

**Learning Design Contingent Valuation (LDCV):
NOAA guidelines, preference learning and coherent arbitrariness**

Ian J. Bateman,¹ Diane Burgess²,
W. George Hutchinson³ and David I. Matthews^{2*}

1st November 2007

JEEM ms #16223

¹Centre for Social and Economic Research on the Global Environment (CSERGE), University of East Anglia, Norwich, NR4 7TJ, United Kingdom. Associate Professor, University of Western Australia.

²Agri-Food and Biosciences Institute (AFBI), Belfast, BT9 5PX, UK.

³ Gibson Institute for Land Food and Environment, Queens University Belfast, BT7 1NN, UK.

Corresponding author: Ian J. Bateman. Email: i.bateman@uea.ac.uk Tel: +44 (0) 1603 593125

* The convention of alphabetic authorship is followed. No seniority of authorship attributed.

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, [3]); (ii) learned or 'discovered' through a process of repetition and experience [37, 43]; (iii) internally coherent but strongly influenced by some initial arbitrary anchor [2]. 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. In so doing we show that such a 'Learning Design Contingent Valuation' method overturns the 'stylised facts' of bias and anchoring within the double bound dichotomous choice elicitation format.

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: Preferences, non-market goods, contingent valuation, learning, coherent arbitrariness, NOAA guidelines, animal welfare, willingness to pay (WTP), similarity.

I. Introduction

The nature of individuals' preferences is fundamentally crucial to the underpinnings of microeconomic theory [52]. However, the process through which such preferences are generated 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 low experience goods, concerns the speed at which individuals can form stable preferences for relatively novel goods presented in unfamiliar markets. This question is important for valuation research 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 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 services 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 [15]. Clearly a key concern here is to use study designs which address the issue of *a-priori* 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 Valdez oil spill [18, 31]; debate which was substantially addressed through the influential NOAA panel report on CV [3]. This report provided guidelines for future applications, a key recommendation being the method through which WTP responses should be elicited. Although a wide variety of elicitation techniques are available [4, 42], the NOAA panel recommended the use of a ‘one-shot’ or single-bound (SB) dichotomous choice referendum style question.

The underlying argument for rejecting all but the SB response format can be traced back to the work of Gibbard [27] and Satterthwaite [46] establishing the *potential* incentive compatibility of one-shot referenda (see also [17]). 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 agencies 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 settings, while some studies find convergence of voting responses between hypothetical markets and those in real, consequential referenda, other studies report divergent results [13, 22, 38, 49]. Given that CV applications typically value novel 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 [43]. 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 [11] refer to as ‘institutional learning’) and discover features of their own preferences (‘value learning’, *ibid*). The first response SB format precludes either institutional or value learningⁱ 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*. 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 incentive compatibility of 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 (e.g. [44]); reduction in the preference reversal anomaly in both real and hypothetical payment formats (e.g. [21]); and, perhaps most pertinently, reduction in preference anomalies amongst more experienced traders or choice makers (e.g. [37]). Evidence for the

positive impact of repetition can also be observed within the stated preference valuation literature. Notably Hu et al., [33] and DeShazo and Fermo [24] speculate that preference consistency may increase as respondents move through the repeated questions of a choice experiment. Similarly some CV studies provide support for the hypothesis that respondents with prior experience of a good have different and more consistent preferences than do inexperienced respondents (e.g. [32])ⁱⁱ.

