

Sequencing Anomalies in Choice Experiments

Brett Day¹ and Jose-Luis Pinto²

¹ Lecturer in Environmental Sciences, University of East Anglia, UK

² Professor of Health Economics, Universidad Sevilla, Spain

Abstract

This paper investigates whether responses to choice experiments (CEs) exhibit the sorts of sequencing anomalies that have been responsible for the widespread discrediting of iterative dichotomous choice contingent valuation methods. Using a novel experimental design that simplifies and clarifies the testing task, our findings categorically reject the assumption that CEs are immune from such anomalies. In particular, we observe sequencing effects operating in both price and commodity dimensions. Moreover, we find robust evidence to suggest that over a series of choice tasks these sequencing effects are cumulative. Our findings cast serious doubt on the current practice of asking each respondent to undertake several choice tasks in a CE whilst treating each response as an independent observation on that individual's preferences.

1. Introduction

Over recent years, techniques of choice modelling such as choice experiments (CEs) have enjoyed a startling rise in popularity amongst the practitioners of non-market valuation (Adamowitz, 2004). The fundamental building block of a CE is a choice task. A choice task confronts subjects with two or more alternative states (one of which may be the status quo), where the states differ both in the qualities of the non-market good (the “commodity dimension”) and in the cost imposed on the subject (the “price dimension”). The usual procedure is to ask subjects to indicate their preferred state in a series of such choice tasks. As such, CEs provide a rich data source from which researchers can deduce how subjects are prepared to trade off between money and the various dimensions of the commodity space.

In contrast, dichotomous-choice (DC) contingent valuation techniques, that had previously enjoyed the status of “most preferred valuation method”, provide a relative paucity of data. In their earliest inception (Bishop and Heberlein, 1979), DC contingent valuation elicitation presented each respondent with just one task; a choice between either the status quo or the provision of the non-market good at a cost. Across a sample of respondents, the commodity dimension was held constant whilst the price dimension was varied so that the willingness to pay (WTP) distribution for that particular manifestation of the non-market good could be estimated. Whilst this ‘single-bounded’ DC (SBDC) elicitation method was strongly endorsed by the US National Oceanographic and Atmospheric Administration’s blue ribbon panel in 1993 (Arrow et al., 1993), practitioners were concerned by its relative inefficiency. In particular, the data from a SBDC cannot provide values for other manifestations of the non-market good that differ along one or more of the various commodity dimensions. Moreover, the fact that each respondent provides just one piece of information requires the costly acquisition of large sample sizes.

In response to the latter criticism, Hanemann, Loomis and Kanninen (1991) proposed the ‘double-bounded’ DC (DBDC) elicitation method. Here, following the initial DC question, a ‘follow-up’ DC question is asked offering the non-market good at a second and different price. The elicitation of a second response yields substantial gains in terms of statistical efficiency. Indeed, even greater gains are possible if further follow-up questions are added (Langford et al., 1996; Scarpa and Bateman, 2000).

Subsequent empirical testing, however, has revealed a robust anomaly in responses to DBDC questions. In particular, numerous studies have observed that the preferences implied by responses to first questions differ systematically from those implied by responses to follow-up questions (McFadden, 1994; Cameron and Quiggen, 1994; Herriges and Shogren, 1996; Bateman et al. 2001; De Shazo, 2002). This observation casts serious doubt on the validity of responses to follow-up questions. So robust are these sequencing anomalies that elicitation formats that present one or more follow-up questions are now widely discredited. Indeed, one might argue that these well-documented problems have precipitated a growing disaffection with DC contingent valuation and contributed to the growing interest in CE methods.

While there are a number of differences between the DC contingent valuation and CE approaches, there are also many similarities. For example, whilst differing in presentation, the SBDC elicitation method is essentially a simple form of CE in which subjects face only one choice task requiring a preference to be stated between the status quo and an alternative in which a non-market good is provided at a price. Likewise, the DBDC elicitation method is a CE with two choice tasks pitting the status quo against an alternative in which the non-market good is provided. In this case, moving from the first choice task to the second, the price dimension of the alternative is altered, but there is no change in the commodity dimension. Given these similarities and given the robust anomalies noted in responses to DBDC elicitation, a fundamental question that must be asked of the CE method is whether it too is subject to sequencing anomalies.

This paper presents the results of an experiment specifically designed to answer that question. Our particular application concerns the valuation of health using a CE with three choice tasks. In each task respondents were asked to choose between two differently priced treatments offering different levels of health benefit. The design is such that there are a relatively small number of particular choice tasks. These choice tasks are presented to independent subsamples in differing sequences. As such, pairwise comparison of responses across samples provides a straightforward test of various hypotheses regarding the existence and nature of sequencing anomalies in CEs.

As explained in detail subsequently, our design allows us to compare worsening price sequences (i.e. the second choice task differs from the first only in that the price of one of the treatments is relatively larger) with improving price sequences. Likewise, we compare worsening and improving commodity sequences (i.e. the second choice task differs from the first only in offering an option with a smaller (greater) health benefit). In addition, we examine the effects of simultaneous price and commodity sequences (i.e. the second choice task differs from the first in offering one option with a worse (improved) price and the other with a worse (improved) health benefit). Finally, our design allows us to consider how sequencing anomalies behave in a series of choice tasks exhibiting a variety of sequences.

The hypotheses we wish to test are similar to those addressed by Holmes and Boyle (2005) though our experimental setup is significantly different. In particular,

Holmes and Boyle use a testing framework that relies on the parametric modeling of the utility function. In contrast, our testing reduces to the straightforward nonparametric comparison of proportions (choosing particular options) across samples. Accordingly, we believe that our experimental design simplifies the testing tasks while increasing the credibility of our findings.

In the next section we describe the sorts of sequencing anomalies previously observed in non-market valuation exercises and consider how these might affect responses to CE-style questions. Section 3 describes the experiment we have designed to test those hypotheses. Section 4 presents the results of our empirical research and Section 5 provides a summary and some concluding remarks.

2. Sequencing Anomalies in Non-Market Valuation

The standard neoclassical economic model asserts axiomatically that individuals have complete and coherent preferences. In the context of a CE, the implication of the standard model is that individuals are able to determine which of any set of options is the one they most prefer. Moreover, provided respondents are motivated to answer the choice tasks truthfully then their responses should not be affected by features of the elicitation procedure that are, according to the theory, decision-irrelevant. For example, according to this model of behaviour the preferred option in a particular choice task should not change according to the nature of the options presented in previous choice tasks. Responses that conform to this prediction are said to demonstrate procedural invariance.

