Sequencing Anomalies in Choice Experiments

Brett Day[1] and Jose-Luis Pinto[2]

[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 widely reported in 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, our data indicate 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. 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). 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.

[INSERT TABLE 1 AROUND HERE]

Now, 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 (which we label Option I and Option II). To simplify further, imagine that the commodity dimension of the good can be described as either small, medium or large (where the increment to 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 this context, we can construct pairs of choice tasks that exactly replicate the sequences of contingent valuation questions of Table 1.

For example, the sequence of choice tasks shown in the upper part of Figure 1 replicates an improving price sequence in a DBDC contingent valuation question. Here Option I is the same in both tasks (the status quo in a DBDC contingent valuation question). Likewise, the commodity dimension of Option II is the same in both tasks (the new level of provision of the non-market good in a DBDC contingent valuation question). All that changes across the choice tasks is that the price dimension of Option II falls (an improving price sequence in a DBDC contingent

[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.

valuation question). Accordingly, we describe Option II in Choice Task 2 as appearing in an *improving price sequence*. Notice that in the context of a CE the sequence is attributed to the particular Option in which it is observed.

[INSERT FIGURE 1 AROUND HERE]

The lower part of Figure 1 shows a sequence of CE questions that replicates the worsening price sequence of a DBDC contingent valuation question. Again Option I is the same in both tasks. Option II also offers the same level of provision but at a low price in the first choice task followed by a high price in the second choice task. Again we attribute the sequence to the option such that we describe Option II in Choice Task 2 as appearing in a *worsening price sequence*.

If the series of choice tasks in Figure 1 exactly replicate improving and worsening price sequences in DBDC contingent valuation questions, then we might expect them to generate the same sequencing anomalies observed by De Shazo. We could simply test that contention by presenting independent samples with the two sequences of choice tasks.

Observe that the first choice task faced by one sample is identical to the second choice task faced by the other sample. Procedural invariance demands that the proportion of subjects favouring Option II 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 II when presented in the second choice task in a 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 Option 1 of the choice tasks following improving and worsening commodity sequences. We say that in the upper part of the diagram of Figure 2 the commodity sequence is improving because Option II does not change and in Option I the price is the same but the amount of good increases. This is then an improving commodity sequence in Option I. The opposite happens in the lower part of Figure II. Again, the first choice task in the upper sequence is identical to the second choice task in the lower sequence. As such, a simple test for procedural invariance would be to present the two sequences of choice tasks to independent samples and compare response proportions for the matched tasks. 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 I when it is presented in an improving commodity sequence in the second choice task will be significantly greater than that observed when the identical choice task is offered to the other sample as the first task. The opposite is expected for the worsening commodity sequence. Here the proportion choosing Option I when presented in a worsening commodity sequence in the second choice task is expected to be significantly less than that observed when the identical choice task is offered to the other sample as the first task.

[INSERT FIGURE 2 AROUND HERE]

While 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 options in 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.

[INSERT FIGURE 3 AROUND HERE]

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 I 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 II in the second question results in no such equivalent bias. This combination of effects would lead to a relatively greater proportion of respondents favouring Option I in this choice task than would do if that same choice task was the first in the 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 effects 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.

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, they find that subsequent prices somehow impact on current choices.

In addition, Holmes and Boyle (2005) test for commodity sequences, However, 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.

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.

Quality of life was measured using the Euroqol (EQ-5D) (Brooks, 1996). EQ-5D is a standardised instrument for use as a measure of health outcome. It describes any health state through five dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) each of which can take one of three responses. The responses record three levels of severity (no problems/some or moderate problems/extreme problems) within a particular EQ-5D dimension. For example, health state 11111 implies full health since it is equivalent to having no problems in each of the five dimensions. However, health state 33333 implies having extreme problems in each of the five dimensions. The advantage of using EQ-5D is that there are clear dominance relations between health states. In this study we used two EQ-5D health states, namely, 21212 and 22223. Cleary, an improvement from 21212 to 11111 (full health) implied a smaller health gain than an improvement from 22223 to full health. A summary of the health state descriptions is provided in Figure 4.

[INSERT FIGURE 4 AROUND HERE]

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. While 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. While 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, health state 22223) or two months of mild ill-health (mild symptoms, health state 21212) 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 Figure 5, 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.

