

Carlsson, Fredrik; Mørkbak, Morten Raun; Olsen, Søren Bøye

## Article

# The first time is the hardest: A test of ordering effects in choice experiments

Journal of Choice Modelling

## Provided in Cooperation with:

Journal of Choice Modelling

*Suggested Citation:* Carlsson, Fredrik; Mørkbak, Morten Raun; Olsen, Søren Bøye (2012) : The first time is the hardest: A test of ordering effects in choice experiments, Journal of Choice Modelling, ISSN 1755-5345, University of Leeds, Institute for Transport Studies, Leeds, Vol. 5, Iss. 2, pp. 19-37

This Version is available at:

<https://hdl.handle.net/10419/66841>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by-nc/2.0/uk/>

## The first time is the hardest: A test of ordering effects in choice experiments

Fredrik Carlsson<sup>1,\*</sup> Morten Raun Mørkbak<sup>2,†</sup> Søren Bøye Olsen<sup>3,‡</sup>

<sup>1</sup> Department of Economics, School of Business, Economics and Law, University of Gothenburg, Box 640, 405 30 Gothenburg, Sweden.

<sup>2</sup> Department of Business and Economics, COHERE, University of Southern Denmark, J.B. Winsløvs vej 9, 5000 Odense, Denmark.

<sup>3</sup> Institute of Food and Resource Economics, University of Copenhagen, Rolighedsvej 25, 1958 Frederiksberg, Denmark.

Received 20 February 2012, revised version received 31 May 2012, accepted 10 June 2012

### Abstract

This paper addresses the issue of ordering effects in choice experiments, and in particular how learning processes potentially affect respondents' stated preferences in a sequence of choice sets. In a case study concerning food quality attributes of chicken breast filets, we find evidence of ordering effects in a sequence of 16 choice sets, where the last eight choice sets are identical to the first eight. We find evidence of changes in preferences. More precisely there are differences in preferences for the price attribute for the two identical sequences. Moreover, we find a reduction in the error variance for the last eight choice sets relative to the first eight choice sets. This is mainly caused by very high error variances in the first two choice sets. These results imply that learning effects in terms of institutional learning as well as – though in our case only to a limited extent – preference learning can indeed be of significant structural importance when conducting CE surveys.

*Keywords: Choice Experiments, Fatigue, Learning, Ordering Effects*

---

\* Corresponding author, T: +46 31 773 41 74, [fredrik.carlsson@economics.gu.se](mailto:fredrik.carlsson@economics.gu.se)  
† T: +45 21 94 93 37, [mrm@sam.sdu.dk](mailto:mrm@sam.sdu.dk)  
‡ T: +45 35 33 36 43, [sobo@foi.dk](mailto:sobo@foi.dk)



## 1 Introduction

Ample experimental evidence in stated preference (SP) surveys suggests that respondents do not always make the coherent choices they are expected to. For instance it has been shown that they are often affected by survey context and various information cues as well as the order of choice tasks that respondents are required to make in Choice Experiments (CE) and in multiple bounded dichotomous choice Contingent Valuation (e.g., Ajzen et al. 1996, Carlsson and Martinsson 2008, Day et al. 2012, Ladenburg and Olsen 2008, Louviere 2006). This is clearly at odds with standard assumptions in economics, e.g., concerning transitivity, rationality, and continuity. In this paper, we are particularly interested in investigating how learning effects potentially affect respondents' choices through a sequence of choice sets in CE. This issue is related to what is known as ordering effects. Day et al. (2012) provide a discussion of different explanations and manifestations of ordering effects. They discuss six different effects that are not necessarily mutually exclusive. The first one is preference learning or the discovered preference hypothesis (DPH), which relates to preference uncertainty (Plott 1996, Plott and Zeiler 2005). The hypothesis states that when respondents are faced with new decisions in unfamiliar environments, initial decisions will exhibit significant randomness. However, as choices are repeated and respondents gain familiarity with their own preferences as well as the decision environment, the decisions progressively become more coherent and less random. The second effect is referred to as institutional learning, which relates to the fact that most respondents participating in SP surveys have never experienced this type of survey before. In experimental economics this has been described as confusion or failure of game form recognition (Andreoni 1995, Chou et al. 2009). Both preference learning and institutional learning suggest that one way to reduce uncertainty in SP surveys is to have respondents make repeated choices (Braga and Starmer 2005). Yet, the third effect discussed by Day et al. (2012) is fatigue. Respondents could get tired of the choice task if it is repeated many times, and thus, their choices may exhibit increasing levels of randomness over the sequence of choice tasks (Swait and Adamowicz 2001). The fourth effect potentially causing ordering effects is the starting point effect, where respondents who are uncertain about their preferences for the good regard a presented price as a cue to the "correct" value for that good, and consequently they anchor their WTP to this value (Kahneman et al. 1982). A preference learning process such as the DPH would lead to elicited preferences moving away from the initial starting point towards the 'true' preferences of the respondent, and, hence the result would be a detectable change in preference structure. However, if respondents are not subject to preference learning but rather try to choose coherently according to an arbitrary starting point as suggested by the coherent arbitrariness view (Ariely et al. 2003), the starting point effect would essentially be undetectable as no change in preference structure would appear. Finally, the fifth and the sixth effects are both related to the fact that respondents may act strategically (Carson and Groves 2007, Day et al. 2012, Day and Pinto 2010)<sup>1</sup>. In the case of provision of private goods, the effect of strategic behavior is not clear-cut since it depends on what assumptions one makes about the respondent's cognitive capacity. However, one typical behavior could be that

---

<sup>1</sup> Day et al. (2012) distinguish between strategic behavior that is based upon a full recall of the presented prices (strategic behavior) and a recall that is weighted toward the recent prices seen (reference price behavior).

respondents reject alternatives if they have already had an opportunity to obtain a similar alternative at a lower cost in a previous choice set.

In the present paper we investigate to what extent preferences are stable in a choice experiment (CE) concerning food safety attributes of chicken breast filets. However, instead of varying the position of a given choice set, we apply an experimental setup using 16 choice sets per respondent where the last sequence of eight choice sets is an exact copy of the first sequence of 8 choice sets. We estimate separate models based on the two sequences. The null hypothesis is that the preferences and the error variance do not differ between the two sequences.

Preference stability in stated preference surveys with repeated questions has received considerable attention in the literature. However, perhaps as could be expected, no consistent pattern has emerged. The majority of the studies have focused on and examined learning and fatigue effects by comparing the choices made in identical choice sets presented at different positions in a sequence of choice sets. A number of studies find no or small differences in preferences based on choices made in the beginning and choices made at the end of a sequence (e.g., Carlsson and Martinsson 2001, Johnson and Bingham 2001, Hanley et al. 2002, Brouwer et al. 2010). Other studies find that stated preferences do depend on the sequence of preference-eliciting choice questions. For example, Day et al. (2012) test this by repeating the first choice set at the end of the sequence of choice sets. They find that respondents do not necessarily choose the same alternative when the choice set is placed at the end of a choice set sequence instead of in the beginning of the sequence. It has also been found that respondents may suffer from starting point bias, or coherent arbitrariness, but that learning seems to reduce this bias (Ladenburg and Olsen 2008). A number of studies also show that a respondent's consistency in choices depends on the complexity of the task, on the positions of the choice sets in the sequence, and on his/her cognitive capability (e.g. Dellaert et al. 1999, DeShazo and Fermo 2002, DeSarbo et al. 2004, Lagerkvist et al. 2006, Brown et al. 2008, Savage and Waldman 2008, Day et al. 2012, Day and Pinto 2010).

