

Bateman, Ian J.; Burgess, Diane; Hutchinson, W. George; Matthews, David I.

Working Paper

Preference learning versus coherent arbitrariness: NOAA guidelines or a learning design contingent valuation (LDCV)

CSERGE Working Paper EDM, No. 06-18

Provided in Cooperation with:

The Centre for Social and Economic Research on the Global
Environment (CSERGE), University of East Anglia

Suggested Citation: Bateman, Ian J.; Burgess, Diane; Hutchinson, W. George; Matthews, David I. (2006) : Preference learning versus coherent arbitrariness: NOAA guidelines or a learning design contingent valuation (LDCV), CSERGE Working Paper EDM, No. 06-18, University of East Anglia, The Centre for Social and Economic Research on the Global Environment (CSERGE), Norwich

This Version is available at:

<http://hdl.handle.net/10419/80293>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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

**Preference Learning versus Coherent
Arbitrariness: NOAA Guidelines or a Learning
Design Contingent Valuation (LDCV)ⁱ**

by

**Ian J. Bateman, Diane Burgess, W. George
Hutchinson and David I. Matthews**

CSERGE Working Paper EDM 06-18

Preference Learning versus Coherent Arbitrariness: NOAA Guidelines or a Learning Design Contingent Valuation (LDCV)

by

Ian J. Bateman¹, Diane Burgess², W. George Hutchinson³ and David I. Matthews⁴

**1. Centre for Social and Economic Research on the Global Environment (CSERGE),
School of Environmental Sciences, University of East Anglia,
Norwich, NR4 7TJ, United Kingdom**

**2. Agri-Food and Bioscience Institute (AFBI), Newforge Lane,
Belfast BT 9 5PX.**

**3. Department of Agricultural and Food Economics, Queens University
Belfast, BT9 5PX, UK.**

**4. Survey Unit, Biometrics Branch. Agri-food and Biosciences
Institute (AFBI), Belfast, BT9 5PX, UK.**

Author contact details:

**Ian Bateman
Tel: ++44 (0) 1603 593125
Fax: ++44 (0) 1603 593739
Email: i.bateman@uea.ac.uk**

Acknowledgements

The authors gratefully acknowledge comments from Alistair Munro and others at *The Royal Economics Society 2004 Annual Conference*, University of Swansea, 5th – 7th April 2004. We are also grateful for funding support from the UK Economic and Social Research Council via their Programme in Environmental Decision Making at CSERGE; the Economics for the Environment Consultancy Ltd (EFTEC); and studentship funding from MAFF/DEFRA/DARDNI.

The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. This work was part of the interdisciplinary research programme of the ESRC Research Centre for Social and Economic Research on the Global Environment (CSERGE).

Abstract

We extend the contingent valuation (CV) method to test three differing conceptions of individuals' preferences as either: (i) *a-priori* well-formed or readily divined and revealed through a single dichotomous choice question (as per the NOAA CV guidelines; Arrow et al., 1993); (ii) learned or 'discovered' through a process of repetition and experience [Plott, 1996; List, 2003]; (iii) internally coherent but strongly influenced by some initial arbitrary anchor [Ariely et al., 2003]. Findings reject both the first and last of these conceptions in favour of a model in which preferences converge towards standard expectations through a process of repetition and learning.

JEL codes: D6 (Welfare Economics), D12 (Consumer Economics: Empirical Analysis), Q51 (Valuation of Environmental Effects), C51 (Model Construction and Estimation), Q18 (Agricultural Policy; Food Policy).

Keywords: Preference formation; discovered preferences; learning; coherent arbitrariness; contingent valuation.

1. INTRODUCTION

The nature of individuals' preferences is fundamentally crucial to the underpinnings of microeconomic theory [Varian, 1999]. However, the process through which such preferences are generated is less well proscribed and remains a matter of debate. The present paper seeks to comment upon both the formation and nature of preferences by addressing two questions. The first of these, which is of particular importance to the valuation of novel or low experience goods (such as new products, healthcare and, as in this case, environmental benefits), concerns the speed at which individuals can form stable preferences for relatively novel goods presented in unfamiliar markets. This question is important to this field in that it dictates the appropriate methodology for valuing such goods. The second question is of general interest and asks whether those stable preferences, once formed, are consistent or at variance with standard theory. As such this addresses a fundamental challenge to economics which, if sustained, requires a radical reconception of its essential underpinnings.

The bulk of applied microeconomics addresses well-formed preferences for high experience goods traded in familiar market institutions. Such applications are not typically concerned with the process through which such preferences are formed or the speed of that process. However, the rapidity of this process is a major concern for studies of low-experience goods and/or unfamiliar markets where the individual may not come to the transaction point with prior, well formed preferences. Examples of such occurrences include certain non-market goods, such as public health goods or those provided by the environment, valued through unfamiliar, often hypothetical markets. The contingent valuation (CV) method is by far the most commonly applied of all the methods available for valuing preferences for such non-market goods with thousands of applications conducted to date [Carson, forthcoming]. Clearly a key concern here is to use study designs which address the issue of a-prior poorly formed or even non-existent preferences for such goods. Failing to successfully tackle such problems is likely to result in uncertain, high variance, willingness to pay (WTP) estimates. This issue was brought into sharp focus by debate regarding the CV estimation of damages arising from the Exxon Valdes oil spill [Carson *et al.*, 1992, 1994, 2003; Hausman, 1993]; debate which was substantially addressed through the influential NOAA panel report on CV [Arrow *et al.*, 1993] which provided guidelines for future applications. A key recommendation of this report concerned the method through which WTP responses should be elicited. Although a wide variety of elicitation techniques are available [Mitchell and Carson, 1989; Bateman *et al.*, 2002], the NOAA panel recommended the use of a 'one-shot' or single-bound (SB) dichotomous choice referendum style question. Here a CV survey respondent is presented with a simple choice between either supporting a given policy program at a specified price (known as the bid-level), or rejecting this opportunity. By varying the bid level across a survey sample estimates of summary statistics such as mean WTP may be obtained for policy purposes.

The underlying argument for rejecting all but the SB response format can be traced back to the work of Gibbard [1973] and Satterthwaite [1975] establishing the *potential* incentive compatibility of one-shot referendaⁱ. However, this work applies to binding referenda involving real payments where the consequences of the referendum vote on agency action is clearly demonstrated. Whether respondents view the consequences of the vote outcome in hypothetical CV referenda as similarly binding upon either themselves or affecting agency action is open to question. Testing of this issue is problematic within a hypothetical CV setting, and advocates of the SB approach tend to refute the evidence of subsequent questioning as violating the incentive compatible framework. However, evidence from economic experiments concerning the incentive compatibility of hypothetical single referenda is decidedly mixed. Even when using common private goods in familiar market setting, while some studies find convergence of voting responses with those in real consequential referenda, other studies report divergent results [see Cummings *et al.* 1997, Taylor *et al.* 2001 and Burton *et al.* 2001]. For example, using a SB approach to value a private good

within a CV exercise Loomis et al [1997] record values which are roughly twice as high as subsequently revealed in real sales data. Given that CV applications typically value novel (and often public) goods presented in unfamiliar, hypothetical markets, the concern is that the uncertain nature of preferences for such goods may overwhelm the already questionable incentive compatibility properties of the SB elicitation format in CV studies. In such cases, residual preference uncertainty seems, at best, likely to yield high variance in WTP estimates while at worst (for reasons discussed subsequently) they may also be systematically biased.

A more fundamental critique of the 'one shot' nature of the SB approach is provided by the Discovered Preference Hypothesis (DPH) proposed by Plott [1996]. The DPH argues that stable and theoretically consistent preferences are typically the product of experience gained through practice and repetition. Plott notes that markets provide an ideal environment for such repetition and learning through which individuals can discover both how best to achieve goals within the operating rules of that real or hypothetical market (a process which Braga and Starmer, 2005, refer to as 'institutional learning') and discover features of their own preferences ('value learning', *ibid*). The first response SB precludes either institutional or value learning and is in direct conflict with the DPH which would suggest that it is the last response in a series of valuations which should be attended to, rather than the first [see also Binmore, 1994; 1999]. This, together with the empirical questioning of whether incentive compatibility arguments from binding referenda can indeed be extended to hypothetical CV studies, raises significant questions regarding the common presupposition in favour of the SB elicitation method.

