ELSEVIER

Ecological Economics 12 (1995) 161–179

ECOLOGICAL ECONOMICS

Elicitation and truncation effects in contingent valuation studies

Ian J. Bateman [a,b,*], Ian H. Langford [a,b], R. Kerry Turner [a,b], Ken G. Willis [c], Guy D. Garrod [d]

[a] *School of Environmental Science, University of East Anglia, NR4 7TJ Norwich, UK*
[b] *Centre for Social and Economic Research on the Global Environment University of East Anglia and University College, London, UK*
[c] *Dept. of Town and Country Planning, University of Newcastle Upon Tyne, Newcastle, UK*
[d] *Dept. of Agricultural Economics and Food Marketing, University of Newcastle Upon Tyne, Newcastle, UK*

Received 14 December 1993; accepted 30 June 1994

Abstract

The contingent valuation method (CVM) uses surveys of expressed preferences to evaluate willingness to pay for (generally) non-market, environmental goods. This approach gives the method theoretical applicability to an extensive range of use and passive-use values associated with such goods. However, recent years have seen the method come under sustained empirical and theoretical attack by critics who claim that the expressed preference statements given by respondents to CVM questions are subject to a variety of biases to the extent that "true" valuations cannot be inferred. This debate was reviewed and assessed in the recent report of the US, NOAA "blue-ribbon" panel which gave cautious approval to the method subject to adherence to a rigorous testing protocol. This paper reports findings from the first UK CVM study to generally conform to those guidelines. The major objective of the research reported on here is the analysis of the effects of altering the method of eliciting willingness to pay (WTP) responses. Three WTP elicitation methods are employed: open-ended questions (where the respondent is free to give any answer); dichotomous choice questions (requiring a yes/no response regarding a set WTP bid level); and iterative bidding questions (where a respondent is free to move up or down from a given WTP starting point). Results indicate that respondents experience significant uncertainty in answering open-ended questions and may exhibit free-riding or strategic overbidding tendencies (although this is less certain). When answering dichotomous choice questions respondents seem to experience much less uncertainty although the suggestion that bid levels affect responses cannot be ruled out, and it is clear that respondents behave somewhat differently to dichotomous choice as opposed to open-ended formats. The iterative bidding approach appears to provide a halfway house with respondents exhibiting certain of the characteristics of both the other formats. We concluded that the level of uncertainty induced by open-ended formats is a major concern, and that further research into the microeconomic motivations of individuals responding to iterative bidding and dichotomous choice CV surveys is high priority. A further aim of the analysis was to test for changes in estimated mean WTP induced by the application of different

* Correspondence to: I.J. Bateman, School of Environmental Sciences, University of East Anglia, NR4 7TJ, Norwich, UK.

SSDI 0921-8009(94)00044-1

forms of truncation across all elicitation methods. Recommendations are made on appropriate truncation strategies for each elicitation method.

Keywords: Contingent valuation; Willingness to pay

1. Introduction

Litigation over natural resource damages, particularly oil spill damage cases, in the USA has most recently led to a further upsurge of interest in, use of and debate over the contingent valuation method (CVM). Following CV assessment of passive-use loss incurred by the Exxon Valdez oil spill (Carson et al., 1992) in 1992, the Exxon company commissioned a number of CV studies and organised a public seminar which heavily criticised both the reliability and validity of the method (Cambridge Economics, 1992). A week later the US Department of Commerce's National Oceanic and Atmospheric Administration (NOAA), which has responsibility for promulgating the regulations for the assessment of natural resource damages under the Oil Pollution Act of 1990, announced that it was setting up a "blue-ribbon" panel under the joint Chairmanship of Kenneth Arrow and Robert Solow to advise on the use of CV in natural resource damage assessments for oil spills. The NOAA panel concluded that CV studies "can produce estimates reliable enough to be the starting point for a judicial or administrative determination of natural resource damages... To be acceptable for this purpose, though, such studies should adhere closely to the guidelines described in this report" (Arrow et al., 1993). Thus the testing protocol has been given prime importance in the Panel's qualified acceptance of CVM.

1.1. Blue-ribbon panel's testing protocol

The following main guidelines, not in any rank order, have been suggested by the Panel (Arrow et al., 1993):

1. For a single dichotomous question (yes–no type) format, a total sample size of at least 1000 respondents is required. Clustering and stratification issues should be accounted for and random sub-sampling will be required to obtain a bid curve. Testing for interviewer and wording biases is also recommended.
2. High non-response rates would render the survey unreliable.
3. Face-to-face interviewing is likely to yield the most reliable results.
4. Full reporting of data and questionnaires is required for good practice.
5. Pilot surveying and pretesting are essential elements in any CV study.
6. A conservative design more likely to underestimate WTP is to be preferred to one likely to overestimate WTP.
7. WTP format is preferred.
8. The valuation question should be posed as a vote on a referendum, i.e., a dichotomous choice question related to the payment of a particular level of taxation.
9. Accurate information on the valuation situation must be presented to respondents; particular care is required over the use of photographs.
10. Respondents must be reminded of the status of any undamaged possible substitute commodities.
11. Time-dependent measurement noise should be reduced by averaging across independently drawn samples taken at different points in time.
12. A "no-answer" option should be explicitly allowed in addition to the "yes" and "no" vote options on the main valuation question.
13. Yes and no responses should be followed up by the open-ended question: "why did you vote yes/no?"
14. Cross-tabulations: the survey should include a variety of other questions that help to interpret the responses to the primary valuation question, i.e., income, distance to the site, prior knowledge of the site, etc.

15. Respondents must be reminded of alternative expenditure possibilities, especially when "warm-glow" effects are likely to be prevalent (i.e., purchase of moral satisfaction through the act of charitable giving).

In this paper we report on selected aspects of a recent large sample (nearly 3000 interviews) CV study of an internationally important UK wetland (the Norfolk Broads) under increasing threat from North Sea salt-water intrusion and requiring new flood protection investments. We will briefly compare the wetland CV study design and method against the Blue-Ribbon Panel's protocol, but devote most of the analysis to an assessment of the elicitation format used. In a briefer discussion we assess the impact of various sample truncation strategies upon each format.

The major research focus of this study concerns guideline 8 above regarding WTP elicitation method. The following three approaches (including the NOAA panel's preferred dichotomous choice format) were tested:

i. Open ended (OE); here the respondent is asked "How much are you willing to pay?" The respondent is therefore free to state any amount (cf. Brookshire et al., 1983).
ii. Dichotomous choice (DC); here respondents are asked "Are you willing to pay £X?" with the bid level X being systematically varied across the sample (cf. Bishop and Heberlein, 1979).
iii. Iterative bidding (IB); here the interviewer states an initial WTP bid level which may be fixed across the sample or varied systematically as in the DC approach. The respondent is then asked if they are WTP this amount. If they reply positively, then the bid level is increased, while if they reply negatively, then the bid level is decreased, and the WTP question is asked again (cf. Randall et al., 1974). Precise details of the particular IB game used are given below.

Before we turn to consider the detail of this study, we wish to emphasise that we are addressing only certain of the issues raised by use of the CVM and recognise that such an analysis cannot of itself provide a sufficient basis for validation of the method.

2. Norfolk Broads wetland CV

The Norfolk Broads is a unique wetland area located in the East Anglian area of England which, due to the age of existing defences, is under increasing threat of saline flooding from the North Sea. A number of flood alleviation schemes have been suggested, including the strengthening of existing river walls and/or the construction of a movable tidal barrier. As part of a cost–benefit study of the available options, a CV study was undertaken in order to estimate a user value for the conservation of the area in its current state. [1]

Our wetland CV study meets the requirements of thirteen of the fifteen guidelines set out by the Blue-Ribbon Panel. In the case of guidelines ten and eleven we would argue that the unique nature of Broadland precludes the use of the former and, the latter is not particularly relevant to the valuation context under test, i.e., unlike the Exxon Valdez oil spill incident where a response-decay over time is likely, deterioration in the Norfolk Broads flood defences is a gradual and ongoing process (although we recognise the validity and desirability of retesting).

A large-scale (2897 useable interviews), on-site, face-to-face CVM survey was undertaken during August and September of 1991. Information regarding the present and projected flooded nature of Broadland [2] was presented and standardised via a "constant information statement" which was read out to all respondents. This also informed respondents about the feasibility of restoring flood defences so as to alleviate flood risk. The questionnaire was extensively piloted (see discussion below) and was designed to minimise sources of bias whilst obtaining a full database regarding

[1] In addition to this user value, a mail survey of non-users throughout the UK was also undertaken (see Bateman et al., 1992).

