UNIVERSITY OF WAIKATO

Hamilton
New Zealand

Investigating the Characteristics of Stated Preferences for Reducing the Impacts of Air Pollution: A Contingent Valuation Experiment

Ian J. Bateman
School of Environmental Sciences, University of East Anglia

Michael P. Cameron
Economics Department, University of Waikato

Antreas Tsoumas
Department of Environmental Studies, University of the Aegean

Department of Economics

Working Paper in Economics 08/06

May 2006

Corresponding author

Ian J. Bateman
School of Environmental Sciences
University of East Anglia
Norwich NR4 7TJ
United Kingdom

Tel: +44 (0) 1603 593125
Email: i.bateman@uea.ac.uk

Abstract

This paper investigates the nature of stated preferences for reducing air pollution impacts. Specifically a contingent valuation (CV) experiment is designed to elicit individuals' values for reducing these impacts and to examine how these may change when multiple schemes for reducing differing impacts are valued. The novel survey design allows simultaneous testing for the presence of several anomalies reported in the CV literature within the same context, including (i) scope sensitivity (ii) part-whole or substitution effects (iii) ordering effects and (iv) visible choice set effects. Results indicate some scope sensitivity and interaction between ordering effects and visible choice set effects, as well as substantial part-whole or substitution effects between two exclusive schemes. A practical consequence of these findings is that estimates of the value of combined programmes may not readily be obtained by summing the values of their constituent parts obtained using the CV method.

Keywords

air pollution
contingent valuation
stated preferences
part-whole effect
experimental surveys

JEL Classification

C42, C90, Q51, Q53

Acknowledgements

The support of the Economic and Social Research Council (ESRC) and the Economics for the Environment Consultancy (EFTEC) is gratefully acknowledged. The authors are grateful to Philip Cooper for detailed comments on an earlier version of this paper. Remaining errors are the responsibility of the authors alone.

1. Introduction

Airborne pollutants impact upon a variety of receptors including humans, animals, plants, buildings and materials. Individuals who are aware of and concerned by such impacts may value their reduction. This paper presents the findings of an experiment designed to investigate the nature of stated preferences for reducing air pollution impacts obtained using the contingent valuation (CV) method.

The CV method is a technique for assigning monetary values to individual preferences for changes in the provision of a good or set of goods (for a review of the CV method see Mitchell and Carson, 1989, and for recent debate see Bateman and Willis, 1999). The method typically operates through surveys of individuals in which respondents are presented with a hypothetical or contingent market for a good and asked to state either their willingness to pay (WTP) or willingness to accept (WTA) compensation for either a gain or loss of that good. CV has been extensively used to assess preferences for non-market goods such as those provided by the environment.

A key objective of our research was to examine the extent to which values derived by CV were consistent with economic theory or exhibited certain anomalies reported in the literature. By anomalies we mean results that appear to be inconsistent with the expectations of economic theory as set out in many standard texts (for example Varian, 1992). We apply the CV method through a split-sample experimental design that for the first time allows investigation of the presence or absence of these anomalous results within the context of the same valuation exercise. The financial confines of the present research precluded investigation of the *origin* of any observed anomalies. To do so would have required a switch away from the hypothetical contingent market which underpins the CV method to the use of real-payment approaches such as those used in Bateman et al (1997a; 1997b).

1.1 Emissions and impacts

In this study we focus upon the impacts of air pollutants rather than the emissions themselves[1]. In order to motivate the empirical study two hypothetical schemes for reducing air pollution impacts were derived as follows:

- Scheme H: Reduction of the impacts of toxic vehicle emissions upon human health

- Scheme P: Reduction of the impacts of acidic power station emissions upon plant life

In order to implement our research design these were supplemented by a further combined scheme as follows:

- Scheme A = Scheme H + Scheme P

[1] Arguably individuals may hold values for reducing emissions which have no discernible impact (e.g. colourless, odourless gasses which have no effect upon any receptor) if they object to the fact that these are non-natural. For simplicity we assume that an individual's values will be driven by impacts rather than emissions.

The goods described by these three schemes provided the basic building blocks for constructing valuation scenarios. In Section 2 we briefly review theoretical expectations regarding CV values for public goods such as these. Specifically we consider four inter-related issues which have been the focus of recent research concerning arguably anomalous results derived from CV studies. These issues are:

(i) Scope sensitivity

(ii) Part-whole / substitution effects

(iii) Ordering effects and

(iv) Visible choice set effects.

In Section 3 we describe our novel experimental design for testing for the presence of such effects in values for the three impact reduction schemes mentioned above. Section 4 presents our experimental results. This opens by providing sampling details and sample socio-economic and demographic characteristics. Valuation results are then presented and a set of hypotheses regarding theoretical expectations (and hence anomalies) are formulated and tested. Finally, Section 5 summarises our findings and presents conclusions.

2. Theoretical Expectations and Anomalies

The basic tenet of welfare economics is that individuals maximise their utility by choosing what they prefer, and preferring what they perceive as yielding maximum utility[2] (Varian, 1992). The preferences underpinning these choices can be expressed as values which in turn may be assessed through measures such as WTP for a particular good. Economic theory says very little about the psychological processes which form preferences[3], but does assume a form of rationality and consistency of preferences from which certain testable hypotheses may be derived. In this section we review the four issues identified previously, describing theoretical predictions and how anomalous responses may cause deviations between predicted and observed value relationships.

2.1 Scope sensitivity

Scope sensitivity describes the extent to which stated values are sensitive to changes in various dimensions of the good under investigation (Carson, Flores and Meade, 2001). For example, it may be that values rise with increases in the physical scale of an impact reduction scheme. However, while standard economic theory suggests that values should not fall as scope increases, it does not require that values rise with scope. For instance, an individual

[2] This is a positive rather than normative theory in which the individual is the sole arbiter of what they feel maximises their own utility. So, for example, despite the associated health risks, smoking cigarettes can contribute to maximising a particular individual's utility.

[3] Indeed as Varian (1992) notes, "A utility function is often a very convenient way to describe preferences, but it should not be given any psychological interpretation" (p.95).

may have a positive WTP for setting up a recreational woodland but, once that is provided, be unwilling to pay for a second such woodland.

The issue of whether scope sensitivity should be and is observed in a given application is essentially an open empirical question dependent upon the nature of the good and change in provision concerned. Nevertheless, since publication of the US NOAA Panel report on the validity of CV (Arrow et al, 1993), scope sensitivity has been viewed (arguably with dubious justification) as a key indicator of study quality and has generated a substantial empirical and theoretical literature (Goodstein, 1995).

