Ordering effects in nested ‘top-down’ and ‘bottom-up’
contingent valuation designs

Powe, N.A.\textsuperscript{a} and Bateman, I.J.\textsuperscript{b}

\textsuperscript{a}Centre for Research in Environmental Appraisal and Management (CREAM), School of Architecture, Planning and Landscape, University of Newcastle upon Tyne, NR1 7RU.

\textsuperscript{b}Programme on Environmental Decision Making (PEDM), Centre for Social and Economic Research on the Global Environment (CSERGE), School of Environmental Sciences, University of East Anglia, NR4 7TJ

\textsuperscript{*} Corresponding author. Tel +44-191-2227570; Fax: +44-191-2228811. E-mail address: n.a.powe@ncl.ac.uk
Abstract

The effect of question order upon willingness to pay for nested goods is investigated using a split-sample contingent valuation study. In one ‘top-down’ design values are first elicited for protecting the ‘whole’ of an endangered area after which values for a ‘part’ sub-area are elicited. A second ‘bottom-up’ design reverses the order of valuation tasks with the ‘part’ scheme valued prior to the ‘whole’. Results show that ‘whole’ and ‘part’ valuation responses are logically ordered and determined by a consistent set of factors irrespective of question order. Variance to question order was observed in both the ‘part’ and ‘whole’ valuations. The correspondence of these results to standard and non-standard preference theories is explored and the difficulties and practical issues concerning valuing nested goods are highlighted.

Key Words: Contingent Valuation method, nested goods, question order effect, split-sample.
1. Introduction

The contingent valuation (CV) method typically employs survey techniques to estimate values for non-market goods (Mitchell and Carson, 1989; Bateman and Willis, 1999). In theory it is a highly flexible method applicable to a wide variety of goods and scenarios. However, this very flexibility invites application to complex problems intractable to other techniques. This in turn has highlighted a host of practical and theoretical problems reflected in voluminous research literature (Arrow et al., 1993; Hanemann, 1994; Diamond and Hausman, 1994; Bateman and Willis, 1999; Carson et al., 2001).

One of the principal foci of the CV debate concerns the simultaneous valuation of differing levels of nested goods, that is goods which either embrace or are embraced by other goods which are themselves objects of the same valuation exercise. Much of this debate has centred upon the sensitivity of welfare estimates to the scope of the goods considered\(^1\), where scope can be defined in terms of changes in quantity and/or quality (Carson and Mitchell, 1995). However, an equally important aspect of this issue concerns the impact of varying question order when valuing nested goods.

Commentators have noted a number of economic-theoretic expectations which may arise when the order of such questions is varied (Hoehn, 1991, Hoehn and Loomis,

\[^1\] Concern over scope sensitivity was formally recognized in a report by the National Oceanic and Atmospheric Administration (NOAA) (Arrow et al., 1993). Following its publication a heated empirical debate has permeated the environmental economics literature: while some studies have shown split-sample scope sensitivity (e.g. Carson and Mitchell, 1993a; Carson and Mitchell, 1993b; Carson, 1997; Hoevenagel, 1996; Smith and Osbourne, 1996), others have not (e.g. Boyle et al., 1994; Schkade and Payne, 1994), and still others show that it is possible to observe both split-sample scope sensitivity and insensitivity within the same study (e.g. Bateman et al., 2001a; Hammitt and Graham, 1999; Loomis et al. 1993; Schulze et al. 1998). This literature suggests that, although split-sample scope insensitivity can be a real problem, it is not inevitable.
1993, Randall and Hoehn, 1996; Carson et al., 1998) while others have highlighted potential psychological effects and biases (Kahneman and Knetsch, 1992). However, in contrast to the substantial number of studies investigating the scope sensitivity issue, the empirical literature examining the presence of ordering effects in CV valuations of nested goods is noticeably thin. This paper sets out to address this deficiency.

In the following section we discuss the design elements of any nested goods valuation study after which we briefly review the pertinent literature. Drawing upon these discussions we formulate an empirical design to assess the issue of ordering in nested goods valuations via a split sample CV study. Here, in one ‘top-down’ treatment, values for an embracing ‘whole’ good (W; the preservation of the entirety of an endangered area of high environmental value) are elicited prior to respondents being asked to value a nested ‘part’ good (P; a sub-area of the whole). Conversely, in the other ‘bottom-up’ treatment, the P good is valued prior to the W good. As discussed subsequently, analysis of the valuation responses obtained from this design reveals a significant ordering effect in most values elicited. Reflections upon these findings and their implications provide a conclusion to the paper.

2. Design parameters, expectations and evidence regarding nested good valuations

The valuation of nested goods raises a complex array of issues for which commentators have suggested a litany of terms such as perfect embedding, part-whole
bias, etc. Given this, an initial task is to clarify a set of terms through which the characteristics of any nested good valuation study can be defined. In this section we build upon the seminal work of Carson and Mitchell (1995) and subsequent extensions by Bateman et al. (2001a) to define a typology of nested good valuations from which our particular area of interest can be isolated and our empirical study defined. Drawing upon this previous work we can identify four characteristics through which nested good valuation studies may be defined:

(i) **Quantitative or categorical nesting**
Goods which differ in terms of a single attribute alone, say their physical size, are considered to be quantitatively nested, whereas those which differ along a number of scales are referred to as being categorically nested (Carson and Mitchell, 1995).

