Journal of Environmental Management (2001) **62**, 255–269
doi:10.1006/jema.2001.0439, available online at http://www.idealibrary.com on IDEAL®

A comparison of alternative contingent valuation elicitation treatments for the evaluation of complex environmental policy

M. Christie

Attempts to evaluate the economic benefits associated with complex environmental policies, using the contingent valuation (CV) method, have been dogged by controversy. In particular, debate has centred on the influence of embedding and related effects on the validity of CV willingness-to-pay (WTP) estimates. This paper discusses these effects in the context of identifying the most appropriate WTP elicitation treatment to evaluate the willingness to pay for the various elements of a multi-dimensional environmental policy. The findings of an empirical experiment demonstrate that a top-down allocation treatment, which uses independent sub-samples for individual policy components, provides the most reliable treatment to value all aspects of a multi-dimensional policy, whilst also overcoming embedding bias problems.

Keywords: contingent valuation, WTP elicitation strategies, embedding bias, complex environmental policy.

Introduction

Contingent valuation (CV) studies have traditionally focused on the evaluation of willingness-to-pay (WTP) for either a single policy programme or a composite package of policy programmes. Although the results of these studies can provide useful information for policy formulation, it is argued that better policy decisions can be made if data on the economic value of both the policy package and its component programmes are available (Christie, 1999; Hoehn, 1991). However, attempts to simultaneously evaluate both of these dimensions of complex, multi-dimensional policies within a single CV instrument have often been influenced by embedding bias and related effects (Brown and Duffield, 1995; Hoehn, 1991). Embedding effects have been the subject of debate and controversy over recent years; both in terms of the exact definition of the term 'embedding' (Carson and Mitchell, 1995; Hanemann, 1994) and its influence on the validity of the CV method (Kahneman and Knetsch, 1992; Hanemann, 1994). This paper aims to investigate further the influence of embedding and related effects on the validity of three WTP elicitation treatments that are used to evaluate a multi-dimensional policy that provides improvements to recreational opportunities in the Scottish countryside. Here, we first identify three alternative WTP elicitation approaches to the valuation of multi-dimensional policies. We then attempt to clarify some of the confusion surrounding the embedding debate and propose a series of hypotheses that are then used to test for these effects. Based on the findings of the empirical experiment, recommendations regarding the validity and reliability of three elicitation strategies for the evaluation of multi-dimension policies are made.

Email of author: mec@aber.ac.uk

Elicitation strategies for multi-dimensional policies

Multi-dimensional environmental policies may be considered at two distinct levels: the

Institute of Rural Studies, University of Wales, Aberystwyth, Llanbadarn campus, Aberystwyth, SY23 3AL Wales, UK

Received 20 March 2000; accepted 31 January 2001

0301–4797/01/070255+15 $35.00/0

overall policy *package*[1] and its component *programmes*, where each programme is a nested subset of the package (Kahneman and Knetsch, 1992). Policy-makers may be interested in the overall impacts of a policy package, as well as the impacts of individual component programmes. Attempts to elicit values for both levels of multi-dimensional policies have predominantly required CV researchers to undertake several studies to evaluated each component separately (Boyle *et al*., 1990). It is clear that the requirement to undertake multiple surveys is expensive and time-consuming (Hoehn, 1991). A more efficient option is to undertake a single study that evaluates both the package and component programmes (Hoehn, 1991; Brown and Duffield, 1995).

Carson and Mitchell (1995) identify two basic WTP elicitation strategies that enable simultaneous evaluation of all elements of multi-dimensional policies. The first strategy involves the evaluation of the benefits associated with the individual component programmes, which are then aggregated to estimate the value for the policy package, i.e. a *bottom-up* (BU) approach.[2] The second strategy involves the evaluation of the total benefits of the policy package, which is then disaggregated into the component values of each of the policy programmes, i.e. a *top-down* approach. Two discrete variants of the top-down approach can be identified. The first variant asks respondents to state their WTP for the composite policy package, followed by their WTP for only one of the component programmes. Independent sub-samples are used to value each component programme. This approach will be referred to as an *allocation using sub-samples* (AS) method. The second variant also asks respondents to state their WTP for the composite policy package. However, it differs from the AS method in that the respondents are then asked to disaggregate their WTP for the composite among *all* of the component programmes. This approach will be referred to as a *disaggregative allocation* (DA) method. The empirical work undertaken in this research examines the validity of the bottom-up and the two top-down elicitation strategies.