This experimental and stated preference 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 would suggest that such responses are prone to be influenced by a variety of choice heuristics and framing effects resulting in apparently anomalous preferences [51]. For example, more recent work defining out the ‘focusing illusion’ [48]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 become linked to any available ‘anchor’, such as the initial SB bid-level itself, which may be taken as some clue to the ‘correct’ value of the good in question [28]^{iv}. In an innovative recent paper, Ariely, Loewenstein and Prelec [2] refine the anchoring argument. Through experimental investigations they show that, while an individual’s choices are typically internally coherent, nevertheless they can also be strongly anchored to some initial, demonstrably arbitrary starting point (discussed further in Section II). In such cases values can be manipulated up or down by altering the 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, at least to some degree, 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 [17]; (ii) learned or 'discovered' through a process of repetition and experience [37,43]; (iii) internally coherent but liable to be strongly influenced by some initial arbitrary anchor [2]. 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 within valuation research. 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 view (DPH) 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 favours a repeated questioning methodology which encourages learning regarding both the market institution and preferences themselves. Both the first and second views 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' CV approaches 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 stated values both before and after such learning opportunities. Further design features and our empirical testing protocol allow us to examine findings against the predictions of (1) *a-priori* well formed preferences (2) the DPH and (3) 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^v. 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 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.

II. LDCV Research Design

II.A. Testing for Institutional Learning and Consistent Values in LDCV

Repetition is the seedcorn of learning and experience within the DPH. Consequently 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; a repeated examination of value coherence within the double bound (DB) dichotomous choice format.

The DB approach was introduced by Hanemann *et al.* [30] 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 between the estimated mean WTP derived from SB responses and that obtained from the first and second responses combined within a DB analysis. A number of studies report such results [7, 14, 23, 41], typical of which are the findings of McFadden [41] 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). This weight of evidence of inconsistent values from SB and DB valuations of a single good, or of a first good in a sequence of valuations, together with the imprecise estimates in initial valuations appears inconsistent with *a-priori* well formed preferences.

Some commentators have argued that the DB internal inconsistency anomaly arises from changes in incentive compatibility between the first and second response [1, 13, 17]. 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 [7, 20] and, given further multiple valuation experience, are liable to revise their responses if given the opportunity [5]. Given such findings 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 through repeat valuations, respondents can receive feedback about the DB institution, then we might expect the anomalous discrepancy between SB and DB responses to diminish with such 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. 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 of the hypothesis that the difference between SB and DB values does not decline across successive goods valued. Our test focus therefore becomes an examination of the SB-DB difference, with the DPH expectation being that there will be a significant decline in this difference as we compare its initial level (from the first good valued) and its level for subsequent goods. In contrast, if preferences are well formed from the outset then we should not expect any substantial decline in this difference as additional goods are considered^{vi}. We test this by examining both the absolute level of difference and trends in that level across goods.

II.B. Testing for Value Learning and Non Arbitrary Values in LDCV

While our institutional learning test focuses upon SB and DB coherence, the findings of Ariely et al [2] demonstrate that such coherence is a necessary but not, on its own, sufficient condition for the identification of theoretically consistent values. As they argue:

“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” [2, 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 economic 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), 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 is developed in LDCV along the lines suggested by Herriges and Shogren [32] which provides a strong test for discriminating between coherent arbitrariness and value learning in CV responses. In this approach the initial bid-level presented in the SB question regarding a given good provides an arbitrary, initial stimuli. By varying these bids both across respondents and across goods, the LDCV allows us to estimate an ‘anchoring parameter’ (γ) between the initial bid and response to the second bid. We can

then examine trends in the degree of anchoring observed across a sequence of goods valued in the LDCV design.. The coherent arbitrariness expectation is that γ should be initially significant and not decrease significantly as successive goods are valued. If however, we observe a significant decline in the level of anchoring across valuation tasks then this argues against the coherent arbitrariness view and mitigates in favour of the DPH argument that repetition and learning in LDCV designs will improve both the theoretical consistency and non-arbitrariness of values.

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. Therefore we reports tests concerning both the absolute value of γ and trends in those values.