Bateman et al (2004) describe a number of tests including examinations of the consistency of scope sensitivity across valuations of nested goods, i.e. where the scope of one 'inclusive' good entirely comprises and exceeds that of another subset good. In the present paper we adopt a straightforward approach to testing for scope sensitivity; specifically that values for an inclusive good should not be less than values for a subset good. Considering the three air pollution impact reduction schemes this equates to theoretical expectations given in Equations (1) and (2) that:

$$\text{WTP (Scheme H)} \leq \text{WTP (Scheme A)} \tag{1}$$

and

$$\text{WTP (Scheme P)} \leq \text{WTP (Scheme A)} \tag{2}$$

Satisfaction of these tests is insufficient to prove the theoretical consistency of our contingent values. As Svedsater (2000) points out, scope sensitivity might be observed when respondents are asked to value nested schemes simply because the respondent is influenced by their previously stated values and attempt to act in an internally consistent way. Failure of these tests, however, would be a strong indication of anomalous stated preferences.

2.2 Part-whole/substitution effects

The 'part-whole[4] phenomena' occurs in the context of CV studies when it appears that the sum of the valuations placed by an individual on the parts of a good is larger than the valuation placed on the good as a whole (i.e. the sum of the part values exceeds that stated for the whole). In the wake of the Exxon Valdez oil spill, part-whole effects emerged as principal focus of debate regarding the validity of the CV method[5]. The occurrence of part-whole effects within CV studies was (and still is) seen by critics as a major challenge to the validity of the CV method. However, Bateman et al (1997a) demonstrate that part-whole effects can be observed in consumers real-money purchases of private goods. This suggests that such effects may constitute a true anomaly and shortcoming of standard theory.

[4] The terms 'part-whole' and 'embedding' are employed in the cognitive psychology literature dealing with the perception of visual parts and wholes, where evidence suggests that one hemisphere of the brain is responsible for perception of wholes, while another deals with the parts of an object (Robertson and Lamb, 1991; Tversky and Hemenway, 1984).

[5] For example see Kahneman and Knetsch (1992), Smith (1992), Harrison (1992), Carson and Mitchell (1993), Boyle et al (1994), as well as through the interchanges in Hausman (1993) and between Hanemann (1994) and Diamond and Hausman (1994).

However, substitution effects mean that the presence of part-whole phenomena for certain goods need not necessarily constitute a theoretical anomaly (Carson et al, 1998). For example, two ‘part’ goods might be regarded as substitutes for each other and then the value of the ‘whole’ bundle consisting of both goods might be less than the sum of the constituent parts[6]. In our application we have chosen goods which individuals may or may not consider as substitutes for each other. It may or may not be that the reduction of air pollution impacts upon plants (Scheme P) is a substitute, or partial substitute, for relieving air pollution impacts upon human health (Scheme H). Therefore we cannot distinguish between the part-whole phenomena (a theoretical anomaly) and a substitution effect (a finding which is entirely consistent with theory). However, this paper is constrained to an empirical investigation of whether part-whole/substitution effects are observed rather than in disentangling the precise cause of such an effect.

Considering our elicited values and remembering that Scheme A involves the joint implementation of Schemes H and P, then part-whole/substitution effects would be observed if Equation (3) holds:

$$[\text{WTP(Scheme H)} + \text{WTP(Scheme P)}] > \text{WTP(Scheme A)} \tag{3}$$

For convenience we will refer to the sum [WTP(Scheme H) + WTP(Scheme P)] as the ‘calculated’ value of Scheme A and contrast this with the amount WTP (Scheme A) which we refer to as the ‘stated’ value of Scheme A.

2.3 Ordering effects, list direction, and list length

One of the earliest findings of empirical CV research is that when respondents are presented with a list of goods and asked to provide values for each of those goods, then the stated value for any given good is dependent upon its position such that the nearer to the start of the list that the good is positioned, the higher is the stated value it is accorded (Randall, Hoehn and Tolley, 1981; Hoehn and Randall, 1982; Hoehn, 1983; Tolley et al, 1983). In a recent reassessment of this issue, Bateman et al (2004) show that whether or not such results are anomalous depends in part upon the type of list in which goods are presented.

In an *exclusive list,* which is the kind of list that choice theory typically addresses, goods are presented as alternatives to any other goods given in that list, with the level of other goods held constant across valuation tasks[7]. Here the stated value for a good valued at any position in such a list always refers to the same unit of that good irrespective of its position in that list.

[6] It is also theoretically possible that goods are viewed as complements. In such an instance, the sum of the parts would add up to less than the value of the whole.

[7] By comparison, in an *inclusive list* goods are presented as additions to (or subtractions from) any good(s) presented previously in that list. Carson and Mitchell (1995) show that in such lists since the value stated by a respondent for any given good is dependent upon their current endowment of private and public goods, the value for a good as the first good presented to an individual will be different from the value stated when the same good appears later in the list. Such *sequencing effects* are an expected prediction of economic theory (Carson and Mitchell, 1995; Randall and Hoehn, 1996), and can apply to both nested and non-nested goods (for example see Carson, Flores and Hanemann, 1998).

Provided that the CV respondent adjusts their perceived holdings of goods back to the initial status quo between valuation tasks, any residual variation associated with presentation is therefore an anomaly and can be termed an *ordering effect*. Empirical evidence of the presence of such effects in CV studies is mixed (Boyle, Welsh and Bishop, 1993).

We can further characterise lists in terms of their ‘direction’, i.e. whether they progress from ‘smaller’ to ‘larger’ goods, which we term a ‘bottom-up’ list, or from ‘larger’ to ‘smaller’ to yield a ‘top-down’ list. Typically for nested goods list direction can be determined through inspection of how goods are nested. In our experiment we have clear nesting of Schemes H and P within Scheme A. However, without strong priors regarding expected values list direction is only obvious ex-post for non-nested goods, e.g. the relationship of Schemes H and P to each other are not, a-priori, obvious (although an anthropocentric world view might suggest that relieving impacts upon humans is more valuable than relieving impacts upon plants). Nevertheless we shall make use of this list direction terminology in discussing our results.

A final permutation concerning list definition concerns the length of lists. Evidence exists that raising awareness of all the constituent parts of a good may increase stated values for that good; a phenomena known as event-splitting (Starmer and Sugden, 1993; Humphrey, 1995; 1996). In our experiment we vary list length between two or three goods, always including Schemes H and A and either including or excluding scheme P. By always presenting Scheme A (which embraces Schemes H and P) as the final valuation object we attempt to see whether prior inclusion of Scheme P results in an event-splitting effect, raising the value of Scheme A. As conjectured in Bateman et al (2004), list length may also have an effect on stated values if warm glow (individual value associated with the act of giving rather than the value of the good (Andreoni, 1990) or other-regarding behaviour (Ferraro et al, 2003) is somehow partitioned across all the valuation tasks that an individual understands that they will be asked to complete.