[2] All impacts were detailed, however, normative statements such as "losses" or "gains" were avoided.

respondents' characteristics so as to permit thorough analysis of WTP responses. [3]

The WTP question was phrased so as to present a readily understood, clearcut scenario: how much would the respondent be WTP to prevent flooding? This provides an equivalent surplus welfare measure of WTP to avoid loss. The final (post-pilot) wording of the OE WTP question was as follows:

> "Now, remembering that any money which you spend on the Broads you cannot spend elsewhere, please consider carefully how much more in taxes you would be willing to pay over the coming year in order to prevent flooding and preserve the Broads in their present state during that time? Any amount which you offer will only be spent on preserving the Broads, nothing else."

The WTP question for IB and DC formats used a similar preamble to the above.

2.1. Elicitation method testing: theoretical expectations and potential biases

The number and level of bids used for the DC experiment were set using data obtained from an OE pilot study (detailed subsequently). The IB experiment was applied to all those respondents initially presented with a DC bid level. If a respondent answered positively to the DC bid, this bid was then doubled and the WTP question asked again. However, if the response to the initial DC question was negative, then the bid amount was halved and the WTP question asked again. The procedure was then repeated given the answer to the secondary questions to provide in total a triple bounded set of yes/no answers. Finally respondents were asked an open-ended maximum WTP question. The full set of possible IB questions emanating from one DC starting level (here £100) is illustrated in Fig. 1. In this paper we confine our discussion of the IB procedure to the maximum WTP reported in the final IB question. [4]

Recent empirical studies allow the formulation of a number of hypotheses regarding the effects of using these various elicitation methods. [5] Given that respondents believe in the credibility of the hypothetical market presented to them (an assertion which is itself the subject of validity analysis throughout the paper), three potential outcomes are feasible given the OE approach:

i. Truth-telling; the respondent may state his/her true maximum WTP.
ii. Understatement; this may occur for differing reasons. If the respondent feels that personal payments may be linked to stated bids, but that the good in question is likely to be provided irrespective of this, a respondent may free-ride and "pretend to have less interest in a given collective activity than he really has" (Samuelson, 1954), thereby gaining a public-funded benefit for which he/she has a higher true WTP than the amount stated (Marwell and Ames, 1981; Brubaker, 1982). Alternatively, if respondents feel that costs will be shared on a per capita basis, then they will not wish to pay more than this amount and

[3] The questionnaire was designed to permit the following analyses: (i) categorisation and frequency of visit types (holiday, day trip, worker, resident); (ii) party and household composition; (iii) location of home/origin of trip; (iv) primary and subsidiary recreational activities; (v) subjective assessments of environmental quality; (vi) preferences regarding change in the natural/man-made Broadland environment; (vii) annual recreational/environmental budgets; (viii) WTP anything at all; (ix) annual WTP question (either OE or DC/IB); (x) length of commitment (years) to stated WTP sum; (xi) lump sum WTP questions (all OE); (xii) reasons for refusal to pay (where relevant); (xiii) predicted visitation/enjoyment under the flooded scenario; (xiv) income; (xv) Age; (xvi) membership of organisations. Most of the above are self-explanatory, however: (vii) was designed to check on potential mental accounting problems as well as raising respondents' recognition of income constraints; (x) checks whether respondents truly perceive WTP questions as being annual payments (as asked) or as once and for all lump sums; (xi) allows comparison with annual sum and calculation of implicit discount rate. In addition records of the interviewers (and sex), interview day (i.e., week/week-end effect), date/season effects and weather conditions were also analysed. Further details are given in Bateman et al. (1993).

[4] Analysis of the triple bounded dichotomous choice responses produced in the IB game is given in Langford et al. (1994).

[5] For a review see Mitchell and Carson (1989) or Bateman and Turner (1993).

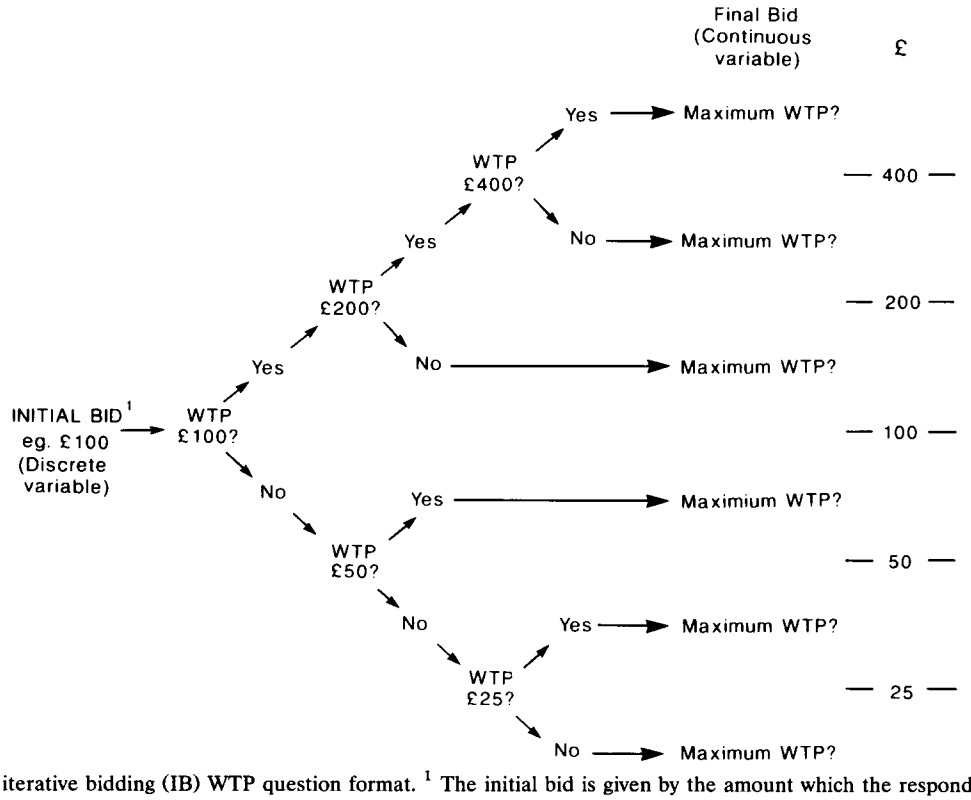

Fig. 1. The iterative bidding (IB) WTP question format. [1] The initial bid is given by the amount which the respondent is asked to pay in the dichotomous choice format. This amount is varied systematically across the dichotomous choice/iterative bidding sample.

will respond by stating the expected cost of the program if this is below WTP and zero otherwise. Furthermore, unfamiliarity with OE questions may lead respondents to adopt risk-averse strategies placing downward pressure upon stated WTP. In their theoretical analysis, Hoehn and Randall (1987) argue that, in an OE format, the respondents' lack of knowledge regarding aggregate costs and benefits gives no incentive for overstatement while CV surveys which provide imperfect information and place respondents under time constraints (both factors which are almost inevitable to at least some extent) are likely to result in understatement of true WTP. All of the above strategies therefore have some basis in economic theory.

iii. Overstatement; given a positive preference towards provision of the good in question, then if the respondent realises that the decision regarding provision depends upon mean WTP, he/she may overstate true WTP in an effort to raise mean WTP and thereby improve the likelihood of provision (Bohm, 1972). Such strategic behaviour has its roots in psychological rather than economic theory and requires an assumption that the individual does not believe that the stated bid corresponds directly to the actual payments.

While OE formats were common in early CV work, the DC method has become increasingly popular in recent years. Hoehn and Randall (1987) provide a theoretical framework to show that, given appropriate conditions (respondents believe that if plurality favours the project will proceed; that approval is conditional on a level of individual cost specified in the question; and that they will pay that specified cost), then an individ-

ual's optimal strategy is truth-telling within a DC questionnaire. Furthermore, Kriström (1990) points out that the DC take-it-or-leave-it approach more accurately reflects a market-decision, thus heightening respondent familiarity and accuracy in comparison with the OE method.