(ii) **Inclusive vs. Exclusive lists.**
Bateman et al. (2001a) identify two distinct types of list within which nested goods may be presented to CV participants. In an **inclusive list** goods are presented as additions to (or subtractions from) any good(s) presented previously in that list. Conversely, in an **exclusive list** goods are presented as **alternatives to** any other good(s) given in that list.

(iii) **List Direction**
Carson and Mitchell (1995) make a directional distinction between list presentations. A **bottom-up** list presents individuals with a succession of nested goods from the ‘smallest’ to the ‘largest’. A **top-down** list reverses the presentation of goods from that
detailed above. Such a presentation is logical for exclusive lists but, as Carson and Mitchell note, it becomes poorly defined for inclusive lists and can only be partially accomplished by using a value partitioning approach such as that adopted by Kahneman and Knetsch (1992) and Brown et al. (1995).

(iv) Visible choice set

Bateman et al. (2001a) define the visible choice set (VCS) as ‘that set of goods which, at any given point in a valuation exercise, the respondent perceives as being the full extent of the purchase options which will be made available in the course of the exercise’ (page 4). As discussed previously, in our empirical study the full set of purchase options are W and P. These can be introduced either at the start of the valuation exercise (referred to as ‘advance disclosure’), or only as each valuation task is undertaken (‘stepwise disclosure’).

Expectations regarding these various design parameters are complex and inter-related. The issue of whether goods are quantitatively or categorically nested is of key importance to the scope debate where substitution effects may explain apparent scope insensitivity within categorically but not quantitatively nested goods. However, for the purposes of our study this issue is of secondary importance as we are concerned here with relative changes in values according to the order of their presentation, not in the absolute magnitude of those values.

Turning to consider the issue of list type, within an inclusive list the value stated by a respondent for any given good is dependent upon their current endowment of private
and public goods. Therefore it is expected that the value for a particular good will depend upon its position in an inclusive list. Such *sequencing effects* are one of the earliest findings of empirical CV research (Randall, Hoehn and Tolley, 1981; Hoehn and Randall, 1982; Hoehn, 1983; Tolley *et al.*, 1983). Indeed it is this sequencing effect, which Kahneman and Knetsch (1992) term an embedding effect\(^2\). However, it has been theoretically demonstrated that such context dependence is to be expected in inclusive lists (e.g., Smith, 1992; Harrison, 1992; Carson and Mitchell, 1995; Carson *et al.*, 1998), as the holdings of public goods varies as one progresses through an inclusive list.

While inclusive lists have been subject to extensive research, less attention has been paid to exclusive lists. Here, the reference income, prices, level of private and public goods and utility level remain constant at the initial status quo across valuation questions. Assuming that exclusive valuation tasks are seen as independent (an assumption to which we return in conclusions to this paper), any residual variation in values associated with presentation is therefore unexpected and can be termed an *ordering effect*. Few studies have examined the possibility of ordering effects in exclusive list design valuations of nested goods. Two exceptions are the studies by Boyle *et al.* (1990) and Boyle *et al.* (1993). Here ordering is tested by allowing list direction to vary such that goods can appear in what equates to top-down or bottom–up lists. Both studies provide some evidence of ordering effects although in the latter study these effects were not significant for survey respondents who had relatively high experience of the resource in question.

\(^2\) See also Kemp and Maxwell (1993), Loomis *et al.* (1993) and Hoevenagel (1996).
A recent study by Bateman et al. (2001a) combines top-down and bottom-up lists with both advance and stepwise disclosure VCS. Findings suggested significant ordering effects in stepwise designs. Given that the majority of CV use stepwise approaches for revealing choice sets to survey respondents this is clearly an area which requires better understanding. However, the Bateman et al., study employed open-ended questions for eliciting willingness to pay responses. These have been the subject of some criticism on the grounds that they have weak incentive properties (Arrow et al., 1993; Carson et al., 1999). Consequently, a reinvestigation of this issue using an incentive compatible dichotomous choice (DC) elicitation technique seems justified and constitutes one of the drivers underpinning the present study.

3. Design and expectations

Translating the above requirements into valuation tasks concerning the W and P goods described previously, implies the split sample design illustrated in Table 1. Here a stepwise presentation exclusive list is used to elicit valuations of both goods from each respondent. Samples are defined according to the order in which goods are valued with the top-down order denoted by the italicised label W/P indicating that good W is valued prior to good P. Similarly the bottom-up order is denoted P/W. The cells of the table indicate the various valuations elicited from this design. Here W/p denotes the valuation of good W as the first good valued by the W/P sample (i.e. the bold character indicates the good valued while the overall label denotes the ordering from which it was derived). Similarly, p/W denotes the value of the same good W obtained as the second good valued by the P/W sample.
TABLE 1 ABOUT HERE

Expectations are provided in terms of directional effects. Based on the discussion above, Table 3 presents a summary of theoretically derived expectations for our study design. These are contrasted with empirical regularities derived from the literature. In both cases the relevant lines of explanatory reasoning are also summarised.