2.4 Visible choice set effects

Bateman et al (2004) define a new dimension through which CV study design may influence scope sensitivity; the *visible choice set*. Reflecting recent theoretical developments by Cubitt and Sugden (2001), they define the visible choice set as that set of goods which, at any given point in a valuation exercise, the respondent perceives as being the full extent of purchase options which will be made available in the course of that exercise. For example, prior to any values being elicited respondents might be told that they are going to be presented with three goods, C, B and A and asked to value each in turn; an approach which Bateman et al (2004) term an *advance disclosure* visible choice set. Conversely, respondents may be presented initially with only good C and a value elicited on the basis of that visible choice set alone; then they are told about good B (i.e., the visible choice set changes relative to that held at the initial valuation) and a further valuation elicited; finally they are presented with good A and a value elicited. Bateman et al (2004) characterise such approaches as exhibiting a *stepwise disclosure* visible choice set. Note that in the stepwise approach each valuation task is undertaken in ignorance of the subsequent expansion of the choice set.

Evidence for the occurrence of such effects is presented in Bateman et al (2004), who analysed visible choice set and list direction effects within a nested set of improvements to an open-access lake in Norfolk, UK. They found that within each treatment increases in the scope of goods are synonymous with rises in WTP, and that a treatment presenting respondents with the lowest value good first and where they are at that time unaware of a wider choice set, yields higher values both for that initial good and for those presented subsequently. This interaction of visible choice set and ordering effects is explicitly tested for in the experimental design used in the present analysis, with visible choice set effects appearing if values for the same good differed according to whether they were obtained from stepwise or advance disclosure treatments.

Whether or not such effects constitute theoretical anomalies is a debateable point. For private goods, choice theory states that preferences are independent of the choice set and therefore we should expect no difference in stated values elicited from either a stepwise or advance disclosure choice set. Yet, for public goods, choice theory predicts that strategic incentives *may* affect stated values where the visible choice set contains more than one such good. Because such strategies could be complex and vary across individuals, we will proceed with the assumption that respondents treat the choices offered as independent. This allows us to test the hypothesis that WTP responses will be invariant to visible choice set type.

3. Study Design

3.1 Scenarios: air pollution impact reduction schemes

A study design was defined to examine whether the various anomalies and effects under investigation were present within a CV study focussing upon values for the reduction of air pollution impacts. The various anomalies were assessed through a split sample design with each sub-sample being presented with a somewhat different questionnaire (full questionnaires for all design permutations may be obtained from the corresponding author).

The objective of this research was purely to investigate the *relative* nature of values for reducing air pollution impacts. Resources were insufficient to investigate the absolute level of those values within an incentive compatible structure. Given these constraints we adopted a simple open-ended response format for eliciting WTP answers. It is recognised that the open-ended format is liable to strategic behaviour by respondents (Carson et al, 1999) with under-representation of true WTP being a frequently cited strategy[8]. However, in a split sample context, such as adopted in this study, the open-ended approach is acceptable for detecting *differences* in WTP responses between treatments (see for example, Bateman and Langford, 1997). The open-ended method is also highly statistically efficient in that each respondent is asked to state their maximum WTP which in turn dramatically reduces sample size requirements relative to the more incentive compatible dichotomous choice approach (Hanemann and Kanninen, 1999), thus facilitating a sufficient sample size within the confines of the available research budget.

Given our focus upon differences in WTP between treatments, rather than a concern for the validity or defensibility of absolute WTP values, efforts were made to simplify the cognitive task faced by respondents. Providing the level of information is kept constant across treatments, any significant difference between sub-samples (other than those due to sample characteristics) may indicate the presence of anomalies. Given this we were able to justify reliance upon respondents' prior levels of information, assuming that this is randomly distributed across sub-samples. This was clarified to respondents in the opening statement of all questionnaires which also introduced the subject of air pollution impacts.

[8] Although over-statement is an equally plausible strategy (see Bateman et al, 2004).

Respondents were then appraised of the valuation tasks before them by informing them that they would be presented with details regarding one or more air quality improvement schemes and that they would be asked to value the implementation of these schemes. Respondents facing advance disclosure visible choice sets were told from the outset the number of air quality improvement schemes (two or three) that they would be presented with during the entire course of the experiment. However, respondents facing the stepwise information treatment were only told of the first scheme that they would face. Respondents were then presented with various combinations of scheme details and valuation tasks in exclusive list formats. The various combinations employed over the split sample design are detailed in Section 3.2.

3.2 Split sample design and corresponding tests

Investigation of the various anomalies discussed previously dictated the various treatments which together define the study design. Combining the tests suggested in sections 2.1 and 2.2 with tests for ordering effects and visible choice set effects led us to devise a study design consisting of five sub-samples of respondents, described in points (i) to (v) below:

(i) Here a stepwise disclosure approach was adopted. Respondents were presented with Scheme H and asked to value it. Respondents were then presented with Scheme A and asked to value that. Comparison of these values provides a simple scope test. We label this sub-sample SHA.

(ii) Here a stepwise disclosure approach was again adopted. Respondents were presented with Scheme H and asked to value it. This process was then repeated for Scheme P and finally for Scheme A. Comparison of these values provides a further simple scope test. Furthermore, the derived values for Scheme H, P and A allow us to conduct a part-whole test for a stepwise treatment. Comparison with sub-sample SHA allows us to see if there is an ordering effect with regard to the value of Scheme A. We label this sub-sample SHPA.

(iii) Here an advance disclosure approach was adopted. Respondents were presented with Scheme H and Scheme A before being asked to value both in turn. Comparison of these values provides a further simple scope test. Comparison with sub-sample SHA allows us to see if there is a visible choice set effect with regard to the value of Schemes H and A. We label this sub-sample AHA.

(iv) Here an advance disclosure approach was again adopted. Respondents were presented with Schemes H, P and A before being asked to value each in turn. Comparison of these values provides a further simple scope test. Furthermore, the derived values for Scheme H, P and A allow us to conduct a part-whole test for an advance information treatment. Comparison with sub-sample SHPA allows us to see if there is a visible choice set effect with regard to the values of Schemes H, P and A. We label this sub-sample AHPA.

(v) Here a stepwise information approach was adopted. Respondents were presented with Scheme P and asked to value it. Respondents were then presented with Scheme H and asked to value that. Comparison of these values with those for the same schemes elicited from sub-samples SHPA and SHA provide tests of ordering effects for these values. We label this sub-sample the label SPH.

Table 1 summarises the split sample design discussed above. Here **bold** type indicates the choice set visible to participants prior to the initial valuation task, while *italic* type shows the subsequent expansion of the visible choice set just prior to the second valuation task for participants in stepwise treatments. Finally, normal type indicates the further expansion of the visible choice set experienced by participants in the SHPA treatment just prior to their third and last valuation task. The fourth column provides labels for the various values directly stated by respondents in each treatment, indicating both the sub-sample from which that value was obtained and, in subscripts, the scheme valued. For example the value stated by respondents in sub-sample SHA for Scheme H is denoted SHA_H. Calculated values are labelled in a similar manner but include a subscript c. Therefore while $SHPA_A$ indicates the stated value for Scheme A derived from the SHPA sub-sample, the calculated value of Scheme A (calculated by summing the stated values for Schemes H and P from the same sub-sample) is denoted $SHPA_{cA}$.

