

Learning Effects in Repeated Dichotomous Choice Contingent Valuation Questions

Ian J. Bateman, Diane Burgess,
W. George Hutchinson and David I. Matthews¹

Abstract

The highly influential NOAA guidelines (Arrow et al., 1993) for the conduct of contingent valuation (CV) studies follows incentive compatibility arguments to recommend the use of a single bound dichotomous choice (SBDC) question as the basis for eliciting willingness to pay (WTP) estimates for non-market goods. However, the ‘one shot’, first response nature of the SBDC approach is directly at odds with the Discovered Preference Hypothesis (DPH) proposed by Plott (1996). The DPH, which is supported by a considerable body of experimental evidence (see, for example, List, 2003), argues that stable and theoretically consistent preferences are typically a product of experience gained through practice and repetition. Under such a hypothesis it is the last response in a series of valuations which should be attended to, rather than the first as emphasized by the NOAA guidelines. We develop a unique study design to permit the repetition of valuation tasks both with the same and across goods so facilitating the first field based test of the DPH using the CV technique. Results are that while significant discrepancies are observed between initial and subsequent WTP responses, these anomalies become insignificant in later responses. These findings raise fundamental questions regarding accepted standard practice for CV studies.

Affiliations:

Bateman and Burgess are, respectively, Professor of Environmental Economics and Senior Research Associate at the Centre for Social and Economic Research on the Global Environment (CSERGE), University of East Anglia, Norwich, NR4 7TJ, United Kingdom. i.bateman@uea.ac.uk [+44 (0) 1603 593125] and diane.burgess@uea.ac.uk [+44 (0) 1603 593743].

Hutchinson is Senior Lecturer at the Department of Agricultural and Food Economics, Queens University Belfast, UK BT9 5PX. george.hutchinson@dardni.gov.uk [+44 (0) 28 90255211]

Matthews is Scientific Officer at the Survey Unit, Biometrics Division, Department of Agriculture and Rural Development, Belfast UK BT9 5PX. dave.matthews@dardni.gov.uk [+44 (0) 28 90255627]

¹ The convention of alphabetic authorship is followed. No seniority of authorship attributed.

Introduction

The contingent valuation (CV) method is by far the most commonly applied of all the methods available for valuing preferences for non-market goods with literally thousands of applications conducted to date (Carson, forthcoming). An early concern regarding such studies was the multitude of design permutations which characterized this literature and the impact which these variations might have upon 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 (Carson and Mitchell, 1989; Bateman *et al.*, 2002), mindful in particular of incentive compatibility arguments, the NOAA panel recommended the use of a ‘one-shot’ or single-bound dichotomous choice (SBDC) 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, or rejecting this opportunity. By varying the specified price across a survey sample estimates of summary statistics such as mean WTP may be obtained for policy purposes.

The theoretical case for rejecting all but the SBDC response format can be traced back to the work of Gibbard (1973) and Satterthwaite (1975) establishing the potential incentive compatibility of one-shot referenda². 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. Evidence from economic experiments concerning the incentive compatibility of hypothetical single referenda is decidedly mixed. Some studies find convergence of voting responses with voting responses in real consequential referenda while other studies report divergent results (see Cummings *et al.* 1997, Taylor *et al.* 2001 and Burton *et al.* 2001). Given this, the claimed incentive compatibility properties of the SBDC elicitation format in CV studies appear questionable.

A more fundamental critique of the ‘one shot’ nature of the SBDC 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 that environment (a process which Braga and Starmer (forthcoming) refer to as ‘institutional learning’) and discover features of their own preferences (‘value learning’, *ibid*). The first response SBDC precludes such learning and is in direct conflict with the DPH under which it is the last response in a series of valuations which should be attended to, rather than the first. This and 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 SBDC elicitation method.

Central to the DPH then is the role of repetition within the formation of stable and theoretically consistent preferences. Empirical evidence for this learning process is provided by a considerable body of experimental evidence gathered across a variety of contexts, examples of which include:

² Within the CV context this argument is developed through Hoehn and Randall (1987) and Carson *et al.*, (1999).

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; 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, 2003); and, perhaps most pertinently, reduction in preference anomalies amongst more experienced traders or choice makers (List and Lucking-Reiley, 2000; List, 2001, 2002a, 2002b).