Brown and Duffield (1995) argue that the approach used to pose the WTP elicitation question within a CV survey may influence the instrument's susceptibility to part-whole bias, which in turn will impact the resultant evaluations of the policy package and its component programmes. They identify two main causes of part-whole bias. Value allocation bias relates to the situation where '*respondents simply misinterpret the description of the good*' (Brown and Duffield, 1995:2342). Although being an issue of concern for an individual CV study, Brown and Duffield argue that such bias effects may be overcome by more clearly specifying the valuation path, and therefore they argue that value allocation bias does not present a major threat to the validity of the CV technique. The second type of part-whole bias identified by Brown and Duffield is embedding bias. Embedding bias is defined as the differences in the scale or scope of the policy that occur '*when the respondent values a larger entity symbolised by the specific good described in the survey*' (Brown and Duffield, 1995:2342). Kahneman and Knetsch (1992) argue that embedding bias effects may threaten the validity of an individual CV study, or indeed the CV method as a whole. It is therefore apparent that investigations of alternative elicitation approaches for the evaluation of multi-dimensional policies are required to consider the influence of embedding bias effects.

Embedding and related effects

The term embedding was first introduced by Kahneman and Knetsch (1992) to describe the phenomenon where the WTP for a particular good may vary '*over a wide range depending on whether the good is assessed on its own or embedded as part of a more inclusive package*' (Kahneman and Knetsch, 1992:58). Kahneman and Knetsch argue that such effects pose a major threat to the validity of CV studies; a conclusion supported by a number of other CV researchers (Desvousges *et al*., 1993; Diamond *et al*., 1993; Hausman and Diamond, 1994). Furthermore, the NOAA panel, in its review of the CV technique, concluded that embedding poses potentially '*the*

[1] Other authors have referred to the overall package of policy programmes as the policy agenda (Hoehn and Loomis, 1993) or the composite good (Brown and Duffield, 1995). In this paper, we shall refer to this grouping of policy programmes as the policy package.

[2] Hoehn (1991) refers to this conventional approach as *independent valuation and summation*.

most important internal argument against the reliability of the CV approach' (Arrow *et al*., 1993:4607).

However, other researchers argue that Kahneman and Knetsch's definition of the term embedding was loosely defined and that the Kahneman and Knetsch's (1992) results were influenced by other related effects which do not pose such a major threat to the validity of the CV technique (Hanemann, 1994; Carson and Mitchell, 1995). Hanemann (1994) argues that the term *embedding* has come confusingly to mean several different things. To clarify the situation, he identifies three distinct notions related to embedding effects: '*scope*' effects relate to situations where the extent of a change in WTP values do not correspond with changes in scale or scope of the item being valued; '*sequencing*' effects where the WTP of a good when placed first in a sequence will be higher than when placed later in the sequence; and '*sub-additivity*' effects where the WTP for a composite change in a group of public goods may be less than the sum of the WTPs for the individual changes valued separately. With regard to scope effects, Hanemann argues that Kahneman and Knetsch's findings are the exception to the norm and that a large number of other CV studies have shown sensitivity to scope effects. Hanemann continues by arguing that the perceived scope insensitivity found by Kahneman and Knetsch may have, alternatively, been induced by poor survey design. Hanemann also suggests that the other two concepts (sequencing and sub-additivity) are to be expected and can be explained by economic theory in terms of substitution effects and diminishing marginal rates of substitution. Therefore it is argued that sequencing and sub-additivity effects pose no real threat to the validity of the CV method. Further support for the argument that the sub-additivity effect is consistent with economic theory is provided in (Brown and Duffield, 1995; Hoehn, 1991; Hoehn and Loomis, 1993).

Carson and Mitchell (1995) also attempt to clarify the embedding confusion by proposing a framework where true embedding and sequencing effects can be separated from amenity misspecification bias effects. They suggest that amenity misspecification biases, which include symbolic, part-whole and probability of provision biases, may produce an embedding-like effect as a result of poor survey design. They argue that such effects may be overcome by improving survey design and have proposed a set of standards to minimise these biases (Carson and Mitchell, 1995:171). To establish the presence of true embedding and sequencing bias effects Carson and Mitchell propose three tests: an embedding effect test; a nested sequence test; and a test of component sensitivity (Carson and Mitchell, 1995:158–160). The adoption of these tests enables a systematic assessment of the sensitivity of a CV instrument to the scope of an environmental policy, and therefore provides a useful assessment of the validity of that particular instrument.

The research aims

The above review has brought together much of the recent literature relating to embedding bias. In doing so, it has attempted to clarify the definition of embedding bias and related effects. Through the identification of the causes of these bias effects, conclusions may be made regarded the impact of these effects on the validity of the CV instrument. Although significant progress has been made to clarify this issue during recent years, one question that remains unanswered relates to which WTP elicitation strategy is least likely to be affected by embedding effects, and therefore which strategy produces the most valid WTP estimates for multi-dimensional policies. It is this question that the research reported here addresses.

Research test hypothesis

The validity of the three alternative elicitation treatments are tested in this research through a series of seven hypotheses that examine the consistency of WTP estimates between treatments and also tests for the presence of embedding and related effects. However, before examining these hypothesis, it is first useful to outline the nomenclature that is used in this paper. The value of the policy package is denoted as $v(P)$, while three nested programmes are denoted by $v(a)$, $v(b)$ and $v(c)$. Where required, a subscript is used to indicate the elicitation treatment used, while a superscript is used

to indicate the position of the WTP elicitation question within the elicitation sequence.