In the majority of applications of the CE methodology procedural invariance is assumed without testing. Violations of procedural invariance, however, present a profound problem for the elicitation of preferences using CEs. If, for example, expressed preferences differed systematically according to the sequence in which choice tasks were presented, then which should be taken as representing true preferences; those expressed at the beginning of the sequence or those at the end?

Of course, violations of procedural invariance are well-documented in iterative DC contingent valuation exercises. In particular, several studies have reported *price sequencing* anomalies in DBDC elicitation (e.g. McFadden, 1994; Cameron and Quiggen, 1994; Carson *et al.*, 1994; Herriges and Shogren, 1996; Alberini, Kanninen

and Carson, 1997; Bateman et al. 2001) possibly the most detailed of which being that of De Shazo (2002). De Shazo (2002) took as a baseline the proportion of individuals in a sample stating that they would be willing to pay a certain price, $\$X$, for a non-market good in the first question and compared this acceptance rate to that observed when $\$X$ was presented as the price in a second question. He observed that when $\$X$ represented a lower price than that presented in the first question (what we shall term an *improving* price sequence) its acceptance rate did not differ systematically from the baseline. In contrast when $\$X$ represented an increase from the price presented in the first question (what we shall term a *worsening* price sequence) its acceptance rate was significantly lower than the baseline.

De Shazo argues that the asymmetric pattern of price sequencing effects can be explained through a framing model derived from prospect theory and reference-dependent utility theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991; Bateman et al., 1997, forthcoming). Key to De Shazo's framing model is the assertion that in accepting an offered price, respondents assume that an informal exchange has been concluded. As a result, a respondent answering "Yes" to the first question forms a 'reference point' that includes the surplus that they expect to enjoy from the exchange made at that price. From this reference point, the subsequent presentation of a higher price is regarded by the respondent as precipitating a loss in surplus. De Shazo argues that this negative framing will tend to bias down acceptance rates for options presented in a worsening price sequence. In contrast, a respondent answering "No" to the first price amount effectively refuses the trade. In this case, no new reference point is formed, such that in an improving price sequence, the second (and lower) price offered in the follow-up question is neither negatively, nor positively framed.

If price sequencing affects acceptance rates in iterative DC contingent valuation exercises then an obvious question would be to ask if *commodity sequencing* also precipitates violations of procedural invariance. This question has received much less attention in the contingent valuation literature. In fact, as far as we can ascertain, only Bateman and Brouwer (2006) have considered this issue previously and then only indirectly whilst investigating the question of scope sensitivity. Bateman and Brouwer confront respondents with two SBDC questions concerning two levels of provision of the same good such that the smaller level of

provision is completely dominated by the larger level. They compare the median value estimated from a sample that answer in an improving commodity sequence (low provision first followed by high provision) with those estimated from a sample answering in a worsening commodity sequence (high provision first followed by low provision). They observe what they describe as “some fanning out of estimates as we move from first to second responses” (p.207). That is to say, the implied values of the large and small levels of provision are relatively more similar when calculated from the first question responses than when calculated from the follow-up responses. This observation is consistent with a commodity sequencing anomaly in which a good is regarded more favourably when preceded by a question offering a relatively smaller level of provision (an improving commodity sequence) whilst being regarded less favourably if preceded by a relatively larger level of provision (a worsening commodity sequence).¹

We summarise the evidence for sequencing effects from the DC contingent valuation literature in Table 1.

Table 1: Impact of sequence on option acceptance rates as observed in dichotomous choice contingent valuation studies

		Sequence	
		Worsening	Improving
Dimension	Price	–	None
	Commodity	–	+

A CE can also present respondents with sequences that replicate those recorded in iterative DC contingent valuation surveys. Consider a simple CE in which the respondent is faced by a series of tasks each requiring a choice to be made between just two options. To simplify further, imagine that the commodity dimension of the good can be described as either small, medium or large (where the increment to

¹ A subsequent parametric analysis of the data reports that these affects are not statistically significant, though this is not altogether surprising given the very small samples sizes used in the study.

a higher level implies complete dominance of the lower levels). Likewise the prices can take the values €0, €Cheap and €Dear (where again the increment is always such that the option becomes unequivocally more expensive). In that case, an improving price sequence can be generated by asking two consecutive choice task questions in which one option in each task remains the same but the price of the other option falls from the first to the second question (holding the commodity dimension constant). Similarly, a worsening price sequence is generated by holding one option constant whilst increasing the price of the other option from first to second question, (again, holding its commodity dimension constant). Examples, of such sequences are illustrated in Figure 1.

Figure 1: Examples of Price Sequencing in a Choice Experiment

Improving Price Sequence:

Question 1	
Option 1	Option 2
Small	Large
€0	€Dear

⇒

Question 2	
Option 1	Option 2
Small	Large
€0	€Cheap

Worsening Price Sequence:

Question 1	
Option 1	Option 2
Small	Large
€0	€Cheap

⇒

Question 2	
Option 1	Option 2
Small	Large
€0	€Dear

If De Shazo’s observations from DC contingent valuation carry over to the CE framework, then we would expect to observe violations of procedural invariance to be induced by the worsening price sequence. To illustrate, imagine presenting independent samples with the two sequences of questions shown in Figure 1. Observe that the first question faced by one sample is identical to the second question faced by the other sample. Procedural invariance demands that the proportion of subjects favouring Option 2 in response to these choice tasks will be roughly similar across the two samples. In contrast, the framing hypothesis proposed by De Shazo only predicts

procedural invariance to characterise the improving price sequence. The worsening price sequence, it is predicted, will induce anomalous behaviour. In particular, the proportion favouring Option 2 in the second question of the worsening price sequence will be significantly less than the proportion favouring that Option when presented in the first task to the other sample.

Figure 2 presents a similar construction but this time illustrating improving and worsening commodity sequences. Again the first question faced by one sample is identical to that faced by the other sample and procedural invariance requires that the proportion choosing Option 2 in these choice tasks remains roughly similar across samples. If the observations of Bateman and Brouwer (2006) carry over to the CE framework, then we expect procedural invariance to be violated in both improving and worsening commodity sequences. In particular, the proportion choosing Option 2 in the second question of the improving commodity sequence will be significantly greater than that observed when offered to the other sample in the identical choice task but as a first question. The opposite is expected in the worsening commodity sequence. Here the proportion choosing Option 2 in the second question of the worsening commodity sequence is expected to be significantly less than that observed when offered to the other sample in the identical choice task as a first question.