This evidence suggests that initial valuations of unfamiliar goods (such as those encountered in CV surveys) are liable to be based upon poorly formed preferences and therefore prone to be influenced by a variety of frequently observed choice heuristics and framing effects resulting in apparently anomalous preferences. Under this argument, preference consistency is more likely to be observed after repeated valuation trials (Binmore, 1994; 1999), i.e. the initial valuation response is viewed as the least rather than the most relevant for welfare estimation and subsequent decision making purposes.

The DPH, together with this wealth of evidence of learning effects arising from repeated valuation trials, stands in stark contrast with the NOAA panel endorsement of the first response SBDC elicitation format. Given this obvious and potentially important conflict, this paper sets out to provide a first empirical investigation of the possibility of learning effects within repeated DC format CV studies. In so doing we conduct the first field-based test of the DPH using the CV technique. This repeats valuation tasks both within and across goods. Results are that while significant discrepancies are observed between the initial and subsequent valuation responses, these anomalies become insignificant in later responses. These findings raise fundamental questions regarding accepted standard practice for CV studies.

The paper is organized as follows. In section 2 we outline a method for conducting repeated DC response valuations for the same good while in the following section we extend this design to permit repeated valuations across goods thereby building up a series of DC responses allowing respondents the opportunity for preference discovery necessary for the DPH to be invoked. Formal testing hypotheses are then stated. In Section 4 we outline sampling and analytical methodology introducing a novel Monte Carlo based approach to allow testing of learning effects across valuation tasks. Section 5 reports results while Section 6 discusses the implications of these findings and concludes.

2. Research Design (i): Repeated DC valuations within goods.

The SBDC is clearly at odds with the ‘learning through repetition’ argument underpinning the DPH. Furthermore, its one-shot nature precludes the repeated trials preference consistency testing which is prescribed by the DPH. However, this can readily be addressed by the addition of a follow-up DC question to yield the ‘double bound’ DC (DBDC) elicitation format pioneered by Hanemann *et al.* (1991) and Welsh and Bishop (1993). In such DBDC designs the amount 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 amount results in a higher value being presented at the second bound.

Although DBDC designs permit a substantial improvement in the statistical efficiency of a given sample relative to that provided by applying a SBDC format, nevertheless the validity of the DBDC approach has been questioned through several studies which have reported significant

differences in the estimated WTP derived from first and second responses (Cameron and Quiggin, 1994; Bateman et al., 2001). McFadden (1994) reports results, 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 this disparity 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, given the evidence cited above regarding whether or not even the first response is indeed incentive compatible, contrasted with the substantial evidence of consistent preference formation through repetition, then a DPH reading of the DBDC literature might be that these studies provide the first (if inadequate) evidence of CV respondents refining their preferences through repetition. Given this, we test such a reading by repeating DBDC format studies across a number of goods as described in the following section.

3. Research Design (ii): Repeated DC valuations across goods.

From a DPH perspective, the two valuation questions presented in a standard DBDC experiment provide inadequate opportunities for respondents to discover their preferences. In particular, such a test provides no opportunity for repeat exposure to the entire valuation institution. To permit such institutional learning (Braga and Starmer, forthcoming) we repeat the DBDC process across a series of four goods. The DPH is then tested in two principal ways. First evidence of increased preference consistency (a direct prediction of Plott’s hypothesis) is tested by examination of the first versus second response disparity found in previous DBDC studies. The DPH expectation being that this disparity should diminish as the valuation tasks are repeated for successive goods. To control for the possibility that any increase in preference consistency is a by-product of the type of goods valued or their order of presentation, a second test uses a split sample methodology in which a second group of respondents (Sample 2) are asked DBDC questions regarding the fourth a final good valued by the initial group (Sample 1). The DPH expectation here is that, controlling for the good, first and second response disparities should be smaller within Sample 1 (who had been exposed to institutional learning) than amongst Sample 2 (for whom this was the only good valued, thereby negating the possibility of institutional learning).

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 (were $i = 1,2$) 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 $i = 2$). The goods were non-nested schemes to improve animal welfare to be paid for via a compulsory tax on all foodstuffs. Details of the goods presented to each sample are as follows:

Sample 1: Here respondents were presented with detailed CV scenarios on the following list of goods, always given in this order:

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

In order to control for possible sequencing effects (Carson and Mitchell, 1995; Carson et al, 1998) goods were presented within an ‘exclusive list’ framework (Bateman et al., 2004) where goods are presented as alternatives to any other goods given in that list. To reinforce this condition, valuations of a given good were elicited prior to valuation tasks concerning any subsequent good.