Our design offers some additional insights compared to previous studies. First of all, by repeating exactly the same choice sets, the influence of exact sequencing of the choice sets is limited. Secondly, by repeating only the first choice set in the end of the sequence as previous studies have mainly done, it is only possible to check for discrepancies between the first and the last choice set. By repeating the entire sequence of choice sets we can make this check for all choice sets in the experimental design. Furthermore, we can compare the full preference structure as well as individual preference parameters and WTP estimates in order to test for ordering effects. Though our experimental design does not allow us to fully discriminate between all the different potential effects of ordering mentioned above, we argue that our results provide evidence mainly for learning effects taking place. We find evidence of changes in preferences. More precisely there are differences in preferences for the price attribute for the two identical sequences. This will translate into a difference in WTP estimates between the two sequences. At the same time, there are no significant differences in preferences for the non-monetary attribute between the two sequences. We find significant reductions in unexplained error variance for the last sequence of eight choice sets relative to the first sequence of 8 choice sets. Elaborating on this, we find that it is especially in the first two choice sets that the error variance is high (compared to the remaining choice sets across the two sequences) and it is also for these two choice sets, in particular the first, that we find the largest share of respondents changing their choices in the identical choice sets

given later in the sequence. We suggest that this observed ordering effect is primarily caused by institutional learning. Our results imply that learning effects in terms of institutional learning as well as – though in our case only to a limited extent – preference learning can indeed be of significant structural importance when conducting CE surveys.

The paper is organized as follows. Section 2 describes the econometric approach and Section 3 the design of the survey and the data. Section 4 presents the results and a discussion and Section 5 concludes the paper.

## 2 Econometric approach

The experimental design consists of 16 choice sets, where the same eight sets are given in two sequences. The last sequence of eight sets is an exact copy of the first sequence of eight sets. In the analysis, we apply a standard random utility model (McFadden 1974), where the utility of alternative  $j$  for individual  $i$  in choice set  $k$  in sequence  $S$  is specified as

$$U_{ijk}^S = v_{ijk}^S + \varepsilon_{ijk}^S = \beta_i^S a_{jk} + \varepsilon_{ijk}^S, \quad (1)$$

where  $a$  is a vector of attributes,  $\beta$  is the corresponding parameter, and  $\varepsilon_{ijk}$  is an error term. If the error terms are iid extreme value distributed with variance  $\pi^2/(6\mu^2)$ , the standard logit model choice probability that individual  $i$  chooses alternative  $j$  is

$$P_{ijk}^S = \frac{\exp(\mu^S v_{ijk}^S)}{\sum_{m \in k} \exp(\mu^S v_{im}^S)}, \quad (2)$$

where  $\mu$  is a scale parameter that is inversely proportional to the error variance. The coefficients ( $\beta$ ) in the econometric models are usually expressed in their scaled form ( $\beta = \mu\beta^*$ ), where the scale parameter  $\mu$  and the ‘true’ coefficients  $\beta^*$  are confounded. Hence, the estimated parameter  $\beta$  indicates the effect of each observed variable relative to the variance of the unobserved factors (Train 2003).

Comparing the estimated models from the first and the last sequence, there are thus two elements involved: the parameter vector and the scale parameters. The problem, though, is that the scaling factor cannot be identified in any particular set of empirical data. Instead, the ratio of the scale factor of one data set relative to another can be identified by normalizing one of them to unity and then defining a range of values of the other scale factor, within which the log likelihood function is expected to be maximized (Swait and Louviere 1993). The scale ratio is then identified either through a grid search procedure or by full maximization of the likelihood. In the present paper we have used the latter approach. We draw on the notion used by, e.g., Holmes and Boyle (2005) and Savage and Waldman (2008) that learning processes can lead to increased consistency in choice, which implies reduced error variance and a higher degree of estimation precision (Heiner 1983, de Palma 1994, Hole 2007), whereas fatigue effects will have the opposite implication (Day et al. 2012). These effects are of course in addition to any potential effects of learning or fatigue on the vector of attribute parameter estimates.

In the analysis, we estimate random parameter models where we assume that all non-price attributes are normally distributed, thereby allowing consumers to place positive as well as negative values on the non-price attributes and the alternative specific constant. Focus group interviews indicated that such heterogeneity could be expected. The price coefficient is assumed to follow a constrained symmetric triangular distribution where the mean equals the spread. From an economic theory point of view this ensures a behaviorally plausible sign for the estimated price coefficient (Greene and Hensher 2007), and moreover we avoid the problem of the non-existence of mean WTP, when using a distribution where the cost coefficient straddles zero (e.g., an unconstrained triangular distribution) (Daly et al. 2011). We have used the software package Biogeme (Bierlaire 2003) to estimate the econometric models. In all models we control for individual level heterogeneity through the use of a panel specification capturing the repeated choice nature of the data<sup>2</sup>. The models are estimated with simulated maximum likelihood using Halton draws with 300 replications; see Train (2003) for details on simulated maximum likelihood and Halton draws. Specifically, we compare a model based on the first sequence of eight sets with a model based on the second sequence of 8 sets. Furthermore, in a model covering all 16 choice sets we introduce a dummy variable for the last sequence which is interacted with all parameter estimates in order to allow for changes in the preference parameters across the two sequences. In addition to comparing the estimated parameters and relative scale parameters between models, we also compute and compare unconditional marginal WTPs for each attribute by applying a parametric bootstrapping procedure using 10,000 replications. Finally, using a panel probit model, we identify variables that explain the probability of making identical choices in identical choice sets across the two sequences.

### 3 The choice experiment

The choice experiment concerned food safety attributes of chicken breast filets. Prior to the design of the CE, we performed three focus group interviews. In the focus groups, the following attributes were identified as being important in relation to the choice of chicken breast filets: type of production, country of origin, and, to some extent, food safety (mainly related to Salmonella). Although food safety did not appear to be of great concern to consumers, the original purpose of the present study was to elicit the relative weighting of food safety. Consequently, we included two food safety attributes associated with the chicken products: “Salmonella-free” and “Campylobacter-free”. These attributes were chosen because of their relevance to product and also because they were judged as representing an increasingly important issue from a scientific as well as a political perspective. Table 1 presents the attributes and their associated levels.