Economic theory therefore suggests that the DC approach may be superior to the OE method. However, psychological arguments have been put forward to suggest that the inherent characteristics of the DC approach induce bias into responses. Kahneman et al. (1982), among others, have argued that respondents faced with an unfamiliar situation (particularly where the good is also not well described) will interpret the bid level to be indicative of the true value of the good in question (Kahneman and Tversky, 1982; Roberts et al., 1985; Kahneman, 1986; Harris et al., 1989). Here the introduction of a specific bid level raises the probability of the respondent accepting that bid. This "framing" or "anchoring" effect may arise where a respondent has not previously considered his/her WTP for a resource (which is likely with regard to public or quasi-public goods) and/or is very unclear in their own mind about their true valuation. In such cases the proposed bid level may provide the most readily available point of reference onto which the respondent latches. There is no a priori presumption about the direction of such an anchoring effect. Positioning a bid-vector such that it has more bid levels on the upper tail of the true WTP distribution should lead to anchoring increasing mean WTP. Conversely, positioning the bid vector so as to emphasise the lower tail of the distribution should depress mean WTP. In this experiment it was decided to structure and position the bid vector in the light of information from the OE pilot. [6]

A further reason for anchoring behaviour was first described by Orne (1962) who proposed the existence of the phenomenon of the "good respondent". Here the interview relationship between analyst and respondent is portrayed as an interactive process in which the respondent seeks clues as to the purpose of the experiment. If this purpose is inadequately conveyed, whether because of poor survey design or inadequate interviewer training, then, Orne claims, respondents may attempt to give the analyst those answers which the respondent feels are wanted (i.e., tries to be "good"). Although "good respondent" and related problems (for example, "yea-sayers" or "nay-sayers") may afflict all the elicitation methods discussed in this paper, with regard to the DC approach, if the interviewer is held in high esteem by the respondent, the latter may feel that the DC bid level proposed is somehow "correct" and thereby respond positively to it (Harris et al., 1989). However, it should be stressed that "good respondent" effects are most generally confined to knowledge-based questions and may be less of a problem with attitude or behavioural intention surveys.

Two other psychological biases may operate to inflate WTP if the respondent does not fully believe the payment obligation. In cases where this is of a minor degree, some rounding-up of bids (from formulated to stated WTP) may occur, while in extreme cases strategic overbidding may occur as discussed above. Clearly, if disbelief in the payment obligation is widespread and severe, then the findings of the study will be invalid.

The DC approach is therefore not immune from potential bias problems. Given this situation and the possible problems inherent in the OE approach, we can see that simple bias tests, such as the comparison of means derived from the two formats, do not provide a very strong validity test. Nevertheless it is interesting to note that whilst Kealy et al. (1988) report no significant difference between OE and DC means, most comparisons (Sellar et al., 1985; Walsh et al., 1989; [7] Kriström, 1990) report DC means in excess of OE means, and none report the reverse result. This would suggest that either the sum total of DC biases serves to inflate WTP; and/or, that the OE ap-

[6] A related problem may occur where respondents feel that bid levels are unrealistically low or high, and therefore the instrument may not appear credible.

[7] Walsh et al. (1989) adopt a metal-analysis approach based on a number of studies.

proach results in a net underestimate of true WTP.

Expectations regarding the IB approach are, a priori, even less certain. The bidding procedure is begun at the particular DC bid level presented to the respondent in question. Therefore, if the DC method exhibits an upward anchoring effect we might expect that this is to some extent reflected in IB responses. However, the IB procedure, particularly with respect to the final maximum WTP bid, introduces elements of respondent control which are similar to those of the OE approach i.e., possibilities for over- or understatement of true WTP are introduced.

In summary, we can identify a variety of possible biases potentially affecting each elicitation format. The primary objective of this study was to examine the differences that actually occur as a result of utilising these different formats.

2.2. Truncation effects

A second objective of the study was to examine the impact of alternative sample truncation strategies upon estimates of mean WTP across different elicitation methods. In the case of the OE and IB (final bid) WTP data, truncation generally refers to the omission of a certain number or percentage of the highest bids. Such a strategy may be defensible where a respondent states a WTP bid which is unreasonably high given their disposable income. Such behaviour may occur where respondents exhibit mental accounting problems (an inability to recognise their many existing, and infinitely possible, purchases/donations given their finite income), are "yea-sayers", or are prone to strategic overstatement. However, such exclusion has been criticised by some commentators (Sagoff, 1988) who claim that such outliers represent "protest" bids made by respondents objecting to the valuation procedure either by refusing to give answers or, as in this case, by stating infeasibly high bids. We report the effects upon mean OE and IB WTP of a variety of truncation options and comment upon the latter.

When dealing with DC data, mean WTP is given by the expected value (E(WTP)) of the

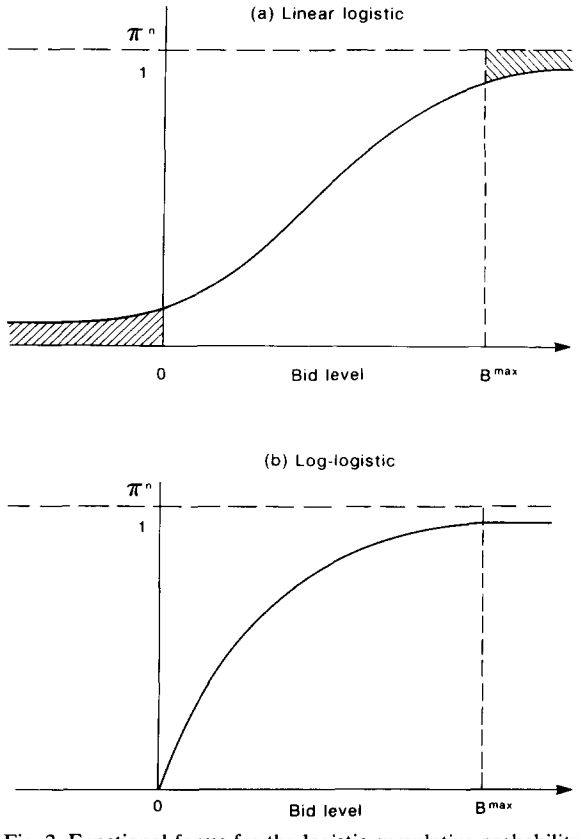

Fig. 2. Functional forms for the logistic cumulative probability distribution curve (source: Langford and Bateman, 1993). B_{max} = maximum bid level used in DC study π_n = probability of a "no" response.

cumulative probability distribution (CPD) curve linking the probability of a "no" response (π^n) to the level of the bid. [8] Langford and Bateman (1993) discuss in detail estimation of both logistic and log-logistic functional forms for the CPD and these are illustrated in Fig. 2. Here we consider three DC truncation options: [9]

[8] It is equal to the area enclosed by the CPD evaluated with respect to the bid level holding any other explanatory variables at their mean values.

[9] Further upper tail truncation options include the limit of respondents' relevant (disposable) income, and omission of a certain percentage of responses (where "yea-saying" is suspected).

i. All real numbers. This follows the reasoning of Johansson et al. (1989) that, given the observed portion of the CPD ($0 \leqslant B_i \leqslant B_{max}$ in panel (a) of Fig. 2), it is reasonable to extend the limits of integration both forwards to ∞ and backwards to $-\infty$. The argument for extending to ∞ is that some respondents are still WTP more than the upper bid level (B_{max}). The argument for extending integration backwards to $-\infty$ is that some respondents were not WTP the lowest bid level and so it is arguable that some of these would, rather than pay for an increase in provision, prefer to receive an increase in their income in exchange for a reduction in provision. We denote E(WTP) for $-\infty \leqslant B_i \leqslant \infty$ as C^* (note that, for a log-logistic functional form C^* is a measure of median WTP only, evaluated from 0 to ∞). [10]

ii. Non-negative, truncated; Sellar et al. (1986) argue that integration of the mean should be truncated to the limits of observable data, the main argument being that beyond the upper bid level extrapolation is dependent upon the distributional assumption being made. [11] We denote E(WTP) for $0 \leqslant B_i \leqslant B_{max}$ as C^{**}.

iii. Non-negative, untruncated; Hanemann (1984) argues against truncation at the upper bid level noting that a "yes" response does not indicate a maximum WTP, but rather a lower bound on that maximum. Evaluation to ∞ allows valid extrapolation to higher amounts which certain individuals would be WTP. Furthermore, in subsequent work Hanemann (1989) argues against integration across negative bid levels, noting that WTP studies are poor approximators of the willingness to accept (WTA) compensation amounts implicit in such an integration. [12] By confining the integral to $0 \leqslant B_i \leqslant \infty$, we are estimating mean WTP for a specified gain in provision. This does not tell us about the welfare change implications of any reduction in provision. We denote E(WTP) for $0 \leqslant B_i \leqslant \infty$ as C^{***}.