The first hypothesis (Equation (1)) examines the equality of WTP estimates for the policy package $v(P)$ between the AS and DA treatments.[3] The null hypothesis (H_0) is that there is no significant difference in the WTP for the policy package between treatments. *A priori* it is expected that the null hypothesis will not be rejected since the two CV questionnaires were identical up until this point:

$$H_0: \quad v(P)_{\mathrm{AS}} = v(P)_{\mathrm{DA}} \tag{1}$$

Where $v(P)$ represents the WTP for the policy package measured using AS and DA elicitation treatments.

The second set of consistency tests (Equation (2)) examines the equality of WTP for the policy programmes $v(a)$, $v(b)$ and $v(c)$ measured using the different elicitation treatments:

$$H_0: \quad v(a)_{\mathrm{I}} = v(a)_{\mathrm{II}} \tag{2}$$

Where $v(a)$ represents WTP for a policy programme measured using two alternative elicitation treatments I and II.

Although it is difficult to speculate the exact outcome of these tests, the following predictions can be made. Where the elicitation questions are posed at the same position in the elicitation sequence, i.e. as in the AS and DA treatments, it is expected that the equality null hypothesis would not be rejected. However, where the elicitation question is embedded at different positions within the elicitation sequence, as is the case in the comparison of the AS and DA treatments with the BU treatment, rejection of the null hypothesis may demonstrate the existence of an embedding effect. Details of this embedding effect test are discussed later in Equation (5).

The third proposed test examines the consistency of the WTP for the policy package $v(P)$, with the aggregate of the WTPs for the component policy programmes; $v(a)+v(b)+v(c)$. The null hypothesis (Equation (3)) indicates that there is no difference between the WTP for the policy package and the aggregate value of the individual programmes.

$$H_0: \quad v(P) = v(a)+v(b)+v(c) \tag{3}$$

Where $v(P)$ and $v(a)$, $v(b)$ and $v(c)$ are the values of the policy package and three component programmes measured using the same elicitation strategy.

The alternative hypothesis to Equation (3) can be used to predict the presence of sub-additivity effects if it is demonstrated that the aggregate WTP value of the three policy programmes is significantly greater than the value of the policy package (Hanemann, 1994).

The final consistency test examines whether the proportion of total WTP allocated to an individual programme is consistent between treatments. The null hypothesis (Equation (4)) predicts equality in the relative allocation of WTP to an individual programme between treatments. Rejection of the null hypothesis would indicate that the treatments were inconsistent with respect to the allocation of WTP among the policy programmes:

$$H_0: \frac{v(a)_{\mathrm{I}}}{v(a)_{\mathrm{I}}+v(b)_{\mathrm{I}}+v(c)_{\mathrm{I}}} = \frac{v(a)_{\mathrm{II}}}{v(a)_{\mathrm{II}}+v(b)_{\mathrm{II}}+v(c)_{\mathrm{II}}} \tag{4}$$

Although no predictions can be made with regard to the outcome of this test, failure to reject the null hypothesis would indicate that the various elicitation treatments were consistent in the way in which they proportionate WTP among the component policy programmes.

In addition to the consistency tests outlined above, three further tests (proposed by Carson and Mitchell, 1995) were conducted on the survey data to establish the sensitivity of the alternative elicitation treatments to the individual components of the multi-dimensional policy.

Carson and Mitchell's (1995) '*embedding effect*' test (Equation (5)) is an external test that examines whether the WTP for a policy programme is reduced as that programme is embedded further into the sequence of WTP elicitation questions. The null hypothesis, H_0, is that there is no significant difference in the WTP estimates. The alternative hypothesis, H_1, indicates that an embedding effect exists

[3] $v(P)$ was not estimated using the BU treatment, and therefore no comparisons of $v(P)$ were made against this treatment.

when the inequalities described in Equation 5 are found:

$$H_1: \quad v(a)^2_{\mathrm{I}} < v(a)^1_{\mathrm{II}} \tag{5}$$

where: $v(a)$ represents the WTP for a specific policy programme, the superscript denotes the position of the elicitation question within the sequence, the subscript represents the sub-sample used.

Carson and Mitchell's (1995) '*nested sequence test*' is an internal test of consistency within the valuation of nested goods (Equation (6)). The alternative hypothesis, H_1, predicts the following inequalities:

$$H_1: \quad v(P)^1_{\mathrm{I}} > v(a)^2_{\mathrm{I}} \tag{6}$$

where: $v(a)$ represents the WTP for a policy programme that is nested within the policy package $v(P)$. The superscript and subscript numbers are the same as in Equation (5).

The presence of the inequality shown in Equation (6) indicates that responses are consistent with economic theory in that respondents were willing to pay more for the policy package $v(P)$ than for the smaller nested policy programme $v(a)$. Therefore acceptance of H_1 demonstrates that the elicitation treatment under investigation is internally consistent.