The two food safety attributes in the survey, Salmonella-free and Campylobacter-free, are quite similar. They both exhibit private good characteristics to a large extent, and both give rise to more or less the same course of illness. The main difference is that the current risk of getting infected by Campylobacter is much higher than the risk of contracting a Salmonella infection. Consequently, our a priori expectation of the

---

<sup>2</sup> Individual level heterogeneity was controlled for in each of the two sequences, implying that tastes were fixed within respondents, but not across choice set sequences. Though, in model (v) table 3 we control for individual level heterogeneity across all 16 choice sets.

**Table 1: The attributes and their levels in the choice experiment for the chicken breast file.**

Attributes	Levels
Type of production	Conventional (indoor), organic (outdoor)
Country of origin	Domestic (Danish), non-domestic
Campylobacter-free	Not labeled Campylobacter-free, Campylobacter-free
Salmonella-free	Not labeled Salmonella-free, Salmonella-free
Price (DKK)	25, 28, 33, 40, 50, 65, 85, 115

Note: DKK 10 ~ EUR 1.34.

value of a Salmonella-free chicken is that respondents will not value it as highly as the Campylobacter-free characteristic.

A fractional factorial design resulting in a sequence of eight sets with three alternatives was used for the experimental design (two generic alternatives plus a status quo). The design was evaluated employing a d-error measure. At the end of the sequence, the entire sequence of 8 sets was repeated, resulting in a total of 16 choice sets per respondent. Hence, the fractional factorial design of 8 sets was essentially presented twice to each respondent. Respondents were not made explicitly aware of this feature of the design and we had no indication from open ended follow-up questions that any respondents realized it<sup>3</sup>. In each choice set, the respondents were faced with two alternative chicken breast file products plus a third status quo alternative, all specified as packages of 500 grams. The latter characterized the respondents' usual purchase, which was identified earlier in the questionnaire. This approach of using the respondents' "own" status quo values has been recommended and used in other studies to mimic the actual purchasing situation as closely as possible (Ruby et al. 1998, Kontoleon and Yabe 2003). The use of individual specific status quo alternatives in the design procedure of CE has been further developed by Rose et al. (2008), in what they refer to as segment-specific efficient designs, two-stage process designs, and individual efficient designs.

## 4 Results and discussion

The CE survey was conducted using an internet panel, and the sample was obtained from Nielsen's online database. In Denmark, there are approximately 2.4 million private households, of which 87 percent have access to the internet. All panel members are 15 years old or older and reside in a household with internet access, yet in the present survey only respondents above age 18 were allowed to participate. The final sample consisted of 389 respondents, each answering 16 choice sets, resulting in a total of 6,224 choice observations. The respondents were not able to go back and look at or change their responses to prior questions. The response rate was 26 percent. To begin with, we split the data and estimate the models based on the two identical sequences, i.e. model (i) for the first 8 choice sets and model (ii) for the last 8 choice sets. Next, we pool the responses from the two sequences and estimate two additional

<sup>3</sup> There was no systematic test of whether respondents were aware of the duplication. If respondents became aware of this during the experiment, it potentially could lead to protest answers or response anomalies. One such indication would be that respondents would choose the status quo more often in the last sequence. However, a Pearson chi-square-test showed that respondents did not consistently choose the status quo more often in the last sequence relative to the first.

models, namely model (iii) using all 16 choice sets, but not accounting for potential difference in scale between the first and last 8 sets, and model (iv) using all 16 sets again and now accounting for potential difference in scale between the first and last eight sets.<sup>4</sup> Table 2 displays the results obtained in the four different models.<sup>5</sup>

**Table 2: RPL models for the first 8 and last 8 sets, and for pooled models; standard errors in parentheses**

	Model(i) First 8 CS		Model(ii) Last 8 CS		Model(iii) All 16 CS - not corrected for scale		Model(iv) All 16 CS – corrected for scale	
Parameter estimates	Coefficients (std. err.)	<i>t</i> -value	Coefficients (std. err.)	<i>t</i> -value	Coefficients (std. err.)	<i>t</i> -value	Coefficients (std. err.)	<i>t</i> -value
<i>Mean estimates</i>								
Campylobacter-free	0.427 (0.038)	11.22	0.488 (0.048)	10.10	0.451 (0.030)	15.29	0.149 (0.011)	13.48
Salmonella-free	0.354 (0.041)	8.54	0.660 (0.063)	10.46	0.450 (0.034)	13.29	0.152 (0.012)	12.74
Domestic produce	0.668 (0.051)	13.03	0.811 (0.069)	11.76	0.704 (0.041)	17.31	0.114 (0.013)	8.58
Outdoor production	0.275 (0.050)	5.47	0.494 (0.067)	7.38	0.343 (0.039)	8.89	0.234 (0.015)	15.20
ASC (Status quo)	0.059 (0.126)	0.47	0.294 (0.142)	2.07	0.158 (0.092)	1.71	0.047 (0.031)	1.55
Price	-0.025 (0.002)	-16.40	-0.036 (0.003)	-14.95	-0.028 (0.001)	-23.43	-0.009 (0.0005)	-19.16
<i>Standard Deviations</i>								
Campylobacter-free	0.093 (0.115)	0.81	0.283 (0.082)	3.43	0.160 (0.078)	2.04	0.061 (0.024)	2.51
Salmonella-free	0.262 (0.083)	315	0.489 (0.084)	5.82	0.368 (0.055)	6.74	0.126 (0.018)	6.99
Domestic produce	0.409 (0.071)	5.75	0.666 (0.083)	7.99	0.524 (0.054)	9.63	0.179 (0.018)	9.78
Outdoor production	0.625 (0.068)	9.22	0.755 (0.086)	8.77	0.700 (0.054)	12.91	0.235 (0.019)	12.35
ASC (Status quo)	2.050 (0.133)	15.43	2.470 (0.171)	14.50	2.240 (0.108)	20.70	0.742 (0.042)	17.54
Price	0.025 (0.002)	16.40	0.036 (0.002)	14.95	0.028 (0.001)	23.43	0.009 (0.0005)	19.16
Scale ratio, $\mu^a$							1.23 (0.069)	17.86
LL	-2483		-2339		-4827		-4821	
Adj. $\rho^2$	0.270		0.312		0.292		0.295	

<sup>a</sup> The scale ratio is the estimated scale factor for the last eight choice sets relative to the scale factor for the first eight choice sets which is normalized to one. Hence, the *t*-value for the scale factor is testing the null hypothesis  $H_0: \mu = 1$ .

<sup>4</sup> Note that the pooled model still assumes independence between the first eight and last eight choice sets.

<sup>5</sup> Note that all non-monetary attributes are effects coded.



Nearly all attribute coefficients are statistically significant at the standard five percent level, with the exception of the ASC in models (i), (iii) and (iv), and the standard deviation of the Campylobacter-free label in model (i).