Table 1: Experimental design and sub-sample structure

Group	Disclosure type	Design (ordering of information provision and valuation questions)	Stated values	Calculated values
SHA (n=40)	Stepwise	**Information: Scheme H** **WTP Scheme H** *Information: Scheme A* *WTP Scheme A*	SHA_H SHA_A	SHA_{cP} = (SHA_A - SHA_H)
SHPA (n=40)	Stepwise	**Information: Scheme H** **WTP Scheme H** *Information: Scheme P* *WTP Scheme P* Information: Scheme A WTP Scheme A	$SHPA_H$ $SHPA_P$ $SHPA_A$	$SHPA_{cA}$ = ($SHPA_H$ + $SHPA_P$)
AHA (n=40)	Advance	**Information: Scheme H** **Information: Scheme A** **WTP Scheme H** **WTP Scheme A**	AHA_H AHA_A	AHA_{cP} = (AHA_A - AHA_H)
AHPA (n=28)	Advance	**Information: Scheme H** **Information: Scheme P** **Information: Scheme A** **WTP Scheme H** **WTP Scheme P** **WTP Scheme A**	$AHPA_H$ $AHPA_P$ $AHPA_A$	$AHPA_{cA}$ = ($AHPA_H$ + $AHPA_P$)
SPH (n=40)	Stepwise	**Information: Scheme P** **WTP Scheme P** *Information: Scheme H* *WTP Scheme H*	SPH_P SPH_H	SPH_{cA} = (SPH_P + SPH_H)

4. Results

Data were collected through one-to-one, in-person surveys of students at their residential addresses at the University of East Anglia, addresses being selected at random. A total sample of 238 respondents was collected of which 50 were used in a pilot survey refining the wording of questionnaires. As wording was substantially simplified following the pilot survey, those presented with the pilot questionnaire are excluded from our analysis.

4.1 Sub-sample demographic characteristics

All respondents were asked a number of socio-economic and demographic questions. These were used to examine possible differences between sub-samples which may complicate our subsequent analyses. Summary statistics for key variables within and across sub-samples are presented in Table 2.

Table 2: Socio-economic and demographic profile of sub-samples

Group	Gross expected income in the next 12 months (£)			Number of non-UK respondents	Gender		Age last birthday			Previously studied economics		Total sub-sample size
	mean	s.e.	median		male	female	mean	s.e.	median	No	Yes	
SHA	6138	676	5000	3	16	24	20.1	0.40	19.0	30	10	40
SHPA	6411	732	4500	2	16	24	20.3	0.24	20.0	33	7	40
AHA	7171	956	5000	2	20	20	21.7	0.33	21.5	31	9	40
AHPA	9059	1007	6650	0	14	14	23.3	0.73	23.0	21	6	28
SPH	6581	895	4000	1	19	21	20.8	0.49	20.0	32	8	40
All sub-samples	6977	393	5000	8	85	103	21.1	0.20	20.0	147*	40*	188
S.o.D.**	0.054			0.446	0.189		0.000			0.945		

* One missing value.
** Significance of differences

Considering respondents expected income over the next 12 months, while sub-sample AHPA appears to have a somewhat higher income than other sub-samples these differences proved to be barely insignificant. Similarly no significant differences were found either in gender or in the number of non-UK respondents in each sub-sample (who arguably would be less likely to receive the long term benefits of any air pollution impact reduction scheme). Considering respondent age, while the descriptive statistics shown in Table 2 show that mean age for all sub-samples was within the range 20 to 24 years, nevertheless significant differences were found with sub-sample AHPA again appearing to be the most different to other sub-samples.

Taking into account that this is also the sub-sample with the smallest number of respondents, it seems likely that there are a few older (and probably higher income) respondents within this sub-sample. Although these are not substantial differences they are worth keeping in mind when we consider our subsequent valuation results.

4.2 WTP for air pollution impact reduction schemes by sub-sample

Descriptive statistics for the various stated and calculated WTP measures obtained from each sub-sample are detailed in Table 3. Examining these we can see that WTP stated values for Scheme H are relatively stable between sub-samples, with mean measures ranging from about £72-£85 and median values being between £50-£70 (notice that the highest median values are obtained from sub-sample SPH, which is the only one where Scheme H is not presented first). Stated values for Scheme P are also relatively stable, with means ranging from £44-£54 and medians varying from £30-£47. However, these values differ substantially from the calculated values for Scheme P (found by subtracting stated values for Scheme H from those for Scheme A), with mean values ranging from £18-£28 and medians of £5-£10[9]. This large excess of stated over calculated values suggest either strong part-whole effects or that Schemes H and P are at least partial substitutes for each other. Stated values for Scheme A are also relatively similar across treatments with means ranging from £100-£113 and medians varying from £70-£100. Calculated values for Scheme A are consistently above their stated equivalents with means from £117-£131 and medians from £90-£120. Again this would be expected if we were either witnessing part-whole or substitution effects.

These results have some important messages for regulatory policy assessment. First, given the lack of incentive compatibility in our study these results tentatively suggest that values for air pollution impact reduction schemes may be significant. Second, the findings suggest that these values may be reasonably robust (although we investigate this issue further below). Thirdly, and perhaps most importantly, these findings suggest the presence of significant part-whole or substitution effects. This suggests that simply adding across schemes to obtain estimates of the value of wider schemes ignores the substitution effects which may exist between schemes and therefore risks the likelihood that the value of wider schemes may be over-estimated.

Finally inspection of the distributional information contained in Table 3 suggests that, as often observed in CV studies, distributions of WTP responses are positively skewed. The final column of the table reports a formal test for normality indicating that, in every case bar one of the calculated measures, normality is rejected at $p < 0.1$. This indicates that parametric tests relying upon such normality assumptions may be unreliable. Given this, in our analysis we employ non-parametric techniques for testing relationships between the measures collected.

[9] Note that the lower end of the distribution of calculated values for Scheme P includes a number of negative values derived from cases where WTP for Scheme H exceeds that for Scheme A. This may be a cause for some concern and possible explanations for such responses are considered in Bateman et al (2001).

Table 3: Descriptive WTP statistics by sub-sample and Scheme

Scheme	Measure	Count**	Mean	s.e. mean	Std Deviation	Median	p^*
H	SHA_H	38	79.50	11.81	72.79	52	.015
	$SHPA_H$	38	84.61	17.39	107.20	55	.010
	AHA_H	39	81.41	11.05	68.98	50	.010
	$AHPA_H$	28	72.18	11.96	63.31	50	.020
	SPH_H	40	81.53	7.70	48.72	70	.073
P	$SHPA_P$	38	54.16	15.89	97.95	30	.010
	$AHPA_P$	28	44.50	6.81	36.02	45	.035
	SPH_P	40	49.85	4.64	29.32	47.5	.010
	SHA_{cP}	38	27.63	11.15	68.71	5	.010
	AHA_{cP}	39	18.31	3.66	22.89	10	.010
A	SHA_A	38	107.13	17.72	109.21	70	.015
	$SHPA_A$	38	113.29	25.42	156.68	77.5	.010
	AHA_A	39	99.72	11.48	71.71	100	.081
	$AHPA_A$	28	104.18	15.95	84.42	90	.010
	$SHPA_{cA}$	38	138.76	32.96	203.17	90	.010
	$AHPA_{cA}$	28	116.68	17.37	91.89	100	.054
	SPH_{cA}	40	131.38	11.77	74.46	120	.348

* Shapiro-Wilk test of normality. Here *p* denotes the probability that the difference between a normal distribution and that of the observed WTP values is due to random chance.