The final test proposed by Carson and Mitchell (1995) is the '*test of component sensitivity*'. This test examines whether the CV instrument is sensitive to different levels of the good (quantitative nesting) or the level of inclusion (categorical nesting). Essentially, this test is similar to the nested sequence test, the main difference being that the test of component sensitivity is an external test, in that the comparison is made between sub-samples rather than within the same sub-sample:

$$H_1: \quad v(P)^1_{\mathrm{I}} > v(a)^1_{\mathrm{II}} \tag{7}$$

Acceptance of the test of component sensitivity (Equation (7)) can be used to support the validity of a CV instrument since it demonstrates that the instrument is sensitive to the scope of the environmental policy. The null hypothesis, which predicts no difference in WTP values, is the test of component insensitivity. Carson and Mitchell (1995) argue that failure to reject the null hypothesis may indicate that the CV instrument used is invalid as it is not sensitive to the scope of the environmental policy.

The empirical study

The multi-dimensional policy investigated in the empirical experiment examined public WTP for improvements to recreation opportunities in the countryside of Grampian Region, Scotland. The recreation improvement policy package $v(P)$ comprised three nested programmes: a path improvement programme $v(a)$, a path creation programme $v(b)$, and a provision of facilities programme $v(c)$. The specifications of the policy package and the three nested programmes (see Table 1 for a summary description of recreation improvement scenarios) were constructed following

Table 1. Summary descriptions of recreation improvement programmes

	Repair of countryside paths programme $v(a)$	Creation of new paths programme $v(b)$	Provision of countryside facilities programme $v(c)$
Policy programmes	Repair the surface of paths and the soil and plants next to the path	Creation of short distance circular paths in popular recreation areas	Provision of more car parks, bins, seats, information boards and sign posting where appropriate
	The use of steps, wooden boardwalks, seats and signs to upgrade paths where appropriate	Creation of long distance routes that link popular areas together	Provision of more toilets, picnic areas, children's play areas and visitor centres where appropriate
Policy package		The policy package includes all components of the above three improvement programmes	

focus group discussions with members of the public and experts in countryside recreation.

The CV questionnaire comprised three sections. The first section aimed to establish respondents' use of the countryside, the problems that they had encountered and the potential solutions to these problems. Standard checklists were used within this section to ensure that respondents were made aware of the full range of recreational opportunities available in Grampian Region.

The second section dealt with the WTP elicitation question. Here, respondents were split among the three alternative elicitation treatments. Respondents were presented with descriptions of the policy package and component programmes relevant to the treatment investigated, before being asked to state their WTP for the policy components. The three WTP elicitation treatments used in the research were the *bottom-up*, the *allocation using sub-samples* and the *disaggregative allocation* treatment (Figure 1). The BU treatment aimed solely to estimate respondents' WTP for a single improvement programme. Respondents of the BU treatment were therefore split into three independent sub-samples, with each sub-sample being presented with only one of the three improvement programmes. Throughout the CV questionnaire, the respondents' attention was focused entirely on the single programme under investigation. This treatment, therefore, aimed to provide an unbiased, separately assessed WTP estimate for each of the three policy programmes. Respondents of the AS treatment were also split into three sub-samples. Initially all respondents were asked to state their WTP for the composite policy package, before being split into discrete sub-samples and asked to indicate their WTP for only one of the policy programmes. Only one sample was used for the DA treatment. Here, survey respondents were asked to indicate their WTP for the policy package and then asked to disaggregate this value among all three component programmes. It should be noted that the main difference between the two top-down treatments was that the AS treatment allowed respondents to freely state their WTP for both the policy package and the component programmes, whereas in the DA treatment, the respondents' WTP bids for the three programmes were restricted by the budget constraints set in the valuation of the policy package. The DA treatment, however, had the additional advantage of requiring a much smaller sample, since only one sample of respondents was required.

The actual WTP elicitation questions used in the study were posed in two stages. First, respondents were asked '*would your household be prepared to pay towards a* [specified] *programme*.' If the respondent stated 'yes' to this question, they were then asked to state '*What is the maximum amount your household would be prepared to pay as increases in Council Tax over a five year period*' for the improvement policy under consideration.[4] An open-ended elicitation format was adopted since this enabled a series of elicitation questions to be posed to a single respondent without inducing implied value cues such as starting point bias (Mitchell and Carson, 1989). Respondents were made aware of their personal budget constraints and also asked to state how they came to their WTP value, or in the case of zero bids the reason for non-contribution towards the improvement policy. The responses to these questions were used to assess the validity of WTP bids and also enabled protest bids to be identified. Protest bids were subsequently removed from the analysis.

The final section of the CV questionnaire established respondents' socio-economic details. This information was used to assess the extent to which the survey sample was representative of the Grampian population and also to analyse of the validity of WTP responses.

The CV questionnaire was developed and tested through a series of pilot studies that examined the questionnaire for comprehension, biases and reliability with respect to the description of the improvement programmes and the elicitation treatments. The actual CV survey was implemented using a postal format, with respondents being chosen randomly from the electoral roll. Dillman's (1987) 'Total Design Method', which advocates good practice in questionnaire design and employs a series of follow-up mailings, was employed to maximise survey response. A total of

[4] The Council Tax is a local government tax that is levied, partly relating to property but independent of income, to fund public service provision.