The difference in responses between the two sequences is initially examined through a likelihood ratio (LR) test for equality of all model parameters (Swait and Louviere 1993). This test involves Models (i) - (iii). Comparing the pooled model in (iii) with the two separate models (i) and (ii), the log likelihood ratio is 5 – the value of the chi-squared test statistic is thus 10 – which means that we cannot reject the hypothesis of equal parameters at the standard five percent level of statistical significance (critical value at five percent and 12 d.f. is 21.03). We therefore proceed with a pooled model where we account for a difference in scale parameters across the two sequences, but where the preference parameters are restricted to be the same across the two sequences. This is model (iv) in Table 2. Recall that the scale parameter is inversely proportional to the standard deviation of the error term in our specification (Swait and Louviere 1993). The estimated relative scale factor of 1.23 in Table 2 implies that the variance of the error term or “noise” in the model based on the last sequence is only 66 percent of the variance of the model based on the first sequence<sup>6</sup>. This is also in accordance with the comparison of model fit between Models (i) and (ii), where the model based on the last sequence clearly provides a better fit to the data. The finding of a reduced error variance in the last sequence is equivalent to the findings by Holmes and Boyle (2005). Overall, the LR test results reject an overall change in the preferences for the attributes when moving from the first to the second sequence, but that error variance is reduced by almost 35 percent in sequence B. This would suggest that there is learning in the choice experiment, but not in terms of changes in preferences, but in terms of less noise, i.e., institutional learning. Do note that any attempt of respondents to be coherent goes in the opposite direction as learning. This could either be the case when there is no preference uncertainty, i.e., the respondent knows his preferences and chooses coherently according to these, or if there is preference uncertainty but the respondent chooses coherently in accordance with an arbitrary starting point. While these two types of behavior clearly have very different implications for interpretation of preference parameter estimates, it is difficult for the analyst to disentangle the two as data in both cases will suggest that no preference learning is taking place.

The above tests are joint tests that demonstrate that overall there appears to be no difference in preferences but in error variance. In the next step we estimate a pooled model accounting for differences in both random taste parameters and scale across the two sequences and allowing for dependence between the two sequences. This is done by pooling the data and estimate a panel structure of all 16 choice sets, and interacting the attribute variables with a dummy variable,  $d_{last8}$ , equal to one if the choice set is in the last sequence. The results are presented in Table 3.

---

<sup>6</sup>  $(1/1.23^2) = 0.661$ . Hence, there is a 34 percent difference in the overall unexplained variance between the two models.

**Table 3: RPL pooled model accounting for differences in preferences and scale between the two sequences**

Parameter estimates	Coefficients	Std. err.	<i>t</i> -value
<i>Mean estimates</i>			
Campylobacter-free label	0.208	0.017	12.02
Campylobacter-free label $\times d_{last8}$	-0.004	0.016	0.25
Salmonella-free label	0.176	0.018	9.70
Salmonella-free label $\times d_{last8}$	0.016	0.016	0.99
Domestic produce	0.307	0.022	14.27
Domestic produce $\times d_{last8}$	-0.003	0.021	0.02
Outdoor production	0.189	0.026	7.35
Outdoor production $\times d_{last8}$	0.018	0.017	1.10
ASC (Status quo)	-0.164	0.047	3.47
ASC (Status quo) $\times d_{last8}$	0.115	0.030	3.82
Price	-0.013	0.006	22.64
Price $\times d_{last8}$	0.002	0.0004	4.88
<i>Standard Deviations</i>			
Campylobacter-free label	0.188	0.019	9.74
Campylobacter-free label $\times d_{last8}$	0.037	0.019	1.95
Salmonella-free label	0.222	0.018	12.43
Salmonella-free label $\times d_{last8}$	0.002	0.020	0.11
Domestic produce	0.295	0.021	13.96
Domestic produce $\times d_{last8}$	0.0003	0.021	0.02
Outdoor production	0.384	0.029	13.36
Outdoor production $\times d_{last8}$	0.024	0.017	1.38
ASC (Status quo)	0.946	0.052	18.04
ASC (Status quo) $\times d_{last8}$	0.017	0.030	0.57
Price	0.013	0.006	22.64
Price $\times d_{last8}$	0.002	0.0004	4.88
Scale ratio, $\mu^a$	1.180	0.061	21.49
LL	-4136		
Adj. $\rho^2$	0.391		

<sup>a</sup> The scale ratio is the estimated scale factor for the last eight choice sets relative to the scale factor for the first eight choice sets which is normalized to one. Hence,

the *t*-value for the scale factor is testing the null hypothesis  $H_0: \mu = 1$ .

The pooled model in table 3 basically confirms the results of the LR test, with one important exception<sup>7</sup>. There are no significant differences in preferences for the non-monetary attributes between the first and last sequence. However, respondents are less price-sensitive in the last sequence as revealed by the significant  $d_{last8}$  interaction term, although the size of the interaction term is relatively small in comparison with the price attribute coefficient. The change in preferences for the price attribute

<sup>7</sup> Comparing the model fit of this interaction model to models (i) and (ii) shows a significant improvement in model fit of almost 100 likelihood units. As mentioned in footnote 2, we control for individual level heterogeneity within each choice sequence of eight choice sets in models (i) and (ii), whereas in the model presented in table 3, individual level heterogeneity is controlled for across all 16 choice sets. This is the main reason for this rather large improvement in fit.

suggests that preference learning is taking place as well.<sup>8</sup> There is also a change in the mean parameter estimate for the ASC which would imply that respondents in the last sequence have opted for the status quo alternative more often – despite the indications of the opposite in the test mentioned in footnote 3. While learning might explain this, it could also be caused by respondents increasingly using a heuristic decision strategy as a consequence of becoming bored or fatigued. The level of heterogeneity in preferences is also similar across the two sequences as none of the interactions with the standard deviation parameters are significant, apart from the price attribute. Finally, it is apparent that the last sequence still exhibits a significantly lower error variance.

The fact that there is a shift in preferences for the price attribute implies that the estimate of WTP will depend on the sequence as well. Since respondents become less price sensitive in the second sequence, mean WTP will be higher using the estimates based on the second sequence. While these results are in line with some previous studies looking at the stability of preferences (e.g., Holmes and Boyle 2005, Day et al. 2012), they are in contrast to others (e.g., Carlsson and Martinsson 2001, Hanley et al. 2002). However, the majority of these previous studies testing for differences in WTP are between-sample tests. To our knowledge, we are the first to provide this particular type of within-sample test for differences in WTP in identical choice sets. In addition, the difference in WTP will in our case only depend on differences in price sensitivity, the trade-off among the non-monetary attributes are actually stable across the two sequences. In order to illustrate the effect of the difference in price sensitivity we report the unconditional marginal mean WTP for each attribute and the ASC obtained on the basis of the indirect utility parameter estimates in table 3. Standard errors are estimated with the Krinsky-Robb method, using 10,000 replications. To compare the different WTP estimates across the two sequences, we have applied the Complete Combinatorial test (CC test) suggested by Poe et al. (2005). For the CC-test we use 1,000 unconditional WTP estimates for each sequence, which implies 1,000,000 differences. Table 4 presents the results.