Sample 2: Here respondents were presented with just one good to value, namely that good which was presented last (fourth) to Sample 1. We denote this good as $PIGS_1^2$.

Remember that the DPH speculates that the divergence in mean WTP between SBDC responses (or just the first response from a DBDC study) and that obtained by all responses to a DBDC exercise for the same good is due in part to inexperience with CV questions and can be attenuated by valuation experience. Given our study design we can now formulate these DPH speculations into the following testable hypotheses:

H_o^1 : For each good we compare mean willingness to pay from first responses (denoted μ_{SBi}) with those from first and second responses modelled as DBDC (denoted μ_{DBi}) data and test the hypothesis $H_o^1: \mu_{SBi} = \mu_{DBi}$. The DPH expectation is that H_o is less likely to hold for the first good valued than for the last good valued in a series of valuation tasks.

H_o^2 : The difference ($\mu_{SBi} - \mu_{DBi}$) will be larger for a good if valued first in a list (i.e. good $PIGS_1^2$ being the only good valued by Sample 2) than when that good is presented after other valuation tasks in a list (e.g. good $PIGS_4^1$ being the fourth good valued by Sample 1). To establish this effect of learning we test the hypothesis, $H_o^2: PIGS_1^2 (\mu_{SB1} - \mu_{DB1}) = PIGS_4^1 (\mu_{SB4} - \mu_{DB4})$.

4. Sampling and econometric methodology

The CV questionnaire was administered by face-to face 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.

The data generated by this survey was analyzed using SBDC methods (applied to the first DC response for each good valued) and DBDC methods (applied to both first and second DC responses for each good valued) as per Hanemann and Kanninen (1999). Details of our modelling strategy are presented through the remainder of this section.

Single Bounded Dichotomous Choice Model

In a SBDC exercise 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 single bid b_1 for a certain good can be expressed as follows:

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

where b_1 is the value of the bid amount present to the respondent, \mathbf{x} represents a set of covariates including the bid and $\boldsymbol{\beta}$ 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. In the single valuation dichotomous choice method there are just two possible outcomes, 'yes' and 'no' so the sample log-likelihood can be expressed as :

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

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 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 value the probability of a yes response π_{yes} will occur when the WTP exceeds the bid offered. This can be expressed as:

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

Where $\alpha_{\text{SB}}, \beta_{\text{SB}}$ are the coefficients of constant and bid respectively. The method of maximum likelihood is used to estimate the coefficients $\alpha_{\text{SB}}, \beta_{\text{SB}}$. In line with many previous studies (e.g. Langford et al, 1998) this study uses a single bounded Logit model.

Double Bounded Dichotomous Choice Model

The DBDC approach extends an initial SBDC-style question with a supplementary DC valuation task. If the individual agreed to pay the initial 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*); (*no,no*) is given by :

$$\begin{aligned} \pi_{yy} &= 1 - H(\boldsymbol{\beta}' \mathbf{x}_h) \\ \pi_{yn} &= H(\boldsymbol{\beta}' \mathbf{x}_h) - H(\boldsymbol{\beta}' \mathbf{x}) \\ \pi_{ny} &= H(\boldsymbol{\beta}' \mathbf{x}) - H(\boldsymbol{\beta}' \mathbf{x}_l) \\ \pi_{nn} &= H(\boldsymbol{\beta}' \mathbf{x}_l) \end{aligned} \quad (4)$$

Where $\mathbf{x}, \mathbf{x}_h, \mathbf{x}_l$ are the vectors of covariates respectively associated with first bid, second bid higher and second bid lower. Where H is the cdf function chosen. This gives the following log likelihood for the sample.

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

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 double-bounded format is often preferred over single-bounded as the latter is statistically inefficient requiring relatively large sample sizes to precisely characterize a population's WTP (Herriges and Shogren, 1996). The additional information provided by the follow-up question makes the DBDC asymptotically efficient relative to the SBDC. A further advantage of the former model is that it is fairly robust with respect to poor bid designs resulting from initial parameter misspecification (Hanemann *et al.*, 1991). 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).