** Note that counts are different to the sub-sample sizes from Table 2 due to five cases of unusable responses to the WTP questions (two cases from each of the SHA and SHPA sub-samples, and one case from the AHA sub-sample).

4.3 Tests of scope sensitivity and value consistency

In order to examine differences between WTP measures for Schemes both within and across treatments, a series of non-parametric tests were conducted[10]. A summary of findings is presented in Table 4. The numbers given in the cells of Table 4 indicate the number of tests which show either a significant (sig) or non-significant (ns) difference between the WTP values concerned[11]. Shaded cells are tests between comparable WTP sums for identical schemes elicited from different treatments. Here theoretical expectations are that all treatments should yield similar values[12]. Of the 23 such tests reported, only the test between the values for $SHPA_P$ and SPH_P show a significant difference in values (i.e. $p>0.05$). This suggests there is strong valuation consistency across treatments.

[10] Where the two samples were drawn from the same treatment a Wilcoxon signed-rank test was employed; where the two samples were drawn from different treatments a Mann-Whitney U test was employed.

[11] The values obtained from all treatments for one scheme were tested against the values obtained from all treatments for the other scheme. For example, when testing WTP for Scheme H against WTP for Scheme P, all five sets of values obtained for Scheme H (see table 3) were tested against all three sets of values obtained for Scheme P (see table 3). This resulted in a total of 15 tests. Three of these tests were between values obtained from the same treatment ($SHPA_H$ and $SHPA_P$ for example) and the results of these within-sample tests are listed in brackets.

[12] Assuming that schemes are seen as independent.

Considering comparisons between different value measures (non-shaded cells), numbers in brackets are within-sample (internal) tests, while numbers outside brackets are between-sample (external) tests. All of the former internal tests hold treatment constant and show consistently significant differences between measures. This confirms that scope sensitivity is indeed statistically significant and that, within any given treatment, values for Scheme H are significant larger than those for Scheme P and significantly smaller than those for Scheme A. This supports the anthropocentric prior that individuals value reduction of air pollution impacts upon human health more than the reduction of impacts upon plants.

The external tests, shown by the figures outside brackets in unshaded cells, are considerably less consistent and indicate that treatment differences across sub-samples do have significant impacts upon WTP values. It is to an analysis of these treatment differences that we now turn.

Table 4: Significance of differences in WTP values for Schemes*

WTP for Scheme		H		P		cP		A		cA	
Significance		sig	ns	sig	ns	sig	ns	sig	ns	sig	ns
WTP for Scheme	H	0(0)	10(0)								
	P	7(3)	5(0)	1(0)	2(0)						
	cP	8(2)	0(0)	6(0)	0(0)	0(0)	1(0)				
	A	1(4)	15(0)	10(2)	0(0)	6(2)	0(0)	0(0)	6(0)		
	cA	4(3)	8(0)	6(3)	0(0)	6(0)	0(0)	3(2)	7(0)	0(0)	3(0)

* Numbers in cells indicate the number of tests which show either a significant (sig) with $p \leq 0.05$ or non-significant (ns) difference between the WTP values concerned Shaded cells indicate tests between comparable WTP sums for identical schemes elicited from different treatments. Numbers in brackets are within-sample (internal) tests, while numbers outside brackets are between-sample (external) tests.

4.4 Tests of treatment effects

We have five distinct treatments, each of which yields stated and/or calculated values for each of the three schemes under consideration. This experimental design permits inspection of the impact of the dimensions discussed in section 2. First, in order to test whether the values varied across treatments[13], each response was categorised by the treatment type and scheme as follows:

Type 1 = stated values from stepwise disclosure treatments with Scheme H valued first (SHA & SHPA)[14]

[13] Only if significant differences exist between treatments can our analysis then examine whether there are significant part-whole effects, visible choice set effects, ordering effects, or event-splitting effects.

[14] Non-parametric tests clearly fail to reject hypothesis of no significant difference between measures included within this category (for SHA_A v. $SHPA_A$ $p = 0.896$; for SHA_H v. $SHPA_H$ $p = 0.888$)

Type 2 = calculated values from stepwise disclosure treatments with Scheme H valued first (SHPA)

Type 3 = stated values from advance disclosure treatments with Scheme H valued first (AHA & AHPA)[15]

Type 4 = calculated values from advance disclosure treatments with Scheme H valued first (AHPA)

Type 5 = stated values (Schemes P and H) or calculated value (Scheme A) from stepwise disclosure treatments with Scheme P valued first (SPH)

Mean and median WTP values for all three schemes across all five treatments are given in Table 5 together with non-parametric tests of the null hypothesis that values for a given scheme do not vary across the levels of the Type variable[16]. Here shaded cells indicated calculated values while unshaded cells indicate those which were directly stated by respondents.

Table 5: Mean and median WTP (£) for three air pollution impact reduction schemes, by five treatments

	Scheme H			Scheme P			Scheme A		
TYPE	n**	Mean WTP (s.e.)	Median WTP	n**	Mean WTP (s.e.)	Median WTP	n**	Mean WTP (s.e.)	Median WTP
1	80	82.05 (10.44)	52.00	38	54.16 (15.89)	30.00	76	110.21 (15.39)	72.50
2	40	59.13 (13.91)	40.00	38	27.63 (11.15)	5.00	38	138.76 (32.96)	90.00
3	68	77.55 (8.10)	50.00	28	44.50 (6.81)	45.00	67	101.58 (9.37)	100.00
4	28	59.68 (10.67)	35.00	39	18.31 (3.66)	10.00	28	116.68 (17.37)	100.00
5	40	81.53 (7.70)	70.00	40	49.85 (4.64)	47.50	40	131.38 (11.77)	120.00
Total	256	74.74 (4.74)	50.00	183	38.59 (4.43)	25.00	249	116.37 (7.79)	95.00
Diff*	$p = 0.026$			$p < 0.001$			$p = 0.057$		

* Diff = Kruskal-Wallis test of the null hypothesis that values for a given scheme do not vary across the various levels of the TYPE variable.

** Here n refers to the number of estimates, not to the number of respondents.

Inspecting Table 5 a number of clear messages can be seen. First, values for Scheme H are consistently higher than within treatment values for Scheme P and, as noted previously, there

15 Non-parametric tests clearly fail to reject hypothesis of no significant difference between measures included within this category (for AHA_A v. $AHPA_A$ $p = 0.888$; for AHA_H v. $AHPA_H$ $p = 0.473$).