[INSERT FIGURE 5 AROUND HERE]

In order to simplify the presentation of our experimental design, it is expedient to further condense each choice task into the simple schematic shown on the right hand side of Figure 5. Here the top box represents Option I where the cost, treatment duration and symptom's duration dimensions have all been suppressed since these were always €0, 12 months and 2 months respectively. Likewise the bottom box represents Option II 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 6. We label the six samples, A_1, A_2, B_1, B_2, C_1 and C_2.

The central tests employed in our analysis are those described in relation to Figures 1, 2 and 3. The comparison of response proportions for the first two choice tasks of samples A_1 and A_2 provide tests of price sequencing (as per Figure 1). The comparison of response proportions for the first two choice tasks of samples B_1 and B_2 provide tests of commodity sequencing (as per Figure 2). Whilst the comparison of response proportions for the first two choice tasks of samples C_1 and C_2 provide tests of mixed price and commodity sequencing (as per Figure 3). Observe that the commodity on offer in our experiment is a bad. As such, an improving commodity sequence is one that moves from a worse health state (severe) to a better health state (mild).

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 proportions choosing each option 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 effects (should they exist) without imposing assumptions regarding the structure of preferences. Our nonparametric approach is both simple and transparent and as such, we believe, increases the credibility of our findings.

In addition to the sequencing tests our design allows us to test for respondent consistency. For example, consider an individual in sample A_1 answering the first choice task. 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 choice task.

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 choice task 3.

Finally, 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 choice task 3 might also show evidence of being influenced by the attributes of the options in the first choice task.

[INSERT FIGURE 6 AROUND HERE]

4. Results

The CE 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. Across-Sample Consistency Checks

Each of the six samples faces the first choice task having received identical information such that responses to these tasks are unaffected by potential sequencing biases. Accordingly, an initial set of tests seeks to confirm whether the responses to the first choice tasks are consistent with each sample being an independent observation of the same underlying population of preferences.

For ease of exposition, we refer to the first question faced by sample A_1 as A_{11}, likewise the second question faced by this sample is A_{12}, and so on. Using this notation, observe that choice tasks A_{11}, B_{21} and C_{21} are identical such that we expect to see the same response proportions in all three samples. What we observe is 25.3% (A_{11}), 25.9% (B_{21}), and 28.9% (C_{21}) choosing Option I. As shown in Tests 1 to 3 of Table 2, 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. 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 I in the combined sample of 187 individuals is some 25.5%.

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. 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 I in which they suffer a severe ill-health event is 25.5% (A_{11}, B_{21} and C_{21}). When the full health treatment has the same cost but Option 1 now only requires enduring a mild ill-health event we observe that the proportion choosing Option I rises to 37.3% (B_{11}). As shown in Test 4 of Table 2, a one-tailed *z*-test of differences in proportions confirms this to be a statistically significant difference (*p*-value: 0.0190).

[INSERT TABLE 2 AROUND HERE]

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_{11}, B_{21} and C_{21}) but this falls to 67.5% when the price of Option II is increased to the high price (A_{21}). 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_{11}). This proportion falls to just to 62.7% when the cost of the full-health treatment increases to the medium price (B_{11}). 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).

2. Within-Person Consistency Checks

A further set of tests allow us to examine whether individuals respond to the CE tasks according to a consistent set of preferences.

First, we compare respondents' answers to the first and second choice tasks. For responses to be internally consistent, it must be the case that if an individual prefers one option to another in the first choice task and either that option gets unequivocally better in the second choice task 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 choice task. In this case, we have four possible tests.

For example, the 21 indivdiuals that opt for the severe ill-health event (Option I) in A_{11} as opposed to the medium-priced treatment (Option II) should also choose the severe ill-health event (Option I) when the alternative treatment is offered at the high-price (Option II) in A_{12}. As shown in Test 7 of table 3, we find that all those individuals make the consistent choice. Similarly, in A_{21} we observe 56 individuals choosing to pay for the high cost treatment (Option II) rather than endure the severe ill-health event (Option I). Clearly, in A_{22} those individuals should also choose that treatment in preference to the severe ill-health (Option I) when it is available to them at the medium price (Option II). Test 8 of Table 3 shows that 100% of those respondents are consistent in their choices. Indeed, the design allows two further comparisons of responses to first and second tasks and, as shown in Tests 9 and 10 of Table 3, in both cases we find responses to be internally consistent.