**Table 4: Mean WTP estimates in DKK per 500g package of chicken breast filet.**

	First 8 CS		Last 8 CS		CC-test p-value <sup>a</sup>
	Mean WTP	St. dev.	Mean WTP	St. dev.	
Campylobacter-free label	32.7	2.7	37.4	3.2	0.13
Salmonella-free label	27.7	2.9	35.2	3.2	0.04
Domestic produce	48.3	3.4	49.5	3.8	0.42
Outdoor production	29.7	4.0	38.0	4.6	0.09
ASC (Status quo)	-12.94	3.8	-4.5	4.2	0.07

Note: DKK 10 ~ EUR 1.34. WTPs and standard deviations are obtained using a Krinsky-Robb bootstrapping procedure with 10,000 draws. The ‘first 8 CS’ WTPs are based only on the main parameter estimates from the pooled model in table 3 whereas the ‘last 8 CS’ WTPs are also taking the interaction parameter estimates into account.

<sup>a</sup> P-values report the proportion of estimates in the complete combinatorial where WTP from the model obtaining the highest mean WTP is higher than WTP from the other model.

<sup>8</sup> It should be noted that the type of preference learning that we find partly depends on model specification. For example, if we instead estimate a model with an unconstrained triangular distribution, the interaction term for the price attribute becomes insignificant, while three other interaction terms become significant but again these are very small relative to the attribute coefficients.

As expected, the mean WTPs are higher for the second sequence. However, the results from the complete combinatorial testing for equality of WTP estimates imply that there is only a significant difference for one of the attribute at the five percent-level: the ‘Salmonella-free label’. At the same time the differences in mean WTP is sometimes substantial.

Another interesting question is whether the change in preferences can be traced to certain choice sets or a certain part of the order of the choice sets. Given our within-subject design, we can make a number of comparisons. Following Rigby and Burton (2011) we use a panel probit model in order to explain ‘hit success’, where the dependent dummy variable takes a value of one if the respondent makes the same choice in the identical choice sets in the first and the last sequence, and zero otherwise. As explanatory variables in the panel probit model we use: 1) Socio-demographic characteristics (gender, age, household income, and education), 2) Respondents self-reported certainty statements<sup>9</sup>, 3) Utility difference at choice set level as a proxy for the complexity of a choice set<sup>10</sup>, and finally 4) An ordering effect which is captured by a series of choice set dummies with the initial choice set as the baseline. The results are presented in Table 5.

There are no significant differences with respect to the socio-demographic variables. The results further show that respondents who state that they are certain with respect to their preferences (value certainty) have higher probability of making the same choice in the last sequence. This is in accordance with our expectations, that respondents with well known preferences do not exhibit value learning. In addition, the significant impact of utility difference reveals that the larger the utility difference in a choice task, the larger the probability that the respondent will choose the same alternative in the last sequence of choice sets. This is also as expected as making a choice in a choice set with large utility difference will, *ceteris paribus*, be easier than choosing from a choice set where utility is more balanced. Finally, we find evidence of choice set order effects. In particular, all seven choice set dummies show a significantly<sup>11</sup> positive effect relative to the baseline dummy for choice sets 1 and 9. In other words, the probability of observing identical choices in choice sets 1 and 9 is lower than in all the following choice sets.

Since we have two observations of choices for the same choice set, we can conduct symmetry tests to further examine this issue (see e.g., Stata 2007). When looking at the choices made in each choice set, we can reject the hypothesis of no difference in responses between the two sequences for four out of eight choice sets at the five percent level. The pairs of choice sets where there is a significant difference are 1 and 9; 2 and 10; 5 and 13; and 8 and 16. However, it is for the first choice set

---

<sup>9</sup> Respondents were asked to state how certain they are after each choice set with respect to institutional and value uncertainty respectively. The two certainty questions were as follows (translated from Danish): Institutional uncertainty statement: ‘On a scale from 1-5 (where 1 is ‘totally incomprehensible’ and 5 is ‘easily comprehensible’), how comprehensible was the construction of the choice task shown above?’ Value uncertainty statement: ‘On a scale from 1-5 (where 1 is ‘very uncertain’ and 5 is ‘very certain’), how certain are you that your chosen alternative is the alternative which contributes with the largest value to you?’

<sup>10</sup> Estimation of the utility differences is done in line with Olsen et al. (2011): Hence, we use the estimates from the model presented in table 3 to calculate the expected aggregate utility of each alternative in each choice set for each individual and then calculate the expected utility difference between the chosen alternative, and the best alternative to that.

<sup>11</sup> Though for choice sets 3 and 11 and 5 and 13 the cutoff level for statistical significance needs to be relaxed to six percent and nine percent, respectively.

**Table 5: Panel probit model explaining identical choices in identical choice sets across the two sequences.**

	Description (mean value)	Marginal effect	Std. Err.	t-value
Female	= 1 if respondent is a female (0.55)	-0.008	0.017	0.47
Age	Age in years (51.32)	-0.001	0.006	0.60
Household income	= 1 if below DKK 100,000; and = 10 if above DKK 900,000 (5.38)	-0.004	0.003	1.24
Low Education	= 1 if respondent only has basic education (0.14)	-0.005	0.024	0.23
High Education	= 1 if respondent has university education (0.46)	-0.018	0.017	1.00
Institutional uncertainty	= 1 if totally incomprehensible; and = 5 if easily comprehensible (4.09)	-0.009	0.010	0.97
Value uncertainty	= 1 if very uncertain; and = 5 if very certain (3.82)	0.043	0,009	4.69
Utility Diff.	Estimated average utility difference (0.005)	0.088	0.010	9.17
Status Quo	= 1 if respondent chose the status quo in the first sequence (0.47)	0.121	0.012	9.56
CS 2 and 10	= 1 if choice set 2 or 10 (0.125)	0.067	0.026	2.57
CS 3 and 11	= 1 if choice set 3 or 11 (0.125)	0.046	0.024	1.94
CS 4 and 12	= 1 if choice set 4 or 12 (0.125)	0.094	0.025	3.76
CS 5 and 13	= 1 if choice set 5 or 13 (0.125)	0.041	0.024	1.71
CS 6 and 14	= 1 if choice set 6 or 14 (0.125)	0.096	0.027	3.58
CS 7 and 15	= 1 if choice set 7 or 15 (0.125)	0.085	0.026	3.26
CS 8 and 16	= 1 if choice set 8 or 16 (0.125)	0.069	0.024	2.85
Constant		0.266	0.311	0.85
Rho		0.123	0.029	4.31

that the largest difference in choices is observed – as is also evident from Table 5. Comparing Set 1 with Set 9, almost 27 percent of the respondents change their answer, while for the other comparisons the share varies between 11 and 20 percent. This suggests that there is something particular with the first choice set. In order to further explore this, we re-estimate the pooled random parameter logit model, yet this time we estimate separate scale factors for choice sets 2 to 16 relative to the normalized scale factor in choice set 1, which is fixed to unity. Table 6 presents the results.