Central to the DPH then is the role of repetition within the formation of stable and theoretically consistent preferences. Whereas the experimental literature questions the importance of incentive compatibility in determining SB responses within CV studies, the same literature provides considerable support for the argument that learning through repetition and experience are important requirements for the revelation of theoretically consistent and stable preferences. Examples of experiments in which learning opportunities appear to lead to a reduction in preference anomalies include: diminution of the WTP/WTA gap and endowment effects over repeated trials [Coursey, Hovis and Schulze, 1987; List, 2003; List and Shogren, 1999; Loomes and Sugden, 1998; Plott and Zeiler, 2003, 2005; Shogren *et al.*, 1994 and 2001]; reducing hypothetical bias through learning [Bjornstad et al., 1997]; reduction in the preference reversal anomaly in both real and hypothetical payment formats [Cox and Grether, 1996; Cherry et al., 2002; Braga and Starmer, 2005]; and, perhaps most pertinently, reduction in preference anomalies amongst more experienced traders or choice makers [List and Lucking-Reiley, 2000; List, 2001, 2002a, 2002b, 2004]ⁱⁱ.

This experimental evidence suggests that when unfamiliar goods are presented in previously unencountered hypothetical market institutions (such as often occurs in CV surveys) resulting initial valuations are liable to be based upon poorly formed preferences. In such situations the 'constructed preference' literature [Slovic, 1995; Hoeffler and Ariely, 1999] would suggest that such responses are prone to be influenced by a variety of choice heuristics and framing effects resulting in apparently anomalous preferences [Tversky and Kahneman, 1974]. For example, recent work defining out the 'focusing illusion' [Schkade and Kahneman, 1998]ⁱⁱⁱ suggests that concentrating on just a single good, presented in a single response framework, is liable to inflate respondents perceptions of the importance of that good and hence raise stated WTP. A further effect of the SB approach is that such initial responses may be responsive to any available 'anchor' such as the SB bid-level itself which may be taken as some clue to the 'correct' value of the good in question [Johnson and Schkade, 1989; Green, Jacowitz, Kahneman and McFadden, 1998]^{iv}. In an important and highly innovative recent paper, Ariely, Loewenstein and Prelec [2003] refine the anchoring argument noting that in their experimental investigations, while an individual's choices were typically internally coherent, nevertheless they could also be strongly anchored to some

initial, demonstrably arbitrary starting point (discussed further in Section II), resultant values being manipulated up or down by altering this starting point. Such behaviour, which Ariely et al., term 'coherent arbitrariness', is a challenge not only to the discovery of stable preferences envisaged by the DPH, but more fundamentally questions the underpinnings of standard microeconomic theory, in effect suggesting that prices determine values rather than vice versa.

In summary, we can identify three important yet different conceptions of individuals' preferences as being either: (i) *a-priori* well-formed or readily divined through a single incentive compatible question [Carson et al., 2000]; (ii) learned or 'discovered' through a process of repetition and experience [Plott, 1996; List, 2003]; (iii) internally coherent but liable to be strongly influenced by some initial arbitrary anchor [Ariely et al., 2003]. The first two of these views differ only in terms of the preference formation process rather than the process outcome (stable, theoretically consistent preferences). Yet the issue of the speed of the preference formation process is vital to the choice of appropriate methodology. The first view stresses incentive compatibility over the prior establishment of preference stability and consistency and hence leads to the NOAA panel recommendation of the SB technique for CV studies. However, the second, DPH, view argues that the single question approach of the SB format is highly liable to result in individuals responding upon the basis of poorly defined preferences resulting in very uncertain, high variance estimates of WTP. The DPH view therefore mitigates in favour of a repeated questioning methodology which encourages learning regarding both the market institution and preferences themselves. However, both approaches and with them the basis of standard microeconomic theory, are fundamentally challenged by the third 'coherent arbitrariness' view which argues that preferences are anchored from the initial starting point, with an individual's desire to maintain internal consistency within responses preserving this anchoring effect through subsequent choices and values.

Given this obvious and potentially important conflict, this paper sets out to provide the first field-based CV study designed to contrast the 'standard' approach with a novel repeated valuation technique consistent with the DPH. We accordingly develop the 'Learning Design Contingent Valuation' (LDCV) method which allows survey respondents to engage in repeated valuation tasks from which we separately test for both institutional and value learning processes and through this evaluate the theoretical consistency of expressed values both before and after such learning opportunities. Further design features and a novel empirical testing protocol allow us to examine findings against the predictions of standard theory and those derived from the 'coherent arbitrariness' hypothesis. In summary, results are that while valuations of an initial good fail tests of both institutional and value learning, responses to subsequent valuation tasks strongly conform to the expectations of the DPH such that valuations of the final good considered pass both types of testing. Furthermore, tests of both the final good valued and the preference formation process itself fail to support the 'coherent arbitrariness' hypothesis. Taken together these results seriously question the standard reliance upon the single-shot SB approach to CV, suggesting instead that a DPH inspired repeated valuation approach can yield measures consistent with standard theory. More fundamentally our results fail to provide convincing evidence rejecting the preference model underpinning that standard theory.

The paper is organized as follows. In the following section we outline our LDCV method for conducting repeated valuations both within and across goods allowing us to formulate tests for both institutional and value learning. We also briefly discuss the empirical case study used to provide data for these tests. In Section III we outline our analytical methodology introducing a novel Monte Carlo based approach to allow testing of learning effects across valuation tasks. Section IV reports results while Section V discusses the implications of these findings and concludes.

2. RESEARCH DESIGN

2.1 Testing for Institutional Learning

Repetition is the seedcorn of experience within the DPH. Given this, we sought to construct a study design which would facilitate repetition of valuation tasks both across and within goods. The within-goods aspect of the resultant LDCV design allows us to formulate arguably one of the stiffest tests of institutional learning possible within a CV study; an examination of value coherence within the double bound (DB) dichotomous choice format.

The DB approach was introduced by Hanemann *et al.* [1991] and Welsh and Bishop [1993] and is simply an SB format supplemented by a follow-up dichotomous choice question asked after the initial response is received. Here the bid-level offered in the second question is determined in part by the response given to the first question such that a positive response to an initial WTP bid-level results in a higher amount being presented at the second bound. Because value estimates are obtained by combining both the first and second response, DB designs permit a substantial improvement in the statistical efficiency of a given sample relative to that provided by applying a SB format. As a result they have risen in popularity and application to become one of the most prevalent of all CV designs. However, despite this popularity, in practice DB studies have consistently reported an anomalous, non-zero difference (which we denote Δ) between the estimated mean WTP derived from SB responses (denoted μ_{SB}) and that obtained from the first and second responses combined within a DB analysis (μ_{DB}). Examples of such results include the studies by Cameron and Quiggin [1994], McFadden [1994], Bateman *et al.*, [2001] and DeShazo [2002]. Typical amongst these are the findings of McFadden [1994] which “reject at the 1% level the hypothesis that first and second responses in the double referendum experiment are drawn from the same distribution” (pp705-706).

Some commentators have argued that the DB anomaly arises from changes in incentive compatibility between the first and second response [Carson *et al.*, 1994; Alberini *et al.*, 1997; Carson *et al.*, 1999]. However, others highlight evidence showing that unfamiliarity with the institutional procedures of the DB means that respondents do not anticipate follow-up questions, are surprised by them [Bateman *et al.*, 2001] and, given further multiple valuation experience, are liable to revise their responses if given the opportunity [Bateman *et al.*, 2004]. Given this and the experimental evidence cited above suggesting both that initial responses may themselves be anomalous and that learning effects may arise through repetition, then a DPH reading of these findings might be that they provide the first (if inadequate) evidence of survey respondents beginning to learn about the previously unencountered hypothetical market institution through which CV responses are elicited. If, given sufficient experience, respondents can learn about the DB institution then we might expect the anomalous discrepancy between SB and DB responses to diminish with repetition. Such a finding would of itself be notable given the persistence of this anomaly across previous studies and the ongoing popularity of the DB approach. Given this, we test such a reading by designing our LDCV to repeat DB format valuation tasks across a number of goods. This provides the data for our institutional learning test which examines whether first and second bound valuations become more consistent and hence coherent across successive goods. Our test focus therefore becomes an examination of the SB-DB difference (Δ), with the DPH expectation being that there will be a decline in Δ between its initial level (from the first good valued) and its level for subsequent goods (with no increase between values of Δ for any good and any other subsequently presented good). Notice that the focus of testing is upon the significance of trends in the value of Δ , rather than upon the absolute value of this difference. These trends allow us to discriminate between the DPH and competing hypotheses. Nevertheless, for comment we report absolute values of Δ and tests of whether these significantly differ from zero.

2.2 Testing for Value Learning

While our institutional learning test is essentially one of within-subject coherence, the findings of Ariely, Loewenstein and Prelec [2003] demonstrate that such coherence is a necessary but not, on its own, sufficient condition for the identification of theoretically consistent values. Rather they argue that:

“valuations are initially malleable but become “imprinted” (i.e. precisely defined and largely invariant), after the individual is called upon to make an initial decision. Prior to imprinting, valuations have a large arbitrary component, meaning that they are highly responsive to both normative and non-normative influences. Following imprinting, valuations become locally coherent as the consumer attempts to reconcile future decisions of a “similar kind” with the initial one. This creates an illusion of order, because consumers’ coherent responses to subsequent changes in conditions disguise the arbitrary nature of the initial, foundational, choice” (pp 74-75).