Next, we check that subjects' responses to the third choice task remain internally consistent with their responses to the first choice tasks. (Our design does not allow us to compare consistency between second and third choice tasks since these always form a mixed sequence; that is, both options change). As documented in Tests 11 to 15 of Table 3, we can make five possible comparisons. In those five comparisons we find at most one individual making the inconsistent choice.

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, respondents achieve a high level of internal consistency in their responses.

[INSERT TABLE 3 AROUND HERE]

3. 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 II in the second question of the worsening price sequence (A_1) will be depressed, whilst the proportion choosing Option II in the improving price sequence (A_2) will be unchanged. This is exactly what we observe in the data. Those choosing Option II (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_{21}) but only 57.8% of the sample that answer this as a second question in a worsening price sequence (A_{12}). As shown in Test 16 of Table 4, 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 II (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_{22}) as compared to when it is presented as the first question (74.5% in A_{11}, B_{21} and C_{21}).

As such our CE data exhibit exactly the same pattern of sequencing effects as observed by De Shazo (2002) in DBDC contingent valuation data. We find a significant reduction in the frequency with which respondents choose an option if that option is in a worsening price sequence. In contrast, no systematic impact on choice probabilities are observed for options in improving price sequences.

[INSERT TABLE 4 AROUND HERE]

4. 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 response proportions in independent samples should not be systematically distorted by the sequence of choice tasks. As recorded in Tests 18 and 19 of Table 2, this is not what we see in the data. Rather we observe that 25.5% choose to suffer the severe ill-health event (Option I) rather than pay for the medium-priced treatment (Option II) when this choice is presented in the first choice task (A_{11}, B_{21} and C_{21}) but that this proportion collapses to only 8.4% when Option I is in a worsening commodity sequence (B_{12}). 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 I (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 choice task is only 37.7%. However, when this same choice task is presented second with Option I in an improving commodity sequence, we observe the proportion choosing Option I leap up 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.

5. 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_{11} and C_{22}. 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_{11} is the first choice task addressed by the C_1 sample, C_{22} is the second choice task such that Option 1 is in an improving commodity sequence (the previous task offering a severe ill-health event) and Option 2 is in an improving price sequence (the previous task offering a medium price). Given our earlier findings, our expectation is that respondents will be much better disposed to Option 1 because it is an the improving commodity sequence but that the improving price sequence will have no impact on preferences for Option 2. As shown in Test 20 of Table 2, this is what we observe; 22.9% prefer Option 1 in C_{11} but this increases to 41% in C_{22}. 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_{21} to C_{12}. 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 in both tasks the proportion favouring Option 1 is identically 32.5%. As shown in Test 21 of Table 4, there are no statistically significant differences in these observations.

6. Cumulative Sequence Effects

Our data tells a very coherent story with regards to the impacts of price and commodity sequencing on responses to the second choice task. 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 tasks (what we shall we refer to as the *initial sequence*) are also precipitated by the sequences implied by the second and third tasks (what we shall refer to as the *secondary sequence*).

Consider responses to A_{13}, the secondary sequence of which 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 5). Indeed, if we compare responses from A_{13} 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_{12}) were also biased in favour of Option 1 (as indicated in the second column of Table 5). 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_{23}. As before, the secondary sequence faced by this sample is an improving mixed sequence that we suspect will bias responses to question A_{23} 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_{23} cannot be driven by biases induced by the initial sequence. Question C_{11} 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 36.1%. A one-tailed test shows this to be a highly significant difference (p-value 0.0306).