16 Outlier sensitivity analysis confirmed that parametric tests are sensitive to outliers while non-parametric tests are stable. Details are available from the corresponding author.

is clear evidence of scope sensitivity with Scheme A values consistently higher than those for other schemes. Second, calculated values for both Scheme H and Scheme P are consistently and substantially below stated values, and this result is generally reversed for Scheme A. Remembering that, for Scheme A, calculated values are obtained by adding together stated values for Scheme H and Scheme P, overall this pattern provides some evidence for a part-whole/substitution effect with the sum of parts exceeding the stated value of the whole.

The last row of Table 5 gives a formal test of the null hypothesis of the equality of values for given schemes across the levels of the TYPE variable. Equality is clearly rejected for both Scheme H and Scheme P. For Scheme A the test statistic falls just outside the conventional 5% significance level, although it is clearly significant at the 10% level[17].

[17] This result becomes significant if the three highest values stated by respondents are omitted (p = 0.033).

Table 6: Treatment effects

Design Variable	Scheme H				Scheme P				Scheme A			
	n	Mean WTP (s.e.)	Median	Sig. Diff. (p)*	n	Mean WTP (s.e.)	Median	Sig. Diff. (p)*	n	Mean WTP (s.e.)	Median	Sig. Diff. (p)*
Stated values	188	80.29 (5.49)	60.00		106	49.98 (6.18)	40.00		143	106.17 (9.26)	80.00	
Calculated values	68	59.36 (9.14)	40.00	0.005	77	22.91 (5.79)	10.00	0.000	106	130.14 (13.34)	100.00	0.025
Advance disclosure	96	72.28 (6.55)	50.00		67	29.25 (3.87)	20.00		95	106.03 (8.34)	100.00	
Stepwise disclosure	160	76.26 (6.52)	51.00	0.933	116	43.98 (6.59)	30.00	0.097	154	122.75 (11.50)	92.50	0.736
Not SPH design	216	73.44 (5.45)	50.00		143	35.44 (5.50)	20.00		209	113.50 (9.01)	85.00	
SPH design**	40	81.53 (7.70)	70.00	0.039	40	49.85 (4.64)	47.50	0.000	40	131.38 (11.77)	120.00	0.006
Only Scheme H									77	103.38 (10.44)	90.00	
Both Schemes H/P***									172	122.19 (10.26)	100.00	0.243

* Mann Whitney U test of the null hypothesis of no significant difference in values across the two levels of the variable in question

** To test for the presence of an ordering effect

*** To test for the presence of the event splitting effect

Table 6 presents details of the influence of various experimental design variables upon WTP values for the three schemes. First, results show that when values are stated rather than calculated mean WTP for Scheme H is £80.29 compared to £59.36 for calculated values. Corresponding median values are £60 and £40 respectively, and the non-parametric test statistic shows that this difference is clearly statistically significant (p=0.005). Results for Scheme P conform to the same pattern. However, as discussed previously, this implies the opposite change in values for Scheme A with calculated mean WTP being £106.17 for stated values compared to £130.14 for stated values (with an increase in medians from £80 to £100). Our test statistic (p=0.025) shows that the part-whole/substitution effect suggested by these findings is clearly significant.

Second, mean WTP values are higher from stepwise than advance disclosure treatments. However, examination of medians shows that this effect is not clear-cut and tests only indicate significance at the 10% level for Scheme P and no significance for either of the other schemes. Table 6 shows that the impact of adopting the SPH design is to significantly raise the WTP statements for the initially valued good (Scheme P) relative to the comparatively low values accorded to this good under other treatments. This effect is carried over into the values for Scheme H elicited from treatment SPH which are again significantly higher than those for other treatments. Unsurprisingly this means that the calculated values for Scheme A from treatment SPH are also significantly higher than those from other sub-samples. Given that we now have clear evidence from previous tables that Scheme P is considered to be the lowest value of the goods presented to participants, this ordering effect finding is consistent with that of Bateman et al (2004) described in Section 2.3.

Finally Table 6 also reports results from our test of event splitting effects in values for Scheme A. Here, respondents asked to value both the constituent parts of Scheme A rather than only valuing Scheme H (and not Scheme P) had higher calculated WTP values for Scheme A with a mean of £122.19 compared to £103.38 (and corresponding median values £100 and £90 respectively). However, while this difference is in accordance with event-splitting expectations, tests show that this effect is not statistically significant in this instance[18].

Summarising Table 6 it appears that the within scheme variation in values is driven by whether a value is calculated or stated, and whether it was derived from the SPH treatment. However, these variables overlap significantly in that all of the values for Scheme A derived from treatment SPH are calculated rather than stated. As an additional test of the treatment effects, regression analysis was used with the WTP as the dependent variable, and each of the design variables from Table 6 as explanatory variables. This allows us to test for the significance of the effect of each design variables while controlling for the effects of the other design variables. Given the sensitivity of linear regression to the presence of outliers, these

[18] We also examined the possibility that the list length seen by respondents in the initial visible choice set may impact upon values for other schemes even after controlling for position within a list. Some, albeit weak, evidence for a list length effect can be gleaned by examining the stated values for Scheme H obtaining from advance disclosure treatments AHA and AHPA. Here Scheme H is always valued first and the only difference between the treatments is in terms of list length. From Table 3 we can see that mean stated WTP for Scheme H from treatment AHA (i.e. AHA_H) is £81.41 while the comparable value for Scheme H from treatment AHPA (i.e.) is £72.18. This suggests that values might decline as list length increases. However, these values are not significantly different yielding equal medians. Nevertheless, we believe that this might be a fertile area for future research.

regressions were performed at the median rather than the mean. The results of these median regressions are presented in table 7.

Table 7: Results of median regressions for design variable effects

	WTP for Scheme H		WTP for Scheme P		WTP for Scheme A	
Variable	Coefficient	*t*-value	Coefficient	*t*-value	Coefficient	*t*-value
Calculated	2	0.28	-5	-0.78	-15	-1.64
Stepwise	-12	-1.55	-25	-3.88***	5	0.52
SPH design	18	1.82*	20	2.58**	35	2.60**
Constant	50	9.21***	35	5.72***	95	12.75***
Pseudo R^2	0.0235		0.1029		0.0227	

* weakly significant at $p < 0.1$; ** significant at $p < 0.05$; *** significant at $p < 0.01$

From the regression analysis it appears that, after controlling for the effects of other design variables, that calculated values are not significantly different from stated values. It appears that the previously-noted difference between stated and calculated values was in fact driven by the SPH design effect, where the stated values for Schemes P and H were significantly higher and hence so were the calculated values for Scheme A. This suggests that the ordering effect (of eliciting the value for the lowest value good first, or otherwise) is the most significant of the design effects for all schemes. The stepwise design appears to have no effect on Schemes H or A, but has a significant negative effect on the median value elicited for Scheme P, i.e. where respondents were advised in advance of the schemes they would value, the values elicited for Scheme P were significantly lower, which provides additional support for the anthropocentric prior suggested in Sections 2.3 and 4.3.