Ariely et al., test for this state by presenting different samples of respondents with differing initial stimuli (sometimes chosen by the analysts, on other occasions being blatantly random by using the last digits of respondents’ social security number) and examining whether these are used as anchors by respondents for their subsequent valuation responses. Through a series of experiments Ariely et al., report that while responses from repeated valuation tasks tended to be internally coherent, with respondents typically ordering values in a consistent manner (for example by always valuing “rare” wines above “average” wines), when viewed across respondents these valuations were found to be strongly anchored to the initial, arbitrarily determined stimuli. Here Ariely et al., interpret the finding of significant anchoring effects as showing that such preferences reflect coherent arbitrariness rather than theoretically consistent “fundamental” values.

The Ariely et al., anchoring approach provides a strong test for discriminating between coherent arbitrariness and value learning in CV responses. The initial bid-level presented in the SB question regarding a given good provides our arbitrary, initial stimuli. By varying this systematically both across respondents and across goods and employing the method developed by Herriges and Shogren [1996]’ for estimating an ‘anchoring parameter’ (γ) between SB and DB responses we can examine trends in the degree of any anchoring observed across repeated valuation tasks. The coherent arbitrariness expectation is that γ should be initially significant and will not decrease significantly (and might even increase) as successive goods are valued. If however, we observe a significant decline in the level of anchoring across valuation tasks then this rejects the coherent arbitrariness view and mitigates in favour of the DPH argument that repetition and learning will yield theoretically consistent preferences.

As in the case of our institutional learning test, we have no a-priori expectation of how many valuation repetitions are required to drive the absolute level of anchoring to non-significant levels. Rather our focus is upon trends in γ as these are the vital discriminators here. That said, in keep with our previous analysis, we also report absolute values of γ and tests of whether these significantly differ from zero.

2.3 Choice of goods and survey implementation details

In order to perform both our institutional and value learning tests we therefore require repetition of valuation tasks both within goods and across goods. Both the DPH and coherent arbitrariness arguments would mitigate in favour of using goods which are, to quote Ariely et al., of a “similar kind” (p.75). Using goods which are formally distinct yet cognitively similar avoids preclusion of the behavioural processes underpinning both learning and

coherent arbitrariness. A further requirement was that, given the CV focus of this research, these should be public rather than private goods; the typical target of such studies.

Given the above constraints it was determined that the empirical case study should present survey respondents with a set of animal welfare improvement goods, each of which improved the farm living conditions for a different species (namely; laying hens, chickens, cows and pigs) to be paid for via a compulsory tax on all foodstuffs. Each animal welfare good was presented as mutually exclusive, thereby avoiding substitution and allied sequencing effects [Carson et al., 1998; Bateman et al., 2004]. Each good was valued using a DB WTP elicitation format. An initial sample (Sample 1) was presented with all four goods, given in the order shown below, thus permitting the repetition of valuation tasks necessary to facilitate either learning or arbitrary coherence. To permit further testing of potential learning effects and to control for the possibility that any observed increase in preference consistency is a by-product of the order in which goods are presented, a second sample of respondents (Sample 2) were asked DB questions solely regarding the good which was valued last by Sample 2 (the pig welfare good). The DPH expectation here is that, controlling for the good, the degree of any institutional anomalies (disparities between in values derived from the SB and DB procedures) and value learning anomalies (coherently arbitrary anchoring effects) should decline across the valuation tasks faced by Sample 1. Contrasting the characteristics of preferences for the common good, presented fourth to Sample 1 and the only good valued by Sample 2, the DPH leads us to expect a lower level of anomaly amongst the former than the latter, although whether such anomalies will have become statistically insignificant is an open-empirical question.

For notational purposes we denote any good presented to a respondent as X_j^i where X denotes the good in question, i refers to the sample providing the valuation (where $i = 1, 2$) and j denotes the order of presentation of that good within the overall list of goods given to that sample (therefore $j = 1, 2, 3, 4$ for $i = 1$ and $j = 1$ for $i = 2$). Therefore for Sample 1 the following goods were valued in the order shown:

- (i) Improving living conditions for laying hens ($HENS_1^1$)
- (ii) Improving living conditions for chickens ($CHICKS_2^1$)
- (iii) Improving living conditions for diary cows ($COWS_3^1$)
- (iv) Improving living conditions for pigs ($PIGS_4^1$)

Whereas respondents in Sample 2 were presented with the good improving living conditions for pigs, denoted $PIGS_1^2$, i.e. that good which was presented last (fourth) to Sample 1.

The vector of bid-level values was determined in accordance with Boyle and Bishop [1988] as refined by Hanemann and Kanninen [1999] through the administration of a prior pilot survey. This suggested a vector with four bid-levels at the first response question, supplemented by a further two (one above and the other below these initial four) at the second response question^{vi}. As indicated above, in the subsequent main survey, these amounts were randomly allocated by good and respondent such that any given respondent faced with any given good had an equal probability of being presented with any one of the initial bid-levels for that good.

The final CV questionnaire was administered by face-to-face, at-home interviews with 400 respondents selected by a random sampling process based on the electoral register of Northern Ireland. Respondents were randomly allocated to the two treatments such that sample size was 200 respondents for both samples.

3. ECONOMETRIC METHODOLOGY

In order to identify potential learning effects both within and across goods the data generated by the survey was analyzed using both SB models (applied to the first response for each good valued) and DB models (applied to both first and second responses for each good valued) as per Hanemann and Kanninen [1999]. Note that it would be inappropriate to model second bound responses in isolation of first bound responses as the bid levels offered in the second question are conditional upon the bid levels and response from the first bound. Again using methods detailed by Hanemann and Kanninen [1999], these analyses allow us to calculate and compare mean willingness to pay from first responses (denoted μ_{SBj}) with those from first and second responses modelled as DB data (μ_{DBj}). Details of our modelling strategy are presented through the remainder of this section.

3.1 Single Bounded (SB) Dichotomous Choice Model

In a SB exercise (or equivalently within our study, the first response of a DB design) CV respondents indicate their WTP by answering *yes* or *no* to a set offered price. For an individual the probability π_{yes} of responding *yes* to an offered bid-level b_1 for a certain good can be expressed as follows:

$$\pi_{yes}(yes : \beta'x > b_1) = H(\beta'x) + e \quad (1)$$

where b_1 is the value of the bid-level presented to the respondent, x represents a set of covariates including the bid-level amount and β is a vector of parameters to be estimated from the sample data. H is a function which expresses the probability and hence the function must return a value between zero and one, and it must sum to unity over all possible outcomes. A cumulative probability distribution (cdf) function is used for this purpose. The SB allows just two possible outcomes, 'yes' and 'no' so the sample log-likelihood can be expressed as:

$$\text{LogLike} = \sum_{i=1}^{n_{yes}} y \log \pi_{yes} + \sum_{i=1}^{n_{no}} (1-y) \log(1 - \pi_{yes}) \quad (2)$$

where n_{yes} and n_{no} are the numbers of respondents replying *yes* and *no* respectively to the bid values offered and y is a dummy variable indicting an individual's choice, being coded 1 for *yes* and 0 for *no*.

When a linear model is used and H is a logistic cdf using only covariates for the offered bid-level value then the probability of a *yes* response π_{yes} will occur when the respondent's WTP exceeds the bid-level. This can be expressed as:

$$\pi_{yes} = 1/\{1+\exp(-\alpha_{SB} - \beta_{SB}b_1)\} \quad (3)$$

Where α_{SB} and β_{SB} are the coefficients of the constant and bid-level respectively. The method of maximum likelihood is used to estimate the coefficients α_{SB} and β_{SB} . In line with many previous studies [e.g. Langford et al, 1998] this study uses a single bounded Logit model.