Estimates of WTP can be computed from the SBDC and DBDC models. e.g. mean WTP can be calculated for the SBDC model specified in equation (3) using:

$$\mu = -\alpha/\beta \tag{6}$$

The standard errors for μ can be obtained from the variance of $\text{Var}(\mu) = \text{Var}(\alpha/\beta)$ which can be calculated using the Delta method first order approximation of variance using Taylor's when the coefficient variance/covariance estimates are known.

Tests for differences in mean WTP

Here we outline methods to test H_0^1 and H_0^2 that WTP divergence between SBDC and DBDC is due to inexperience with CV questions and will diminish for experienced respondents. To show this we test the proposition that differences in welfare estimates of mean WTP between SBDC and DBDC models are zero (i.e. $H_0:(\mu_1-\mu_2)=0$). To test this a t-statistic can be calculated as follows:

$$t = (\mu_1 - \mu_2) / \sqrt{\text{Var}(\mu_1 - \mu_2)} \tag{7}$$

When μ_1, μ_2 are valued using two independent samples the estimates of the means are uncorrelated ($\text{Covar}(\mu_1, \mu_2) = 0$) 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 independent means μ_1, μ_2 can be computed as:

$$\text{Var}(\mu_1 - \mu_2) = \text{Var}(\mu_1) + \text{Var}(\mu_2) \tag{8}$$

This is appropriate when comparing differences in mean WTP for independent samples. When comparing difference between the first response data used in an SBDC model and the same first responses supplemented by follow-up question responses, as per a DBDC exercise, the samples can no longer be considered independent since both estimates are computed using the same initial responses from the same individual. Hence when estimating $\text{Var}(\mu_{\text{SB}} - \mu_{\text{DB}})$ the $\text{Covar}(\mu_{\text{SB}}, \mu_{\text{DB}})$ cannot be ignored as in (8) above. Due to these difficulties with SBDC and DBDC comparisons, the $\text{Var}(\mu_{\text{SB}} - \mu_{\text{DB}})$ cannot be obtained from a known closed form solution. An alternative approach is presented in this paper.

To obtain the distribution of differences in the sample means, and to compute an estimate of the $\text{Var}(\mu_{\text{SB}} - \mu_{\text{DB}})$ Monte Carlo techniques can be employed (Efron and Tibshirani, 1993). These methods use resampling techniques 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, an estimate $(\mu_{SB} - \mu_{DB})$. The Jack-knife method is chosen to estimate the variance $\text{Var}_J(\mu_{SB} - \mu_{DB})$.

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 multiple 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 sample s is θ_s . Thus a distribution of that statistic which reflects the original true distribution is created. Here θ_s is computed as the difference in mean WTP estimates obtained from the SBDC and DBDC models.

The variance for the Jack-knife estimate is obtained from:

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

$$\text{Var}_J(\theta) = \{ (n-1)/n \} (\Sigma(\theta_s - \theta_m))^2 \quad (9)$$

Here the estimate for the difference in θ_s is :

$$\theta_s = (\mu_{SBs} - \mu_{DBs})$$

$$\theta_m = \Sigma (\mu_{SBs} - \mu_{DBs})/n$$

Applying these techniques to our survey data we obtain our study results and significance tests, which we now present.

5. Results

Following Hanemann et al (1991), we estimate parsimoniously specified SBDC and DBDC models for each good as presented in Table 1. The models provide parameter estimates of the coefficients α , β for the constant and bid for SBDC and DBDC logistic models. 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.

Table 1 SBDC and DBDC models of WTP to pay for specified animal welfare improvement schemes.