TABLE 2 ABOUT HERE

Table 2 is subdivided into three types of expectation/empirical regularity which we discuss as follows:

(i) Scope sensitivity

Although not the main issue considered in this study, the outcome of a split-sample comparison of scope sensitivity is likely to influence the sensitivity of valuations to the order adopted. Assuming positive marginal utility and areas protected differing only in terms of scope, economic theory would suggest that values for the W scheme would be greater than or equal to those for the P scheme. A weaker expectation is for non-negative marginal utility allowing values for goods W and P to be insignificantly different.

(ii) Sub-component valuations

Using an exclusive ordering, the valuations P/w and w/P theoretically have the same reference levels and should yield the same values. However, as discussed previously,
within a stepwise design, if valuation tasks are not viewed as independent then the variation in VCS as a respondent progresses through a valuation exercise may equate to changes in strategy space. A finding that $P/w$ and $w/P$ do differ could therefore be interpreted as indicating that the assumption of independence does not hold. However, alternative explanations can also be put forward suggesting that such a difference constitute an anomaly. Here initially stated values are driven by warm glow (Andreoni, 1990) or some form of other-regarding behaviour (Goerre, Holt and Laury, 1999) while subsequent valuations are set purely to be consistent with these such that smaller goods are accorded smaller willingness to pay and vice versa.

(iii) Inclusive good valuations

Again for exclusive lists where valuation tasks are perceived as independent, economic theory expects that $W/p = p/W$. As before, relaxing the independence assumption implies that we can no longer be sure of such expectations. Furthermore, as per (ii), competing explanations can be put forward wherein initial valuations are driven by warm glow motives and subsequent values are adjusted up or down from this to ensure internal consistency. Both these and violation of the independence assumption result in $W/p < p/W$. However, note that respondents are more likely to encounter budget constraints with larger goods. Such constraints would act to reduce the magnitude of this difference relative to that described for the P goods under expectation (ii) above.

3 Note that, strictly speaking, this interpretation run counter to theory relying as it does on mental accounting theories as discussed in the context of CV studies by Sugden (1999).
4. Case study

The case study employed to motivate this analysis concerns a riverine saline flood alleviation scheme (FAS) in Broadland, an area that stretches across parts of Norfolk and Suffolk in Eastern England as illustrated in Figure 1 (versions of which were presented to survey respondents depending upon the scheme valued).

FIGURE 1 ABOUT HERE

The Broadland area consists of 200 km of waterways and almost 50 small lakes (known locally as Broads), which make the network one of the largest water based recreation areas in Europe, and unique within the United Kingdom. The character of the low-lying Broadland landscape is dependent on river embankments that protect some 21,300 hectares of the area from tidal water. At the time of the survey\(^4\), these river embankments were old and generally of poor quality. Continual erosion from boat traffic, tidal forces and sinking of the general ground level had caused the embankments to deteriorate. Consequently, return periods for overtopping were declining, whilst the probability of a major fracture in the flood defences had become significant. The magnitude of the predicted saline flooding in terms of the likely minimum and the possible extent are also shown in Figure 1 (Turner and Brooke, 1988). Such flooding would have caused large-scale changes to the area in terms of ecology, landscape and recreation. The Broadland survey assessed the equivalent loss of recreational visitors in the absence of the FAS.

\(^4\) Following the Broadland survey the W FAS was commissioned and the river embankments discussed have been improved.
Given the current level of saline flood protection in Broadland, which is regarded as insufficient, two alternatives were considered within this analysis: a series of ‘part’ area flood alleviation schemes (a set of $P$ goods) to upgrade the protection of either area A, B, C or D in Figure 1; or a ‘whole’ area flood alleviation scheme (the $W$ good) to upgrade the protection of the entire threatened extent of Broadland (shown as the likely minimum and possible maximum threatened extent in Figure 1). The resultant bundles of goods and services are nested, with $P^A$, $P^B$, $P^C$ or $P^D$ being subsets of $W$.

The $P$ areas considered differ in a number of respects, most notably size. However, as they are not perfectly homogeneous in non-size attributes they cannot be considered as perfectly nested. An indication of the multi-attribute characteristics of the various $P$ and $W$ goods is given in Table 3.

TABLE 3 ABOUT HERE

The survey was conducted in Broadland using face-to-face interviews with a randomly selected sample of visitors. Interviews were undertaken at a number of locations around the study area and respondents were presented with one of the $P$ goods (chosen to include the area in which the particular interview was being undertaken) and the $W$ good. Exclusive list formats were adopted throughout and the test of ordering was facilitated by employing either a top-down (i.e. valuing $W$ before $P$) or bottom-up (i.e. valuing $P$ before $W$) presentation order.