3.2 Double Bounded (DB) Dichotomous Choice Model

The DB approach extends an initial SB-style question with a follow-up dichotomous choice valuation task. If the individual agreed to pay the initial bid amount then the follow-up question posits a higher amount. Conversely if the initial amount is rejected, then the

proffered follow-up concerns a lower bid level. The probability of a respondent choosing each of the four possible responses (*yes,yes*), (*yes,no*), (*no,yes*) and (*no,no*) is given by:

$$\begin{aligned}
\pi_{yy} &= 1 - H(\beta'x_h) \\
\pi_{yn} &= H(\beta'x_h) - H(\beta'x) \\
\pi_{ny} &= H(\beta'x) - H(\beta'x_l) \\
\pi_{nn} &= H(\beta'x_l)
\end{aligned} \tag{4}$$

where x , x_h , x_l are the vectors of covariates respectively associated with first bid, second bid higher and second bid lower, and H is the cdf function chosen. This gives the following log likelihood for the sample:

$$\text{LogLike} = \sum_{i=1}^{n_{yy}} yy \log(\pi_{yy}) + \sum_{i=1}^{n_{yn}} yn \log(\pi_{yn}) + \sum_{i=1}^{n_{ny}} ny \log(\pi_{ny}) + \sum_{i=1}^{n_{nn}} nn \log(\pi_{nn}) \tag{5}$$

where n_{yy} , n_{yn} , n_{ny} and n_{nn} are the number of occurrences in the sample of each of the four outcomes and yy, yn, ny, nn are dummy variables indicating the choice for each individual. Following Hanemann et al., [1991] we use a logistic cdf for H . This then becomes a double bounded logit model.

The additional information provided by the follow-up question makes the DB asymptotically efficient relative to the SB. A further advantage of the DB model is that it is fairly robust with respect to poor bid designs resulting from initial parameter misspecification [Hanemann *et al.*, 1991; Scarpa and Bateman, 2000]. In effect, the higher second bid provides insurance against too *low* a choice for the initial bid and the lower second bid provides insurance against too *high* a choice for the initial bid [Hanemann and Kanninen, 1999].

3.3 Testing for Institutional Learning: Consistency of mean WTP μ_{SBi} and μ_{DBi}

Estimates of WTP can be computed from the SB and DB models. For example, mean WTP can be calculated for the SB model specified in Equation (3) using:

$$\mu_{SBj} = -\alpha/\beta \tag{6}$$

The standard errors for μ_{SBj} can be obtained from the variance of $\text{Var}(\mu_{SBj}) = \text{Var}(\alpha/\beta)$ which can be calculated using the Delta method (Taylor's first order approximation of variance when the coefficient covariance estimates are known). We can now begin to calculate a measure of the difference between μ_{SB} and μ_{DB} for sample i facing good j (i.e. the value Δ_j^i) which will in turn be used to examine trends in this difference across valuation tasks (i.e. across goods). However, the estimation of absolute values of Δ_j^i requires the development of novel econometric techniques controlling for the non-independent (within-respondent) nature of the first and second bound responses which provide data for this exercise. Appendix 1 to this paper discusses the techniques developed for this purpose.

As discussed previously, once values for Δ_j^i are established, the major focus of our testing is to examine *trends* within the statistical significance of this variable across valuation tasks (i.e. across goods). The DPH recognises that initial values of Δ_j^i may be high but suggests that these values will decline significantly with the increased opportunity for institutional learning afforded by successively repeating the valuation process across goods. This

expectation will apply at any point across the valuation sequence (i.e. $\Delta_j^i \geq \Delta_{j+k}^i$ for any $k > 0$) setting up some six tests within Sample 1.

In addition to these central tests we can also examine whether the absolute value of Δ_j^i is statistically significant. Although neither the DPH nor coherent arbitrariness say anything about such absolute values (focussing instead upon trends across valuation tasks), the comparison of such values for the same good presented either at the end of a sequence or as the first (and for Sample 2, only) good (i.e. comparing Δ_4^1 with Δ_1^2) is revealing and provides a further distinguishing test between these hypotheses.

3.4 Testing for Value Learning: Robustness against Anchoring Effects

Following established methods outlined in Herriges and Shogren [1996] and Hanemann and Kanninen [1999] we apply an econometric test for whether responses and resultant values obtained from DB data are significantly anchored on the value of the initial bid-level. This test adds an anchoring parameter (γ) into the DB model of Equation (5). According to Herriges and Shogren [1996] the revised WTP in response to the second bid is:

$$WTP_r = (1-\gamma)WTP_o + \gamma b_1 \quad (7)$$

where WTP_o is the prior WTP and WTP_r is the revised WTP following any anchoring effect induced by the initial bid-level b_1 . From the above the effective bid-level for b_2 becomes b_{2r} as follows:

$$b_{2r} = (b_2 - \gamma b_1)/(1-\gamma) \quad (8)$$

The Log Likelihood function for the anchoring model is obtained by substituting the b_{2r} value into the standard DB likelihood function given in Equation (4).

Calculating γ for responses to each good (i.e. γ_j^i) the empirical significance of the parameter provides a test of whether our DB responses are anchored by the SB bid-level. Following the coherent arbitrariness hypothesis, initial anchoring is expected to be significant and persistent across goods. In contrast, following the DPH any initial anchoring effects should decay away across successive goods as the valuation of similar goods allows respondents time to consider and learn about their preferences. Again, while neither hypothesis is definitive regarding the absolute value of γ , comparison of the levels associated with the same good, presented either first or last (i.e. comparing γ_4^1 with γ_1^2) should reveal a further insight into the validity of these competing hypotheses in this context.

4. RESULTS

4.1 SB and DB models

Table I presents parsimoniously specified logistic SB and DB models for each good estimated as per Hanemann et al [1991]. The models provide parameter estimates of the coefficients α , β for the constant and bid-level respectively for the four goods valued by Sample 1 ($HENS_1^1$, $CHICK_2^1$, $COWS_3^1$ and $PIGS_4^1$) and the single good valued by Sample 2 ($PIGS_1^2$). While desirable in benefit transfer and policy analysis, additional socio-economic and attitudinal covariates are not needed to test for the effects of learning on these welfare estimates. Other columns report the standard error and t-value associated with each

parameter estimate and the log-likelihood of the model. All coefficients have expected signs and are highly significant as are the overall models.

Table I: SB and DB models of WTP for specified animal welfare improvement goods.

Good	Single Bounded (SB) Models					Double Bounded (DB) Models				
	Estimate	Coeff.	Std. Err.	t-ratio	Log Like.	Estimate	Coeff.	Std. Err.	t-ratio	-Log Like.
<i>Sample 1</i>										
<i>HENS</i> ₁ ¹	α_{SB}	0.92	0.27	3.40	-131.54	α_{DB}	1.54	0.20	7.70	266.61
	β_{SB}	-0.19	0.09	-2.10		β_{DB}	-0.56	0.05	-11.20	
<i>CHICK</i> ₂ ¹	α_{SB}	1.44	0.28	5.10	-122.62	α_{DB}	1.57	0.20	7.89	252.61
	β_{SB}	-0.54	0.11	-5.07		β_{DB}	-0.63	0.05	-12.10	
<i>COWS</i> ₃ ¹	α_{SB}	1.32	0.28	4.81	-125.55	α_{DB}	1.57	0.20	7.70	274.18
	β_{SB}	-0.43	0.10	-4.38		β_{DB}	-0.55	0.05	-11.80	
<i>PIGS</i> ₄ ¹	α_{SB}	1.29	0.29	4.47	-120.64	α_{DB}	1.39	0.20	7.07	263.57
	β_{SB}	-0.62	0.12	-5.15		β_{DB}	-0.68	0.05	-12.94	
<i>Sample 2</i>										
<i>PIGS</i> ₁ ²	α_{SB}	1.25	0.28	4.49	-126.01	α_{DB}	1.61	0.20	7.98	271.88
	β_{SB}	-0.42	0.10	-4.23		β_{DB}	-0.68	0.05	-13.15	

4.2 Results from the Institutional Learning Test

Our institutional learning test examines whether, as respondents value successive goods, their increasing familiarity with the contingent market results in greater consistency of valuation responses between the two bounds of the DB format. In order to undertake this test we first need to estimate the mean values ($\mu_{SBi} - \mu_{DBi}$) and hence evaluate Δ_j^i . Table II details results from this analysis test. Standard errors for Δ_j^i are calculated using the Jack-knife method so as to control for intra-respondent correlation between first and second responses for each good. Corresponding t-statistic and probability levels are also reported in the final two columns of the table.

Table II: Differences between mean WTP for SB and DB estimates for each good, where $\Delta_j^i = \mu_{SB} - \mu_{DB}$ for good X_j^i

Scheme	Estimate	Value	Std.Er. ¹	t-ratio	H ₀ : $\mu_{SB} = \mu_{DB}$ (Prob.)
<i>Sample 1</i>					
<i>HENS</i> ₁ ¹	μ_{SB}	£4.72			
	μ_{DB}	£2.74			
	Δ_1^1	£1.98	£1.21	1.64	0.10
<i>CHICK</i> ₂ ¹	μ_{SB}	£2.68			
	μ_{DB}	£2.51			
	Δ_2^1	£0.17	£0.17	1.00	0.32
<i>COWS</i> ₃ ¹	μ_{SB}	£3.10			
	μ_{DB}	£2.87			
	Δ_3^1	£0.23	£0.26	0.88	0.38
<i>PIGS</i> ₄ ¹	μ_{SB}	£2.07			
	μ_{DB}	£2.06			
	Δ_4^1	£0.01	£0.15	0.07	0.95
<i>Sample 2</i>					
<i>PIGS</i> ₁ ²	μ_{SB}	£2.98			
	μ_{DB}	£2.38			
	Δ_1^2	£0.60	£0.25	2.40	0.02

Note: 1. Standard errors are computed using the Jack-knife method to take account of intra-respondent correlation of responses.