The impacts of all three DC truncation options upon estimates of mean WTP are reported and contrasted with those obtained from the various truncations of OE and IB data.

3. The pilot study

In order to test whether the questionnaire was adequate, a face-to-face, on-site pilot survey of 433 respondents was undertaken prior to the main study. [13] Pilot study findings resulted in a number of minor adjustments to the questionnaire, mostly changes in the wording of some questions. However, two issues of major interest were highlighted: firstly, what would be the most appropriate payment vehicle, and secondly, what would be the most suitable number and level of bids for use in the DC experiment. Following Boyle and Bishop (1988), it was decided that, in the absence of any a priori expectations, the pilot survey should be undertaken using an OE approach and that bid levels for the DC experiment should be based upon those received in the OE pilot.

Three payment vehicles were employed in the pilot survey:

1. An unspecified charitable donation (DONATE).
2. Payments to a specified charitable fund set up to facilitate flood defence work in Broadland (FUND).
3. Payments via direct taxation (TAX).

Other alternatives were considered to lack realism. In particular, entrance fees were not thought to be credible because of the nature of the resource (large area; considerable resident population; no UK precedent).

The DONATE vehicle suffered disproportionately from zero WTP bids (46.5%) compared to either of the other vehicles. It was felt that the vague definition of this vehicle led respondents to be uncertain that their donations would be effec-

[10] See Langford and Bateman (1993).

[11] C^{**} is in fact evaluated using a normalised truncated bid function following Boyle et al. (1988).

[12] See also Hanemann (1991).

[13] Full details in Bateman et al. (1993).

tively used. In short the vehicle did not engender credibility in the hypothetical market and was therefore rejected. The FUND vehicle also performed badly in terms of a high zero bid rate (23.1%) and also produced much higher bid variation than other vehicles. The FUND vehicle was therefore also rejected. The TAX vehicle produced by far the lowest zero-bid rate (11.8%) and also performed better in terms of bid variability than the FUND vehicle and about as well as the DONATE vehicle. It was also supported by the fact that, if flood defence works were to be built, such works would in reality be paid for out of taxes rather than trust-fund donations. The TAX vehicle therefore had the advantages of realism and immediate applicability and was adopted for the main survey.

All respondents were asked why they had responded in the way they had. Many of those presented with the FUND and DONATE vehicles commented that they were not confident that payments via such vehicles would be fully channelled towards preservation work (trust funds were not to be trusted!). Furthermore, many of those responding to the TAX vehicle commented that, while they disliked paying extra taxes, they had confidence that such money would be spent efficiently upon any flood defence scheme. The different payment vehicles therefore seem to have differing perceived probabilities that the good will be provided (Mitchell and Carson, 1989) and, as expected, as this probability falls so does WTP.

The main OE survey was begun in early August 1991. However, it was decided to suspend the start of the parallel DC/IB survey until sufficient OE responses had been collected using the finalised questionnaire to provide an adequate basis for determination of the DC bid-vector. If there was no elicitation effect related to changing from the OE to the DC approach, then any respondent stating a particular maximum WTP of £X in an OE survey would logically have responded negatively to a DC WTP question with a bid level higher than £X. In such a situation the probability of accepting various DC bid levels can be predicted on the basis of OE data. If a subsequent actual DC experiment yields results which are significantly different from this prediction, then, given that we have not incurred any sampling bias, such a difference is indicative of some bias in one or both of the elicitation methods. DC bid levels were eventually set on the basis of the first 427 OE responses collected. As was expected, the actual OE bids received were not smoothly nor normally distributed, but grouped around certain category values with a significant positive skew. Given this, it was decided to fix the DC bid levels in the following way:

i. The use of technically precise amounts such as £3, £7, £12, etc. was considered likely to cause uncertainty and confusion amongst respondents. Category values such as £5, £10, £20, etc. were considered to be those most readily acceptable and relevant to respondents.
ii. In order to aid accurate positioning of extremities on the cumulative probability distribution curve, the lower DC bid was chosen so as to be almost universally acceptable. Given constraint (i), this was fixed at £1.
iii. Similarly, the upper bid level was chosen so as to be almost universally rejected. Based upon initial OE responses, this was fixed at £500.
iv. The number of bid level categories was influenced by the following criteria:
    a. Expected sample size; this should not only permit analysis across bid levels, but also be divided between bid levels such that robust within-bid level analysis is feasible.
    b. The number of bid levels and their positioning should enhance robust analysis of the cumulative probability distribution curve.

Given that a parametric estimator is to be used, there is a clear tradeoff between the number of bid amounts to be chosen and the resulting precision of the estimates to be obtained. In the event, given the above criteria, eight bid levels were chosen (distributed so as to reflect the observed positive skew): £1, £5, £10, £20, £50, £100, £200 and £500.

4. Factors enhancing study validity

Given the ongoing nature and tone of the academic debate regarding the CVM, we claim

only to have addressed certain limited issues using a large sample and a policy-relevant environmental issue. We have attempted to minimise bias wherever possible recognising that minimisation is not the same as reduction to insignificance. However, the study does conform to the main guidelines of the blue ribbon panel's testing protocol, and we would highlight several factors which do appear to enhance the relative validity of this particular application.

Respondents are likely to have possessed very high prior information levels regarding the resource. Whilst only 16% were local residents (or were working in the area) the vast majority of visitors were on a repeat trip (72%). Furthermore, the majority of visits were holidays rather than day trips, indicating that respondents understood the nature of the good under investigation.

Extensive piloting preceded a large sample main survey. Although adequate sample size is a necessary factor for any valid social survey, [14] CVM studies in the UK to date have been characterised by small sample size. [15] Indeed, to our knowledge, this study constitutes the largest of its kind carried out in Europe to date.

It may be that the character of the area itself led to a minimisation of the embedding problems highlighted by Kahneman and Knetsch (1992). The Norfolk Broads is in many ways a unique resource. Its landscape, ecology and many of its recreational facilities have no adequate national substitute. This means that the confusion of part and whole described by Kahneman and Knetsch is less likely to be a problem; in effect the part and the whole are one. [16]

Finally we believe that the scenario presented to respondents was easily understood and highly believable. The low-lying nature of the area, its proximity to the sea and the readily observable dilapidation of the existing flood defences make the possibility of flooding appear realistic. The feasibility of a technical solution to such a problem is also high since the prevailing character of the area is due to existing flood defences. Furthermore, in the UK context, the only realistic agency for such an undertaking would be one which was centrally funded from taxes. We believe that these factors combine to considerably enhance the credibility (and therefore relative validity) of this study.

5. The OE experiment: Results and discussion

In both the OE and DC/IB surveys, respondents were asked, prior to the WTP question, whether they would be in favour of increased personal taxation in order to preserve the Broads. The main objective of such a question was to validate a zero WTP response. It was recognised that if this was omitted from the questionnaire then asking people directly to state a WTP sum may make respondents feel obliged to state some non-zero amount. [17] For the OE sample, factors such as the respondents income; age; visitor type (on holiday, daytripper, resident, worker); interest in boating; first time/repeat visitor, etc., were all shown to be insignificant in determining answers to this "will/won't" pay question. However, two factors were significant predictors (at the 99% confidence level) of a respondent agreeing to pay namely: (i) where the respondent identified "relaxing/enjoying scenery" as the main reason for the visit and; (ii) where the respondent was a member of an environmental group.

In total 862 interviews were completed using the OE elicitation method. Of these some 131

[14] See discussion in Mitchell and Carson (1989).

[15] For a review of UK studies, see Turner et al. (1992).

[16] A related aspect of this problem, the inability of respondents to recognise other competing demands upon a finite recreational/environmental budget, was addressed by asking respondents to calculate this budget and take it into consideration when determining WTP. In both the OE and IB experiment WTP was approximately 16% of the budget, an amount which appears reasonable (although we admit that this does not constitute incontrovertible proof that no embedding problem exists).

[17] However, some respondents with a low but non-zero WTP may reply negatively to the "WTP anything at all?" question if they feel that their true WTP is well below the level required for provision. Such respondents will consequently be treated as a spike at zero, but such an approach conforms to the NOAA panel's recommendations 6 and 12 reported at the start of this paper.