Figure 2: Examples of Commodity Sequencing in a Choice Experiment

Improving Commodity Sequence:

Question 1	
Option 1	Option 2
Medium	Large
€0	€Dear

⇒

Question 2	
Option 1	Option 2
Small	Large
€0	€Dear

Worsening Commodity Sequence:

Question 1	
Option 1	Option 2
Small	Large
€0	€Dear

⇒

Question 2	
Option 1	Option 2
Medium	FH
€0	€Dear

Whilst our primary interest is in the possibility of observing sequencing anomalies in CE, we are aware that in the majority of applications it is unusual to have only one of the options changing from choice task to choice task. The possibility exists that sequencing anomalies might arise in this context also. Figure 3, for example, illustrates choice tasks that present respondents with mixed commodity and price sequences; that is to say, one option presents an improving (worsening) price sequence whilst the other an improving (worsening) commodity sequence.

Figure 3: Examples of Mixed Pricing and Commodity Sequencing in a Choice Experiment

Improving Commodity (Option 1) and Price (Option 2) Sequences:

Question 1	
Option 1	Option 2
Small	Large
€0	€Dear

⇒

Question 2	
Option 1	Option 2
Medium	Large
€0	€Cheap

Worsening Commodity (Option 1) and Price (Option 2) Sequences:

Question 1	
Option 1	Option 2
Medium	Large
€0	€Cheap

⇒

Question 2	
Option 1	Option 2
Small	Large
€0	€Dear

If the patterns of behaviour observed in DC contingent valuation studies carry over to these more complex choice situations, then the improving mixed sequence should result in violations of procedural invariance. In particular, in the improving sequence the relatively larger commodity offered by Option 1 in the second question leads respondents to regard this option more favourably, whilst De Shazo's framing hypothesis indicates that the improved price offered by Option 2 in the second question results in no such equivalent bias. As a result, we might expect to observe a relatively greater proportion favouring Option 1 in a choice task presented in an improving mixed sequence.

In the worsening mixed sequence, the two sequencing biases work in the same direction; the relatively smaller commodity offered by Option 1 makes this option appear less favourable but the relatively greater price offered by Option 2 makes this also appear less favourable. Since we are unable to determine in advance which of the two sequencing affects will dominate, it is not possible to make predictions concerning violations of procedural invariance in this case.

As far as we are aware, the only previous study to examine sequencing anomalies in CEs is that of Holmes and Boyle (2005). As we discuss subsequently, they employ a very different and considerably less direct testing framework to that employed in our research. All the same, they find evidence of both price and commodity sequencing. With regards to pricing sequences they observe some limited evidence to suggest that worsening price sequences reduce the likelihood of choosing an option. However, they also find somewhat stronger evidence that improving price sequences increase the likelihood of choosing an option; a finding inconsistent with the evidence from DBDC contingent valuation studies and De Shazo's framing hypothesis.

Somewhat surprisingly Holmes and Boyle (2005) also test to see if the prices offered by options in subsequent choice tasks impact on choices in the current task; though it is not at all clear how subjects are made aware of the prices offered by these subsequent tasks. Nonetheless, and perhaps worryingly so far as the credibility of their findings is concerned, they find that these subsequent prices may also impact on current choices.

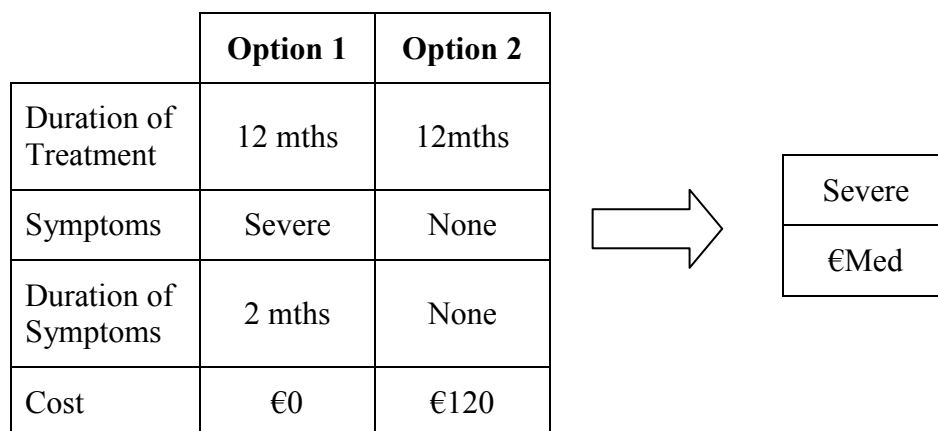
In addition, Holmes and Boyle (2005) test for commodity sequences, though their findings are very hard to interpret. In particular, they make no differentiation between worsening and improving sequences. Accordingly, their results merely inform us that if an option in the previous task offered a certain level of the commodity then this will have some fixed impact on choices in the current task; that is to say, an impact that is independent of whether that level is more or less than that offered by the same option in the current choice task. The behaviour implied by this result is difficult to reconcile with our intuition concerning how sequencing effects responses and is out of kilter with the limited evidence from the DC contingent valuation literature.

3. Experimental Design and Testing Framework

Our application concerns the valuation of health using a CE with three choice tasks. In each choice task respondents were asked to imagine that they had been diagnosed with a medical problem that would result in a considerable deterioration in their quality of life. Respondents were informed that the problem was treatable and that following treatment they would be returned to full health within one year. In each choice task, subjects were asked to choose between two treatment options. One option was a medicine provided by the hospital. Whilst this option was free of charge it meant that the subject would still experience quality of life-reducing symptoms for the first 2 months of treatment. The second option in each choice task was to purchase an alternative medicine from a pharmacy. Whilst this treatment was costly, it meant that the subject would avoid any symptoms whilst being treated thereby enjoying a quality of life that was equivalent to full health.

The commodity dimension of options in our CE described one of three states of health; two months of severe ill-health (severe symptoms) or two months of mild ill-health (mild symptoms) or on-going full health (no symptoms). In addition, the price dimension of options in our CE described one of four levels of treatment cost; €0, €60 (€Low), €120 (€Med) or €240 (€High). As shown in the left hand panel of Figure 4, a typical choice task pitted a zero cost treatment with 2 months of reduced quality of life, against a costly treatment that returned the subject to immediate full-health.