Since we have two observations of choices for the same choice set, we can conduct symmetry tests to further examine this issue (see e.g., Stata 2007). When looking at the choices made in each choice set, we can reject the hypothesis of no difference in responses between the two sequences for four out of eight choice sets at the five percent level. The pairs of choice sets where there is a significant difference are 1 and 9; 2 and 10; 5 and 13; and 8 and 16. However, it is for the first choice set that the largest difference in choices is observed – as is also evident from table 5. Comparing Set 1 with Set 9, almost 27 percent of the respondents change their answer, while for the other comparisons the share varies between 11 and 20 percent. This suggests that there is something particular with the first choice set. In order to further explore this, we re-estimate the pooled random parameter logit model, yet this time we estimate separate scale factors for choice sets 2 to 16 relative to the normalized scale factor in choice set 1, which is fixed to unity. Table 6 presents the results.

**Table 6: Estimation and comparison of choice set-specific scale factors in the pooled RPL model, t-values for test of  $H_0: CS(t) = 1$** 

<b>CS<sub>i</sub> (First 8 CS)</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>
Scale factor $\mu$	1	1.28	1.92	1.54	1.86	1.54	1.10	2.26
Std. Err.	0	0.132	0.154	0.144	0.174	0.144	0.155	0.175
t-values,		2.16	6.08	3.75	5.09	3.89	0.66	7.41
<b>CS<sub>i</sub> (Last 8 CS)</b>	<b>9(1)</b>	<b>10(2)</b>	<b>11(3)</b>	<b>12(4)</b>	<b>13(5)</b>	<b>14(6)</b>	<b>15(7)</b>	<b>16(8)</b>
Scale factor $\mu$	1.54	1.69	2.23	1.65	1.70	1.75	1.02	2.00
Std. Err.	0.131	0.144	0.175	0.144	0.158	0.258	0.150	0.156
t-values	4.23	4.84	6.98	4.60	4.60	4.85	0.14	6.60
t-values (CS(t) vs CS(t+8))	4.28	3.03	1.83	0.79	1.06	1.46	0.54	1.43

Note: The vertical alignment of choice sets in the table corresponds to the choice sets that are identical across the two sequences.

As Table 6 shows, almost all of the estimated scale factors for sets 2 to 16 differ significantly from the normalized scale factor in the first set. The exceptions are sets 7 and 15, which, interestingly, are the same choice tasks. Furthermore, they all are larger than one, and hence the error variance is smaller in these sets relative to the first set. This is further evidence that the behavior in the first choice set is different from the behavior in the other sets 2 to 16 and that the noise in most cases is significantly larger in the first set. Moreover, there are no significant differences in scale factors across the identical pairs of choice sets except between sets 1 and 9 and sets 2 and 10. Our results are largely in accordance with Bateman et al. (2008), who found evidence of institutional learning in a repeated double bounded contingent valuation survey. Similarly, both DeSarbo et al. (2004) and Brown et al. (2008) find evidence of learning effects in terms of increasing choice consistency as respondents progress through a sequence of choices.

While we interpret our results as being clearly in support of learning effects, there might be precedent-dependent effects that could explain why respondent choices differ in otherwise identical choice sets (Day and Pinto 2010, Day et al. 2012). Particularly, starting point effects and strategic behavior could affect respondent behavior. If respondent uncertainty, be it institutional or preference, introduces a true random component as a psychological view on the random utility model would prescribe (Thurstone 1927), we would expect some deviations between choices made in identical choice sets that only differ by order (Day et al. 2012). This, being a purely random component, would not affect or bias the preference parameter estimates in any way, only the noise-to-signal ratio, i.e. the scale factor and standard errors associated with the parameter estimates. However, if the respondent anchors his choice on some arbitrary starting point due to such initial uncertainty, a structural component will be introduced, likely to bias the preference parameters in the direction of the starting point. Such starting point effects could potentially explain the indications we have of at least some preference learning going on in our data (Carlsson and Martinsson 2008; Ladenburg and Olsen 2008). However, the fact that the good being valued here is a very common market good, namely chicken breast filet with which the majority of the respondents have extensive consumer experience, we would expect respondents to be relatively more certain about their preferences than for instance for a pure non-market good with which they have no previous consumer experience (Plott 1996, List 2003,

Day et al. 2012). As a consequence, we argue that the relatively high error variance in the first choice sets is mainly indicative of institutional uncertainty. This is further supported by the fact that we see a fast learning process where the differences in unexplained error variance across the two sequences reduces drastically after the first couple of choice sets. Bateman et al. (2008) and Swait and Adamowicz (2001) suggest that such a fast learning process is indicative of institutional learning. Preference learning would arguably be a slower process as for instance in Ladenburg and Olsen (2008), where a starting point bias is found to persist for a larger number of choice sets in an environmental non-market good case. In the same vein, Czajkowski and Giergiczny (2011) in a case considering a non-market good find increasing scale parameters over a larger number of choice sets. However, Plott (1996) and Hanley et al. (2009) suggest that the speed of preference learning would be positively correlated with the level of experience with the good, and Nunes and Boatwright (2004) find some evidence that respondents also experience preference uncertainty over well-known, common market goods. In our case, the preference learning is centered on a change in the preferences for the price attribute, while the trade-offs among the non-monetary attributes are constant. Coherent arbitrariness could explain why we find no differences in the trade-offs between the non-monetary attributes. Nevertheless, we find this explanation less likely since, first of all, we would expect a low degree of preference uncertainty for this well-known marketed good, and, secondly, we do find some evidence of preference learning effects which may be just enough to offset a low degree of preference uncertainty. This, however, remains speculative as we have no built-in test for coherent arbitrariness in the current dataset. Furthermore, the shift in preferences is most likely not a result of strategically behavior either. As we have discussed in the introduction, a typical strategic behavior would be that respondents reject alternatives that they have already had an opportunity to obtain at a lower cost in a previous choice set (Day et al. 2012). Finally, even though we rule out fatigue effects in terms of increasing variance, our results concerning changes in status quo choice propensity as a consequence of fatigue are ambiguous. Additionally, we cannot rule out the possibility that the apparent change in preferences for the price attribute is actually caused by respondents increasingly using some other type of simplifying decision heuristic – e.g., attribute non-attendance – as a consequence of becoming fatigued. While we have no way of testing for this in the current dataset, the string of research looking into decision heuristics at the choice task level (e.g., Scarpa et al. 2010, Meyerhoff and Liebe 2009 and Puckett and Hensher 2009) would seem a promising avenue ahead to investigate this in future research.