None of this is inconsistent with the suggestion that these results suggest a part-whole/substitution bias. When two independent nested schemes are presented separately, the good valued is the first and only good valued. In this case, when Scheme P is valued first in a stepwise treatment, it is significantly higher than when it is valued in an advance disclosure. This suggests significantly higher calculated values for Scheme A when first response values for Scheme P are added to first response values for Scheme H.

With this in mind, a test was performed which permits examination of the crucial question of whether values derived from single ‘part’ good valuation studies can be added to those for other ‘parts’ to correctly estimate values for embracing ‘whole’ goods. To test this we examine whether the sum of values for Scheme H and Scheme P, presented as the first good encountered by respondents in designs where they are unaware of any subsequent valuation possibilities, yield a calculated value for A which is similar to that obtained from stated values for Scheme A.

Here we have two values for Scheme H which are both the first good valued by respondents and where those respondents faced stepwise designs and were unaware of the subsequent opportunities to value other goods (values SHA_H or $SHPA_H$). In contrast we have just one such value for Scheme P (value SPH_P). By adding the latter value for Scheme P with each in turn of the values for Scheme H we obtain estimates of calculated value for Scheme A based exclusively upon first response values from stepwise designs. Note that there are several ways in which these values could be aggregated, and a priori we have no theoretical expectations about how Scheme H and Scheme P should be aggregated. We have chosen two methods – firstly aggregating the highest values of Scheme H with the highest values of Scheme P[19], and secondly aggregating the lowest values of Scheme H with the highest values of Scheme P (and vice versa)[20]. In reality, the 'true' aggregation scheme is likely to lie somewhere between these two extremes. These aggregations mimic the estimated value of Scheme A which would typically be obtained from combining values from most conventional CV studies of the constituent parts of this good. These values can be contrasted with those stated values for Scheme A derived from our design (values SHA_A, $SHPA_A$, AHA_A, $AHPA_A$).

Table 8 compares measures for the stated values for Scheme A with those obtained by summing first response stepwise values[21].

Table 8: Comparing stated WTP for Scheme A with values calculated from stepwise first responses for Scheme H and Scheme P

Scheme A measure	Count	Mean (s.e.)	Median	*p*-value (Mann-Whitney U test)
Stated Values for A	143	106.17 (9.26)	80.00	–
<u>High-High aggregations</u>				
$SHA_H + SPH_P$	38	131.58 (16.30)	102.00	0.0526
$SHPA_H + SPH_P$	38	136.68 (21.35)	105.00	0.0558
<u>High-Low aggregations</u>				
$SHA_H + SPH_P$	38	125.66 (8.86)	110.00	0.0018
$SHPA_H + SPH_P$	38	130.76 (15.13)	115.00	0.0013

19 This assumes a positive association between values given for Scheme P with those given for Scheme H.

20 This assumes a negative association between values given for Scheme P with those given for Scheme H.

21 Note that there is a difference in the sample size between the SHA_H (38), $SHPA_H$ (38) and SPH_P (40) sub-samples. This was resolved by omitting the two lowest values of SPH_P in the high-high aggregation, and the two highest values of SPH_P in the high-low aggregation.

Inspecting Table 8 we can see that calculated values obtained by summing first response values for the constituent parts of Scheme A substantially overestimates the stated values of the Scheme. Overestimates of mean WTP values range from 18-29% while overestimates of median values range from 28-44%. Non parametric tests confirm that for the high-low aggregation the differences are highly significant and for the high-high aggregation the differences only just fail to achieve significance at the 5% level although it is clearly significant at the 10% level. Given that the 'true' aggregation lies somewhere between these two extremes it is highly likely that single good (part) valuations, added together, result in substantial and significant overestimates of combined good (whole) values.

5. Summary and Conclusions

This study reports an analysis of certain characteristics of values for the reduction of air pollution impacts as estimated using the CV method. We investigated a number of issues and potential anomalies which have been highlighted in the CV and experimental economics literature, including (i) scope sensitivity; (ii) part-whole / substitution effects; (iii) ordering effects; and (iv) visible choice set effects. A novel split sample experimental design for the first time allowed simultaneous investigation of all these anomalies within the context of the same valuation exercise. Values were elicited for three Schemes to reduce the impacts of air pollution upon (i) human health (Scheme H); (ii) plants (Scheme P); and (iii) human health and plants (Scheme A; which combined the effects of Scheme H and Scheme P). *Stated values* were obtained for each of these schemes across various treatments which define our study design. In addition to these *calculated values* were obtained, implicitly assuming the absence of part-whole/substitution effects between the values of Schemes H and P. By comparing stated and calculated values for different treatments and Schemes we can test these assumptions and for the presence of the other anomalies.

Our experiments yielded a number of findings. There was considerable value consistency within stated values for each scheme suggesting that respondents were referring to some underlying (although not necessarily theoretically consistent) preferences or valuation process. Furthermore, no anomalies were found regarding sensitivity to the scope of schemes, instead general evidence of significant scope sensitivity was observed.

However, the use of stepwise designs which present participants with low value goods first (i.e. our SPH treatment) appears to generate significantly different values than do other approaches. Specifically, when a good which is valued at a relatively low level in other treatments is presented at the beginning of a stepwise list its value is elevated. This finding could be interpreted as either a theoretically consistent substitution effect (Carson et al, 1998) or as the impact of a theoretically inconsistent 'moral satisfaction of giving' to a good cause being attached to first responses (Kahneman and Knetsch, 1992). Disentangling the different potential drivers of an identical effect is problematic and would require a considerable 'verbal protocol' extension to our design (Schkade and Payne, 1992). However, the consequent effect upon Scheme H values in the SPH treatment cannot be explained by economic theory which would expect that the movement from first position in all other treatments to second position in the SPH list would result in a reduction in stated values arising from substitution effects. Instead, as shown in Tables 6 and 7, values for Scheme H from the SPH treatment are significantly higher than those in other treatments. While this is inconsistent with economic theory it does conform to psychological expectations based on an 'anchoring and adjustment' heuristic (Tversky and Kahneman, 1974) wherein the high values stated previously for Scheme P feed through into elevated values for Scheme H stated subsequently.

While the latter finding is of most concern from a theoretical and methodological perspective, perhaps the most important practical finding concerned the clear evidence found of significant part-whole/substitution effects. In particular, we found that summing the values obtained from several single good valuation exercises (i.e. corresponding to first responses in our stepwise disclosure designs) to calculate estimates for wider goods risks the likelihood of significantly overestimating the value of the latter wider goods. Policymakers need to be aware of the potential for such relationships when assessing valuation evidence as part of efforts to design appropriate economic instruments for regulatory purposes.

In summary, our findings raise a number of theoretical and methodological and applications issues which need to be borne in mind when undertaking valuation work regarding air pollution externalities. Indeed we might expect that a number of these concerns may well apply to many public good valuation exercises. However, in conclusion we should remember that this was a relatively simple exercise dictated by resource constraints which precluded the use of incentive compatible designs. Therefore its findings should be treated with caution. Nevertheless the fundamental nature of the concerns raised suggest that these issues are worthy of further investigation within a more controlled and incentive compatible framework.