In line with the NOAA Blue-Ribbon panel recommendations (Arrow et al., 1993) a dichotomous choice (DC) format willingness to pay question was employed where
respondents are asked to state ‘yes’ or ‘no’ to a researcher specified bid level (BL) amount for provision of the flood alleviation strategy described. By varying the BL across the sample the researcher can infer summary willingness to pay statistics. This approach was extended using a second DC question as per the double-bound approach formulated by Hanemann et al. (1991) and discussed in detail by Hanemann and Kanninen (1999). Here positive responses to the initial BL determine a second, higher, BL which is presented to the respondent. Similarly, a negative response to the initial BL determines a second lower BL which is presented to the respondent. This approach significantly boosts the statistical efficiency of the DC approach while combining this with the novel use of exclusive lists ensured that (assuming that valuation tasks are seen as independent) a consistent, status quo, endowment point is used for all valuations thus enhancing comparability of valuations.

The questionnaire was refined using a ‘mixed methodology’ approach combining to focus group analyses and an open-ended pilot exercise involving a sample of 175 respondents. This process was useful in increasing the clarity and relevance of the information provided to respondents and provided confirmation of other survey design choices. The revised final survey questionnaire opened by asking respondents about their visit frequency, travel time, sites visited, activities they undertake and aspects they like or dislike about Broadland. Next, the possibility of saline flooding was presented by the interviewer through an information statement and a storyboard

---

5 Note that such a willingness to pay to avoid loss format corresponds to the equivalent loss measure proposed by Hicks (1943;1956). For further discussion of this and other welfare measures see Bateman et al. (1997) and Scarpa and Bateman (2000).

6 However, a number of commentators have questioned the strategic incentives of the double-bounded approach (Carson et al., 1999). For a recent review and empirical investigation see Bateman et al. (2001b).
containing further text and photographs depicting the landscape and ecology changes after persistent saline flooding. Areas at or below sea level were illustrated on a map and pointed out prior to any valuation task so as to ensure that all respondents were aware of the full extent of the possible flooding problem in the study area.

Within either presentation ordering (W/P or P/W) the first good to be valued was introduced by superimposing on the map a transparency of the area which would be protected under the scheme in question. Respondents were then informed about the payment vehicle (increases in general taxation\(^7\)) and asked whether they agreed in principle with an increase in their taxes (irrespective of the exact amount) to ensure that the improved FAS (P or W) would be implemented. This is referred to here as the ‘payment principle question’ and was included to reduce any feelings of moral obligation to pay and thereby justify non-payment. Those responding positively to this question were then presented with the double bounded DC format questions outlined above. Irrespective of their response to the initial good, the second good was then presented for valuation with an identical procedure being employed to that described above. As any difference may reduce the effectiveness of the comparison, the same starting BLs were used for both the first and second good valued.

At the end of the questionnaire respondents were asked a variety of motivational and scenario perception questions including reasons for their valuation response, attitudes towards the payment vehicle, trust in the authority responsible for implementing the

\(^7\) This payment vehicle was chosen to be comparable with a preceding study of the Broadland area described in Bateman et al. (1995) wherein this vehicle was shown to work well for flood alleviation schemes.
scheme, the perceived realism of the proposed scheme etc. Demographic and socio-economic characteristics of the respondent’s household were also elicited.

5. Results

A total of 1592 interviews were completed (1063 W/P and 529 P/W)\(^8\), with approximately 68% of those approached completing the questionnaire. As detailed in Table 4, potentially important differences were observed in the perceived credibility of schemes. Here those respondents facing the top-down W/P order consistently found all schemes significantly less realistic (\(\alpha = 5\%\)) than those facing the bottom-up P/W ordering. Similarly, the level of credibility was not constant across all goods with the P goods consistently rated as more realistic than the W goods irrespective of the presentation ordering. Together these findings suggest that the W good is seen as less realistic than the P good and that the presentation of good W at the start of the W/P valuation ordering undermines the credibility of subsequent P good relative to that observed in the P/W ordering.

TABLE 4 ABOUT HERE

5.1. Ordering effects upon values

The welfare measures were estimated using conventional methods (Hanemann and Kanninen, 1999) from log-logistic models which were found to provide a superior functional form than other models tested. As the coefficients on the log of BL were

---

\(^8\) The larger number of respondents in the W/P was due to the requirement for a parallel study comparing the W FAS valuations to those from a previous survey by Bateman et al. (1995). Given the size and likelihood of flooding a priority was given to collecting observations for Area D. The survey for the present study was conducted in the summer of 1996.
found to be between 0 and −1, the mean was estimated by integrating the area under the cumulative distribution function modelled (ibid.). The mean estimates were calculated using the Simpson method (Duffield and Patterson, 1991), with the distribution censored at the highest BL offered. This approximation implies that the estimates represent a lower bound of the mean (LBM). Alternatives to censoring at the maximum BL are available, however they all depend on an arbitrary choice of the upper truncation point and their adoption generally increases the standard error of the estimates (Ready and Hu, 1995; Haab and McConnell, 1998).

A non-parametric bootstrapping approach (Efron and Tibshirani, 1993) using 1000 iterations was adopted to estimate confidence intervals for the mean and median values. This approach avoids parametric assumptions and is versatile in that it can be used to provide estimated distributions for any welfare measure required and so provides the basis for examining differences between measures, summaries of which are given in Table 5.