Although our principle tests concern trends in Δ_j^i across valuation tasks, nevertheless the estimates reported in Table II are worthy of comment. Recall that the literature on previous DB applications has resulted in acknowledgement as a stylised fact that such studies invariably yield a disparity between μ_{SBi} and μ_{DBi} [Carson, Groves and Machina, 2000]. Given this, the findings set out in Table II are remarkable. Considering the first good valued by both Sample 1 and 2 (*HENS*₁¹ and *PIGS*₁² respectively) we obtain the standard stylised result of a significant difference in mean WTP as calculated from the SB and DB models (μ_{SBi} and μ_{DBi}). These differences are not only statistically significant but also highly substantial. For example, for Sample 1 (valuing *HENS*₁¹) we have $\mu_{SBi} = £4.72$ while $\mu_{DBi} = £2.74$.

However, when these same Sample 1 respondents are presented with a second good to value ($CHICKS_2^1$) the disparity in SB and DB means becomes much smaller and proves statistically insignificant. Indeed there is a clear pattern of significance running across successive Sample 1 valuation tasks with $\mu_{SBi} - \mu_{DBi}$ differences (Δ_j^i) becoming steadily less significant until for the fourth and final good ($PIGS_4^1$) this difference is just one penny. The cross-sample, within-good test comparing Δ_4^1 (from $PIGS_4^1$) with Δ_1^2 (from $PIGS_1^2$) is also revealing. Results show that the experienced respondents in Sample 1 do indeed generate significantly lower ($\mu_{SBi} - \mu_{DBi}$) differences than do inexperienced respondents in Sample 2 ($p=0.04$) for this common good.

One final result of Table II is worthy of note. Returning to the value estimates for the first good seen by Sample 1 ($HENS_1^1$) we can note two observations. First, the SB estimate is very high compared to its DB counterpart, indeed it is substantially greater than any of the other values elicited for any of the other goods. This result recalls the focussing illusion expectation that the first response for the first good will be inflated. As noted previously, it is impossible to disprove the Carson et al., [2000] argument in favour of the SB format within a hypothetical CV study as no criterion (demonstrably correct) value is available within such a context. However, the weight of evidence from experimental studies (discussed in our opening remarks) is now reinforced by the present results and clear reduction in values across Sample 1 (discussed further below). Together this suggests that the initial values provided by the SB approach are substantially out of line (and inflated upward) compared to those values elicited at the end of the LDCV process. Secondly, the standard error around Δ_1^1 is nearly five times larger than that for any of the other goods valued by Sample 1. Arguably this reflects uncertainty in underlying preferences for this group when faced with this initial task; a degree of uncertainty which is not repeated in subsequent valuations.

The results of Table II do suggest that the learning opportunities inherent in the LDCV approach do indeed yield greater theoretical consistency in preferences, in this case completely removing one of the best documented and most persistent anomalies in the CV literature. However, it is our tests regarding trends in Δ_j^i across goods which are key. Results from these tests are presented in Table III.

Table III: Hypothesis tests of differences in Δ_j^i across the sequence of goods valued.

Hypothesis test	(Value) ¹	Std Error ²	t test Result ³
<i>Across Goods (Sample 1)</i>			
$(\Delta_1^1 - \Delta_2^1) < 0$	1.81	1.23	Sig. <0.10
$(\Delta_1^1 - \Delta_3^1) < 0$	1.75	1.22	Sig. <0.10
$(\Delta_1^1 - \Delta_4^1) < 0$	1.96	1.18	Sig < 0.05
$(\Delta_2^1 - \Delta_3^1) < 0$	-0.06	0.32	NS
$(\Delta_2^1 - \Delta_4^1) < 0$	0.22	0.30	NS
$(\Delta_3^1 - \Delta_4^1) < 0$	0.16	0.23	NS
<i>Within Goods (Sample 2 vs. Sample 1)</i>			
$(\Delta_1^2 - \Delta_4^1) < 0$	0.59	0.29	Sig < 0.05

Notes:

1. Refers to the sum in parentheses in the first column of the table.
2. Estimated using first order Jackknife procedure (as per the Appendix; Equation A3).
3. The DPH, which underpins the tests reported here, gives a clear directional expectation. Hence a 1 tailed test is appropriate (n=200 in all cases).

The upper panel of Table III reports Δ_j^i trend findings for Sample 1. Within this the first three rows detail the significance of differences between the value of Δ for the first good (Δ_1^1) those for and subsequent goods (Δ_2^1 , Δ_3^1 and Δ_4^1). Finding show a clear and statistically significant downward trend in Δ as subjects value successive goods. This pattern runs counter to that proposed under the coherent arbitrariness hypothesis and directly in line with the expectations of the DPH. Furthermore, if anything the reduction in anomaly predicted by the DPH appears to be getting stronger as respondents pass through the LDCV repeated valuation design.

The next two rows of Table III change the baseline comparison from the initial good to the second good, comparing Δ_2^1 with Δ_3^1 and Δ_4^1 . Here no further significant increase or decrease in Δ are observed; a result which holds for the comparison of Δ_3^1 and Δ_4^1 given in the penultimate row of the table. The Δ trend results from the repeated valuation test of Sample 1 shows that respondents very rapidly learn how the CV market works, such that even after answering the DB questions for just the initial good the associated anomaly, prevalent though it is in all previous DB studies, quickly evaporates in the face of that learning. Such a result strongly supports the DPH. Given this clear evidence of greater theoretical consistency within the final values elicited from the LDCV process, re-inspection of the high value elicited from the SB question for the first good suggests that the latter is providing an upwardly biased estimate of WTP.

The final row of Table III compares values of Δ for the same good (*PIGS*) presented as either the first or fourth good valued by separate samples of respondents. Here the high value of Δ_1^2 (showing the standard DB anomaly associated with inexperienced respondents) is found to be significantly greater than the low value of Δ_4^1 given by experienced respondents facing their fourth valuation task.

Overall then, the results reported in Tables II and III strongly support the DPH expectation of institutional learning arising from increased familiarity and experience with the contingent market.

4.3 Results from the Value Learning Test

Although our institutional learning test indicates that while the valuations of a good expressed by individuals become more internally consistent as they become more familiar with the operating rules of the contingent market, we need to determine whether this consistency is procedurally invariant and therefore conforms to standard theory, or whether it is subject to the cross-individual anchoring effects symptomatic of coherent arbitrariness.

Table IV: Estimates of DB model with anchoring coefficient γ , testing whether second response is anchored on the first bid level.

Good	Variable	Coeff.	Std.Err.	t-ratio	p
<i>Sample 1</i>					
<i>HENS</i> ₁ ¹	α_{DB}	0.786	(0.22)	3.57	
	β_{DB}	-0.224	(0.11)	-2.04	
	γ_1^1	0.670	(0.17)	4.03	<0.001
<i>CHICK</i> ₂ ¹	α_{DB}	1.392	(0.28)	4.94	
	β_{DB}	-0.551	(0.12)	-4.65	
	γ_2^1	0.146	(0.15)	0.98	0.329
<i>COWS</i> ₃ ¹	α_{DB}	1.198	(0.37)	3.26	
	β_{DB}	-0.391	(0.13)	-2.92	
	γ_3^1	0.334	(0.17)	2.00	0.047
<i>PIGS</i> ₄ ¹	α_{DB}	1.427	(0.30)	4.76	
	β_{DB}	-0.691	(0.14)	-4.94	
	γ_4^1	-0.026	(0.18)	-0.14	0.886
<i>Sample 2</i>					
<i>PIGS</i> ₁ ²	α_{DB}	1.194	(0.26)	4.59	
	β_{DB}	-0.497	(0.13)	-3.82	
	γ_1^2	0.315	(0.15)	2.07	0.040

By estimating the anchoring model specified in Equations (7) and (8) we obtain a series of models which allow for the presence of anchoring within DB responses. These are reported in Table IV. Here coefficients on α_{DB} and β_{DB} are consistently in accord with prior expectation and statistically significant throughout. However, these are of secondary interest and therefore are omitted from the p-values reported in the final column of the table, which instead focuses upon the anchoring parameter γ_j^i . As noted previously, neither the coherent arbitrariness hypothesis nor the DPH have any expectations regarding the absolute size or significance of any given γ_j^i . Nevertheless, as before, certain of these findings are worthy of comment. In particular while there is highly significant anchoring in responses regarding the first good valued (in both samples) this becomes entirely non-significant by the time the final good is valued. This result strongly accords with the predictions of the DPH. However, as before it is our trend analysis which is crucial here, results for which are presented in Table V.