II.C. Choice of goods and survey implementation for the LDCV

In order to perform both our institutional and value learning tests we require repetition of valuation tasks both within and across goods. Both the DPH and coherent arbitrariness arguments would mitigate in favour of using goods which are of a “similar kind” [3, 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 related goods, each of which improved the farm living conditions for a different species of animal to be paid for via a compulsory tax on all foodstuffs. This type of good allowed us to experiment with the issue of ‘similarity’ within the learning process. Specifically we chose two initial goods which were highly similar (living conditions for laying hens and living conditions for chickens) the idea being that this

would maximise the potential for value learning. However, for the third good we chose a species which, although still a common farm animal, was sufficiently distinct to potentially disrupt or partially restart the value learning process (living conditions for cows). This was followed by a fourth and final species similar to the third (living conditions for pigs), the reasoning being that this would allow any restarted value learning process to continue. Therefore, while all goods are reasonably similar, similarity is greatest between the first pair (small birds) and last pair of goods (large animals) and somewhat less across these two pairs. This permits an ancillary examination of the link between the degree of similarity and the extent of the value learning process^{vii}. Each animal welfare good was presented as mutually exclusive, thereby avoiding substitution and allied sequencing effects [5, 6, 16]^{viii}.

Each good was valued using a DB elicitation format. An initial sample (Sample 1) was presented with all four goods, given in the order discussed, thus permitting the repetition of valuation tasks necessary to test for 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 1 (living conditions for pigs). The DPH expectation here is that, controlling for the good, the degree of any institutional anomalies (disparities between 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 as 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 non-significant after valuing several goods is an open empirical question.

For notational purposes we denote any good presented to a respondent in the LDCV 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 the LDCV 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 ($PIGS_4^1$)

In contrast, respondents in the DB Sample 2 were only 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 [10] as refined by Hanemann and Kanninen [29] 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.^{ix}

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. Subsequent testing confirmed that the two samples did not differ significantly in terms of any of the considerable number of socio-economic or demographic characteristic variables collected as part of the survey (including

gender, age, income, educational background, employment status, food purchasing frequency, etc.).

III. 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 [29]. 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}). Following Hanemann *et al.*, [30] we use a logistic cumulative distribution function to model response data.

III.A. Testing for Institutional Learning: Consistency of mean WTP μ_{SBj} and μ_{DBj} in LDCV

Estimates of mean WTP (μ_{SB} , μ_{DB}) are computed for SB and DB models for each good j following Hanemann *et al.*, [30]. A measure of the difference between μ_{SB} and μ_{DB} for good j in sample i (denoted Δ_j^i) is expressed as:

$$\Delta_j^i = \mu_{SBj} - \mu_{DBj}$$

We wish to test the proposition that differences in estimates of mean WTP between SB and DB models are zero, i.e. $H_0: (\Delta_j^i = 0)$, repeating this test for each good in turn. Differences in these statistics will also be used to examine trends in the difference across valuation tasks (i.e. across goods, examining whether $\Delta_j^i \geq \Delta_{j+m}^i$ for $m > 0$).

Testing H_0 above requires an econometric technique that controls for use of the same sample and the non-independence (within-respondent) of the first and second bound responses. When testing the significance of differences between estimates using 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 an estimate of the variance $\text{Var}(\Delta_j^i)$ cannot be obtained from a known closed-form solution. Accordingly a non-parametric resampling approach was used to estimate $\text{Var}(\Delta_j^i)$ and hence test H_0 . Monte Carlo methods can be used to obtain an estimate of the sampling distribution of differences Δ_j^i and thus estimate $\text{Var}(\Delta_j^i)$ for each good.

The jackknife variance estimator [25] is used here to estimate $\text{Var}(\Delta_j^i)$. For each jackknife sample k the value θ_k is estimated as the difference in mean WTP estimates obtained from the SB and DB models, i.e. Δ_j^i . The estimated variance $\text{Var}(\Delta_j^i)$ is then obtained from:

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

where n is the sample size and the θ_k is the estimate of Δ_j^i using the k^{th} Jackknife sample:

$$\theta_k = (\Delta_j^i)_k$$

The estimate of variance is obtained using all jackknife samples thus:

$$\text{Var}(\Delta_j^i) = [(n-1)/n] \sum_{k=1}^n (\theta_k - \Delta_j^i)^2 \quad (1)$$

Hence H_0 can be tested using the t statistic obtained from the jackknife estimate of variance;

$$t = \Delta_j^i / [\text{Var}(\Delta_j^i)]^{1/2}$$

As discussed previously, once values for Δ_j^i are established, the focus of our testing extends to examine *trends* within the statistical significance of this variable across valuation tasks (i.e. across goods), for which again our jackknife variance estimates are employed^x. 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. $\Delta_j^i \geq \Delta_{j+m}^i$) defining six tests within Sample 1.