[INSERT TABLE 5 AROUND HERE]

Sample B_1 face the same sort of improving mixed sequence in choice tasks 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_{11} where 22.9% favour Option 1 which is little different from the 24.1% selecting that option in B_{13}. 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_{23}. 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 choice tasks 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_{21} where we observe 32.5% selecting Option 1. In choice task B_{23} 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 choice tasks 2 and 3 that we expect to bias up the proportion favouring Option 1 in C_{13}. The initial pair of choice tasks 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_{12}. All the same, it is not appropriate to use that information in testing for bias in responses to C_{13}. As a result, we use a two-tailed test to compare the proportion favouring Option 1 in C_{13} (49.4%) to the appropriate comparison task in B_{11} (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 is consistent 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 choice tasks asked of sample C_2. Here C_{23} 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_{22}. Whilst we have no prior expectations as to how the worsening mixed sequence in the second and third choice tasks will bias responses, observe that this is the exact same sequence of tasks as presented in C_{11} and C_{12}. 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_{23}. This is exactly what we observe. Comparing responses to C_{23} to the appropriate comparison task in A_{21} 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_{23} and C_{13}). 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_{23}). 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_{13}). 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_{13})

5. Summary and Concluding Remarks

Despite the rapid uptake in CE over recent years, there are reasons to suspect that the method may suffer from the same sorts of sequencing anomaly that have been observed in iterative DC contingent valuation studies. 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 findings categorically reject the assumption that CEs are immune from sequencing anomalies. In particular, our results indicate that the frequency with which respondents choose options in worsening price sequences is significantly reduced though the opposite anomaly is not observed for options in improving price sequences (a pattern observed in DC contingent valuation data by De Shazo, 2002). In addition, we find highly significant anomalies resulting from commodity sequences. In this case, the effects work in both directions with significant reductions (increases) in choice probabilities observed for options in worsening (improving) commodity sequences. An open question raised by our research is why significant sequencing effects are observed in improving commodity sequences but not improving price sequences.

The patterns of choice suggested by these sequencing behaviours completely organize our data, explaining the choice probabilities observed to choice tasks in mixed sequences; that is, when one option of a choice task is presented in an improving (worsening) price sequence, while the other is presented in an improving (worsening) commodity sequence. 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.

One special feature of our data is that the two options presented in the choice experiment are manifestations of identifiable products; one option describes the outcome of healthcare provided free of charge by the public sector, the other the outcome of healthcare provided at a cost by the private sector. In this case, a price sequence is clearly identifiable as a situation in which the price of the private sector product changes from one choice task to the next. Likewise, a commodity sequence occurs when the health benefit offered by the public sector product changes from one choice task to the next. Of course, in many choice experiments, the options offered to respondents are generic manifestations of products defined only by the levels of the various product characteristics. In that case, options do not have specific identities, and respondents may have less reason to directly compare the levels of attributes offered by particular options from one choice task to the next. If that is so, and again this is an open research question, CEs using generic option descriptions may be less susceptible to the sorts of sequencing anomalies observed in our data.

All the same, 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.

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.

Brooks, R., (1996). "EuroQol: The current state of play", *Health Policy*, 37(1), pp. 53-72.

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.

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

		Sequence	
		Worsening	**Improving**
Dimension	**Price**	–	None
	Commodity	–	+

Table 2: Tests of across-sample consistency in responses to Choice Task 1

Test	Test Case	Comparison Case	Expected Difference in Response Proportions	Signed *p*-value
Response proportions the same for identical choice tasks:				
Test 1:	A_{11}	B_{21}	None	+0.9312[b]
Test 2:	A_{11}	C_{21}	None	+0.6004[b]
Test 3:	B_{21}	C_{21}	None	+0.6593[b]
More is better:				
Test 4:	A_{11}, B_{21} & C_{21}	B_{11}	+ Option I	+0.0190[a]
Higher prices are worse:				
Test 5:	A_{11}, B_{21} & C_{21}	A_{21}	– Option II	-0.1061[a]
Test 6:	C_{11}	B_{11}	– Option II	-0.0212[a]

Notes:

[a] From one-tailed test of equality of proportions

[b] From two-tailed test of equality of proportions

Table 3: Tests of within-person consistency

Test	Initial Choice	Consistent Subsequent Choice	% Consistent
Consistency between 1st and 2nd Choice Tasks:			
Test 7:	A_{11}: Option I	A_{12}: Option I	100% (21 out of 21)
Test 8:	A_{21}: Option II	A_{22}: Option II	100% (56 out of 56)
Test 9:	B_{11}: Option II	B_{12}: Option II	100% (52 out of 52)
Test 10:	B_{21}: Option I	B_{22}: Option I	100% (22 out of 22)
Consistency between 1st and 3rd Choice Tasks:			
Test 11:	A_{11}: Option I	A_{13}: Option I	95% (20 out of 21)
Test 12:	B_{11}: Option II	B_{13}: Option II	96% (51 out of 52)
Test 13:	B_{21}: Option II	B_{23}: Option II	95% (21 out of 22)
Test 14:	C_{11}: Option I	A_{13}: Option I	95% (18 out of 19)
Test 15:	C_{21}: Option I	A_{23}: Option I	100% (24 out of 24)