Table V: Hypothesis tests on differences in anchoring coefficients γ_j^i across the sequence of goods valued.

Hypothesis test	(Value) ¹	Std Error ²	Test result
<i>Across Goods (Sample 1)</i>			
$(\gamma_1^1 - \gamma_2^1) < 0$	0.49	0.23	Sig. <0.05
$(\gamma_1^1 - \gamma_3^1) < 0$	0.30	0.23	Sig. <0.10
$(\gamma_1^1 - \gamma_4^1) < 0$	0.66	0.24	Sig. <0.01
$(\gamma_2^1 - \gamma_3^1) < 0$	-0.19	0.29	NS
$(\gamma_2^1 - \gamma_4^1) < 0$	0.18	0.23	NS
$(\gamma_3^1 - \gamma_4^1) < 0$	0.36	0.26	Sig. <0.10
<i>Within Goods (Sample 2 vs. Sample 1)</i>			
$(\gamma_1^2 - \gamma_4^1) < 0$	0.36	0.20	Sig. < 0.05

Notes:

1. Refers to the sum in parentheses in the first column of the table.
2. Estimated using jack-knife procedure (as per the Appendix; Equation A3).

As per our previous trend analysis, the upper panel of Table V reports findings regarding the trend in anchoring across the various goods valued by Sample 1. Within this the first three rows compare anchoring within the first good with that observed for subsequent goods. As before these results are unequivocal in conforming to the expectations of the DPH. While Table IV showed that the first good considered by Sample 1 had valuations which were strongly anchored to the randomly chosen bid initial level for that good, Table V shows that this degree of anchoring falls significantly in responses to subsequent goods (Table IV showing that anchoring becomes non-significant in absolute terms by the time the final good is valued). The next three rows of Table V show that (as in our analysis of Δ levels) there is relatively little further reduction in anchoring once the initial good has been considered, although in this case we do see a marginally significant reduction between the third and

fourth good, suggesting that full learning effects may not have been completed within the four good design adopted.

The final row of Table V shows that anchoring in responses to a good presented as the first one valued is significantly greater the level of anchoring observed when the same good is presented fourth in a series of valuation tasks.

Overall then these results again strongly support the DPH assertion that repetition and the learning opportunities it affords results in a significant improvement in the theoretical consistency of preferences. By contrast these findings contrast markedly with the predictions of the coherent arbitrariness hypothesis which suggested that there should be no significant reduction in anchoring across repeated valuation tasks.

5. CONCLUSIONS

We have developed a new approach to eliciting stated preferences for non-market goods; the LDCV. Employing this we have found clear evidence of both institutional learning and value learning in repeated responses to CV questions. Valuations of an initial good exhibited typical anomalies, namely within-person inconsistencies in valuations of that good and across-person anchoring effects. In particular analysis of trends in within-good valuation differences and in anchoring show significant reductions in both anomalies. Indeed by the time respondents have undertaken a number of CV valuations both anomalies disappear. The tests applied are far from trivial, indeed the existing literature shows that they are rarely satisfied. Indeed our test for institutional learning provides what is to our knowledge the first instance of coherence between first and second response distributions recorded by any DB study to date; a result which defies what had become accepted as a stylised finding for such studies. Here we find evidence that individuals quickly learn the operating rules of a contingent market and yield internally consistent valuations once they have gained this experience. Similarly our value learning test concerns one of the most persistent anomalies identified in a host of economic and psychological studies; the anchoring effect. Here our results suggest that value learning occurs as subjects gain experience through repeated valuation tasks.

These findings appear to strongly support the DPH, suggesting that CV respondents require experience of both the operating rules of the contingent market and of the type of goods in question before they can provide theoretically consistent valuation responses. Such findings seem to be in accordance with the growing body of experimental results highlighting the importance of learning effects and consequent experience as vital precursors to the revelation of robust preferences. This in turn highlights a central dilemma for CV research. As a result of our findings we question the standard presumption in favour of the first response SB design applied to the valuation of a single good. On reflection it may seem unreasonable that inexperienced members of the public can provide robust valuations of goods without prior experience of both the CV institution itself and the task of valuing similar goods. The SB format fails to offer the repetition, learning and experience possibilities of real markets and is therefore particularly prone to framing effects such as anchoring and focussing illusion which appear to have upwardly biased SB values for the initial good.

We feel that these are significant failings which should be addressed through improved elicitation techniques. However, these failings should not be seen as an excuse to ignore issues of incentive compatibility. Specifically we feel that an ideal elicitation format should use repetition and exposure to allow respondents the opportunity to gain experience of the valuation mechanism (institutional learning) and experience of the good under investigations (value learning) prior to the use of an incentive compatible valuation question. One simple innovation would be to use 'practice' questions (such as those described by Plott and Zeiler,

2005, or the increasingly common choice experiment method discussed by Adamowicz, et al., 1998, and Louviere, et al., 2000] to develop institutional and value learning. This exercise could then be followed by a single, overtly incentive-compatible contingent valuation question. Such an approach, we suspect, would address much of the preference malleability and consequent anomalies observed in many prior CV studies.

Finally, considering the more general and fundamental focus of this paper, our results find no evidence to support the contention that the stable preferences formed through repetition and experience (even within a contingent market) are at variance with standard theory. Indeed, the plethora of tests presented describe trends which clearly show a movement towards theoretically consistent preferences and away from those characterised by coherent arbitrariness. This, fundamental finding suggests that a radical reconception of underlying theory in this respect is not, at present, clearly necessary although we acknowledge that this is a single study and further testing of this challenge remains a research priority.

References

- Adamowicz, W., Boxall, P., Williams, M., Louviere, J., 1998, Stated preference approaches for measuring passive use values: choice experiments and contingent valuation, *American Journal of Agricultural Economics*, 80(1): 64-75.
- Alberini, A., Kanninen, B. and Carson, R.T., 1997, Modelling response incentive effects in dichotomous choice contingent valuation data, *Land Economics*, 73: 309-324.
- Ariely, D., Loewenstein, G., and Prelec, D., 2003 "Coherent arbitrariness": Stable demand curves without stable preferences, *Quarterly Journal of Economics*, 118 (1) : 73-105.
- Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R. and Schuman, H., 1993, Report of the NOAA panel on contingent valuation, *Federal Register*, 58: 4601-4614.
- Bateman, I.J., Carson, R.T., Day, B., Hanemann, W.M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Özdemiroğlu, E., Pearce, D.W., Sugden, R. and Swanson, J., 2002, *Economic Valuation with Stated Preference Techniques: A Manual*, Edward Elgar Publishing, Cheltenham.
- Bateman, I.J., Cole, M., Cooper, P., Georgiou, S., Hadley, D. and Poe, G.L., 2004, On visible choice sets and scope sensitivity, *Journal of Environmental Economics and Management*, 47: 71-93.
- Bateman, I.J., Langford, I.H., Jones, A.P. and Kerr, G.N., 2001, Bound and path effects in multiple-bound dichotomous choice contingent valuation, *Resource and Energy Economics*, 23(3): 191-213.
- Bateman, I.J., Langford, I.H., Turner, R.K., Willis, K.G. and Garrod, G.D., 1995, Elicitation and truncation effects in contingent valuation studies, *Ecological Economics*, 12(2):161-179.
- Bateman, I.J., Munro, A., Rhodes, B., Starmer, C. and Sugden, R., 1997, A test of the theory of reference-dependent preferences, *Quarterly Journal of Economics*, 112(2): 479-505.
- Binmore, K., 1994, *Playing Fair*, MIT Press.
- Binmore, K., 1999, Why Experiment in Economics?, *The Economic Journal* 109, F16-F24.
- Bjornstad, D., Cummings, R., Osborne L., 1997, A Learning Design for Reducing Hypothetical Bias in the Contingent Valuation Method, *Environmental and Resource Economics*, 10: 207-221.
- Boyle, K.J. and Bishop, R.C., 1988, Welfare measurements using contingent valuation: a comparison of techniques, *Journal of the American Agricultural Association*, February: 20-28.
- Boyle, Kevin. J., Michael P. Welsh and Richard C. Bishop, 1993, The Role of Question Order and Respondent Experience in Contingent Valuation Studies, *Journal of Environmental Economics and Management*, 25(1), 45–55.
- Braga, Jacinto and Chris Starmer, 2005, Preference Anomalies, Preference Elicitation and the Discovered Preference Hypothesis, *Environmental and Resource Economics*, 32(1): 55-89.
- Burton A.C., Carson K.S., Chilton S.M., Hutchinson, W.G., 2001, Are Hypothetical Referenda Incentive Compatible? Paper Presented at the Association of Environmental and Resource Economists (AERE) Sessions at the *Allied Social Science Association Annual Conference*, New Orleans, Louisiana, January 2001
- Cameron, T.A. and Quiggin, J., 1994, Estimation using contingent valuation data from a "dichotomous choice with follow-up" questionnaire, *Journal of Environmental Economics and Management*, 27, 218-234.
- Carson, Richard T., forthcoming, *Contingent Valuation: A Comprehensive Bibliography and History*. Northampton, MA: Edward Elgar.
- Carson, R.T., Flores, N.E. and Hanemann, W.M., 1998, Sequencing and valuing public goods, *Journal of Environmental Economics and Management*, 36; 314-323.
- Carson, R.T., Groves, T. and Machina, M.J., 2000, Incentive and informational properties of preference questions, *Plenary Address, Ninth Annual Conference of the European Association of Environmental and Resource Economists (EAERE)*, Oslo, Norway.