In addition to these trend analyses 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 at its start (i.e. comparing Δ_4^1 with Δ_1^2) provides a further distinguishing test between these hypotheses and *a-priori* well formed preferences.

III.B. Testing for Value Learning: Robustness against Anchoring Effects in LDCV

Following established methods [29, 32] 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 [32] the revised WTP in response to the second bid is:

$$\text{WTP}_r = (1-\gamma)\text{WTP}_o + \gamma b_1 \quad (2)$$

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 the second response in a DB format (denoted b_2) becomes b_{2r} as follows:

$$b_{2r} = (b_2 - \gamma b_1)/(1-\gamma) \quad (3)$$

The Log Likelihood function for the anchoring model is obtained by substituting the b_{2r} value into the standard DB likelihood.

Calculating γ for responses to each good (i.e. γ_j^i) allows us to test the empirical significance of this parameter and so provide 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 the sequence of goods valued. 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 (although our perturbing of the degree of similarity between the second and third good allows us to test for any re-emergence of anchoring which this might induce). 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.

IV. Results

IV.A. SB and DB models

Table 1 presents parsimoniously specified logistic SB and DB models for each good estimated as per Hanemann *et al.*, [30]. The models provide parameter estimates of the coefficients α and 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 (and as noted, there was no significant difference between the samples in this respect). 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.

INSERT TABLE 1 ABOUT HERE

IV.B. Results from the Institutional Learning Tests

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 SB and DB formats. 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 2 details results from this analysis test. As discussed, standard errors for Δ_j^i are calculated using the Jackknife 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.

INSERT TABLE 2 ABOUT HERE

Considering Table 2, 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} [17]. Given these prior findings, the results set out in Table 2 are remarkable. Considering the first good valued by both Sample 1 and 2 ($HENS_1^1$ and $PIGS_1^2$ 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 $HENS_1^1$) we have $\mu_{SBi} = \text{£}4.72$ while $\mu_{DBi} = \text{£}2.74$. Variance is also correspondingly large reflecting uncertain preferences in the first SB valuation. 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 non-significant. Indeed there is a clear pattern of significance running across successive Sample 1 valuation tasks, with $\mu_{SBi} - \mu_{DBi}$ differences (Δ_j^i) declining in significance 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.

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 and Groves [17] argument in favour of the SB format within a hypothetical CV study as no criterion (demonstrably correct) value is available for such a context. However, the weight of evidence from experimental studies is now reinforced by the present results and clear reduction in values across Sample 1. Together, these remarks suggest 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 2 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. Table 3 analyses these responses further by examining trends in Δ_j^i across goods.

INSERT TABLE 3 ABOUT HERE

The upper panel of Table 3 reports Δ_j^i trend findings for Sample 1. Within this the first three rows test and reject the hypothesis that there is no difference between Δ for the first good and Δ for all the subsequent goods. The trend towards a decreasing Δ and increasing SB-DB coherence is in line with the expectations of the DPH. Furthermore, the reduction in inconsistency predicted by the DPH appears to be getting stronger as respondents pass through the LDCV repeated valuation design. The next three rows of the table test the hypothesis of no significant differences in the Δ for the second and subsequent goods valued. The results show little significant further decreases in Δ once respondents have finished the DB exercise for the first good. Comparison of the estimate of Δ for the first good with the corresponding measure for subsequent goods indicate that respondents very rapidly learn how the DB market works and that the associated anomaly, observed in all previous DB studies, quickly evaporates in the face of learning and experience of the DB mechanism. 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.