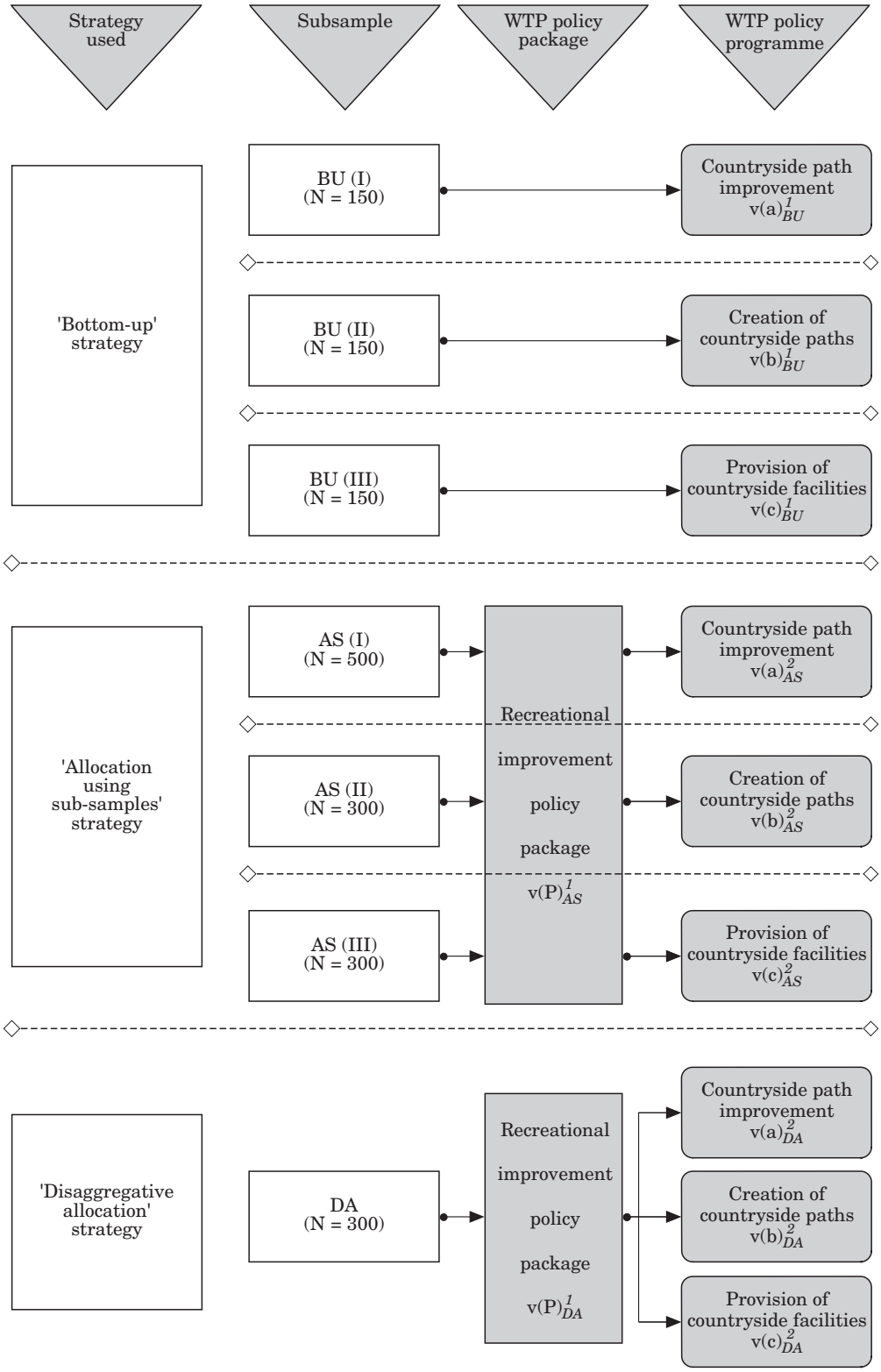

Figure 1. Diagrammatic illustration of the design of the CV experiment.

1850 Grampian households were included in the study. The distribution of respondents between the treatments is outlined in Figure 1. A full account of the development and implementation of the CV questionnaire may be found in Christie (1999).

Results

Of the original 1850 questionnaires mailed to Grampian residents, 931 (50·3%) were returned in a usable format (Table 2). Three types of response to the WTP question were recognised: positive bids, zero bids and protest bids.[5] The distribution of response type within the AS and DA treatments were found to be similar ($\chi^2=0{\cdot}22, P=0{\cdot}89$); however, significant differences were found between the types of response found in the BU treatment than with the DA and AS treatments ($\chi^2=12{\cdot}65, P=0{\cdot}00$; $\chi^2=18{\cdot}43, P=0{\cdot}00$, respectively). Although the data collected did not enable the reason for this difference to be determined, it was considered that the lower level of positive bids found using the BU treatment may be due to the fact that respondents were less willing to contribute towards a single programme compared to the composite package.