## 7 Conclusions

This paper addresses the issues of how ordering effects related to the repeated choices in a choice experiment setting potentially affect respondents' stated preferences. The main focus is on respondent uncertainty and its potential impacts on preference structure, error variance, and willingness-to-pay estimates, and on how such impacts reflect institutional and/or preference learning effects. We use an experimental design where a sequence of choice sets is given twice to the same respondent, allowing for within-sample tests. Comparing the two sequences of identical choice sets we find significant differences in error variances as well as in the preferences for the price attribute. More specifically, the error variance is lower in the last than in the first sequence, suggesting that respondent uncertainty is lower in the second sequence. In the model where we estimate separate scale factors for choice sets 2 to 16 relative to

the normalized scale factor in choice set 1, we find that it is only in the first choice set that the error variance is higher in the first sequence. This is also where we find that the largest share of respondents change their choices when comparing identical choice sets across the first and last sequence. We argue that institutional uncertainty is mainly related to increased error variances whereas preference uncertainty can be associated with increases in error variances as well as biases in the estimated preference structure and WTP estimates, depending on the decision strategy employed by the respondent to overcome such preference uncertainty. In a study considering a well-known market good, namely chicken breast filet, we find evidence that institutional learning as well as preference learning, is indeed present. A very rapid decrease in error variance wears off after the first couple of choice sets suggesting that institutional learning takes place. This is in line with results in Hess et al. (2012) who also find increasing scale after the first choice set. The difference in preferences for the price attribute for the two identical sequences could be evidence of preference learning which translates into a difference in WTP measures. At the same time, there are no significant differences in preference estimates for the non-monetary attribute between the two sequences. We argue that part of our results indicate that fatigue is not present – we do not find a larger error variance in the last sequences nor do we find any evidence on respondents choosing inconsistently in the last sequence compared to the first sequence. Though we cannot rule out the difference in the price attribute being due to fatigue caused by respondents using different decision heuristics like attribute non-attendance. This issue is scope for further research.

Our findings have clear implications for the design of choice experiments. If the goal is to minimize the error variance and, hence, maximize the statistical power of the structural model one should be careful when including responses to the first two choice sets in the dataset. In fact, as suggested by Ladenburg and Olsen (2008), one should consider including one or two examples of choice sets or additional choice sets that are not generated by the statistical design and that are not intended for use in the analysis of preferences. This approach is likely to diminish any institutional uncertainty caused by the CE format and thus reduce potential ordering effects. This is similar to findings in experimental economics, where it is evident that the experimental instructions and the context could play an important role. For example, in the case of public good games, some types of subjects act out of confusion and not kindness (Andreoni 1995) because they fail to understand the incentives, what Chou et al. 2009 denotes ‘failure of game form recognition’. Ferraro and Vossler (2010) find evidence that this confusion can partly be reduced by learning, but more importantly, that the instructions can reduce confusion substantially.

Furthermore, if the data is used for estimating preference parameters and informing decision-makers, our results show that ordering effects potentially cannot be neglected. It should however be noted that there is a risk of starting point bias if preference uncertainty is present, as may be the case especially in non-market good surveys. This suggests that the first choice set should be varied between survey versions in order to test and control for potential starting point bias, and, if present, this first choice set (or potentially more) should be excluded from estimation.

This, as well as the major body of available studies on learning and fatigue effects, has assumed that all respondents follow the same pattern of learning and fatigue. However, recent developments in the literature (e.g., Campbell et al. 2011, Czajkowski and Giergiczny 2011) suggest that the impacts of this assumption are not trivial. Further research in this direction will be relevant in order to explore in more



detail the potential impacts of learning effects, and in particular the extent to which institutional and preference uncertainty as well as learning styles differ across people.

## Acknowledgements

We wish to thank two anonymous reviewers and seminar participants at the University of Gothenburg as well as participants at the 2011 EAERE conference in Rome and the 2011 ICMC conference in Leeds for helpful comments.

## References

- Ajzen, I., T. C., Brown, L. H., Rosenthal and H., Lori, 1996. Information bias in contingent valuation: effects of personal relevance, quality of information and motivational orientation. *Journal of Environmental Economics and Management*, 30(1), 43–57.
- Andreoni, J., 1995. Cooperation in public-goods experiments: Kindness or confusion. *American Economic Review*, 85(4), 891-904.
- Ariely, D., G., Loewenstein and D., Prelec, 2003. Coherent arbitrariness: Stable demand curves without stable preferences. *Quarterly Journal of Economics*, 118(1), 73-105.
- Bateman, I., D., Burgess, G., Hutchinson and D., Matthews, 2008. Learning design contingent valuation (LDCV): NOAA guidelines, preference learning and coherent arbitrariness. *Journal of Environmental Economics and Management*, 55(2), 127-141.
- Bierlaire, M., 2003. BIOGEME: A free package for the estimation of discrete choice models. Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.
- Braga, J. and C., Starmer, 2005. Preference anomalies, preference elicitation and the discovered preference hypothesis. *Environmental and Resource Economics*, 32(1), 55-89.
- Brouwer, R., T., Dekker, J., Rolfe and J., Windle, 2010. Choice certainty and consistency in repeated choice experiments. *Environmental and Resource Economics*, 46(1), 93-109.
- Brown, T., D., Kingsley, G., Peterson, N. E., Flores, A., Clarke and A., Birjulin, 2008. Reliability of individual valuations of public and private goods: Choice consistency, response time, and preference refinement. *Journal of Public Economics*, 92(7), 1595–1606.
- Campbell, D., M., Boeri, E., Doherty and G., Hutchinson, 2011. Learning and fatigue in discrete choice experiments: addressing preference and variance instability. Paper presented at the 18<sup>th</sup> EAERE conference, 29 June – 2 July 2011, Rome.
- Carlsson, F., and P., Martinsson, 2001. Do hypothetical and actual marginal willingness to pay differ in choice experiments? *Journal of Environmental Economics and Management*. 41(2), 179-192.
- Carlsson, F., and P., Martinsson, 2008. How much is too much?—An investigation of the effect of the number of choice sets, starting point and the choice bid vectors in choice experiments. *Environmental and Resource Economics*, 40(2), 165–176.
- Carson, R., and T., Groves, 2007. Incentive and informational properties of preference questions. *Environmental and Resource Economics*, 37(1), 181-210.