Overall then, the results reported in Tables 2 and 3 strongly support the DPH expectation of institutional learning and consistency arising from increased familiarity and experience with the contingent market and mitigate against the hypothesis of *a-priori* consistent preferences.

IV.C. Results from the Value Learning Tests

Although our institutional learning test indicates that while the valuations of a good become more internally consistent as respondents become more familiar with the operating rules of the DB mechanism, we need to determine whether this consistency is procedurally invariant and therefore conforms to standard theory, or whether it is subject to the anchoring

effects symptomatic of coherent arbitrariness. By estimating the anchoring model specified in Equations (2) and (3) we obtain a series of models which allow for the presence of anchoring within DB responses. These are reported in Table 4.

INSERT TABLE 4 ABOUT HERE

In Table 4 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 within-good 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^{xi}. Furthermore, the pattern of significance across the goods valued by sample 1 is interesting. The highly significant anchoring observed for the first good ($HENS_1^1$) becomes clearly insignificant when respondents move on to value the highly similar second good ($CHICKS_2^1$). Equally interesting, when the somewhat less similar third good ($COWS_3^1$) is valued anchoring reappears as a feature of responses (although somewhat less strongly than for the initial good). However, anchoring very clearly disappears (with γ falling to its lowest level) when the final good ($PIGS_4^1$; more similar to the preceding third good) is valued. Similarity of goods does appear to play some logically interpretable role within the learning process shown by these results, which in turn appears to accord strongly with the predictions of the DPH. Our analysis of the trend in anchoring results is presented in Table 5.

INSERT TABLE 5 ABOUT HERE

As per our previous trend analysis, the upper panel of Table 5 reports findings regarding the trend in anchoring across the various goods valued by Sample 1. Within this, the first three rows test the hypothesis that the anchoring parameter γ on the first good is equal to that observed for subsequent goods. This hypothesis is consistently rejected; the degree of anchoring is lower in subsequent goods than in the initial good, a result which supports the expectations of the DPH rather than coherent arbitrariness. Considering the results of the next three rows we see that there is relatively little further reduction in γ once the initial good has been considered, although we do see some further significant reduction in γ between the third and fourth good. This result further strengthens the conjecture that the reduction in good-similarity experienced as respondents move to valuing the third good may have somewhat brought anchoring back into play within the response formation process, but that anchoring again reduced as respondents move on to the similar fourth good.

Overall these findings support the DPH assertion that the repetition and learning opportunities afforded by the LDCV result in a significant reduction in the anchoring of values as repeated valuations are made. These findings contrast with the predictions of the coherent arbitrariness hypothesis which gives no suggestion that there should be any reduction in anchoring across repeated valuations.

V. Conclusions

We have developed a new approach to eliciting stated preferences for non-market goods; the LDCV. Employing this approach we have found evidence of both institutional learning and value learning in repeated responses to CV questions. Valuations of an initial good exhibited typical anomalies, namely inconsistencies between SB and DB valuations of that good and anchoring effects. Analysis of trends in both within-good valuation differences and in anchoring show significant reductions in both anomalies as repeated valuations are made.

Indeed by the time respondents have undertaken a number of CV valuations both anomalies completely disappear. The consistency and anchoring 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 SB and DB response distributions recorded by any DB study to date; a result which defies what had become accepted as a stylised fact regarding 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 feedback on the nature of the DB question format through repeated valuation tasks.

These findings strongly support the DPH as opposed to the competing hypotheses of *a-priori* well formed preferences and coherent arbitrariness. They suggest 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. 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. 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, [44])^{xii} to develop institutional and value learning. This exercise could then be followed by a single, overtly incentive-compatible valuation question, emphasising the binding nature of the decision. Such an approach, we suspect, might 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 are at variance with standard theory. Indeed, the tests presented describe trends which clearly show a movement towards theoretically consistent preferences. This finding suggests that a radical reconception of underlying theory in this respect may not, at present, be clearly necessary although we acknowledge that this is a single study and that further testing of innovative challenges such as coherent arbitrariness remain a research priority.