- Carson, R.T., Mitchell, R.C., Hanemann, W.M., Kopp, R.J., Presser, S. and Ruud, P.A., 1994, Contingent valuation and lost passive use: damages from the Exxon Valdez, *Discussion Paper 94-18*, Resources for the Future, Washington, D.C.
- Carson, Richard T., Robert C. Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud., 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.
- Carson, R.T., Mitchell, R.C., Hanemann, W.M., Kopp, R.J., Presser, S., and Ruud, P.A., (2003) Contingent Valuation and Lost Passive Use: Damages from the Exxon Valdez, *Environmental and Resource Economics*.
- Carson, R.T. and Mitchell, R.C., 1987, *Economic Value of Reliable Water Supplies for Residential Water Users in the State Water Project Service Area*, report to the Metropolitan Water District of Southern California.
- Chapman, Gretchen B., and Eric J. Johnson, 1999, "Anchoring, Activation, and the Construction of Values," *Organizational Behavior and Human Decision Processes*, 79: 115–153.
- Cherry, Todd L., Thomas D. Crocker, and Jason F. Shogren. 2002. "Rationality Spillovers," *Journal of Environmental Economics and Management*.
- Coursey, D.L., Hovis, J.L. and Schulze, W.D., 1987, The disparity between willingness to accept and willingness to pay measures of value, *Quarterly Journal of Economics*, 102: 679-90.
- Cox, J. C. and D. M. Grether, 1996, "The Preference Reversal Phenomenon: Response Mode, Markets and Incentives." *Economic Theory*, 7: 381–405.
- Cummings, R. G., Elliott S., Harrison G. W. and Murphy J., 1997, Are Hypothetical Referenda Incentive Compatible? *Journal of Political Economy*, 105(3): 609-621
- DeShazo, J.R., 2002, "Designing Transactions without Framing Effects in Iterative Question Formats," *Journal of Environmental Economics and Management*, 43, 360-385.
- Epley, Nicholas, and Thomas Gilovich, 2001, "Putting Adjustment Back in the Anchoring and Adjustment Heuristic: Differential Processing of Self-Generated and Experimenter-Provided Anchors," *Psychological Science*, 12: 391–396.
- Gibbard A., 1973, Manipulation of Voting Schemes: A General Result, *Econometrica*, 41: 587-601.
- Green, Donald, Karen E. Jacowitz, Daniel Kahneman and Daniel McFadden, 1998, Referendum contingent valuation, anchoring, and willingness to pay for public goods, *Resource and Energy Economics*, 20: 85-116.
- Greene, W.H., 1993, *Econometric Analysis*, New York: Macmillan.
- Hanemann, W.M. and Kanninen, B. 1999, The statistical analysis of discrete-response CV data, in Bateman, I.J. and Willis, K.G. (eds.) *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, Oxford University Press, pp. 302-441.
- Hanemann, M.; Loomis, J. and Kanninen, B., 1991, Statistical efficiency of double-bounded dichotomous choice contingent valuation, *American Journal of Agricultural Economics*, 73: 1255-1263.
- Harris, C.C., Driver, B.L. and McLaughlin, M.J., 1989, Improving the contingent valuation method: a psychological approach, *Journal of Environmental Economics and Management*, 17:213-229.
- Hausman, J. 1993, ed. *Contingent Valuation: A Critical Assessment*. Amsterdam: North-Holland.
- Herriges, J.A. and Shogren, J. F., 1996, Starting point bias in dichotomous choice valuation with follow-up questioning, *Journal of Environmental Economics and Management*, 30(1): 112-131.
- Hoeffler, Steve and Dan Ariely, 1999, Constructing stable preferences: A look into dimensions of experience and their impact on preference stability, *Journal of Consumer Psychology*, 11: 113-139.

- Hoehn, J.P. and Randall, A., 1987, A satisfactory benefit cost indicator from contingent valuation. *Journal of Environmental Economics and Management* 14(3):226-247.
- Jacowitz, Karen E., and Daniel Kahneman, 1995, "Measures of Anchoring in Estimation Tasks," *Personality and Social Psychology Bulletin*, 21: 1161–1166.
- Johnson, Eric J., and David A. Schkade, 1989, Bias in utility assessments: Further evidence and explanations, *Management Science*, 35: 406-424.
- Kahneman, D., 1986. Comments, in Cummings, R.G., Brookshire, D.S. and Schulze, W.D. (eds.) *Valuing Environmental Goods: A State of the Arts Assessment of the Contingent Method*, Rowman and Allanheld, Totowa, N.J.
- Kahneman, D., Slovic, P. and Tversky, A., 1982. *Judgement Under Uncertainty: Heuristics and Biases*, Cambridge University Press, New York.
- Kahneman, D. and Tversky, A., 1982. The psychology of preferences, *Scientific American*, 246(1):160-173.
- Langford, I.H., Bateman, I.J., Jones, A.P. Langford, H.D. and Georgiou, S., 1998, Improved estimation of willingness to pay in dichotomous choice contingent valuation studies, *Land Economics*, 74(1): 65-75.
- List, John A. 2004, Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace, *Econometrica*, 72(2): 615-625.
- List, John A. 2001, Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards, *American Economic Review*, 91(5); pp. 1498-1507.
- List, J.A., 2002, Neoclassical Theory Versus Prospect Theory: Evidence from the Marketplace, *Working Paper*, University of Maryland
- List, J.A., 2003, Does market experience eliminate market anomalies?, *Quarterly Journal of Economics* (February): 41-72.
- List, J.A., 2002, Preference reversals of a different kind: The more is less phenomenon, *American Economic Review*, (December): 1636-1644.
- List, John A. and David Lucking-Reiley, 2000, Demand Reduction in Multiunit Auctions: Evidence from a Sportscard Field Experiment, *American Economic Review*, 2000, 961-972.
- List, J. and J. Shogren, 1999, Price Information and Bidding Behavior in Repeated Second-Price Auctions, *American Journal of Agricultural Economics*, 81, 942-49.
- Loomes, G. and R Sugden, Testing Different Stochastic Specifications of Risky Choice, *Economica*, 1998, 65: 581-98.
- Loomis, John, Thomas Brown, Beatrice Lucero and George Peterson, 1997, Evaluating the validity of the dichotomous choice question format in contingent valuation, *Environmental and Resource Economics* 10: 109–123.
- Louviere, J, Hensher, D and Swait, J., 2000, *Stated Choice Methods: Analysis and Application*, Cambridge University Press, Cambridge.
- Matthews, David I., Glass C. A. and Hutchinson W. George, 2003, *Use of the Bivariate Probit in the Joint Estimation of Willingness to Pay for Several Animal Welfare Improvements* Paper in Proceedings of the Agricultural Economics Society Annual Conference at University of Plymouth April 2003.
- McCollum, Daniel, W., and Kevin J. Boyle, 2005, The Effect of Respondent Experience/Knowledge in the Elicitation of Contingent Values: An Investigation of Convergent Validity, Procedural Invariance and Reliability, *Environmental & Resource Economics*, 30: 23–33.
- McFadden, D., 1994, Contingent valuation and social choice, *American Journal of Agricultural Economics*, 76: 689-708.
- Mitchell, Robert Cameron, and Richard T. Carson, 1989, *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington: Resources for the Future, 1989.
- Plott, Charles R., 1996, "Rational Individual Behavior in Markets and Social Choice Processes: The Discovered Preference Hypothesis," in *Rational Foundations of Economic Behavior*. K. Arrow, E. Colombatto, M. Perleman, and C. Schmidt, eds. London: Macmillan and NY: St. Martin's, pp. 225–50.