- Chou, E., M., McConnell, R., Nagel and C., Plott, 2009. The control of game form recognition in experiments: understanding dominant strategy failures in a simple two person 'guessing' game. *Experimental Economics*, 12(2), 159-179.
- Czajkowski, M., and M., Giergiczny, 2011. How many choice-situations are optimal? Investigating ordering effects – The impact of the number of choice situations on individual preference and scale heterogeneity using the G-MNL model. Paper presented at the 18<sup>th</sup> EAERE conference, 29 June – 2 July 2011, Rome
- Daly, A., S., Hess and K., Train, 2011. Assuring finite moments for willingness to pay in random coefficient models. *Transportation*, 39(1), 19-31.
- Day, B., I., Bateman, R., Carson, D. Dupont, J., Louviere, S. Morimoto, R., Scarpa, and P., Wang, 2012. Ordering effects and choice set awareness in repeat-response stated preference studies, *Journal of Environmental Economics and Management*, 63(1), 73-91.
- Day, B. and J. L., Pinto, 2010. Ordering anomalies in choice experiments. *Journal of Environmental Economics and Management*, 59(3), 271-285.
- de Palma, A., G., Meyers and Y., Papageorgiou, 1994. Rational choice under an imperfect ability to choose. *American Economic Review*, 84(3), 419-440.
- Dellaert, B., J., Brazell, J., Louviere, 1999. The effect of attribute variation on consumer choice consistency. *Marketing Letters*, 10(2), 139-147.
- DeSarbo, W. S., D. R., Lehmann and F. G. Hollman, 2004. Modeling dynamic effects in repeated-measures experiments involving preference/choice: an illustration involving stated preference analysis. *Applied Psychological Measurement*, 28(3), 186-209.
- DeShazo, J.R., and G., Fermo, 2002. Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management*, 44(1), 123-143.
- Ferraro, P., and C., Vossler, 2010. The source and significance of confusion in public goods experiments. *The B E Journal of Economic Analysis and Policy*, 10(1), Article 53.
- Greene, W. H. and D. A., Hensher, 2007. Heteroscedastic control for random coefficients and error components in mixed logit. *Transportation Research Part E: Logistics and Transportation Review*, 43(5), 610-623.
- Hanley, N., B., Kristrom and J. F., Shogren, 2009. Coherent Arbitrariness: On Value Uncertainty for Environmental Goods. *Land Economics*, 85(1), 41-50.
- Hanley, N., R.E., Wright and G., Koop, 2002. Modelling Recreation Demand Using Choice Experiments: Climbing in Scotland. *Environmental and Resource Economics*, 22(3), 449-466.
- Heiner, R., 1983. The origin of predictive behavior. *American Economic Review*, 73(4), 560-595.
- Hess, S., D., Hensher. and A., Daly, 2012. Not bored yet – revisiting respondent fatigue in Stated Preference experiments. *Transportation Research Part A*, 46(3), 626-644.
- Hole, A. R., 2007. A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health Economics*, 16(8), 827-840.
- Holmes, T. and K., Boyle, 2005. Learning and Context-Dependence in Sequential, Attribute-Based, Stated-Preference Valuation Questions. *Land Economics*, 81(1), 114-126.
- Johnson, F. R. and M., Bingham, 2001. Evaluating the validity of stated-preference estimates of health values. *Swiss Journal Economics and Statistics*, 137(1), 49-63.

- Kahneman, D., P., Slovic and A., Tversky, 1982. *Judgement Under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press
- Kontoleon, A. and M., Yabe, 2003. Assessing the Impacts of Alternative 'Opt-out' Formats in Choice Experiment Studies: Consumer Preferences for Genetically Modified Content and Production Information in Food. *Journal of Agricultural Policy Research*, 5 1-43.
- Ladenburg, J. and S. B., Olsen, 2008. Gender-specific starting point bias in choice experiments: Evidence from an empirical study. *Journal of Environmental Economics and Management*, 56(3), 275-285.
- Lagerkvist, C.J., F., Carlsson and D., Viske, 2006. Swedish Consumer Preferences for Animal Welfare and Biotech: A Choice Experiment, *AgBio Forum*, 9(1), 51-58.
- List, J., 2003. Does Market Experience Eliminate Market Anomalies? *Quarterly Journal of Economics*, 118(1), 41-71.
- Louviere, J., 2006. What you don't know might hurt you: some unresolved issues in the design and analysis of discrete choice experiments. *Environmental and Resource Economics*, 34(16), 173-188.
- McFadden, D., 1974. *Conditional Logit Analysis of Qualitative Choice Behavior*. In: P. Zarembka (ed.) *Frontiers in Econometrics*. Academic, New York.
- Meyerhoff, J. and U., Liebe, 2009. Discontinuous preferences in choice experiments: Evidence at the choice task level. Paper presented at the 2009 EAERE Conference, Amsterdam.
- Nunes, J., and P., Boatwright, 2004. Incidental Prices and Their Effect on Willingness to Pay. *Journal of Marketing Research*, 41(4), 457-466.
- Olsen, S. B., T., Lundhede, J. B., Jacobsen and B. J., Thorsen, 2011. Tough and Easy Choices: Testing the Influence of Utility Difference on Stated Certainty-in-Choice in Choice Experiments. *Environmental and Resource Economics*, 49(1), 491-510.
- Plott, C. R., 1996. *Rational Individual Behaviour in Markets and Social Choice Processes: The Discovered Preference Hypothesis*. In: K. Arrow et al. (ed) *Rational Foundations of Economic Behaviour*. Macmillan Press Ltd., London.
- Plott, C. R., and K., Zeiler, 2005. The Willingness to Pay-Willingness to Accept Gap, the "Endowment effect", Subject Misconceptions, and Experimental Procedures for Eliciting Valuations. *American Economic Review*, 95(3), 530-545.
- Poe, G. L., K. L., Giraud and J. B., Loomis, 2005. Computational methods for measuring the difference of empirical distributions. *American Journal of Agricultural Economics*, 87(2), 353-365.
- Puckett, S., M., and D. A., Hensher, 2009. Revealing the extent of **process** heterogeneity in choice analysis: an empirical assessment. *Transportation Research Part A*, 43(2), 117-126.
- Rigby, D. and M., Burton, 2011. Intertemporal choice consistency and the information sensitivity of welfare estimates in stated preference studies. Paper presented at the 18th Annual Conference of the European Association of Environmental and Resource Economists, Rome, Italy, 29 June - 2 July 2011.
- Rose, J. M., M., Bliemer, D., Hensher, and A. T., Collins, 2008. Designing efficient stated choice experiments in the presence of reference alternatives. *Transportation Research Part B*, 42(4), 395-406.
- Ruby, M. C., F. R., Johnson, and K. E., Mathews, 1998. Just say no: Opt-out alternatives and anglers' stated preferences. TER General Working Paper No.T-9801R.

- Savage, S. J. and D., Waldman, 2008. Learning and Fatigue During Choice Experiments: A Comparison of Online and Mail Survey Modes. *Journal of Applied Econometrics*, 23(3), 351-371.
- Scarpa, R., M., Thiene, and D. A., Hensher, 2010. Monitoring Choice Task Attribute Attendance in Nonmarket Valuation of Multiple Park Management Services: Does It Matter? *Land Economics*, 86(4), 817-839.
- Stata, 2007. *Stata Statistical Software: Release 10*. College Station, TX: StataCorp LP.
- Swait, J., and W., Adamowicz, 2001. The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching. *Journal of Consumer Research*, 28(1), 135-148.
- Swait, J., and J., Louviere, 1993. The Role of the Scale Parameter in the Estimation and Comparison of Multinomial Logit Models. *Journal of Marketing Research*, 30(3), 305-314.
- Thurstone, L.L., 1927. A Law of Comparative Judgement. *Psychological Review*, 34, 273-286.
- Train, K., 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.