References

Andreoni, J., (1990) Impure altruism and donations to public goods, *Economic Journal*, 100: 464-77.

Arrow, K., Solow, R., Portney, P.R., Leamer, E.E., Radner, R. and Schuman, H. (1993) *Report of the NOAA Panel on Contingent Valuation*, Resources for the Future, Washington D.C.

Bateman, I.J., Carson, R.T., Day, B., Hanemann, W.M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozdemirogla, E., Pearce, D.W., Sugden, R. and Swanson, J. (2001). *Guidelines for the Use of Expressed Preference Methods for the Valuation of Preferences for Non-market Goods*, report to the Department of the Environment, Transport and the Regions, Economics for the Environment Consultancy, London.

Bateman, I.J., Cole, M., Cooper, P., Georgiou, S., Hadley, D. and Poe, G.L. (2004). On visible choice sets and scope sensitivity, *Journal of Environmental Economics and Management,* 47, 71-93.

Bateman, I.J. and Langford, I.H. (1997) Budget constraint, temporal and ordering effects in contingent valuation studies, *Environment and Planning A*, 29(7): 1215-1228.

Bateman, I.J., Munro, A., Rhodes, B., Starmer, C. and Sugden, R. (1997a) Does part-whole bias exist? An experimental investigation, *Economic Journal*, 107(441): 322-332.

Bateman, I.J., Munro, A., Rhodes, B., Starmer, C. and Sugden, R. (1997b) A test of the theory of reference-dependent preferences, *Quarterly Journal of Economics,* 112(2): 479-505.

Bateman, I.J. and Willis, K.G. (eds.) (1999) *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*, Oxford University Press, pp645.

Boyle, K., Desvourges, W.H., Johnson, F.R., Dunford, R.W. & S.P. Hudson, (1994), An Investigation of Part-Whole Biases in Contingent Valuation Studies, *Journal of Environmental Economics and Management*, 27, 64-38.

Boyle, K. J., M. P. Welsh, and R. C. Bishop, (1993) The role of question order and respondent experience in contingent-valuation studies, *Journal of Environmental Economics and Management*, 25 S-80-S-99.

Carson, R.T. (1997) Contingent valuation surveys and tests of insensitivity to scope, in Kopp, R.J., Pommerehne, W.W. and Schwarz, N. (eds.) *Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*, Kluwer Academic Publishers, Boston.

Carson, R.T., Flores, N.E. and Hanemann, W.M. (1998) Sequencing and valuing public goods, *Journal of Environmental Economics and Management*, 36; 314-323.

Carson, R.C., N. E. Flores, and N. F. Meade, (2001) 'Contingent valuation: Controversies and evidence', *Environmental and Resource Economics*, 19(2): 173-210.

Carson, R.T. and Mitchell, R.C. (1993), The Issue of Scope in Contingent Valuation Studies, *American Journal of Agricultural Economics*, December: 1265-1267.

Carson, R.T. and Mitchell, R.C. (1995) Sequencing and nesting in contingent valuation surveys, *Journal of Environmental Economics and Management*, 28(2): 155-173.

Cubitt, R.P., and Sugden, R. (2001) On money pumps, *Games and Economic Behavior*, 37(1):121-160.

Diamond, P.A. and Hausman, J., (1994), Contingent Valuation: Is some number Better than No Number?, *Journal of Economic Perspectives*, 8(4), 43-64.

Ferraro, P. J., Rondeau, D. and Poe, G.L. (2003) Detecting other-regarding behavior with virtual players, *Journal of Economic Behavior and Organization*, 51(1): 99-109.

Goodstein, E.S. (1995) *Economics and the Environment*. Prentice Hall, Englewood Cliffs, NJ.

Hanemann, W.M., (1994), Valuing the Environment Through Contingent Valuation, *Journal of Economic Perspectives*, 8(4), 19-43.

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, pp302-442.

Harrison, G.W. (1992), Valuing public goods with the Contingent Valuation Method: A Critique of Kahneman and Knetch, *Journal of Environmental Economics and Management*, 23, 248-57.

Hausman, J.A., (ed.) (1993) *Contingent Valuation: A Critical Assessment*, North-Holland, Amsterdam.

Hoehn, J.P. (1983) The benefits-costs evaluation of multi-part public policy: A theoretical framework and critique of estimation methods, *Ph.D. Dissertation*, University of Kentucky.

Hoehn J. P. and A. Randall (1982). *Aggregation and disaggregation of program benefits in a complex policy environment: A theoretical framework and critique of estimation methods*, paper presented at the annual meetings of the American Agricultural Economics Association, Logan, Utah.

Humphrey, S.J. (1996) Do anchoring effects underlie event-splitting effects? An experimental test, *Economics Letters*, 51: 303-308.

Humphrey, S.J. (1995) Regret-aversion or event-splitting effects: More evidence under risk and uncertainty, *Journal of Risk and Uncertainty*, 11: 263-274.

Kahneman, D. and Knetsch, J.L. (1992) Valuing public goods: the purchase of moral satisfaction, *Journal of Environmental Economics and Management*, 22:57-70.

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

Randall, A., and J. P. Hoehn (1996) Embedding in market demand systems, *Journal of Environmental Economics and Management*, 30:369-380.

Randall, A., J. P. Hoehn, and G. S. Tolley (1981). *The structure of contingent markets: Some empirical results*, paper presented at the Annual Meeting of the American Economic Association, Washington D. C.

Schkade, D.A. and Payne, J.W. (1994). How people respond to contingent valuation questions: a verbal protocol analysis of willingness to pay for an environmental regulation, *Journal of Environmental Economics and Management*, 26: 88-109.

Smith, V.K. (1992), Comment: Arbitrary Values, Good Causes, and Premature Verdicts, *Journal of Environmental Economics and Management*, 22: 71-79.

Starmer, C. and Sugden, R. (1993) Testing for juxtaposition and event-splitting effects, *Journal of Risk and Uncertainty*, 6: 235-254.

Svedsater, H. (2000), Contingent valuation of global environmental resources: Test of perfect and regular embedding, *Journal of Economic Psychology*, 21(6): 605-623.

Thaler, R. (1985). Mental accounting and consumer choice, *Marketing Science*, 4(3): 199-214.

Tolley, G. S., A. Randall, G. Blomquist, R. Fabian, G. Fishelson, A. Frankel, J. Hoehn, R. Krumm, and E. Mensah (1983). *Establishing and valuing the effects of improved visibility in the Eastern United States*, interim report to the U.S. Environmental Protection Agency.

Tversky, A. and Kahneman, D. (1974) Judgment under uncertainty: Heuristics and biases, *Science*, 185: 1124-1130.

Varian, H.R. (1992) *Microeconomic Analysis*, Third edition, University of Michigan.