Chi-square comparisons of the socio-economic characteristics (gender, age, income, employment status and recreation participation behaviour) indicated no significant difference at the 5% level between the respondents of the three treatments, thus supporting the comparability of sub-sample populations. A similar comparison of the survey sample's socio-economic characteristics with those from the Grampian census indicated that the survey sample was representative of the Grampian population as a whole. Details of these test may be found in Christie (1998).

[5] The protest bids were identified, principally, as those respondents who protested about the payment vehicle, i.e. respondents stating that they did not want to pay increases in Council Tax.

Consistency tests

The mean and truncated[6] mean WTP bids for the policy package and associated policy programmes elicited using the three elicitation treatments are summarised in Table 3. The truncated mean WTP values (positive and zero bids) elicited for the composite policy package $v(P)$ were £7·22 and £5·72 per household using the AS and DA treatments, respectively.[7] Regression analysis, which is used to test the validity of WTP responses, indicated that the bid curves behaved as expected, with WTP increasing in line with various socio-economic attributes including income and membership of environmental/recreational organisations. Details of this analysis can be found in Christie (1999). The equality of the AS and DA estimates of WTP for the policy package was tested as proposed in Equation (1). Both a t-test and a Mann-Whitney U test[8] (bottom row in Table 4) demonstrate no significant difference ($P>0{\cdot}05$) between these two estimates of $v(P)$, thus the consistency null hypothesis was not rejected. This finding was expected since the two surveys were identical up to

[6] A 5% fixed factor truncation technique, recommended by Mitchell and Carson (1989:227) was used for the removal of outliers. Although it was evident that truncation reduced the mean WTP, it was considered that the truncated data set was the preferred data set for analysis since it was evident that a small number of very high bids were greatly influencing the resultant mean WTP values. Therefore, the removal of these bids removed the adverse effect of the outliers.

[7] A value for the composite package was not elicited using the bottom-up treatment as this version of the survey aimed only to address individual programmes.

[8] Following other researchers (e.g. Frykblom, 1997), the non-parametric Mann-Whitney U test was also used in the analysis to confirm the results of the t-test since some of the datasets only demonstrated weak normality characteristics.

Table 2. Distribution of response type to the CV surveys

Treatment used	Number of returned usable questionnaires	Positive bids (%)	Zero bids (%)	Protest bids (%)
Bottom-up[a]	173 out of 450 (38·4%)	24·3	67·0	8·6
Allocation using sub-samples[b]	582 out of 1100 (52·9%)	34·0	48·9	17·0
Disaggregative allocation[b]	176 out of 300 (58·7%)	32·4	49·4	18·2

[a] The response type reported here relates to the WTP of the policy programme.
[b] The response type reported here relates to the WTP of the policy package.

Table 3. The mean WTP (positive and zero bids) for the policy package and policy programmes using the three measurement treatments

Policy component	5% truncated			Untruncated		
	Mean WTP (£)	SE	N	Mean WTP (£)	SE	N
A. 'Bottom-up' treatment						
Path improvement programme; $v(a)^1_{BU}$	3·81	1·10	51	7·82	3·00	54
Path creation programme; $v(b)^1_{BU}$	2·42	0·93	49	4·28	1·58	51
Provision of facilities programme; $v(c)^1_{BU}$	5·00	1·61	52	8·98	3·21	55
Aggregate value of programmes; $v(a+b+c)_{BU}$	11·23	3·64	152	21·08	7·79	160
B. 'Allocation using sub-samples' treatment						
Recreation improvement policy package; $v(P)^1_{AS}$	7·22	0·59	416	11·19	0·95	443
Path improvement programme; $v(a)^2_{AS}$	4·31	0·55	181	6·65	1·02	190
Path creation programme; $v(b)^2_{AS}$	2·67	0·48	117	4·56	1·07	129
Provision of facilities programme; $v(c)^2_{AS}$	3·36	0·57	118	5·82	1·39	124
Aggregate value of policy programmes; $v(a+b+c)_{AS}$	10·34	1·60	416	17·03	3·48	443
C. 'Disaggregative allocation' treatment						
Recreation improvement policy package; $v(P)^1_{DA}$	5·72	0·85	131	7·96	1·17	137
Path improvement programme; $v(a)^2_{DA}$	2·08	0·32	131	3·03	0·48	137
Path creation programme; $v(b)^2_{DA}$	1·77	0·31	131	2·50	0·42	137
Provision of facilities programme; $v(c)^2_{DA}$	1·91	0·33	131	2·43	0·38	137
Aggregate value of policy programmes; $v(a+b+c)_{DA}$	5·72	0·96	131	7·96	1·28	137

this point. This consistency helps to support the credibility of the CV questionnaire.