Table 5 presents summary models (full details in Table 6) along with descriptive statistics for median and mean welfare measures for each of the W and various P goods valued. In each case these are reported for both the top-down (W/P) and bottom-

---

9 Those refusing the payment principle question were included by counting them as 'no' DC responses (Hanemann and Kanninen, 1999). ‘Don’t know’ responses (13% part and 12% whole) were excluded, along with a small number of respondents (3% for both part and whole) who, in the debriefing questions at the end of the interview stated that they either did not feel they would actually have to pay for the goods on offer or that they were responding as if they were giving to a good cause rather than the good in question.

10 In a study comparing techniques for calculating confidence intervals, Cooper (1994) found a similar approach to perform as well as three alternative methods. Similar non-parametric bootstrap approaches are applied to CV data by Ready et al. (1996) and Sprent (1998).

11 As the resultant distribution of the means were found to be normal and the standard errors differed only in terms of sample variation, a difference of two means Z-test was performed to compare
up ($P/W$) ordering treatments. The models are in themselves unremarkable apart from the clear difference in intercept and slope estimates across ordering treatments. Turning to consider both the median and mean values we see that the former are considerably lower than the latter indicating the positively skewed response distributions typical of CV data. Within each of these measures we can see that values for the inclusive $W$ good are consistently larger than those for the various $P$ goods. Furthermore, the overlap in the distributions show a mixed picture of scope sensitivity and insensitivity, where scope sensitivity is consistent with positive marginal utility for increased provision of this public good (i.e. satisfying Expectation (i) in Table 3 from an economic-theoretic perspective). Consistent scope sensitivity is observed for Areas A and D.

While the finding of consistent scope sensitivity accords with expectations, the main focus of this study is the examination of possible ordering effects. Table 5 reports ten within-good tests for such effects, eight of which are significant at the 1% level (the two exceptions both concerning good $P^b$). All significant differences lie in the same direction, namely that values derived from the bottom-up $P/W$ ordering exceed those from the top-down $W/P$ ordering. This implies that values for the less inclusive $P$ goods are highest when that good is presented first in a list, while values for the inclusive good $W$ are larger when it is presented second in a list.

### 5.2 Investigating ordering effects
Quantitative investigation of the routes through which ordering effects may be impinging upon valuation responses was undertaken both through regression modelling and analysis of patterns of DC responses across bounds under each ordering treatment. Table 6 details four double bounded log-logistic regression models of valuation responses, which for convenience we label (i) to (iv). The first two of these consider pooled valuations for the W good and are distinguished by the way in which potential ordering effects are modelled. In Model (i) ordering is incorporated by the use of a simple intercept shift and a BL interaction term both of which indicate how the W/P ordering differs from the P/W base case. In Model (ii) the intercept shift is replaced by a series of interaction terms showing how the W/P ordering differs with respect to each of the explanatory variables included in the model and defined in notes to this table. Models (iii) and (iv) repeat this exercise for the P good valuation responses.

**TABLE 6 ABOUT HERE**

The results in Table 6 show the valuation responses to be affected by a number of factors, with all (non-ordering) effects conforming to expectations\(^{12}\). Agreement to pay for a given flood alleviation scheme was positively related to the respondents income, their long term preference towards protecting Broadland (Bropref1), attitude towards the payment vehicle used (Govrank), membership of pro-environmental groups (Green; not significant in two cases). A further interesting relationship was that

\(^{12}\) Several variables that proxy distance were calculated from survey responses but were not found to be statistically significant.
as the perceived realism of a given scheme increased so respondents were more likely to agree to pay for that scheme (Realism), a relationship highlighted by Carson (1997) as a major cause of scope insensitivity in low credibility scenarios.\(^\text{13}\)

Turning to consider the issue of ordering, neither of the interaction terms in Model (i) are statistically significant at the 5% level, suggesting that, when other explanatory variables are controlled for, ordering effects are insignificant within valuations of the W good (although a log-likelihood test for the inclusion of these interaction variables showed a model improvement significant at the 5% level). Considering the fuller specification in Model (ii) a number of interaction terms prove significant suggesting a complex array of forces which together cancel each other out). Turning to consider responses for the P good, Model (iii) shows highly significant ordering interaction terms. The sign on the \(W/P\) (intercept) variable confirms our previous result that valuations for good P are significantly higher when obtained from the \(P/W\) ordering. Comparison with Model (i) provides an interesting extension to the trends suggested in Table 5. When we control for other explanatory variables it seems that ordering effects for the smaller P good are significantly stronger than those for the more inclusive W good. Some indication of the factors driving the elevation of P values is provided by Model (iv), in particular the change in the coefficient on BL suggests that respondents are more likely to agree to pay relatively high amounts for good P when it is presented as the initial as opposed to subsequent good in a stepwise list, even when that list is exclusive in nature. The significantly elevated impact of higher incomes within the P/W presentation of good P is also an interesting feature of these results.

\(^{13}\) In the case of the P FAS valuation responses, the perceived realism variable reflects variation between individuals and P FASs valued.
Conclusions

Despite its use of incentive compatible DC elicitation methods, our study has found clear evidence of ordering effects in values for nested goods presented in top-down and bottom-up stepwise lists. An initial reaction to such findings might be to see them as constituting either a theoretical anomaly or reflecting a methodological flaw in the CV technique. However, economic–theoretic expectations turn on the assumption of whether or not valuation tasks are perceived to be truly independent.