Figure 4: Schematics of typical choice task



In order to simplify the presentation of our experimental design, it is expedient to further condense each choice task into the simple schematic shown in the right hand panel of Figure 4. Here the top box represents Option 1 where the cost, treatment duration and symptoms duration dimensions have all been suppressed since these were always €0, 12 months and 2 months respectively. Likewise the bottom box represents Option 2 with the commodity dimensions suppressed since these were always 12 months of treatment with no symptoms.

Our experimental design involved a six-way split sample with each sample facing a different set of three choice tasks. The tasks and order in which they were presented to each respondent are summarized using the simplifying schematic in Figure 5.

Taking procedural invariance as our null hypothesis our design is intended to examine the possibility of the alternative hypotheses suggested by the sequencing anomalies observed in DC contingent valuation and summarized in Table 1. In particular, procedural invariance requires that given two independent samples the proportion choosing Option 1 (or for that matter Option 2) should be approximately equal and independent of the details of any preceding choice tasks. The essence of our experimental design, therefore, is to confront independent samples with the same choice task but set in differing sequences and to record the proportions choosing each option. In this context, the null hypothesis can be tested using robust nonparametric tests to compare those proportions across the two samples. Since the alternative hypotheses described in Table 1 are, in the main, directional, the significance of those comparisons can be determined with reference to a one-tailed test for equality of proportions in independent samples.

More usually, the testing of elicitation methods in the non-market valuation literature adopts parametric testing techniques. For example, consider the study of sequencing in CEs undertaken by Holmes and Boyle (2005). Their design involved a relatively large number of choice tasks offering options with different combinations of commodity and price attributes. In contrast to our experimental design, the various tasks were presented to respondents in a “completely randomized” (p.120) order. Accordingly, Holmes and Boyle do not control for the nature of sequences presented to respondents such that there is no guarantee that their design has the power to untangle the confounding influences of commodity and price sequences in worsening

and improving directions. In the usual manner, Holmes and Boyle (2005) proceed by fitting a parametric model of preferences that assumes a linear indirect utility function that is additively separable in price and commodity attributes. In addition, they allow for the possibility of sequencing affects by including terms reflecting the values of price and commodity attributes of options in previous (and future) choice tasks. Their specification is completed with an additively separable random error term that they assume to be normally distributed. Whilst following standard practice, each of the modeling assumptions made in developing this specification adds to the untested structure imposed on preferences increasing the likelihood of reaching erroneous conclusions concerning the nature of sequencing effects.

In contrast, our experimental design has the power to identify sequencing affects (should they exist) without imposing untested structure on preferences. Our nonparametric approach is both simple and transparent and as such, we believe, increases the credibility of our findings.

The central tests employed in our analysis are those described in relation to Figure 1, Figure 2 and Figure 3. In particular, observe that the first two questions faced by sample A_1 represent a worsening price sequence for Option 2, whilst the order of those questions is reversed for sample A_2 providing the equivalent improving price sequence for Option 2. For convenience we refer to the first question faced by sample A_1 as A_1-1 , likewise the second question faced by this sample is A_1-2 , and so on. Using this notation, our null hypothesis is that the proportion of sample A_2 choosing Option 2 in A_2-1 should equal the proportion of sample A_1 choosing that option in A_1-2 . Since A_1-2 is the second task in a worsening price sequence our alternative is the directional hypothesis that the proportion choosing Option 2 in A_1-2 will be less than that observed in A_2-1 .

The test for an improving price sequence pits the proportion choosing Option 2 in A_1-1 against the proportion choosing Option 2 in A_2-2 . In this case, our null and alternative hypotheses both predict procedural invariance such that we do not expect to observe significant differences in the behaviours of the two samples.

Tests of commodity sequencing are provided by samples B_1 and B_2 . Option 1 of Question B_2-2 is in a worsening commodity sequence. As such a comparison with B_2-1 allows us to test our expectations that the worsening commodity sequence biases

down the number of respondents choosing an option. Likewise Option 1 of Question B_{1-2} is in an improving commodity sequence so that we can compare this with responses to B_{2-1} in order to ascertain whether the improving commodity sequences biases up the number of respondents choosing an option.

Figure 5: Experimental design

<u>Sample</u>	<u>Question 1</u>	<u>Question 2</u>	<u>Question 3</u>
A_1	Severe €Med	Severe €High	Mild €Med
A_2	Severe €High	Severe €Med	Mild €Low
B_1	Mild €Med	Severe €Med	Mild €Low
B_2	Severe €Med	Mild €Med	Severe €High
C_1	Mild €Low	Severe €High	Mild €Med
C_2	Severe €Med	Mild €Low	Severe €High

Notes:

Top box represents Option 1, bottom box Option 2.

Severe = 2S, Mild = 2Q

High = €240, Medium = €120, Low = €60

Tests of mixed price and commodity sequencing are provided by samples C_1 and C_2 . Here Option 1 of C_2-2 is in an improving commodity sequence whilst Option 2 is in an improving price sequence. Accordingly, a comparison with C_1-1 allows us to test our expectation that the improving price sequence has no impact on responses whilst the improving commodity sequence biases up the number of respondents choosing an option. In that case, we expect to observe the number choosing Option 1 in C_2-2 to be significantly higher than those choosing that option in C_1-1 .

The test for a mixed price and commodity sequence in the worsening direction was supposed to be provided by a comparison of C_1-2 with C_2-1 . Unfortunately the survey company miscoded Option 2 of C_2-1 , presenting the medium price rather than the high price. This was inconvenient since, as we discuss shortly, we had hoped to check for consistency across samples by comparing responses to this task to the identical task presented in A_2-1 . All the same, the design was not completely undermined as we are still able to compare C_1-2 with responses to the identical task in A_2-1 to examine the possible impacts of the mixed worsening sequence. Indeed, since this comparison was the one for which we have no definitive alternative hypothesis (both options should be regarded less favourably in the worsening sequences) we regarded this comparison as the least important of those provided by the original design.

Our design allows for a number of other tests that we hoped would shed further light on the way in which subjects formulate responses to CE questions. Observe first that there are only four different choice tasks in the design. Each of these is presented to at least one sample as the first question. Since (unlike Holmes and Boyle) we do not believe that respondents can be influenced by the attributes of options in choice tasks they have not as yet observed, we take these as providing an unbiased indicator of preferences. If this is true, then responses to questions A_1-1 , B_2-1 and C_2-1 should be approximately the same since these present respondents in independent samples with exactly the same choice task (erroneously, as it happens, in the case of C_2-1).