Scheme	Single Bounded Models (SBDC)					Double Bounded Models (DBDC)				
	Variable	Coeff.	Std.Err.	t-ratio	LL	Variable	Coeff.	Std.Err.	t-ratio	LL
<i>HENS</i> ₁ ¹	α_{SB}	0.92	0.270	3.40	-131.54	α_{DB}	1.54	0.200	7.70	-266.61
	β_{SB}	-0.19	0.093	-2.10		β_{DB}	-0.56	0.050	-11.2	
<i>CHICK</i> ₂ ¹	α_{SB}	1.44	0.283	5.10	-122.62	α_{DB}	1.57	0.199	7.89	-252.61
	β_{SB}	-0.54	0.106	-5.07		β_{DB}	-0.63	0.052	-12.1	
<i>COWS</i> ₃ ¹	α_{SB}	1.32	0.278	4.81	-125.55	α_{DB}	1.57	0.204	7.70	-274.18
	β_{SB}	-0.43	0.099	-4.38		β_{DB}	-0.55	0.047	-11.80	
<i>PIGS</i> ₄ ¹	α_{SB}	1.29	0.289	4.47	-120.64	α_{DB}	1.39	0.197	7.07	-263.57
	β_{SB}	-0.62	0.121	-5.15		β_{DB}	-0.68	0.052	-12.94	
<i>PIGS</i> ₁ ²	α_{SB}	1.25	0.277	4.49	-126.01	α_{DB}	1.61	0.202	7.98	-271.88
	β_{SB}	-0.42	0.099	-4.23		β_{DB}	-0.68	0.051	-13.15	

Table 2 presents the results of testing for equality of mean WTP for the same good using SBDC and DBDC models. The table provides estimates of the mean WTP for μ_{SB} , μ_{DB} (derived from the coefficient estimates shown in Table 1 and calculated as $\mu = -\alpha/\beta$ for each model) and the difference ($\mu_{SB} - \mu_{DB}$) between these estimates. Standard errors of this difference 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 2: Comparison of differences between mean WTP for SBDC and DBDC models for each animal welfare improvement scheme

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

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

Bold type indicates differences in mean WTP which are significant at $p \leq 0.1$

The first good valued by respondents in Sample 1 (the treatment with multiple valuation tasks) is *HENS*₁¹. Comparison of mean WTP for the SBDC and DBDC models for this good shows by far the largest absolute and relative difference in mean WTP of any of the goods values. Although standard error is also large the $\mu_{SB} - \mu_{DB}$ difference is significantly different from zero at the 10% significance level. This anomaly conforms to the existing empirical literature on tests of DBDC designs (Carson et al, 1999). However, this finding contrasts markedly with those for the succeeding goods valued by Sample 1 such that the significance of difference in mean WTP between SBDC and DBDC models falls steadily to $p=0.95$ for the final good valued, *PIGS*₄¹ for which mean WTP from the two model differs by considerably less than one percent. This latter result is in direct contrast with that observed when the same good is presented as the first and only good valued by Sample 2 (*PIGS*₁²). Here the $\mu_{SB} - \mu_{DB}$ difference is significantly different from zero at the 2% significance level. Interestingly the pattern of responses observed for *PIGS*₁² is most similar to that observed for the *HENS*₁¹ good. Both goods are the first valued by their respective samples and both yield significant anomalies which, in the case of Sample 1, disappear when respondents have the opportunity to value subsequent goods. Overall then, the results shown in Table 2 conform well to the learning effect expectations postulated by the DPH.

A further test of the within good contrast can be calculated by looking at whether the difference in mean WTP between SBDC and DBDC for the same good is significantly different for experienced respondents (those in Sample 1 valuing *PIGS*₄¹ as the fourth good they encounter) as opposed to inexperienced respondents (those in Sample 2 valuing *PIGS*₁² as the first and only good they encounter). Results show that experienced respondents do indeed generate significantly lower $\mu_{SB} - \mu_{DB}$ differences than do inexperienced respondents ($p=0.04$).

A final investigation of the $PIGS_4^1$ versus $PIGS_1^2$ contrast can be obtained by examining the four mean WTP measures obtained from applying SBDC and DBDC models to these goods. We can denote mean WTP estimates obtained from applying SBDC models to the $PIGS_1^2$ and $PIGS_4^1$ values as μ_{SB1} and μ_{SB4} respectively. Similarly, the mean WTP obtained from applying DBDC models to the $PIGS_1^2$ and $PIGS_4^1$ values are denoted μ_{DB1} and μ_{DB4} respectively. Table 3 details differences between these means and reports significance tests for these differences.

Table 3 : Comparison of mean WTP for the pig welfare schemes presented as either the first ($PIGS_1^2$) or fourth ($PIGS_4^1$) good valued.