References

- [1] Alberini, A., Kanninen, B. and Carson, R.T., 1997, Modelling response incentive effects in dichotomous choice contingent valuation data, *Land Economics*, 73: 309-324.
- [2] Ariely, D., Loewenstein, G., and Prelec, D., 2003 “Coherent arbitrariness”: Stable demand curves without stable preferences, *Quarterly Journal of Economics*, 118 (1) : 73-105.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] Bateman, I. J., Cole, M. A., Georgiou, S., and Hadley, D. J. (2006). Comparing contingent valuation and contingent ranking: A case study considering the benefits of urban river water quality improvements, *Journal of Environmental Management*, 79: 221-231.
- [7] 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.
- [8] 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.
- [9] 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.

- [10] 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.
- [11] Braga, Jacinto and Chris Starmer, 2005, Preference Anomalies, Preference Elicitation and the Discovered Preference Hypothesis, *Environmental and Resource Economics*, 32(1): 55-89.
- [12] Burgess, D.E., 2001 *Determining Society's values for Programmes to Improve the Welfare of Farm Animals in the UK*. Unpublished PhD Thesis, Queens University Belfast.
- [13] Burton A C, Carson K S, Chilton S M and Hutchinson W G (2007) Resolving Questions about Bias in Real and Hypothetical Referenda, *Environmental and Resource Economics*, in press (DOI 10.1007/s10640-007-9095-6).
- [14] 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.
- [15] Carson, Richard T., 2007, *Contingent Valuation: A Comprehensive Bibliography and History*. Northampton, MA: Edward Elgar.
- [16] 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.
- [17] Carson, R.T., and Groves, T. 2007, Incentive and informational properties of preference questions, *Environmental and Resource Economics*, 37(1): 181-210.
- [18] 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*.

- [19] 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.
- [20] Cooper, J., W. M. Hanemann, and G. Signorello, (2002) One-and-One-Half-Bound Dichotomous Choice Contingent Valuation, *Review of Economics and Statistics* 84(4): 742-750.
- [21] Cox, J. C. and D. M. Grether, 1996, "The Preference Reversal Phenomenon: Response Mode, Markets and Incentives." *Economic Theory*, 7: 381–405.
- [22] 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
- [23] DeShazo, J.R., 2002, "Designing Transactions without Framing Effects in Iterative Question Formats," *Journal of Environmental Economics and Management*, 43, 360-385.
- [24] DeShazo, J. R. and German Fermo (2002) Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency, *Journal of Environmental Economics and Management*, Volume 44, Issue 1, July 2002, Pages 123-143.
- [25] Efron, B. and R.J. Tibshirani, 1993, *An introduction to the bootstrap*, Chapman & Hall.
- [26] Fehr, Ernst and Tyran, Jean-Robert (2005) Individual Irrationality and Aggregate Outcomes, *Journal of Economic Perspectives*, 19(4): 43-66
- [27] Gibbard A., 1973, Manipulation of Voting Schemes: A General Result, *Econometrica*, 41: 587-601.
- [28] 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.
- [29] 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.

- [30] 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.
- [31] Hausman, J. 1993, ed. *Contingent Valuation: A Critical Assessment*. Amsterdam: North-Holland.
- [32] 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.
- [33] Hu, Wuyang, Wiktor, L. Adamowicz and Michele M. Veeman (2006) Labeling context and reference point effects in models of food attribute demand, *American Journal of Agricultural Economics*, 88 (4), 1034–1049.
- [34] Kahneman, D., Slovic, P. and Tversky, A., 1982. *Judgement Under Uncertainty: Heuristics and Biases*, Cambridge University Press, New York.
- [35] Knetsch, Jack L., Fang-Fang Tang and Richard H. Thaler (2001) The Endowment Effect and Repeated Market Trials: Is the Vickrey Auction Demand Revealing?, *Experimental Economics*, 4(3): 257-269
- [36] 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.
- [37] List, J.A., 2003, Does market experience eliminate market anomalies?, *Quarterly Journal of Economics* (February): 41-72.