In a similar manner we can use responses to the first questions to test some simple predictions of economic theory. For example, Option 1 of A_2-1 is identical to that presented in A_1-1 , B_2-1 and C_2-1 , however, Option 2 of A_2-1 offers a price that is categorically higher than the price offered in Option 2 of the other three questions.

Since economic theory assumes that high prices are less favourable to low prices, we would expect to see significantly fewer respondents opting for Option 2 in A_2-1 . A similar comparison can be made between B_1-1 and C_1-1 . Likewise, economic theory maintains that “more is better”. As such we would expect to see a greater proportion of respondents favouring Option 1 in B_1-1 than in A_1-1 , B_2-1 and C_2-1 .

The design also allows inter-personal tests of consistency. For example, consider an individual in sample A_1 answering the first question. If they indicate that they would prefer to suffer the severe ill-health event rather than opt for the treatment offering an immediate return to full health at the medium price, then it follows that they should also favour the severe ill-health event over the costly treatment when that treatment is offered at the higher cost in the second question.

Since these questions are directly juxtaposed, it should be reasonably self-evident to respondents how to answer the questions with internal consistency. That consistency might be less easy to maintain across a series of questions. Take, once again, the example of the individual in sample A_1 favouring the severe ill-health event to the costly treatment. To be internally consistent they should also favour the mild ill-health event over the same-priced costly treatment when that choice is presented to them in question 3.

The inclusion of a third question in the design allows us to consider the extent to which sequencing anomalies are triggered solely by the values of attributes in the directly preceding choice task. Alternatively, answers to question 3 might also show evidence of being influenced by the attributes of the options in the first question. However, we postpone consideration of such effects until we have presented the results of our study.

4. Implementation and Results

The experiment described in the previous section formed part of a larger survey instrument designed to examine the value placed on various aspects of quality life with respect to personal health. After a general introduction and some tasks that familiarised respondents with the various ill-health events that formed the subject matter of the survey, the CE was the first exercise faced by each respondent. Surveying was undertaken by professional interviewers in personal interview

sessions. Each sample was chosen so as to provide a representative sample of the population of Northern Spain that formed the location of interest for our investigations. Each sample consisted of 83 observations except for sample B_2 for which we obtained 85 observations. The data are summarised in Figure 6 which shows the proportions of each sample choosing each option in each choice task.

1. Consistency Checks

Our first set of tests concern the consistency of responses to first questions. Each of the six samples faces the first choice question having received identical information. Provided there are no systematic differences in the characteristics of each sample (as should have been achieved by our sampling design) then economic theory indicates certain relationships should exist in acceptance rates.

First, choice tasks A_1-1 , B_2-1 and C_2-1 are identical such that we expect to see the same proportion choosing Option 2 (or for that matter Option 1) in all three samples. We observe proportions of 74.7% (A_1-1), 74.1% (B_2-1), and 71.1% (C_2-1). A series of pairwise comparisons using a two-tailed z -test of differences in proportions confirms that there are no statistically significant differences in these acceptance rates (p -values; A_1-1 vs. B_2-1 : 0.9312, A_1-1 vs. C_2-1 : 0.6004, B_2-1 vs. C_2-1 : 0.6593). It is perfectly reasonable, therefore, to combine these three observations and treat them as observations pertaining to the same set of underlying preferences. The proportion choosing Option 2 in the combined sample of 187 individuals is some 74.5%.

In addition, standard economic theory indicates that individuals prefer more of good things to less. Accordingly, we would expect more individuals to choose a treatment if the health benefit offered by that treatment increased or its price fell. Indeed, this is what we observe in the data from the first question. As we have seen, when the treatment offering a return to full health has the medium price, the proportion choosing Option 1 in which they suffer a severe ill-health event is 25.5% (A_1-1 , B_2-1 and C_2-1). When the full health treatment has the same cost but Option 1 now only requires enduring a mild ill-health event we observe the proportion choosing Option 1 rise to 37.3% (B_2-1). A one-tailed z -test of differences in proportions confirms this to be a statistically significant difference (p -value: 0.0190).

Figure 6: Observed acceptance rates for options in each task

<u>Sample</u>	<u>Question 1</u>		<u>Question 2</u>		<u>Question 3</u>	
A_1	Severe	25.3	Severe	42.2	Mild	49.4
	€Med	74.7	€High	57.8	€Med	50.6
A_2	Severe	32.5	Severe	24.1	Mild	36.1
	€High	67.5	€Med	75.9	€Low	63.9
B_1	Mild	37.3	Severe	8.4	Mild	24.1
	€Med	62.7	€Med	91.6	€Low	75.9
B_2	Severe	25.9	Mild	64.7	Severe	38.8
	€Med	74.1	€Med	35.3	€High	61.2
C_1	Mild	22.9	Severe	32.5	Mild	49.4
	€Low	77.1	€High	67.5	€Med	50.6
C_2	Severe	28.9	Mild	41.0	Severe	59.0
	€Med	71.1	€Low	59.0	€High	41.0

Notes:

Top box represents Option 1, bottom box Option 2.

Acceptance rates are printed in the box to the right of each Option

Severe = 2S, Mild = 2Q

High = €240, Medium = €120, Low = €60

In a similar vein, economic theory indicates that, all else equal, people prefer to spend less rather than more in purchasing a good. Again, this is what we see in the data. Some 74.5% of respondents choose to pay the medium price in order to return to full health rather than endure a severe ill-health event (from A_1-1 , B_2-1 and C_2-1) but this falls to 67.5% when the price of Option 1 is increased to the high price (A_2-1).

Likewise, 77.1% of respondents are prepared to pay the low price to attain full health rather than suffer the mild ill-health event (C_2-1). This proportion falls to just to 62.7% when the cost of the full-health treatment increases to the medium price (B_1-1). Using a one-tailed z -test, the first of these comparisons is not statistically significant (p -value: 0.1061), the second is highly significant (p -value: 0.0212).

Finally, we check to see whether individuals answer the first and second questions consistently. For this to be true, it must be the case that if an individual prefers one option to another in the first question and either that option gets unequivocally better in the second question whilst the alternative option remains the same or that option remains the same whilst the alternative option gets unequivocally worse, then that individual should express a preference for the same option in the second question. In this case, we have four possible tests. Those that choose Option 1 in A_1-1 should also choose Option 1 in A_1-2 . Those that choose Option 2 in A_2-1 should also choose Option 2 in A_2-2 . Those that choose Option 2 in B_1-1 should choose that same option in B_1-2 . Finally, those that choose Option 1 in B_2-1 should also select Option 1 in B_2-2 . We find that all respondents that fall into these four cases express internally consistent preferences.

