Carlsson, F. and Martinsson, P.(2001). Do hypothetical and actual marginal willingness to pay differ in choice experiments? Application to the valuation of the environment. *Journal of Environmental Economics and Management*, vol. 41, pp. 179-192.

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

Colman, A.M. and Stirk, J.A. (1999). Singleton bias and lexicographic preferences among equally valued alternatives. *Journal of Economic Behaviour and Organization*, vol. 40, pp. 337-351.

Encarnación, J. (1964). A note of lexicographical preferences. *Econometrica*, vol. 32, pp. 215-217.

Encarnación, J. (1990). Consumer choice of qualities. *Economica*, vol. 57, no. 225, pp. 63-72.

Fearnley, N. and Sælensminde, K. (2001). *Tester av stated preference-teknikker og samvalgdesign*. TØI Report 544/2001, Institute of Transport Economics, Oslo.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford University Press, CA.

Foerster, J.F. (1979). Mode choice decision process models: a comparison of compensatory and non-compensatory structures. *Transport Research Part A*, vol. 13, pp. 17-28.

Fosgerau, M. (2006). Investigating the distribution of the value of travel time savings. *Transportation Research Part B*, vol. 40, no. 8, pp. 688-707.

Gigerenzer, G. and Todd, P.M. (1999). *Simple Heuristics That Make Us Smart*. Oxford Univeristy Press, Oxford.

Georgescu-Roegen, N. (1954). Choice, expectations and measurability. *Quarterly Journal of Economics*, vol. 68, pp. 503-534.

Greene, W.G. (1993). *Econometric Analysis*. MacMillan, NY.

Heiner, R.A. 1983. The origin of predictable behaviour. *American Economic Review*, vol. 73, pp. 560-595.

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

Hensher, D.A., Barnard, P.O. and Truong, T.P. (1988). The role of stated preference methods in studies of travel choice. *Journal of Transport Economics and Policy*, vol. 22, pp. 45-58.

Hensher, D.A. and Rose, J.M. (2005). Respondent behaviour in discrete choice modelling with a focus on the valuation of travel time savings. *Journal of Transportation and Statistics*, vol. 8, no. 2, pp. 17-30.

Hojman, P., Ortúzar, J. de D. and Rizzi, L.I. (2004). On the joint valuation of averting fatal victims and severe injuries in highway accidents. *Journal of Safety Research*, vol. 36, no. 4, pp. 377-386.

Johnson, F.R. and Desvousges, W.H. (1997). Estimating stated preferences with rated-pair data: environmental, health, and employment effects of energy programs. *Journal of Environmental Economics and Management*, vol. 34, pp. 79-99.

Johnson, R., Huber, J. and Bacon, L. (2003). Adaptive choice based conjoint analysis. Sawtooth Software, Inc. Sequim, WA.

Kanninen, B.J. (2002). Optimal design for multinomial choice experiments. *Journal of Marketing Research*, vol. 39, pp. 214-227.

Killi, M., Samstad, H. and Sælensminde K. (2002). *Trafikanters verdsetting av trafikkinformasjon. Resultater fra en stated preference pilotstudie*. TØI Report 537/2001, Institute of Transport Economics, Oslo.

Killi, M. and Samstad, H. (2002). *Trafikanters verdsetting av informasjon med utgangspunkt i arbeidsreiser*. TØI Report 620/2002, Institute of Transport Economics, Oslo.

Kjørstad, K.N. (1995). *Kollektivtrafikantenes preferanser. Erfaringer fra Moss, Grenland, Kristiansand, Tromsø og Ålesund*. TØI Report 312/1995, Institute of Transport Economics, Oslo.

Kriström, B. (1997). Spike models in contingent valuation. *American Journal of Agricultural Economics*, vol. 79, pp. 1013-1023.

Lanscar, E. and Louviere, J. (2006). Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? *Health Economics*, vol. 15, pp. 797-811.

Lockwood, M. (1999). Preference structures, property rights, and paired comparisons. *Environmental and Resource Economics*, vol. 13, pp. 107-122.

Mazzotta, M.J. and Opaluch, J.J. (1995). Decision making when choices are complex: a test of Heiner's hypothesis. *Land Economics*, vol. 71, pp. 500-515.

Montgomery, H. (1987). Image theory and dominance search theory: how is decision making actually done? *Acta Psycologica*, vol. 66, pp. 221-224.

Norheim, B. (2003). Stated preference survey on Internet: testing the Oslo survey. Unpublished draft, Institute of Transport Economics, Oslo.