- [38] Loomis, J., Brown, T., Lucero, B. and Peterson, G. (1997) Evaluating the Validity of the Dichotomous Choice Question Format in Contingent Valuation, *Environmental and Resource Economics*, 10: 109–123.
- [39] Lusk J.L., Feldkamp T. and Schroeder T.C. (2004) Experimental Auction Procedure: Impact on Valuation of Quality Differentiated Goods, *American Journal of Agricultural Economics*, 86(2): 389-405.
- [40] 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*, Proceedings of the Agricultural Economics Society Annual Conference, University of Plymouth, April 2003.
- [41] McFadden, D., 1994, Contingent valuation and social choice, *American Journal of Agricultural Economics*, 76: 689-708.
- [42] Mitchell, Robert Cameron, and Richard T. Carson, 1989, *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington: Resources for the Future, 1989.
- [43] 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.
- [44] 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.
- [45] Rubinstein, A. (1988) Similarity and Decision-Making Under Risk, *Journal of Economic Theory*, 46, 145-153.

- [46] 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.
- [47] 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.
- [48] 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.
- [49] 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.
- [50] Tversky, A., 1977. Features of similarity. *Psychological Review* 84, 327–352
- [51] Tversky, Amos and Daniel Kahneman, 1974, Judgement under uncertainty: Heuristics and biases, *Science*, 185: 1124-1131.
- [52] Varian, H. (1999) *Intermediate Microeconomics: A Modern Approach*, (Fifth Edition), Norton, New York.

Acknowledgements: The authors gratefully acknowledge comments from Alistair Munro and the audience at *The Royal Economics Society 2004 Annual Conference*, University of Swansea, 5th – 7th April 2004 and *envecon 2007*, The Royal Society, London, 23rd March 2007. Ian Bateman acknowledges support from the RELU ChREAM project (RES-227-25-0024), the ESRC via CSERGE and EFTEC. He is also Adjunct Professor in the Department of Agricultural and Resource Economics at the University of Western Australia, Perth and the Department of Economics, University of Waikato Management School, New Zealand. Funding for Diane Burgess was provided by MAFF/DEFRA and DARDNI.

Table 1: 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(l).	Estimate	Coeff.	Std. Err.	t-ratio	Log(l)
<i>Sample 1</i>										
<i>HENS</i> ₁ ¹	α_{SB}	0.92	0.27	3.40	-131.5	α_{DB}	1.54	0.20	7.70	-266.6
	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	-122.6	α_{DB}	1.57	0.20	7.89	-252.6
	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	-125.6	α_{DB}	1.57	0.20	7.70	-274.2
	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	-120.6	α_{DB}	1.39	0.20	7.07	-263.6
	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	-126.0	α_{DB}	1.61	0.20	7.98	-271.9
	β_{SB}	-0.42	0.10	-4.23		β_{DB}	-0.68	0.05	-13.15	

Table 2: Institutional learning tests: 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.	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

Table 3: Further institutional learning tests: Tests of no difference in Δ_j^i across the sequence of goods valued.

Hypothesis test	(Value) ¹	Std Error	t-test result ²
<i>Sample 1</i>			
$(\Delta_1^1 - \Delta_2^1) = 0$	1.81	1.23	Reject <0.10
$(\Delta_1^1 - \Delta_3^1) = 0$	1.75	1.22	Reject <0.10
$(\Delta_1^1 - \Delta_4^1) = 0$	1.96	1.18	Reject < 0.05
$(\Delta_2^1 - \Delta_3^1) = 0$	-0.06	0.32	Accept
$(\Delta_2^1 - \Delta_4^1) = 0$	0.22	0.30	Accept
$(\Delta_3^1 - \Delta_4^1) = 0$	0.16	0.23	Accept
<i>Sample 2</i>			
$(\Delta_1^2 - \Delta_4^1) = 0$	0.59	0.29	Reject < 0.05

Notes:

1. Refers to the sum in parentheses in the first column of the table.
2. 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).