So far, the message for CE is quite hopeful. An across-sample test of responses indicates that our samples express a similar distribution of preferences and that these conform to the basic assumptions of economic theory; more of a commodity is better whilst higher prices are worse. In addition, in the first two questions subjects' responses are unerringly internally consistent.

2. Price Sequence Effects

Our tests for price sequencing are contained in a comparison of the responses of samples A_1 and A_2 . As catalogued in the first two rows of Table 2, our expectation is that the proportion choosing Option 2 in the second question of the worsening price sequence (A_1) will be depressed, whilst the proportion choosing Option 2 in the improving price sequence (A_2) will be unchanged. This is exactly what we observe in the data. Those choosing Option 2 (and thereby indicating they would pay for the costly treatment at the high price rather suffer a severe ill-health event) represent 67.5% of the sample that answer this as the first choice task (A_2-1) but only 57.8% of the sample that answer this as a second question in a worsening price sequence (A_1-2).

A one-tailed test reveals this difference to be significant at the 90% level of confidence (p -value: 0.0996). In contrast, the proportions choosing Option 2 (and thereby indicating they would pay for the costly treatment at the medium price rather suffer a severe ill-health event) do not differ significantly (p -value from a two-tailed test: 0.7896) when that choice is presented as the second question in an improving price sequence (75.9% in A_2 -2) as compared to when it is presented as the first question (74.5% in A_1 -1, B_2 -1 and C_2 -1).

Our data provide support De Shazo's framing hypothesis that worsening price sequences exert significant downward pressure on acceptance rates, whilst improving price sequences have no systematic impact on acceptance rates.

Table 2: Tests of Sequencing Anomalies

Sequence	Test Case	Comparison Case	Predicted Sequence Anomaly	Signed p -value
Price Sequence:				
Worsening	A_1 -2	A_2 -1	–	-0.0996 ^a
Improving	A_2 -2	A_1 -1, B_2 -1 and C_2 -1	None	+0.7896 ^b
Commodity Sequence:				
Worsening	B_1 -2	B_2 -1	–	-0.0002 ^a
Improving	B_2 -2	B_1 -1	+	+0.0005 ^a
Mixed Sequence:				
Worsening	C_1 -2	A_2 -1	+ or –	1.000 ^b
Improving	C_2 -2	C_1 -2	+	+0.0063 ^a

Notes:

^a From one-tailed test of equality of proportions

^b From two-tailed test of equality of proportions

3. Commodity Sequence Effects

The responses of samples B_1 and B_2 provide the basis for our tests of commodity sequencing. Again, our null hypothesis is that the proportion of independent samples choosing Option 1 for a particular choice task should be the same independent of sequence. This is not what we see in the data. Rather we observe that 25.5% choose to suffer the severe ill-health event rather than pay for the medium-

priced treatment when this choice is presented as a first question (A_{1-1} , B_{2-1} and C_{2-1}) but that this proportion collapses to only 8.4% in the worsening commodity sequence (B_{1-2}). A one-tailed test reveals this to be a highly significant difference (p -value: 0.0002).

In contrast, in sample B_1 the proportion choosing Option 1 (and thereby indicating that they would endure the mild ill-health event rather than pay for the costly treatment at the medium price) in the first question is only 37.7%. However, when this same choice task is presented as the second question in an improving commodity sequence, this proportion leaps to 64.7% a highly significance difference (p -value from a two-tailed test: 0.0005).

Accordingly, the pattern of responses observed in our data offers strong evidence of commodity sequencing anomalies. Moreover, in contrast to the asymmetry observed in price sequencing, these anomalies function in both improving and worsening directions.

4. Mixed Sequence Effects

The findings of our two core tests confirm that the sorts of sequencing anomalies observed in DC contingent valuation studies (Table 1) persist in CEs. Of course CEs typically vary more than just one attribute of one option from choice task to choice task. Samples C_1 and C_2 allow us to examine whether the insights provided by our price sequencing and commodity sequencing tests carry over to such cases.

Consider first the comparison between responses to questions C_{1-1} and C_{2-2} . Both present a choice between a costless treatment resulting in a mild ill-health event (Option 1) and a low priced treatment that removes all symptoms (Option 2). While C_{1-1} is the first choice task addressed by the C_1 sample, C_{2-2} is the second choice task such that Option 1 shows an improving commodity sequence and Option 2 an improving price sequence. Given our earlier findings, our expectation is that the improving commodity sequence will tend to increase preferences for Option 1 but that the improving price sequence will have no impact on preferences for Option 2. Again this is what we observe; 22.9% prefer Option 1 in C_{1-1} but this increases to 41% in C_{2-2} . A one-tailed test reveals this to be a significant difference (p -value: 0.0063).

The worsening mixed price and commodity sequences are tested by comparing responses to A_{2-1} to C_{1-2} . Here we have no directional hypothesis since both options

should be regarded less favourably in the worsening sequences. The data provides support for this contention as the proportion favouring Option 1 is identically 32.5% in both tasks. Clearly, there are no statistically significant differences in these observations.

5. Responses to the Third Question

Our data tells a very coherent story with regards to the impacts of price and commodity sequencing on responses to the second question. In particular, they corroborate the findings of iterative DC contingent valuation studies following the patterns of behaviour predicted in Table 1. Of course, most CE studies confront each respondent with a whole series of choice tasks. As such, we are interested to see whether the response anomalies that we have observed as a result of the sequence of first and second questions (what we shall refer to as the *initial sequence*) are also precipitated by the sequences implied by the second and third questions (what we shall refer to as the *secondary sequence*).

First, however, we check to see whether subjects' responses to the third questions remain internally consistent with their responses to the first and second questions. Observe from Figure 6 that relative to the second question, the third questions always represent mixed sequences with both options either improving or worsening. In this case, we cannot say a priori how a respondent should choose in the third question knowing their response to the second question. In contrast, a comparison of the first and third choice tasks reveals a number of cases where we can test for internal consistency. For example, in the first question sample A_1 is offered the choice between a free treatment that requires them to suffer a severe ill-health and a treatment that returns them immediately to full health but at the medium cost. Twenty-one individuals (25.3% of the sample) opt for the severe ill-health event. In the third question they are faced by a choice task which pits a mild ill-health event against the same priced treatment offering a return to full health. Clearly, to be internally consistent those 21 individuals should opt for the mild ill-health. In fact, 20 of those individuals make the internally consistent choice but one respondent chooses counter to their originally professed preferences.