Table 1
Truncation effects – open-ended WTP study [a]

No. of upper tail truncated	0	1	8	42	84	126	168	211
% of upper tail truncated	0%	0.1%	1%	5%	10%	15%	20%	25%
N	846 [b]	845	838	804	762	720	678	635
Mean WTP [c]	67.19	65.79	60.89	46.76	37.38	32.57	28.39	25.54
Median WTP	30.00	30.00	30.00	25.00	25.00	20.00	20.00	12.00
St. dev.	113.58	106.10	90.08	55.19	38.64	33.69	30.10	24.41
S.E. mean	3.91	3.65	3.11	1.95	1.40	1.26	1.16	0.97
Maximum bid [d]	1250.00	1000.00	500.00	250.00	150.00	100.00	100.00	100.00
Lower quartile	5.00	5.00	5.00	5.00	5.00	2.13	2.00	1.00
Upper quartile	100.00	100.00	100.00	60.00	50.00	50.00	50.00	50.00

[a] All rows, except the upper three, are measured in £.
[b] Total sample of 862 interviews included 16 incompleted questionnaires (omitted from calculation of mean).
[c] Includes, as zeros, those who refused to pay anything at all.
[d] Minimum bid = zero throughout.

respondents answered "no" to the "will/won't" pay question. All such respondents were then asked to state why they had given such an answer. The most common reasons for non-payment was related to income and existing commitments (almost 40% of non-payers, equivalent to 6% of the total OE sample) followed by the pure free-rider reply that, although the area was valued, someone else (e.g., the government) should pay (almost 25% of non-payers, equivalent to 4% of the total OE sample). Whilst income constraints pose no problem here, the free riding effect does point to a possible downward bias in the OE estimate of WTP. More importantly this small group of extreme free-riders may indicate the existence of a larger group of respondents who, whilst still stating some non-zero sum, nevertheless reduced their stated WTP below true WTP as a result of the free-rider incentive. However, attempts to quantify such a strategy would have required a significant extension to the questionnaire (and possibly laboratory-type controls) and were consequently not undertaken. [18]

[18] It is interesting to note that recent reviews have indicated that free-riding behaviour may result in a reduction of stated WTP to (very approximately) between 60–95% of true WTP, depending upon the strength of the incentive to free ride. See chapters 6 and 7 of Mitchell and Carson (1989) and Milon (1989).

Evidence of "protest bidding", in the sense of a refusal to participate in the valuation process, was conspicuously absent. The possibility of such a response was directly catered for by listing a refusal to value the Broads as an explicit option amongst reasons for refusing to pay. However, only 30 respondents (1% of the total OE + DC/IB sample) gave this as their reason for refusal. This finding strongly contradicts the assertion by some commentators that CVM studies are pervasively invalidated by the prevalence of protest bids (Sagoff, 1988).

Alongside evidence of free-riding, other respondents in the OE sample appeared to exhibit strategic overbidding. Table 1 details univariate OE WTP statistics for a variety of upper tail truncation points. Mean WTP for the entire OE sample of 846 respondents was £67.19 (95% CI = £59.53–£74.86). However, omission of just the single highest bid (0.11% of the OE sample) caused OE mean WTP to fall to £65.79 (a reduction of over 2%). Similarly, truncating the top 1% of bids causes a reduction in the mean of nearly 10%. In themselves such statistical effects are not conclusive proof of strategic overbidding as a skewed distribution may simply reflect the socioeconomic and preference characteristics of the sample. However, upon inspection it was found that the sums stated by those at the upper tail of the bid distribution appeared infeasible given the

ability of these respondents to pay. Several of the highest bidders stated WTP sums which exceeded their entire annual expenditure upon all recreational and environmental goods (in some cases by a factor of 5). We therefore conclude that there is strong evidence for a degree of strategic overstatement by a small number of respondents in the OE experiment.

Validity testing was applied to all three elicitation methods following the criteria set out by Mitchell and Carson (1989). Content validity was, in the main, carried out prior to the survey and consisted of a number of meetings with recognised authorities in the fields of economics, marketing, social surveys and psychology. These consultancies addressed all aspects of the study with particular emphasis on the design of the questionnaire, associated information and survey sampling strategy. Criterion validity testing (comparison with actual WTP for the good) was not feasible and therefore a major effort was made to establish construct validity (i.e., testing whether results conformed to expectations). One simple approach, comparing mean WTP with that of other studies (convergent validity), was only feasible for the OE study as other formats have had few applications in the UK to date. Results from the OE experiment were contrasted with those from 28 comparable UK use-value studies. [19] This analysis showed results to be logically related according to two factors:

i. the number of adequate substitute sites available;
ii. the magnitude of the proposed change in provision.

Most other UK studies have looked at sites with some or many substitutes, facing relatively marginal changes in provision. Accordingly, the fact that this study estimates a mean WTP value higher than most others seems logically correct.

The theoretical validity of OE responses was examined via estimation of the bid function. A full range of explanatory variables was investigated. Functional form was a priori uncertain (although linear forms were theoretically undesirable), but an initial analysis indicated that a high degree of overall explanation was unlikely to be achieved (a characteristic of OE studies). Therefore detailed (e.g., Box-Cox) analysis of functional form was by-passed in favour of using standard forms. The best model was provided by a double log form which is reported in Eq. 1. This model narrowly outperformed a semi-log (dependent) form which contained the same explanatory variables.

$$\underset{}{\text{LWTP(OE)}} = \underset{(0.22)}{0.1934} + \underset{(3.32)}{0.2920}\ \text{LINC} + \underset{(4.15)}{0.2695}\ \text{RELAX} + \underset{(3.93)}{0.2473}\ \text{ENV} \quad (1)$$

where: LWTP(OE) = Natural log of open-ended WTP response; LINC = Natural log of respondents income (continuous variable); RELAX = 1 if respondent often visits area to relax/enjoy scenery (= 0 otherwise); ENV = 1 if respondent is a member of an environmental group (= 0 otherwise). $R^2 = 5.29\%$; total d.f. = 800. [20] Figures in brackets are t statistics.

The explanatory variables given in Eq. 1 are all significant at the 99% level, while no further variables were significant at even the 95% level. The major feature of this "best model" is its very poor overall degree of explanatory power, [21] which although more extreme than usual in this case, is a characteristic trait of many OE studies. Therefore, while the logical ordering of mean responses observed across studies indicates that economic theory is adequate to explain results at such a level, the poor performance of the model given in Eq. 1 suggests that further consideration of the motivations underlying individual responses is required here (see below).

[19] Full results are given in Bateman et al. (1994).

[20] Eq. 1 omits all responses for which information on any explanatory variable was missing.

[21] To ensure that no errors had been made, statistical analysis was carried out independently at the University of East Anglia and at the University of Newcastle-upon-Tyne. Both analyses confirm the weak explanatory power of the "best" model.

6. The DC experiment: Results and discussion

As with the OE survey, those interviewed using the DC/IB questionnaire were asked, prior to the WTP questions, whether or not they were willing to pay any extra taxes. In total 240 of the 2070 DC/IB respondents answered "no" to this question (11.6%). Tests showed there to be only one significant predictor of a positive response to this question, namely, membership of an environmental group. [22] All respondents who refused to pay any extra taxes were asked to specify a reason why. As before, the most common reasons involved income constraints and existing commitments (33% of non-payers; 3.9% of the total sample) closely followed by the pure free-riding response (31.7% of non-payers; 3.7% of the total sample). [23] Analysis failed to reveal any significant factors determining the reason for non-payment.

Those respondents who indicated that they were prepared to pay at least some amount were then asked to pay one of the bid levels, selected at random. The mean DC WTP is calculated by integrating the CPD function between appropriate truncation limits. Accurate estimation of the bid function is therefore vital because an incorrectly fitted function will give a spurious estimate of the mean. Both linear and log models were tested using both logit and probit link functions. Log models gave a markedly better fit than linear specifications. The choice between link functions was more difficult as both logit and probit approaches performed similarly well. [24] However, a log-logistic model gave a marginally better fit and as this has been used extensively elsewhere, it was preferred for further analysis. In all cases the most remarkable feature of the estimated models was the very high explanatory power of the bid level in determining WTP response. Eq. 2 presents the log-logistic model resulting from the single explanatory variable LBID; the natural logarithm of the bid level (£) presented to respondents.

$$\underset{}{\text{LOGIT}}\ (\pi_i) = \underset{(-19.74)}{-4.932} + \underset{(18.39)}{0.9939}\ \text{LBID} \qquad (2)$$

Deviance change = −594.4; residual deviance = 1325.7; d.f. = 1624. Figures in brackets are t values.