- Plott, Charles R. and Kathryn Zeiler, 2003, The willingness to pay / willingness to accept gap, the 'endowment effect' and experimental procedures for eliciting valuations, Social Science Working Paper 1132, Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena.
- Plott, Charles R. and Kathryn Zeiler, 2005. The Willingness to Pay-Willingness to Accept Gap, the "Endowment Effect," Subject Misconceptions, and Experimental Procedures for Eliciting Valuations, *The American Economic Review*, Volume 95, Number 3, Page 530 – 545.
- Roach, B., K. J. Boyle, J. C. Bergstrom and S. D. Reiling (1999), 'The Effect of Instream Flows on Whitewater Visitation and Consumer Surplus: A Contingent Valuation Application to the Dead River, Maine', *Rivers*, 7(1), 11–20.
- Roberts, K.J., Thompson, M.E. and Pawlyk, P.W., 1985. Contingent valuation of recreational diving at petroleum rigs, Gulf of Mexico. *Transactions of the American Fisheries Society*, 114:155-165.
- Satterthwaite, M. A., 1975, Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems of Voting Procedures and Social Welfare functions, *Journal of Economic Theory*, 10: 187- 217.
- Scarpa, R. and Bateman. I.J., 2000 Efficiency gains afforded by improved bid design versus follow-up valuation questions in discrete choice CV studies, *Land Economics*, 76(2): 299-311.
- Schkade, D. A., & Kahneman, D., 1998, Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science*, 9, 340-346.
- Shogren, J., S. Cho, C. Koo, J. List, C. Park, P. Polo and R. Wilhelm, 2001, "Auction mechanisms and the measurement of WTP and WTA", *Resource and Energy Economics*, 23: 97-109.
- Shogren, J.F., Shin, S.Y., Hayes, D.J. and Kliebenstein, J.B., 1994, Resolving differences in willingness to pay and willingness to accept, *American Economic Review*, 84: 255-70.
- Slovic, Paul, 1995, The construction of preferences, *American Psychologist*, 50: 364-371.
- Strack, Fritz, and Thomas Mussweiler, 1997, "Explaining the Enigmatic Anchoring Effect: Mechanisms of Selective Accessibility," *Journal of Personality and Social Psychology*, 73: 437–446.
- Taylor L. O., McKee, M., Laury, S.K. and Cummings, R. G., 2001, Induced Value Tests of the Referendum Voting Mechanism, *Economic Letters*, 71(1): 61-6.
- Tversky, Amos and Daniel Kahneman, 1974, Judgement under uncertainty: Heuristics and biases, *Science*, 185: 1124-1131.
- Hal Varian, H. (1999) *Intermediate Microeconomics: A Modern Approach*, (Fifth Edition), Norton, New York.
- Welsh, M.P. and Bishop, R.C., 1993, Multiple bounded discrete choice models, W-133, *Benefits and Costs Transfer in Natural Resource Planning: Sixth Interim Report*, John C. Bergstrom, compiler, Department of Agricultural and Applied Economics, University of Georgia, Sept. 1993.

Appendix: Developing a within-respondent technique for assessing the statistical significance of Δ_j .

We wish to test the proposition that differences in estimates of mean WTP between SB and DB models are zero (i.e. $H_0: (\Delta_j = [\mu_{SBj} - \mu_{DBj}]) = 0$), repeating this test for each good in turn. To test this a t-statistic can be calculated as follows:

$$t = (\mu_{SBi} - \mu_{DBi}) / \sqrt{\text{Var}(\mu_{SBi} - \mu_{DBi})} \quad (A1)$$

When two means μ_1 and μ_2 are estimated using two independent samples those estimates are uncorrelated, i.e. covariance is zero ($\text{Covar}(\mu_1, \mu_2) = 0$) and the variance of the difference of two such means can be obtained by summing the individual variances. The variance of the differences of two such independent means μ_1, μ_2 can be computed as:

$$\text{Var}(\mu_1 - \mu_2) = \text{Var}(\mu_1) + \text{Var}(\mu_2) \quad (A2)$$

This can be used to test whether differences in mean WTP between independent samples are significantly different from zero. However, when we test the significance of differences between the first response data used in an SB model and the same first responses supplemented by follow-up question responses, as per a DB exercise, the samples can no longer be considered independent since both estimates are computed using the same initial responses from the same individuals. Hence when estimating $\text{Var}(\mu_{SBi} - \mu_{DBi})$ the $\text{Covar}(\mu_{SBi}, \mu_{DBi})$ cannot be ignored as in (6) above. Due to these difficulties no statistical test exist to examine whether $(\mu_{SBi} - \mu_{DBi}) = 0$ as the $\text{Var}(\mu_{SBi} - \mu_{DBi})$ cannot be obtained from a known closed-form solution. A novel Monte Carlo approach proposed by Matthews et al (2003) is developed in the present paper to take account of the missing $\text{Covar}(\mu_{SB}, \mu_{DB})$ and to test whether $(\mu_{SBi} - \mu_{DBi}) = 0$. To the authors knowledge this is the first statistical test within the CV literature which tests the consistency of welfare estimates from SB and DB models for the same sample.

Monte Carlo techniques [Efron and Tibshirani, 1993] are employed to obtain the distribution of differences $(\mu_{SBi} - \mu_{DBi})$ in the sample means and to compute an estimate of the $\text{Var}(\mu_{SBi} - \mu_{DBi})$. These methods use resampling routines to create a number of sample estimates derived from the original sample using the same sampling method as used to obtain the original sample. Here we obtain for each sample and good, an estimate $(\mu_{SBi} - \mu_{DBi})$. The jack-knife method is chosen to estimate the variance $\text{Var}_J(\mu_{SBi} - \mu_{DBi})$.

For an original sample of size n the jack-knife method uses a set of sample estimates θ which are computed from n new samples each of which contain $n-1$ observations taken from the original sample. In this study one observation refers to all data provided by an individual, so when an observation is dropped so are all responses given by that individual, replicating the original sampling method. The set of samples are composed from the original sample with each sample having a different observation removed. The statistic of interest for jack-knife sample k is θ_k . Thus a distribution of that statistic which reflects the original population distribution is created. Here θ_k is computed as the difference in mean WTP estimates obtained from the SB and DB models, i.e. $(\mu_{SBi} - \mu_{DBi})$.

The variance for the jack-knife estimate Var_J is obtained from:

$$\theta = \{\theta_1, \theta_2 \dots \theta_n\}$$

where the estimate for the difference in the k^{th} jack-knife sample θ_k is:

$$\begin{aligned}\theta_k &= (\mu_{SBI} - \mu_{DBi}) \\ \theta_m &= \Sigma (\mu_{SBI} - \mu_{DBi})/n \\ \text{Var}_J(\boldsymbol{\theta}) &= (n-1)/n \Sigma(\theta_i - \theta_m)^2\end{aligned}\tag{A3}$$

ⁱ Within the CV context this argument is developed through Hoehn and Randall [1987] and Carson et al., [1999].

ⁱⁱ Note that a small number of papers also examine whether pre-existing exposure to a good (as opposed to the learning through repetition examined in the present paper) results in more theoretically consistent responses within CV studies [Boyle et al. 1993; Roach et al. 1999; McCollum and Boyle, 2005]. However, these papers yield conflicting findings.

ⁱⁱⁱ The idea behind the focusing illusion can be summarised in the proverb that “Nothing is as important as when you think about it”.

^{iv} Anchoring or starting point effects are one of the most well documented response heuristics, being replicated in a host of CV, economic and psychological studies [Kahneman, Slovic and Tversky, 1982. Kahneman and Tversky, 1982; Roberts et al., 1985; Kahneman, 1986; Harris et al., 1989; Bateman et al., 1995; Jacowitz and Kahneman, 1995; Strack and Mussweiler, 1997; Chapman and Johnson, 1999; Epley and Gilovitch, 2001].

^v The Herriges and Shogren [1996] paper in itself is highly pertinent to the present study in that they find significant anchoring in DB responses given by less experienced visitors to a recreational site, but no significant anchoring amongst responses from more experienced local residents. This provides support for the existence of an experience driven, value learning process.

^{vi} In accordance with Boyle and Bishop [1988] an initial pilot survey asked respondents an open-ended WTP question. Responses were then used to refine a vector of initial bids following the design efficiency advice of Hanemann and Kanninen [1999]. The resultant vector placed bid-levels for the first response question at the 90th percentile (£5.00), the 65th percentile (£2.00), the 35th percentile (£1.50) and the 15th percentile (£1.00). The bid vector for the second response question supplements these with two more extreme bid-levels at the 95th percentile (£10.00) and the 7th percentile (£0.50).