Table 4: Value learning test: 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

Table 5: Further value learning tests: Tests of no differences in anchoring coefficients γ_j^i across the sequence of goods valued.

Hypothesis test	(Value) ¹	Std Error	t- test ² result
<i>Sample 1</i>			
$(\gamma_1^1 - \gamma_2^1) = 0$	0.49	0.23	Reject <0.05
$(\gamma_1^1 - \gamma_3^1) = 0$	0.30	0.23	Reject <0.10
$(\gamma_1^1 - \gamma_4^1) = 0$	0.66	0.24	Reject <0.01
$(\gamma_2^1 - \gamma_3^1) = 0$	-0.19	0.29	Accept
$(\gamma_2^1 - \gamma_4^1) = 0$	0.18	0.23	Accept
$(\gamma_3^1 - \gamma_4^1) = 0$	0.36	0.26	Reject <0.10
<i>Sample 2</i>			
$(\gamma_1^2 - \gamma_4^1) < 0$	0.36	0.20	Reject < 0.05

Notes:

1. Refers to the sum in parentheses in the first column of the table.
2. 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).

ⁱ Note that even following best-practice CV guidelines for the provision of accurate information regarding goods will not give respondents opportunities for such learning.

ⁱⁱ However, certain anomalies, such as those associated with the framing of a question, can be strengthened across rounds in the laboratory [35, 39] while other anomalies appear robust to individual learning in markets [26].

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 economic valuation and psychological studies (e.g. [8, 34]).

^v Note that the nature of information feed-back within the present study differs from that of many previous experiments. Such experiments typically provide formal feedbacks such as auction prices, etc. In contrast, within the present study, the initial valuation task provides hands-on experience of both the relevant contingent market and similar goods to that used in subsequent valuation tasks. While differing from standard experimental feed-backs, we feel such experience provides a powerful base for institutional and value learning.

^{vi} The argument here being that, assuming well formed preferences, if a non-significant SB-DB difference for an initial good was due to the change in incentive compatibility between the first and second bound, then it is not apparent why this should not persist for subsequent goods given the subsequently discussed, exclusive good nature of the contingent market.

^{vii} Although arising in a different context, previous work on similarity is relevant here [45, 50].

^{viii} Full details of all the WTP questions are given in Burgess [12] of which the following is an example (text in parentheses added) “If the government could introduce ONLY the scheme to improve the welfare of all laying hens (via a previously specified route), all other farm

animals remaining in their existing conditions, would you be willing to pay £X as an addition to your weekly food bill to ensure that ONLY this scheme takes place ?”.

^{ix} In accordance with Boyle and Bishop [10] 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 [29]. 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).

^x To the authors knowledge this is the first statistical test within the CV literature examining the consistency of welfare estimates from SB and DB models for the same sample (details presented in [40]).

^{xi} An anonymous referee suggested that an alternative test of Ariely et al’s coherent arbitrariness hypothesis is to examine the extent to which the initial bid level used for the first good ($HENS_1^1$) anchored WTP responses for subsequent goods ($CHICKS_2^1$, $COWS_3^1$ and $PIGS_4^1$). Adapting the approach of Table 4, a version of such an anchoring test was conducted which showed that the corresponding anchoring parameter (akin to γ in Table 4), was not significantly different from zero for all goods other than the first good, $HENS_1^1$ (even using a lower, 90% confidence level), thereby providing further support for the value learning hypothesis.

^{xii} It is interesting that such ‘practice questions’ are a very common, virtually standard, feature of experimental economics, yet have been explicitly spurned by the environmental valuation literature.