$PIGS_1^2$ Mean WTP for pig welfare scheme presented as first good valued			$PIGS_4^1$ Mean WTP for pig welfare scheme presented as fourth good valued			Independent sample t-tests testing equality of mean WTP				
Estimate	Value	Std.Er.	Estimate	Value	Std.Er.	Estimate	Value	Std.Er.	t-ratio	Prob.
μ_{SB1}	£2.98	£0.39	μ_{SB4}	£2.07	£0.25	$\mu_{SB1}-\mu_{SB4}$	£0.91	£0.46	1.96	0.05
μ_{DB1}	£2.38	£0.21	μ_{DB4}	£2.06	£0.21	$\mu_{DB1}-\mu_{DB4}$	£0.32	£0.30	1.08	0.28
						$\mu_{SB1}-\mu_{DB4}$	£0.92	£0.44	2.08	0.04
						$\mu_{DB1}-\mu_{SB4}$	£0.31	£0.33	0.95	0.34

Note: **Bold** type denotes differences which are significant at $p \leq 0.1$

Table 3 reveals a consistent pattern of differences in mean WTP across treatment and elicitation bounds. Where comparison is made between the mean derived from the first response regarding the first presented good (i.e. the mean μ_{SB1}) then this yields a significant difference with either the SBDC or DBDC model based upon responses to the same good presented as the fourth good valued. Conversely the DBDC model (recall that this combines first and second responses) from the first good does not yield significant differences from measure derived from the same good presented as the fourth valued. Overall our results for pig welfare show that it is only the first response value stated by inexperienced respondents, which is significantly out of line with all other values for this good. In our concluding section we consider the implications of these findings.

Conclusions

We contend that this paper highlights a central dilemma for CV research. We question the standard presumption in favour of the first response to a DC design focused upon the claimed incentive compatibility properties of such a format and contrast this with an approach based upon the DPH and supported by a considerable body of experimental evidence to focus instead upon the last response to a series of valuation tasks. We have presented findings from a CV study designed to contrast the above approaches. These findings are that a familiar, well documented and often observed anomaly, the disparity between first and second responses in a double bound DC format, disappears when repeated across valuation tasks. Given the persistent nature of this

anomaly, this in itself is a notable feature of the paper (as is the novel Monte-Carlo based methodology used to analyze these data).

In considering these results, defenders of the traditional, first response approach to CV design would need to dismiss not only our central finding of increased consistency and the disappearance of the double bound DC anomaly across valuation rounds, but also dismiss the DPH itself and the large and steadily increasing body of experimental evidence which supports it and dismiss those studies which find that the claimed incentive compatibility of first response elicitation formats does not hold within hypothetical CV markets. In contrast those who are persuaded by the latter sets of evidence may well view our findings as further support for this view.

Our own conclusions are that these findings should not be seen as an excuse to ignore issues of incentive compatibility. However, we do question the conventional CV view that first-response DC elicitation formats approximate market situations. Importantly such formats do not offer the repetition, learning and experience possibilities of real markets. We feel that this is a significant failing and one which should be addressed through improved elicitation techniques. 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 a incentive compatible valuation question. Such an approach, we contend, would address much of the familiarity, error and consequent anomalies observed in this and many prior CV studies.

References

- Alberini, A., Kanninen, B. and Carson, R.T., 1997, Modelling response incentive effects in dichotomous choice contingent valuation data, *Land Economics*, 73: 309-324.
- 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*
- 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.
- 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.
- Braga, J. and C. Starmer (2003), "Does market experience eliminate preference reversal?", Mimeo, University of Nottingham.

- Braga, J. and Starmer, C. (forthcoming) Protean preferences and public policy, *Environmental and Resource Economics*.
- 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., Groves, T. and Machina, M.J., 1999, Incentive and informational properties of preference questions, *Plenary Address, Ninth Annual Conference of the European Association of Environmental and Resource Economists (EAERE)*, Oslo, Norway, June 1999.
- 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.
- 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
- Gibbard A., 1973, Manipulation of Voting Schemes: A General Result, *Econometrica* 41, 587-601.
- 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.
- Hausman, J. (1993), ed. *Contingent Valuation: A Critical Assessment*. Amsterdam: North-Holland.
- 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.
- 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. "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards," *American Economic Review*, December 2001, 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, "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.
- 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, C. 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, C.R. and Zeiler, K., 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.
- 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.
- Shogren, J., S. Cho, C. Koo, J. List, C. Park, P. Polo and R. Wilhelmi (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.
- 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
- 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.