Where:

$$\text{LOGIT}\ \pi_i = \ln\left[\frac{\pi_i}{1-\pi_i}\right]$$

π_i = probability of an individual saying "no" to the bid level. [25]

As can be seen from Eq. 2, a log-logistic model with the single explanatory variable LBID fits the dichotomous choice dataset extremely well. [26] Further explanatory variables were then added to this model in an attempt to improve the fit. [27] The best log-logistic model is given as Eq. 3.

$$\begin{aligned}\text{LOGIT}\ (\pi_i) &= \underset{(-6.23)}{-3.736} + \underset{(18.40)}{1.026}\,\text{LBID} - \underset{(-1.34)}{0.0907}\ \text{LINC} \\ &\quad \underset{(-3.35)}{-0.5888}\ \text{BOAT} - \underset{(-2.58)}{0.3756}\ \text{RELAX} \\ &\quad \underset{(-2.22)}{-0.3126}\ \text{ENV} \qquad (3)\end{aligned}$$

Deviance change = −622.9; residual deviance = 1297.2; d.f. = 1620. Figures in brackets are t values.

[22] Significant at $\alpha = 1\%$. No other significant factors at $\alpha = 5\%$.

[23] Note that as this question was asked prior to any WTP question, this response does not refute the earlier suggestion that DC formats may inhibit free-riding.

[24] Full details in Bateman et al. (1993).

[25] Readers should be aware that this means that "positive" relationships (e.g. between WTP and income, etc.) will have a negative sign and vice versa.

[26] As an ancillary test of this result, individual models were fitted for the data within each bid level ($214 \leqslant n \leqslant 227$ for each level). All of the eight models produced were exceptionally weak. This is inevitable for the lower bid levels where very few respondents registered refusals i.e., very little variation. However even the best of these models (for the £50 bid level) only recorded a change in deviance of −24.65 with residual deviance being 223.73. These results confirm the key role of the bid level in determining responses.

[27] Alternative models are considered in Bateman et al. (1993).

Table 2
Mean WTP estimates [a]: log-logistic models (£/annum)

Truncation option	Single variable: Eq. 2	Full model: Eq. 3
C**	112 (68–168)	111
C***	144 (75–261)	140

[a] Figures in brackets represent 95% confidence intervals around the mean (not calculated for full models). Full details of all models are given in Bateman et al. (1993). Truncation options as specified previously.

Where
LINC = Natural logarithm of respondents household income (continuous variable); BOAT = 1 if respondent does participate in some boating activity (= 0 otherwise); RELAX = 1 if respondent visits area to relax/enjoy scenery often (= 0 otherwise); ENV = 1 if respondent is a member of an environmental group (= 0 otherwise); other variables as previously defined.

Although not significant at $\alpha = 5\%$, the variable LINC is included in Eq. 3 to underline the finding that, although complying with expectations, the respondents income plays a very weak (statistically insignificant) part in determining response to dichotomous choice WTP questions. While economic theory would lead us to expect the "price" variable (LBID) to be the most significant, its degree of dominance over other variables, particularly income, is of interest. We comment on these findings subsequently.

Table 2 presents mean WTP calculated for both single variable and best fit versions of the log-logistic models. [28] Means are reported for the C** and C*** truncation options discussed in Section 2 of this paper. C* and C*** are essentially the same for the log-logistic model, and so C* has been omitted from these results (see Langford and Bateman (1993) for further details).

[28] The importance of choosing the correct functional form is underlined by the result that mean WTP for comparable (but poorer fitting) linear logistic models, ranges from £238 to £256 (full details in Bateman et al., 1993). Note that, so as to ensure comparability with reported OE means, all those who refused to pay anything at all were included as zeros (randomly distributed across bid levels) in the calculation of DC means.

95% confidence intervals were generated by a Monte Carlo simulation of the bivariate normal distribution of the intercept and slope (a and b) of the regression equation (Langford and Bateman, 1993). For a discussion of estimating variance of WTP around the point estimate of mean or median, see Duffield and Patterson (1991), and Langford (1994).

Table 2 shows that means for each full model are very similar to those for corresponding single variable models, indicating that explanatory variables other than bid level have relatively little influence upon the mean. However, choice of truncation points clearly does have an impact upon estimates of the mean. While the upper truncation point is ∞ for C*** (and C*), it is set equivalent to the upper bid level for C** and can therefore vary across studies if different maximum bid levels are used. The impact of changing the maximum bid level was investigated by comparing estimates of C** and C*** based upon Eq. 2 and calculated across the whole dichotomous choice sample, to those calculated when respondents answering the £500 bid level question were omitted (i.e., changing the maximum bid level and C** truncation point to £200), and again when respondents answering the £200 bid level question were omitted (i.e., maximum bid level = C** truncation point = £100). Results are given in Table 3 below.

As can be seen, while changes to the maximum bid level have little impact upon C***, the C** measure is highly responsive and must therefore be treated with some caution. Hence the authors' preference is for the C*** measure as derived from the full log-logistic model, i.e., mean WTP = £140 as estimated from Eq. 3.

Table 3
Mean WTP measures: impact of changing the maximum bid level/truncation point in Eq. 2

Maximum bid level and C** truncation point	Truncation option	
	C**	C***
£500	112	144
£200	84	150
£100	59	142

7. Comparing OE and DC results

Do our findings from the OE and DC experiments conform to economic theory or is there evidence of the psychological biases discussed in Section 2 of this paper? The most common assessment of elicitation effects has been through comparison of means and our study confirms the general finding of DC mean exceeding that from the OE experiment. However, this result above gives little indication regarding the validity of either approach. Indeed, as Kriström (1993) points out, even if means were the same, this need not imply similarity of distribution.

Fig. 3 presents both DC and OE response distributions in the form of survival functions for those WTP at least some amount (i.e., excluding, for both formats, all those respondents who refused to pay anything at all). Here the proportion of DC respondents giving positive responses at each bid level is compared with the proportion of OE respondents stating WTP sums equivalent or greater than that bid level. In the absence of any elicitation effects these proportions should roughly coincide across the bid vector. However, we can see that the DC format apparently generates a response distribution which is shifted outwards compared to that of the OE approach. [29]

Fig. 3 suggests that it is more likely that a respondent will agree to pay a particular amount X when presented with that amount as a DC bid level, rather than via an OE experiment. This discrepancy can be viewed from either an economic or psychological perspective. Economic theory suggests that the OE format provides no incentive for overstatement (Hoehn and Randall, 1987) and may be subject to free-riding or expected cost effects, both of which will give an incentive to understate WTP. Furthermore, given the necessary assumptions (discussed previously),

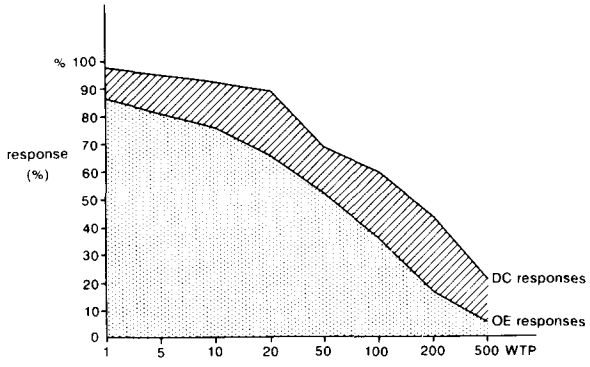

Fig. 3. Survival functions for OE and DC responses.

truth-telling is the optimal strategy in a DC format (ibid). The observed discrepancy between OE and DC results is therefore not inconsistent with economic theory. However, such results can also be explained in terms of certain of the "psychological" biases discussed earlier. Here, commentators have seen OE/DC divergence as evidence of some sort of anchoring effect in the DC responses and we can use this as a generic term for the overall effect of the various potential DC biases identified.