As discussed in Section 2 of this paper, in inclusive lists independence clearly does not hold and hence sequencing effects are to be expected. Conversely, in exclusive lists using private goods task independence clearly does hold and any ordering effect constitutes as the nature of public goods and their modes of funding means that the independence of valuation tasks cannot be taken for granted. While the DC elicitation format is typically cited as incentive compatible, the introduction of a second nested good within a stepwise list may encourage respondents to engage in strategic behaviour. Here valuation tasks are no longer seen as independent and it is therefore possible that the ordering affects seen in this study do not constitute anomalies but are a reflection of changes in strategy space between the valuation of the first and second goods encountered.

While we accept the coherence of the above argument, there are a number of psychological and anomaly based explanations of ordering effects in CV studies which also fit the data well. These include an intersection between the ‘warm glow’
hypothesis of Kahneman and Knetsch (1992a), wherein CV valuations are significantly motivated by the moral satisfaction which respondents obtain from contributing to good causes, and the ‘anchoring and adjustment’ heuristic emphasized by Kahneman et al. (1982), which suggests that an initial valuation provides an anchor for subsequent responses which typically insufficiently adjust from that level. If, as argued by Kahneman and Knetsch, a respondents first valuation response tends to be a dumping ground for moral satisfaction then this might be reflected in the elevation of \( P/w \) over \( w/P \). Furthermore, if the elevated \( P/w \) response provides a relatively ‘high’ anchor for the subsequent \( p/W \) response then this might explain the excess of the latter value over the \( W/p \) response.

Other anomaly-based arguments can be put forward to explain our results. In an experimental context, Cubitt and Sugden (2002) note that preferences seem malleable to surprise induced by unexpected additional choices. From this perspective it is arguable that the second valuation task within the stepwise top-down (\( W/P \)) ordering may pose less of a surprise than does the second question in the bottom-up (\( P/W \)) presentation, the argument being that individuals know that a large good can be divided but do not automatically assume that a small good can be increased (empirical evidence supporting such an interpretation is given by Bateman et al., 2001a). Such reasoning may underpin the ordering difference in perceived scheme realism detailed in Table 4 (responses for which were elicited after the valuation of both goods). However, the models of Table 6 explicitly incorporate this highly significant (and theoretically consistent) factor and this does not seem to be driving the valuation difference between orderings. Therefore we are driven back to our competing
arguments concerning the independence of valuation tasks or moral satisfaction and anchoring.

In conclusion we have a clear result, that ordering effects occur in such designs, but face controversy in distinguishing between two competing explanations for such findings, one of which is compatible with economic theory while the other is not. This study therefore paves the way for further research not into whether ordering effects occur, but why they do. It seems to the authors that qualitative methods using debriefing and think-aloud techniques offer an obvious course for such future developments.
Acknowledgements

The research on which this paper is based was funded jointly by the Economic and Social Research Council and the Environmental Agency. The authors are grateful to Riccardo Scarpa (University of Newcastle) for his Limdep code and to Greg Poe (Cornell University) for discussion of issues regarding nested good. Errors are the sole responsibility of the authors.
References


Table 1: Survey samples and valuation responses

<table>
<thead>
<tr>
<th>Sample sequence</th>
<th>1st valuation response</th>
<th>2nd valuation response</th>
</tr>
</thead>
<tbody>
<tr>
<td>W/P</td>
<td>W/p</td>
<td>w/P</td>
</tr>
<tr>
<td>P/W</td>
<td>P/w</td>
<td>p/W</td>
</tr>
</tbody>
</table>

Note:
1. **Bold**, upper case letters denote the good for which the value is expressed
Table 2: Summary of theoretically and empirically derived expectations for our study design.

<table>
<thead>
<tr>
<th>Type</th>
<th>Expectation/empirical regularity</th>
<th>Economic reasoning</th>
<th>Other reasoning</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) Scope sensitivity</td>
<td>Economic theory and some previous evidence suggests: $W/p &gt; P/w$</td>
<td>Positive marginal utility</td>
<td></td>
</tr>
<tr>
<td></td>
<td>but other previous evidence has found: $W/p = P/w$</td>
<td></td>
<td>Initial responses are warm glows and irrespective of scope</td>
</tr>
<tr>
<td>(ii) Sub-component valuations</td>
<td>Theory suggests: $P/w &gt; w/P$</td>
<td>Awareness of superior substitute in $w/P$ response</td>
<td>Initial responses are scope insensitive warm glows; subsequent values are set to be internally consistent with initial responses.</td>
</tr>
<tr>
<td>(iii) Inclusive good valuations</td>
<td>Theory suggests; $W/p = p/W$</td>
<td>Procedural invariance</td>
<td>Initial responses are warm glows and irrespective of scope Subsequent responses set to be internally consistent, in this case by raising $p/W$ above $W/p$. However, budget constraints may limit this effect</td>
</tr>
<tr>
<td></td>
<td>but other non-economic theories imply that; $W/p &lt; p/W$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3: Flood alleviation scheme characteristics