Four other such comparisons are possible. The 52 individuals opting for a medium cost treatment in B_1-1 should also choose that treatment in B_1-3 . In fact, we

observe one individual that makes the inconsistent choice. Likewise, 22 individuals opt for the severe ill-health event in B_2-1 as opposed to the medium-priced treatment. Clearly, they should also choose the severe ill-health event over the high-priced treatment in B_2-3 and the majority do with just one individual making the inconsistent choice. We observe much the same in the C_1 sample. Here, the 19 individuals choosing Option 1 in C_1-1 should also choose Option 1 in C_1-3 though one individual fails to make the consistent choice. Finally, all 24 of the individuals choosing Option 1 in C_2-1 are observed to make the consistent choice and choose Option 1 in C_2-3 .

In the main, therefore, individuals achieve internal consistency. In comparison to the complete internal consistency exhibited in responses to the adjacent questions 1 and 2, however, our data suggests that respondents are minorly less capable of making consistent choices when those choices are separated by interceding tasks.

Turning to the question of sequencing effects in the third question, consider responses to A_1-3 . The secondary sequence of which this choice task is a part represents an improving mixed sequence. Our expectation is that such a sequence will increase the proportion of respondents favouring Option 1 (as indicated in the third column of Table 3). Indeed, if we compare responses from A_1-3 to those given by sample B_1 when presented with same choice task as a first question, we see exactly that result. The proportion favouring Option 1 increases from 37.7% to 49.4%; a difference that a one-tailed test proves to be significant (p -value 0.0587). Importantly, notice that the initial sequence answered by the A_1 sample was itself in a worsening price sequence. As a result, answers to the preceding question (A_1-2) were also biased in favour of Option 1 (as indicated in the second column of Table 3). In this case both the initial sequence and the secondary sequence serve to make Option 1 appear relatively more favourable; though we are not in a position to determine whether both or just one of those sequences is responsible for the observed anomaly.

More light is shed on this question by the responses to A_2-3 . As before, the secondary sequence faced by this sample is an improving mixed sequence that we suspect will bias responses to question A_2-3 in favour of Option 1. In this case, however, the initial sequence is an improving price sequence that we have already determined results in no sequencing bias. Accordingly, any anomaly observed in the responses to question A_2-3 cannot be driven by biases induced by the initial sequence. Question C_1-1 provides the base case with which to compare. We find that the

expected bias in favour of Option 1 is clearly observable in the data with the proportion choosing this option increasing from 22.9% to 49.4%. A one-tailed test shows this to be a highly significant difference (p -value 0.0306).

Table 3: Expected cumulative sequencing bias in responses to Option 1

Sample	Bias induced in Responses to Option 1			Signed p -value
	Initial Sequence	Secondary Sequence	Cumulative	
A_1	+	+	+	+0.0587 ^b
A_2	None	+	+	+0.0306 ^b
B_1	-	+	+ or -	+0.8547 ^c
B_2	+	+ or -	+ or -	+0.3947 ^c
C_1	(+ or -) ^a	+	+	+0.0854 ^c
C_2	+	(+ or -) ^a	+	+0.0006 ^c

Notes:

^a we observe empirically that the indeterminate impact of this sequence does not, in fact, result in a bias in responses.

^b From one-tailed test of equality of proportions

^c From two-tailed test of equality of proportions

Sample B_1 face the same sort of improving mixed sequence in questions 2 and 3 as we have considered with regards to samples A_1 and A_2 . As such, we once again expect the secondary sequence to exert an upward bias on the proportion favouring Option 1. Recall, that the A_1 sample faced an initial sequence that served to reinforce the bias in favour of Option 1, while sample A_2 faced an initial sequence that induced no sequencing bias. Sample B_1 provides the final possible case in which the initial sequence serves to bias respondents against Option 1. Accordingly, in this case, the bias induced by the initial sequence works in opposition to that induced by the second. Indeed, what we observe is that these two biases appear to cancel each other out. The appropriate comparison choice task is C_1 -1 where 22.9% favour Option 1 which is little different from the 24.1% selecting that option in B_1 -3. It is no surprise that a two-tailed test confirms that these two proportions do not differ significantly (p -value: 0.8547).

It is rather more difficult to predict how sequencing biases might impact on responses to question B_2 -3. Here the secondary sequence is a worsening mixed

sequence; the kind of sequence in which both options look relatively less favourable such that we have no directional hypothesis concerning which affect will dominate. As a matter of interest, the initial pair of questions in this series present an improving commodity sequence that tends to bias up the proportion favouring Option 1. The comparison choice task in this case is A_2-1 where we observe 32.5% selecting Option 1. In choice task B_2-3 the proportion favouring that option is somewhat higher at 38.8% but using a two-tailed test this is not a significant difference (p -value: 0.3947).

Sample C_1 face an improving mixed sequence in questions 2 and 3 that we expect to bias up the proportion favouring Option 1 in C_1-3 . The initial pair of questions in this series present a worsening mixed sequence for which we have no directional hypothesis. As a matter of fact, we have already established that in this particular case the initial sequence does not bias responses to C_1-2 . All the same, it is not appropriate to use that information in testing for bias in responses to C_1-3 . As a result, we use a two-tailed test to compare the proportion favouring Option 1 in C_1-3 (49.4%) to the appropriate comparison task in B_1-1 (37.7%). The test reveals a significant upward bias in preferences for Option 1 (p -value: 0.0854). As recorded in Table 3, this result complies with our expectations given our empirical observation that, the initial sequence does not induce bias (or perhaps, more correctly produces biases that cancel each other out).

The opposite case is observed in the series of questions asked of sample C_2 . Here C_2-3 is in a worsening mixed sequence whilst the initial pair of questions form an improving mixed sequence. Accordingly, we expect (indeed, have already observed) that the proportion favouring Option 1 will be biased up in C_2-2 . Whilst we have no prior expectations as to how the worsening mixed sequence in the second and third questions will bias responses, observe that this is the exact same sequence of questions as presented in C_1-1 and C_1-2 . We have already established that that particular worsening mixed sequence does not result in any bias (or again, more correctly, biases that cancel each other out). Given this information, our expectation is that the upward bias in favour of Option 1 induced by the initial sequence will carry through to responses to C_2-3 . This is exactly what we observe. Comparing responses to C_2-3 to the appropriate comparison task in A_2-1 using a two-tailed test, we find a very significant bias in favour of Option 1 (59.0% as opposed to 32.5% with p -value: 0.0006).