The WTP values estimated for the individual improvement programmes $v(a)$, $v(b)$ and $v(c)$ were also tested for equality between treatments (Equation (2)). No significant differences were found between the WTP values for the three policy programmes elicited using the BU and AS treatments (Table 4). However, some significant differences were found between the DA and BU estimates (Table 4). The implications of these findings are discussed more fully in the analysis of the Carson and Mitchell tests.

The tests analysing the consistency of policy programmes elicited using the AS and DA treatments indicate that significant differences ($P \leq 0{\cdot}05$) were found in almost all of the comparisons.[9] This finding was not predicted and indicates that even although the elicitation questions were posed at the same position in the elicitation sequence, the two treatments appear to produce inconsistent estimates of WTP for nested policy programmes. The additional fact that no difference was found in the same treatment comparison for the policy package $v(P)$ make this finding of particular interest since it indicates that the elicitation treatment adopted in a CV study had a marked effect on the elicited WTP values for the nested programmes.

The findings of the sub-additivity test (Equation (3)) are reported in Table 5. The null hypothesis that predicted equality was not rejected for the DA treatment test. This finding was expected since $v(P)$ was directly allocated among the three component programmes. The null hypothesis was, however, rejected in the AS test, therefore indicating the possible presence of a sub-additivity effect (Hanemann, 1994). Since $v(P)$ was not established for the BU treatment, a direct comparison could not be made. However, a comparison of the aggregate of $v(a)_{BU}+v(b)_{BU}+v(c)_{BU}$ with both $v(P)_{AS}$ and $v(P)_{DA}$ rejected the equality null hypothesis. Again, this suggests the presence of sub-additivity effects. Therefore, it would appear that elicitation treatments that rely on the aggregation of WTP values of policy programmes for the estimation of the WTP of the composite policy package are likely to be affected by sub-additivity effects. Once again these findings support Brown

[9] The only exception to this was in the t-test comparison for the path improvement programme $v(b)$.

Table 4. Tests of consistency of mean WTP estimates for improvement programmes using the different measurement treatments. prob.=probability

Programme	'Bottom-up' treatment (Mean WTP)	'Allocation using sub-samples' treatment (Mean WTP)	t-test[a] t-value (prob.)	Mann-Whitney U test Z-value (prob.)
$v(a)_{BU}=v(a)_{AS}$	3·81	4·31	−0·42 (0·675)	−0·343 (0·732)
$v(b)_{BU}=v(b)_{AS}$	2·42	2·67	−0·26 (0·793)	−0·263 (0·279)
$v(c)_{BU}=v(c)_{AS}$	5·00	3·36	−1·19 (0·234)	−1·769 (0·077)
Programme	'Bottom-up' treatment (Mean WTP)	'Disaggregative allocation' treatment (Mean WTP)	t-test[a] t-value (prob.)	Mann-Whitney U test Z-value (prob.)
$v(a)_{BU}=v(a)_{DA}$	3·81	2·08	0·58 (0·562)	−2·770 (0·006)
$v(b)_{BU}=v(b)_{DA}$	2·42	1·77	0·85 (0·397)	−3·401 (0·001)
$v(c)_{BU}=v(c)_{DA}$	5·00	1·91	2·71 (0·007)	−4·183 (0·000)
Programme	'Allocation using sub-samples' treatment (Mean WTP)	'Disaggregative allocation' treatment (Mean WTP)	t-test[a] t-value (prob.)	Mann-Whitney U test Z-value (prob.)
$v(a)_{AS}=v(a)_{DA}$	4·31	2·08	2·51 (0·012)	−4·834 (0·000)
$v(b)_{AS}=v(b)_{DA}$	2·67	1·77	1·61 (0·110)	−4·289 (0·000)
$v(c)_{AS}=v(c)_{DA}$	3·36	1·91	2·26 (0·025)	−4·876 (0·000)
$v(P)_{AS}=v(P)_{DA}$	7·22	5·72	1·320 (0·193)	0.448 (0.654)

[a]Two-tailed independent samples t-test.

Table 5. Sub-additivity test

Null hypothesis	$v(P)$	$v(a)+v(b)+v(c)$	t-value	Probability
$v(P)_{AS}=v(a+b+c)_{AS}$	7·22	10·34	−2·76	0·007
$v(P)_{DA}=v(a+b+c)_{DA}$	5·72	5·72	0	1·00
$v(P)_{AS}=v(a+b+c)_{BU}$	7·22	11·23	−2·17	0·031
$v(P)_{DA}=v(a+b+c)_{BU}$	5·72	11·23	−2·86	0·005

and Duffield's (1995) argument that the elicitation treatment adopted in a CV study has a significant influence on the resultant WTP values.