Testing for such anchoring is problematic. Kriström (1993) discounts comparisons based upon means because of their implicit assumptions regarding distributional form. He suggests the use of non-parametric tests based upon the distance between the OE and DC survival functions. However, there are various complications associated with the simultaneous application of such an approach to discrete and continuous data. Kriström therefore uses a simple chi-square test to show that the OE and DC responses do not come from the same distribution. This is supplemented by a somewhat unusual test of an anchoring hypothesis in which responses from the OE sample are compared with supplementary OE responses given by the DC sample. Although, as Kriström states, the distance test used is known to have low power, [30] it is interesting to note that the computed statistic actually rejects the anchoring hypothesis ($\alpha = 5\%$).

[29] It is interesting to note that this figure is very similar to that reported by Kriström (1993) in his study of preservation values for Swedish forests. This similarity would be even greater if we were to extend our WTP axis to include OE strategic overbidders/yea-sayers and their expected DC counterparts had even higher bid levels been used.

[30] The Kolomogorov-Smirnoff test; see Kanji (1993).

However, comparative analysis can be used to illustrate the strength of the apparent cognitive difference between responses to OE and DC questions. In order to underline this difference it was decided to treat all the OE responses as if they had come from DC questions, i.e., an OE WTP bid of £10 was taken as a "yes" response to bid levels £1, £5 and £10 and a "no" response to all others. Here the (optimal) log-logistic model gave a very much poorer fit to the data (residual deviance of 227.72 with 6 degrees of freedom) than it did for the genuine DC data (residual deviance of 6.24). This indicates that the format used in DC questioning places respondents within a fixed framework of evaluating their WTP, very well described by the log-logistic model, whereas OE respondents appear to be undergoing significantly different cognitive processes in formulating and stating their responses. [31]

8. The IB experiment: Results and discussion

All respondents initially presented with a DC WTP question were then entered into the IB bidding game. Discussion of "willingness to pay anything at all" and "reasons for refusal" for the IB sample are therefore as for the DC experiment.

The open-ended WTP question presented at the end of the IB procedure gave a mean WTP of £74.91 (95% CI = £69.27; £80.55). However, as in the OE experiment, this amount was highly responsive to the truncation of higher WTP bids. [32] Omission of the upper 5% of responses, for example, resulted in a 30% decline in the mean to £52.41. As in the OE experiment then, the possibility that certain respondents engage in strategic overbidding cannot be ruled out.

As the final WTP bid in the iterative bidding game was given in response to an OE question, the dependent variable in any bid curve estimation will be continuous, but truncated at zero. Bid function analysis quickly revealed a strong positive association between the DC bid level presented to respondents (which constituted the starting point of the IB game) and the final WTP amount stated by respondents at the end of the IB process. This relationship was strongest when both the final WTP response and the initial bid level were expressed as natural logarithms. The optimal model is given as Eq. 4.

$$\begin{aligned}\text{LWTP(IB)} &= \underset{(22.18)}{2.104} + \underset{(19.79)}{0.3733}\ \text{LBID} + \underset{(1.86)}{0.000005}\ \text{INC} \\ &\quad + \underset{(3.67)}{0.1758}\ \text{BOAT} + \underset{(4.70)}{0.1720}\ \text{ENV} \\ &\quad - \underset{(-2.89)}{0.1222}\ \text{FIRST} \qquad (4)\end{aligned}$$

$R^2 = 21.86\%$; total df = 1634. Where: LWTP(IB) = natural log of respondents final WTP statement in the IB game; INC = respondents household income (continuous variable); FIRST = 1 if respondent is on his/her first visit to the area (= 0 otherwise). Other variables as previously defined.

Signs on the explanatory variables of Eq. 4 are as expected. The variable INC is included for interest although it is only significant at the 90% confidence level. Interestingly when tested, the variable LINC was found to be significantly weaker. As expected, ignoring the constant, by far the most powerful explanatory variable was the (log) bid level first presented to respondents. This appears to have strongly anchored respondents into a corresponding range of final WTP bids, i.e., a classic starting point effect.

In their theoretical analysis, Hoehn and Randall (1987, p. 237) appear to imply that DC and IB approaches, when started from identical bid levels, should yield similar mean results. Clearly this has not occurred in this case. Our IB format can be viewed as an amalgam of the DC and OE approaches and as such it is not surprising that we see evidence of several of the characteristics of those formats reflected in IB responses. The power of the initial bid level, so dominant in the DC bid functions, is clearly apparent. However, the IB approach now allows for OE "understatement" traits such as free-riding or expected-cost

[31] Another approach might be to compare common covariates in the OE and DC bid functions.

[32] Full results in Bateman et al. (1993).

strategies to emerge as reflected in the reduced estimate of mean WTP.

A further characteristic of the IB data was identified by comparing responses as they developed across the first, second and third dichotomous bounds. Here it was noticed that, *at all bid levels*, respondents exhibited a certain unwillingness to accept a doubling of a previously accepted amount. This trend was, to varying extents, apparent whether that previous amount was £10 or £500, and continued to appear (and intensify) at successive bounds. [33] At the second (or third) bound respondents appear to view their previous response as more or less representing their total WTP and therefore resist the further doubling of this amount.

8. Summary and conclusions

This analysis was undertaken in order to examine both the effect of varying elicitation formats and different truncation strategies on the willingness to pay results derived from a CVM study of proposed wetland protection expenditure in East Anglia.

Our results show that the precise strategy that is adopted in order to truncate response data ranges will have a significant impact on mean WTP estimates across all elicitation formats. In the case of the OE and IB formats, we have shown that truncation analysis can be utilised as a preliminary test for severe strategic overbidding. In order to guard against possible charges of protest bidding exclusion, we recommend that a sensitivity analysis of several truncation strategies (including no truncation) be undertaken and reported in all OE and IB-based studies. With respect to DC format studies we recommend a truncation approach which eliminates the negative sums implied by the bid function, and integrates beyond the upper accepted bid level to some logical limit (e.g., income constraint or ∞).

In the context of elicitation formats our analysis confirms that while there are some similarities across formats (i.e., optimal bid functions contain a number of common explanatory variables; and the confidence intervals of the mean WTP estimates also overlap) there are more dominant differences. The possibility of conflicting effects such as free-riding and strategic overbidding, as well as considerable uncertainty (high variability) within OE responses, seems likely. However, many, if not all, of these characteristics can be accounted for within economic theory.

The disparity between OE and DC results might also be explained by economic theory. However, psychological arguments may also be valid here. The large influence of the bid level within the DC bid function can be interpreted either as an expected economic price effect, or as an anchoring bias. Because of the number of potential exacerbating, conflicting and confounding effects discussed with respect to both formats, we have doubts about the usefulness of simple comparisons between OE and DC results. Rather, we choose to emphasise the results of the test which treated OE data as if it were derived from DC questioning. This indicates that there is a highly significant cognitive response difference depending on which question format is being used. These differences in interpretation appear to indicate that the mental processes initiated by these questions include certain quite separate elements, probably both economic and psychological.

The IB approach can be seen as a hybrid of both the DC and OE formats and as such demonstrates a mix of the effects associated with both. The dominance of the bid level, so characteristic of the DC approach, is clearly evident as a classic starting point bias (Roberts et al., 1985). It appears that, once the initial (DC) response is elicited, the ensuing respondent control may engender OE-type "understatement" strategies.

In conclusion, we pose the question, does the presence of elicitation effects invalidate the use

[33] This means that, at a given bid level, we are likely to have a lower proportion of recorded refusals at the first bound than at the second (because those doubling up from an initial lower bid level may refuse to pay this amount). This will mean that the discrete variable estimate of mean WTP will fall between bounds. This result accords with the findings of Hanemann et al. (1991). Further details are given in Langford et al. (1994).

of the CVM? We would argue, quite strongly, that they may not. Many commentators have seen differences as indicating that one method is flawed (usually the OE) while another is both reliable and valid (usually the DC). We have attempted to highlight the fact that, in general, the results obtained are not inconsistent with economic theory. In particular the analysis of Hoehn and Randall (1987) has apparently been born out in the observed OE and DC findings. However, we have also acknowledged that competing psychological arguments can often not be ruled out by the same findings. The decisionmakers' question is likely to be more one of degree: if "psychological biases" were adding to expected economic effects, do they do so to such an extent as to preclude the use of CV in a real-world policy setting? Given that much decision making is undertaken without any reference to the value of the public good impacts involved, then it seems to us that these evaluations must be of significant and real informational value. The task for future research in this area must be to refine our understanding of, and control for, these elicitation effects.