Responses to the third questions reveal some extremely interesting patterns. It appears that the sequencing anomalies observed in our data are cumulative. Consider those cases where the two questions in a series of choice tasks do not precipitate sequencing anomalies. In this case, we observe that the biases induced by the second two questions in that series concur exactly with those we have observed for initial sequences (A_{2-3} and C_{1-3}). In other words, subjects appear to be biased in the same manner whether they observe a particular type of sequence in the initial or secondary sequence of questions. Alternatively, consider the situation in which the initial sequence precipitates a bias that tends to favour a particular option but our expectation is that the secondary sequence does not induce a bias. In that case, we observe that the bias from the initial sequence is carried over to responses to the third question (C_{2-3}). To support this assertion we find that when the bias induced by the initial and secondary sequences both favour one option, then that option is significantly more likely to be chosen in the third question (A_{1-3}). Likewise when the biases induced by the initial and secondary sequences work so as to favour opposing options, then these opposing biases appear to cancel each other out (B_{1-3}).

5. Summary and Concluding Remarks

Despite the rapid uptake in CE over recent years, the technique has not been subjected to the same rigorous testing as other non-market valuation methods. For example, iterative dichotomous choice contingent valuation (DC-CV) formats have been shown to suffer from a robust sequencing anomaly; respondents are significantly less willing to pay a particular price for a non-market good in the second, as opposed to first, DC question. Importantly, there are enough similarities between DC-CV and CE formats to suspect that the sequencing anomalies that have widely discredited the former may also afflict responses to the latter. This paper reports on the findings of a study specifically designed to assess that contention.

Our particular application concerns the valuation of health using a CE with three choice tasks. In each task respondents were asked to choose between differently priced treatments offering different levels of health benefit. Our experimental design differs from that of previous studies (e.g. Holmes and Boyle, 2005) in that it was devised specifically for the purpose of examining sequencing anomalies rather than recovering the parameters of a utility function. In particular, we use a six-way split sample in which the pairwise comparison of responses across samples provides a

straightforward test of various hypotheses regarding the existence and nature of sequencing anomalies in CEs. Our first sample responds to questions presenting a worsening price sequence (i.e. the second choice task differs from the first only in that the price of one of the treatments is relatively larger). We compare those responses to those from the second sample that face an improving price sequence. We repeat this basic experimental set up with our third and fourth samples, though here we present worsening and improving commodity sequences (i.e. the second choice task differs from the first only in offering a smaller (greater) health benefit). Finally, our fifth and sixth samples face choice tasks that follow a sequence in which simultaneous changes occur in the attributes of both options as might be more usual in a standard CE application.

Our findings categorically reject the assumption that CEs are immune from sequencing anomalies. In particular, our results indicate that worsening price sequences significantly reduce choice probabilities though the opposite anomaly is not observed in improving price sequences (a pattern observed in DC contingent valuation data by De Shazo, 2003). In addition, we find highly significant anomalies resulting from health-state sequences. In this case, the effects work in both directions with significant reductions (increases) in choice probabilities observed in worsening (improving) health-state sequences. The patterns of choice suggested by these behaviours completely organize our data, explaining the anomalous choice probabilities observed in the fifth and sixth samples as well as responses to the third choice task in all samples. In addition, we find robust evidence to suggest that these sequencing anomalies are cumulative. That is to say, a sequencing anomaly induced by the first pair of choice tasks; remains observable in the data if the second pair of choice tasks does not induce sequencing bias, is mitigated if the second pair of choice tasks induces the opposite sequencing bias, or exacerbated if the second pair of choice tasks induces the same sequencing bias.

The central message of our paper is that CEs are vulnerable to sequencing anomalies. As such, our findings cast serious doubt on the current practice of asking each respondent to undertake several choice tasks in a CE whilst treating each response as an independent observation on that individual's preferences. Indeed, pending further testing, we conclude that the outputs from such analyses be treated with considerable scepticism.

References

- Adamowicz, W. L., (2004). "What's it worth? An examination of historical trends and future directions in environmental valuation", *Australian Journal of Agricultural and Resource Economics*, 48(3), pp. 419–443.
- Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner and H. Schuman (1993). "Report of the NOAA panel on contingent valuation", *Federal Register*, 58, pp.4601-4614.
- Bateman, I. J. and R. Brouwer (2006). "Consistency and construction in stated WTP for health risk reductions: A novel scope-sensitivity test", *Resource and Energy Economics*, 28, pp. 199-214.
- Bishop, R. C., and T. A. Heberlein. (1979). "Measuring values of extramarket goods: Are indirect measures biased?", *American Journal of Agricultural Economics*, 61(5), pp. 926-930.
- Cameron, T.A., and J. Quiggin (1994). "Estimation using contingent valuation data from a dichotomous choice with follow-up questionnaire", *Journal of Environmental Economics and Management*, 27(3), pp. 218-234.
- De Shazo, J.R., (2002). "Designing transactions without framing effects in iterative question formats", *Journal of Environmental Economics and Management*, 43, pp. 360-385.
- Hanemann, W. M., J. Loomis, and B. Kanninen (1991). "Statistical efficiency of double-bounded dichotomous choice contingent valuation", *American Journal of Agricultural Economics*, 73(4), pp. 1255-1263.
- Holmes, T. P., and K. J. Boyle (2005). "Dynamic learning and context-dependence in sequential, attribute-based, stated preference valuation questions", *Land Economics*, 81(1), pp. 114-126.
- Herriges, J., and J. Shogren (1996). "Starting point bias in dichotomous choice valuation with follow-up questioning", *Journal of Environmental Economics and Management*, 30, pp.112-131.
- Langford, I. H., I. J. Bateman and H. D. Langford (1996). "A multilevel modelling approach to triple-bounded dichotomous choice contingent valuation", *Environmental and Resource Economics*, 7(3), pp.197-211.

McFadden, D., (1994). "Contingent valuation and social choice", *American Journal of Agricultural Economics*, 76, pp. 689-708.

Scarpa, R., and I. J. Bateman (2000). "Efficiency gains afforded by improved bid design versus follow-up valuation questions in discrete choice CV studies", *Land Economics*, 76(2), pp. 299-311.