<table>
<thead>
<tr>
<th>Area&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Good to be valued</th>
<th>Total area (ha.)</th>
<th>Proportion of total area which is SSSI&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Does area contain trees or lakes?</th>
<th>Landscape quality relative to Whole area&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Saltwater flooding category&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole W</td>
<td></td>
<td>21300</td>
<td>0.11</td>
<td>Yes</td>
<td>Average</td>
<td>Various&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Area A P&lt;sup&gt;A&lt;/sup&gt;</td>
<td></td>
<td>475</td>
<td>0.36</td>
<td>Yes</td>
<td>Below</td>
<td>Possible</td>
</tr>
<tr>
<td>Area B P&lt;sup&gt;B&lt;/sup&gt;</td>
<td></td>
<td>420</td>
<td>0.17</td>
<td>No</td>
<td>Below</td>
<td>Possible</td>
</tr>
<tr>
<td>Area C P&lt;sup&gt;C&lt;/sup&gt;</td>
<td></td>
<td>1700</td>
<td>0.06</td>
<td>Yes</td>
<td>Above</td>
<td>Likely</td>
</tr>
<tr>
<td>Area D P&lt;sup&gt;D&lt;/sup&gt;</td>
<td></td>
<td>4000</td>
<td>0.30</td>
<td>No</td>
<td>Above</td>
<td>Likely</td>
</tr>
</tbody>
</table>

Notes:

<sup>a</sup> The areas protected by these riverine flood alleviation schemes are shown in Figure 1.

<sup>b</sup> An SSSI is a Site of Special Scientific Interest. Areas estimated from local area maps.

<sup>c</sup> This indicates whether the area is above or below average landscape quality for the Broadland area, where the assessment was based on the extent to which the landscape was characteristic of the area (Turner and Brooke, 1988).

<sup>d</sup> ‘Possible’ relates to the ‘possible extent of saltwater flooding’ and ‘likely’ to the ‘likely minimum extent of saltwater flooding’ in Figure 1. These categories were taken from Turner and Brooke (1988).

<sup>e</sup> Saltwater flood risk varies from unlikely to likely over the Whole area; details in Figure 1 as presented to survey respondents.
Table 4: Response frequency and perceived scheme credibility disaggregated by presentation ordering and the good valued.

<table>
<thead>
<tr>
<th>Ordering →</th>
<th>Number of responses</th>
<th>Scheme credibility (%)&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>W/P</td>
<td>P/W</td>
</tr>
<tr>
<td>W</td>
<td>1063</td>
<td>529</td>
</tr>
<tr>
<td>P&lt;sub&gt;A&lt;/sub&gt;</td>
<td>236</td>
<td>94</td>
</tr>
<tr>
<td>P&lt;sub&gt;B&lt;/sub&gt;</td>
<td>237</td>
<td>120</td>
</tr>
<tr>
<td>P&lt;sub&gt;C&lt;/sub&gt;</td>
<td>230</td>
<td>106</td>
</tr>
<tr>
<td>P&lt;sub&gt;D&lt;/sub&gt;</td>
<td>360</td>
<td>209</td>
</tr>
</tbody>
</table>

Note:

<sup>a</sup> Columns give the percentage responding positively to a question asking whether respondents felt it was realistic to protect the area concerned from flooding.
Table 6: Pooled log-logistic models of double bounded dichotomous choice questions\textsuperscript{a}