Norheim, B. (2004). STATLEX: A model for statistical tests of lexicographic answers in SP-surveys. Unpublished draft, Institute of Transport Economics, Oslo.

Norheim, B., Kjørstad, K.N. and Renolen, H. (1994). *Ny Giv for kollektivtrafikk i Drammensregionen. Hovedresultater fra samvalganalysen*. TØI Report 241/1994, Institute of Transport Economics, Oslo.

Norheim, B. and Stangeby, I. (1993). *Bedre kollektivtransport. Oslo-trafikantenes verdsetting av høyere standard*. TØI Report 167/1993, Institute of Transport Economics, Oslo.

Nossum, Å. (2003). *Kollektivtilbudet i Osloregionen. Trafikantenes verdsetting av tid*. TØI Report 633/2003, Institute of Transport Economics, Oslo.

Ortúzar, J. de D. and Garrido, R.A. (1994). A practical assessment of stated preferences methods. *Transportation*, vol. 21, no. 3, pp. 289-305.

Ortúzar, J. de D. and Rodríguez, G. (2002). Valuing reductions in environmental pollution in a residential location context. *Transportation Research Part D*, vol. 7, pp. 407-427.

Payne, J.W., Bettman, J.R. and Johnson, E.J. (1993). *The Adaptive Decisionmaker*. Cambridge University Press, NY.

Rosenberger, R.S., Peterson, G.L., Clarke, A. and Brown, T.C. (2003). Measuring dispositions for lexicographic preferences of environmental goods: integrating economics, psychology and ethics. *Ecological Economics*, vol. 44, pp. 63-76.

Rouwendal, J. and Blaeij, A.T. de (2004). *Inconsistent and lexicographic choices in stated preference analysis*. TI 2004-038/3 Tinbergen Institute Discussion Paper, Tinbergen Institute, Amsterdam.

Sawtooth Software. (1991). *ACA 5.0 Technical Paper*. Sawtooth Software, Inc. Sequim, WA.

Spash, C.L. (2000). Ethical motives and charitable contributions in contingent valuation: empirical evidence from social psychology and economics. *Environmental Values*, vol. 9, pp. 453-479.

Stern, D.I. (2000). Limits to substitution and irreversibility in production and consumption: a neoclassical interpretation of ecological economics. *Ecological Economics*, vol. 21, pp. 197-215.

Stopher, P.R. and Hensher D.A. (2000). Are more profiles better than fewer? Searching for parsimony and relevance in stated choice experiments. *Transportation Research Record*, vol. 1719, pp. 165-174.

Swait, J. and Adamowicz, W. (1999). Choice task complexity and decision strategy selection. Working paper September 1999, Dept. of Marketing, College of Business Administration, University of Florida, Gainesville, FL.

Sælensminde, K. (1999). Stated choice valuation of urban traffic air pollution and noise. *Transportation Research Part D*, vol. 4, pp. 13-27.

Sælensminde, K. (2000). *Valuation of nonmarket goods for use in cost-benefit analyses: methodological issues*. TØI Report 491/2000, Institute of Transport Economics, Oslo.

Sælensminde, K. (2001). Inconsistent choices in stated choice data: use of the logit scaling approach to handle resulting variance increases. *Transportation*, vol. 28, pp. 269-296.

Sælensminde, K. (2002). The impact of choice inconsistencies in stated choice studies. *Environmental and Resource Economics*, vol. 23, pp. 403-420.

Sælensminde, K. (2006). Causes and consequences of lexicographic choices in stated choice studies. *Ecological Economics*, vol. 59, no. 3, pp. 331-340.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, vol. 76, pp. 31-48.

Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological Review*, vol. 79, pp. 281-299.

Tversky, A., Sattath, S. and Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review*, vol. 95, pp. 371-384.

Varian, H.R. (1992). *Microeconomic analysis*. 3.ed., Norton and Company Inc., NY.

Veisten, K., Navrud, S. and Valen, J.S.Y. (2006). Lexicographic preference in biodiversity valuation: tests of inconsistencies and willingness to pay. *Journal of Environmental Management and Planning*, vol. 49, no. 2, pp. 167-180.

Vibe, N., Kjørstad, K.N., Nossum, Å. and Ruud, A. (2004). *Kollektivalternativene i Tønsbergpakken: TØIs bidrag til konsekvensutredningen*. TØI Report 698/2004, Institute of Transport Economics, Oslo.

Widlert, S. (1998). Stated preference studies: the design affects the results. In: Ortúzar, J. de D., Hensher, D.A. and Jara-Diaz, S.R. (eds.) *Travel Behaviour Research: Updating the State of Play*, Pergamon, Oxford.