Table 4: Tests of Sequencing Anomalies

Sequence	Test Case	Comparison Case	Expected Difference in Response Proportions	Signed *p*-value
Price Sequence:				
Test 16: Worsening	A_{12}	A_{21}	– Option II	-0.0996[a]
Test 17: Improving	A_{22}	A_{11}, B_{21} & C_{21}	None	+0.7896[b]
Commodity Sequence:				
Test 18: Worsening	B_{12}	B_{21}	– Option I	-0.0002[a]
Test 19: Improving	B_{22}	B_{11}	+ Option I	+0.0005[a]
Mixed Sequence:				
Test 20: Improving (price) Improving (commodity)	C_{22}	C_{11}	+ Option I None Option II	+0.0063[a]
Test 21: Worsening (price) Worsening (commodity)	C_{12}	A_{21}	- Option I - Option II	1.000[b]

Notes:

[a] From one-tailed test of equality of proportions

[b] From two-tailed test of equality of proportions

Table 5: 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

Figure 1: Examples of Price Sequencing in a Choice Experiment

Option II in an Improving Price Sequence:

Choice Task 1	
Option I	**Option II**
Small	Large
€0	**€Dear**

⇨

Choice Task 2	
Option I	**Option II**
Small	Large
€0	**€Cheap**

Option II in a Worsening Price Sequence:

Choice Task 1	
Option I	**Option II**
Small	Large
€0	**€Cheap**

⇨

Choice Task 2	
Option I	**Option II**
Small	Large
€0	**€Dear**

Figure 2: Examples of Commodity Sequencing in a Choice Experiment

Option I in an Improving Commodity Sequence:

Choice Task 1	
Option I	**Option II**
Small	Large
€0	€Dear

⇒

Choice Task 2	
Option I	**Option II**
Medium	Large
€0	€Dear

Option I in a Worsening Commodity Sequence:

Choice Task 1	
Option I	**Option II**
Medium	Large
€0	€Dear

⇒

Choice Task 2	
Option I	**Option II**
Small	Large
€0	€Dear

Figure 6: Experimental design and observed acceptance rates for options in each task

<u>**Sample**</u>	<u>**Choice Task 1**</u>			<u>**Choice Task 2**</u>			<u>**Choice Task 3**</u>		
	A_{11}			A_{12}			A_{13}		
A_1	I	Severe	25.3	I	Severe	42.2	I	Mild	49.4
	II	€Med	74.7	II	€High	57.8	II	€Med	50.6
	A_{21}			A_{22}			A_{23}		
A_2	I	Severe	32.5	I	Severe	24.1	I	Mild	36.1
	II	€High	67.5	II	€Med	75.9	II	€Low	63.9
	B_{11}			B_{12}			B_{13}		
B_1	I	Mild	37.3	I	Severe	8.4	I	Mild	24.1
	II	€Med	62.7	II	€Med	91.6	II	€Low	75.9
	B_{21}			B_{22}			B_{23}		
B_2	I	Severe	25.9	I	Mild	64.7	I	Severe	38.8
	II	€Med	74.1	II	€Med	35.3	II	€High	61.2
	C_{11}			C_{12}			C_{13}		
C_1	I	Mild	22.9	I	Severe	32.5	I	Mild	49.4
	II	€Low	77.1	II	€High	67.5	II	€Med	50.6
	C_{21}			C_{22}			C_{23}		
C_2	I	Severe	28.9	I	Mild	41.0	I	Severe	59.0
	II	€Med	71.1	II	€Low	59.0	II	€High	41.0

Notes:

Top box of each choice task represents Option I, bottom box Option II.
Acceptance rates are printed in the box to the right of each Option
Severe = 2 months 22223, Mild = 2 months 21212
High = €240, Medium = €120, Low = €60