<table>
<thead>
<tr>
<th></th>
<th>Pool ‘whole’ (i)</th>
<th>Pooled ‘whole’ (ii)</th>
<th>Pooled ‘part’ (iii)</th>
<th>Pooled ‘part’ (iv)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.3079 (0.01)</td>
<td>-1.5921 (0.06)</td>
<td>-1.6917 (0.05)</td>
<td>-2.6123 (0.01)</td>
</tr>
<tr>
<td>ln (bid level)</td>
<td>-0.4957 (0.00)</td>
<td>-0.4840 (0.00)</td>
<td>-0.6205 (0.00)</td>
<td>-0.6162 (0.00)</td>
</tr>
<tr>
<td>Ln (income)</td>
<td>0.4001 (0.00)</td>
<td>0.3550 (0.00)</td>
<td>0.2589 (0.00)</td>
<td>0.3559 (0.00)</td>
</tr>
<tr>
<td>Bropref\textsuperscript{1}</td>
<td>0.8054 (0.00)</td>
<td>0.4374 (0.02)</td>
<td>0.4641 (0.00)</td>
<td>0.5215 (0.00)</td>
</tr>
<tr>
<td>Govrank\textsuperscript{1}</td>
<td>0.5109 (0.00)</td>
<td>0.5128 (0.01)</td>
<td>0.3076 (0.01)</td>
<td>0.5407 (0.01)</td>
</tr>
<tr>
<td>Realism</td>
<td>0.4959 (0.00)</td>
<td>0.2922 (0.17)</td>
<td>1.5231(0.00)</td>
<td>1.3993 (0.00)</td>
</tr>
<tr>
<td>Green</td>
<td>0.2693 (0.07)</td>
<td>-0.1697 (0.50)</td>
<td>0.3134 (0.03)</td>
<td>-0.1520 (0.56)</td>
</tr>
<tr>
<td>W/P (intercept)</td>
<td>0.1968 (0.51)</td>
<td>-</td>
<td>-1.3374 (0.00)</td>
<td>-</td>
</tr>
<tr>
<td>W/P (Ln(bid level))</td>
<td>-0.0938 (0.18)</td>
<td>-0.1213 (0.08)</td>
<td>0.1842 (0.01)</td>
<td>0.1743 (0.01)</td>
</tr>
<tr>
<td>W/P (Ln(income))</td>
<td>-</td>
<td>0.0928 (0.05)</td>
<td>-</td>
<td>-0.1543 (0.00)</td>
</tr>
<tr>
<td>W/P (Bropref\textsuperscript{1})</td>
<td>-</td>
<td>-0.4819 (0.00)</td>
<td>-</td>
<td>0.5426 (0.72)</td>
</tr>
<tr>
<td>W/P (Govrank\textsuperscript{1})</td>
<td>-</td>
<td>0.0170 (0.95)</td>
<td>-</td>
<td>-0.3636 (0.16)</td>
</tr>
<tr>
<td>W/P (Realism)</td>
<td>-</td>
<td>0.2407 (0.37)</td>
<td>-</td>
<td>0.1828 (0.57)</td>
</tr>
<tr>
<td>W/P (Green)</td>
<td>-</td>
<td>0.6638 (0.04)</td>
<td>-</td>
<td>0.6747 (0.03)</td>
</tr>
<tr>
<td>LRI</td>
<td>0.05</td>
<td>0.06</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td>Sample size</td>
<td>1241</td>
<td>1241</td>
<td>1252</td>
<td>1252</td>
</tr>
</tbody>
</table>

Note:
\textsuperscript{a} ‘yes’ responses coded as 1 and ‘no’ as 0.
\textsuperscript{b} significance level in () brackets based on a t-test.

LRI = likelihood ratio index

ln (bid level) = the natural logarithm of the bid level.

Ln (income) = the natural logarithm of the respondents household income.

Bropref = 1 if the respondents chose the option ‘protect all of Broadland, whatever the cost’ as their first choice from a list of options; = 0 otherwise.

Govrank = 1 if the respondents choose government taxation as their first choose from a series of payment vehicle options.; = 0 otherwise.

Realism = 1 if respondents consider the W or the P FAS to be feasible; = 0 otherwise.

Green = 1 if the respondent is a member of the Royal Society for the Protection of Birds and/or the National Trust; = 0 otherwise

W/P ( ) = interaction terms showing departures for the W/P sequence.
Table 5: Welfare measure estimates using the log-logistic model\textsuperscript{a,b}

<table>
<thead>
<tr>
<th>Good being valued / ordering</th>
<th>Sample size</th>
<th>Models</th>
<th>Medians</th>
<th>Low Bound Means (LBMs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Intercept</td>
<td>BL coef.</td>
<td>LLR\textsuperscript{c}</td>
</tr>
<tr>
<td><strong>Whole Area:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W/p</td>
<td>822</td>
<td>2.7410</td>
<td>-0.5966</td>
<td>0.06</td>
</tr>
<tr>
<td>p/W</td>
<td>419</td>
<td>2.4568</td>
<td>-0.4917</td>
<td>0.04</td>
</tr>
<tr>
<td><strong>Part Area A:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/P</td>
<td>195</td>
<td>0.9253</td>
<td>-0.3862</td>
<td>0.06</td>
</tr>
<tr>
<td>P/w</td>
<td>76</td>
<td>2.9230</td>
<td>-0.6849</td>
<td>0.07</td>
</tr>
<tr>
<td><strong>Part Area B:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/P</td>
<td>169</td>
<td>0.1880</td>
<td>-0.4709</td>
<td>0.19</td>
</tr>
<tr>
<td>P/w</td>
<td>89</td>
<td>2.5558</td>
<td>-0.5753</td>
<td>0.09</td>
</tr>
<tr>
<td><strong>Part Area C:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/P</td>
<td>176</td>
<td>1.6307</td>
<td>-0.6113</td>
<td>0.04</td>
</tr>
<tr>
<td>P/w</td>
<td>77</td>
<td>3.3233</td>
<td>-0.7332</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>Part Area D:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>w/P</td>
<td>293</td>
<td>1.3873</td>
<td>-0.4092</td>
<td>0.09</td>
</tr>
<tr>
<td>P/w</td>
<td>177</td>
<td>2.4439</td>
<td>-0.6820</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Notes:
\textsuperscript{a} ‘yes’ responses coded as 1 and ‘no’ as 0
\textsuperscript{b} The log-logistic models contain Ln(bid level), Ln(income), Bropref, Govrank, Realism and Green. See Table 6 for the meaning of these variables.
\textsuperscript{c} The likelihood ratio index (LRI) is a measure of goodness of fit calculated as one minus the ratio of the restricted and unrestricted log likelihood.