Acknowledgements

The Centre for Social and Economic Research on the Global Environment (CSERGE) is a designated research centre of the UK Economic and Social Research Council (ESRC). This research paper was co-sponsored by ESRC Grant Number W119-25-1014. The authors wish to warmly thank an anonymous referee for a series of helpful suggestions for improving the paper. Remaining errors are, of course, our responsibility.

References

Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R. and Schuman, E.H., 1993. Report of the NOAA Panel on Contingent Valuation. Report to the General Counsel of the US National Oceanic and Atmospheric Administration, Resources for the Future, Washington, DC.

Bateman, I.J. and Turner, R.K., 1993. Valuation of the environment, methods and techniques: the contingent valuation method. In: R.K. Turner (Editor), Sustainable Environmental Economics and Management: Principles and Practice. Belhaven Press, London.

Bateman, I.J., Willis, K.G., Garrod, G.D., Doktor, P., Langford, I.H. and Turner, R.K., 1992. Recreation and environmental preservation value of the Norfolk Broads: a contingent valuation study. Environmental Appraisal Group, University of East Anglia.

Bateman, I.J., Langford, I.H., Willis, K.G., Turner, R.K. and Garrod, G.D., 1993. The impacts of changing willingness to pay question format in contingent valuation studies: An analysis of open-ended, iterative bidding and dichotomous choice formats. Global Environmental Change Working Paper 93-05, Centre for Social and Economic Research on the Global Environment, University of East Anglia, Norwich and University College London.

Bateman, I.J., Willis, K.G. and Garrod, G.D., 1994. Consistency between contingent valuation estimates: a comparison of two studies of UK National Parks. Reg. Stud., 28: 457–474.

Bishop, R.C. and Heberlein, T.A., 1979. Measuring values of extra-market goods: are indirect measures biased? Am. J. Agric. Econ., 61: 926–930.

Bohm, P., 1972. Estimating demand for public goods: an experiment. Eur. Econ. Rev., 3: 111–130.

Boyle, K.J. and Bishop, R.C., 1988. Welfare measurements using contingent valuation: a comparison of techniques. J. Am. Agric. Assoc., February: 20–28.

Boyle, K.J., Welsh, M.P. and Bishop, R.C., 1988. Validation of empirical measures of welfare change: comment. Land Econ., 64(1): 94–98.

Brookshire, D.S., Eubanks, L.S. and Randall, A., 1983. Estimating option price and existence values for wildlife resources. Land Econ., 59: 1–15.

Brubaker, E., 1982. Sixty-eight percent free revelation and thirty-two percent free ride? Demand disclosures under varying conditions of exclusion. In: V.K. Smith (Editor), Research in Experimental Economics, Vol. 2. JAI Press, Greenwich, Ct.

Cambridge Economics, 1992. Contingent Valuation: A Critical Assessment. Cambridge Economics Inc., Massachusetts.

Carson, R.T., Mitchell, R.C., Hanemann, W.M., Kopp, R.J., Presser, S. and Ruud, P.A., 1992. A contingent valuation study of lost passive use values resulting from the Exxon Valdez oil spill. Report to the Attorney General of the State of Alaska, Westat Inc., Rockville, MD.

Duffield, J.W. and Patterson, D.A., 1991. Inference and optimal design for a welfare measure in dichotomous choice contingent valuation. Land Econ., 67: 225–239.

Hanemann, W.M., 1984. Welfare evaluations in contingent valuation experiments with discrete responses. Am. J. Agric. Econ., 66: 332–341.

Hanemann, W.M., 1989. Welfare evaluations in contingent valuation experiments with discrete responses: Reply. Am. J. Agric. Econ., 71: 1057–1061.

Hanemann, W.M., 1991. Willingness to pay and willingness to accept: How much can they differ? Am. Econ. Rev., 81: 635–647.

Hanemann, W.M., Loomis, J. and Kanninen, B., 1991. Statistical efficiency of double-bounded dichotomous choice contingent valuation. Am. J. Agric. Econ., 73: 1255–1263.

Harris, C.C., Driver, B.L. and McLaughlin, M.J., 1989. Improving the contingent valuation method: a psychological approach. J. Environ. Econ. Manage., 17: 213–229.

Hoehn, J.P. and Randall, A., 1987. A satisfactory benefit cost indicator from contingent valuation. J. Environ. Econ. Manage., 14: 226–247.

Johansson, P-O., Kriström, B. and Maler, K.G., 1989. Welfare evaluations in contingent valuation experiments with discrete response data: comment. Am. J. Agric. Econ., 71: 1055–1056.

Kahneman, D., 1986. Comments. In: R.G. Cummings, D.S. Brookshire and W.D. Schulze (Editors), Valuing Environmental Goods: A State of the Arts Assessment of the Contingent Method. Rowman and Allanheld,

Kahneman, D. and Knetsch, J.L., 1992. Valuing public goods: the purchase of moral satisfaction. J. Environ. Econ. Manage., 22: 57–70.

Kahneman, D. and Tversky, A., 1982. The psychology of preferences. Sci. Am., 246: 160–173. Totowa, NJ.

Kahneman, D., Slovic, P. and Tversky, A., 1982. Judgement Under Uncertainty: Heuristics and Biases. Cambridge University Press, New York.

Kanji, G.K., 1993. 100 Statistical Tests. SAGE Publications, London.

Kealy, M.J., Dovidio, J.F. and Rochel, M.L., 1988. Accuracy in valuation is a matter of degree. Land Econ., 65: 158–171.

Kriström, B., 1990. Valuing environmental benefits using the contingent valuation method – an economic analysis. Umeå Economic Studies No. 219, University of Umeå, Sweden.

Kriström, B., 1993. Comparing continuous and discrete valuation questions. Environ. Resour. Econ., 3: 63–71.

Langford, I.H., 1994. Using a generalized linear mixed model to analyze dichotomous choice contingent valuation data. Land Econ., 70: 507–514.

Langford, I.H. and Bateman, I.J., 1993. Estimation and reliability of welfare measures for dichotomous choice and open-ended contingent valuation studies. Global Environmental Change Working Paper 93-04, Centre for Social and Economic Research on the Global Environment, University of East Anglia, Norwich and University College London.

Langford, I.H., Bateman, I.J. and Langford, H.D., 1994. Multilevel modelling and contingent valuation. I. A triple bounded dichotomous choice analysis. CSERGE Working Paper 94-04, Centre for Social and Economic Research on the Global Environment, University of East Anglia, Norwich and University College London.

Marwell, G. and Ames, R.E., 1981. Economists free ride, does anyone else? Experiments on the provision of public goods. J. Public Econ., 15: 295–310.

Milon, J.W., 1989. Contingent valuation experiments for strategic behaviour. J. Environ. Econ. Manage., 17: 293–308.

Mitchell, R.C. and Carson, R.T., 1989. Using Surveys to Value Public Goods: The Contingent Valuation Method. Resources for the Future, Washington, DC.

Orne, M.T., 1962. On the social psychology of the psychological experiment. Am. Psychol., 17: 776–789.

Randall, A., Ives, B.C. and Eastman, C., 1974. Bidding games for valuation of aesthetic environmental improvements. J. Environ. Econ. Manage., 1: 132–149.

Roberts, K.J., Thompson, M.E. and Pawlyk, P.W., 1985. Contingent valuation of recreational diving at petroleum rigs, Gulf of Mexico. Trans. Am. Fish. Soc., 114: 155–165.

Sagoff, M., 1988. Some problems with environmental economics. Environ. Ethics, 10: 55–74.

Samuelson, P., 1954. The pure theory of public expenditure. Rev. Econ. Stat., 36: 387–389.

Sellar, C., Stoll, J.R. and Chavas, J.-P., 1985. Validation of empirical measures of welfare change: A comparison of nonmarket techniques. Land Econ., 61: 156–175.

Sellar, C., Chavas, J.-P. and Stoll, J.R., 1986. Specification of the logit model: The case of valuation of nonmarket goods. J. Environ. Econ. Manage., 13: 382–390.

Turner, R.K., Bateman, I.J. and Pearce, D.W., 1992. United Kingdom. In: S. Navrud (Editor), Pricing the European Environment. Scandinavian University Press.

Walsh, R.G., Johnson, D.M. and McKean, J.R., 1989. Issues in nonmarket valuation and policy application: A retrospective glance. West. J. Agric. Econ., 14: 178–188.