In addition to examining the absolute WTP values, the relative allocation of WTP bids between the three policy programmes was investigated (Equation (4)). This test indicated that the elicitation treatment adopted did not significantly ($P>0·05$) affect the relative allocation of WTP among the three programmes (Table 6). This finding was of particular interest since significant differences in absolute WTP for the policy programmes were found between the DA treatment and that from the BU and AS treatments (Table 4). A possible explanation for this is that both the BU and AS treatments are affected by sub-additivity

Table 6. Allocation of WTP value of the policy package allocated to the improvement programmes

Improvement programme	BU treatment (%)	AS treatment (%)	DA treatment (%)
$v(a)/v(a+b+c)$	33·9	41·7	36·3
$v(b)/v(a+b+c)$	21·5	25·8	30·9
$v(c)/v(a+b+c)$	44·5	32·5	33·4
	100	100	100
Friedman test:	$\chi^2=4.667$	$P=0·097$	

effects that lead to higher absolute WTP values for the policy programmes than the DA treatment which is not affected by such effects. It should, however, be noted that even although the absolute values may differ, the fact that the relative allocation of WTP is similar among all treatments is significant

in that it suggests that the three treatments are capable of producing comparable WTP results.

Embedding, sequencing and component sensitivity tests

To examine further the cause of the inequalities identified above, the embedding, sequencing and component sensitivity tests proposed by Carson and Mitchell (1995) were used to analysis the survey data. The embedding effect test (Equation (5)) examines the influence that the placement of a WTP question within the elicitation sequence has on the resultant WTP values. The null hypothesis, which predicts no significant difference between WTP values, was not rejected ($P > 0{\cdot}05$) in all of the embedding effects tests that examined the WTP values in the BU (first in the sequence) and AS (second in the sequence) treatments (Table 7). These findings infer that embedding bias does not significantly affect valuations obtained using the AS treatment. It also infers that respondents were able to recognise the scope of the individual programmes both with and without reference to the overall policy package. The null hypothesis was, however, rejected by the non-parametric tests in the comparison of the DA and BU treatments.[10] This suggests that embedding-like effects may be induced when the DA treatment is adopted. However, rather than being caused by embedding bias, it is more likely that lower WTP values elicited using the DA treatment were a direct result of the budget constraints associated with the DA treatment which do not allow sub-additivity or substitution to take place.

The equality null hypothesis (Equation (6)) was rejected in all of the nested sequence tests undertaken (Table 7). This finding provides evidence that the respondents of both AS and DA treatments were able to distinguish between the policy package $v(P)$ and the nested programmes $v(a)$, $v(b)$ and

[10] The null hypothesis was also rejected for $v(c)$ using the t-test, although no difference was found for the other two programmes.

Table 7. Embedding, sequencing and component sensitivity tests

Embedding effect test[a]					
$H_0: v(a)^2_{AS} = v(a)^1_{BU}$	$t = 0{\cdot}42$ ($P = 0{\cdot}675$)	$Z = -0{\cdot}343$ ($P = 0{\cdot}732$)	$H_0: v(a)^2_{DA} = v(a)^1_{BU}$	$t = -0{\cdot}58$ ($P = 0{\cdot}562$)	$Z = -2{\cdot}770$ ($P = 0{\cdot}006$)
$H_0: v(b)^2_{AS} = v(b)^1_{BU}$	$t = 0{\cdot}26$ ($P = 0{\cdot}793$)	$Z = -0{\cdot}263$ ($P = 0{\cdot}279$)	$H_0: v(b)^2_{DA} = v(b)^1_{BU}$	$t = -0{\cdot}85$ ($P = 0{\cdot}397$)	$Z = -3{\cdot}401$ ($P = 0{\cdot}001$)
$H_0: v(c)^2_{AS} = v(c)^1_{BU}$	$t = -1{\cdot}19$ ($P = 0{\cdot}234$)	$Z = -1{\cdot}769$ ($P = 0{\cdot}077$)	$H_0: v(c)^2_{DA} = v(c)^1_{BU}$	$t = -2{\cdot}71$ ($P = 0{\cdot}007$)	$Z = -4{\cdot}183$ ($P = 0{\cdot}000$)
Nested sequence test[b]					
$H_0: v(P)^1_{AS} = v(a)^2_{AS}$	$t = 4{\cdot}97$ ($P = 0{\cdot}000$)	$Z = -4{\cdot}763$ ($P = 0{\cdot}000$)	$H_0: v(P)^1_{DA} = v(a)^2_{DA}$	$t = 6{\cdot}27$ ($P = 0{\cdot}000$)	$Z = -3{\cdot}065$ ($P = 0{\cdot}002$)
$H_0: v(P)^1_{AS} = v(b)^2_{AS}$	$t = 2{\cdot}87$ ($P = 0{\cdot}005$)	$Z = -4{\cdot}526$ ($P = 0{\cdot}000$)	$H_0: v(P)^1_{DA} = v(b)^2_{DA}$	$t = 6{\cdot}53$ ($P = 0{\cdot}000$)	$Z = -3{\cdot}624$ ($P = 0{\cdot}000$)
$H_0: v(P)^1_{AS} = v(c)^2_{AS}$	$t = 4{\cdot}142$ ($P = 0{\cdot}000$)	$Z = -4{\cdot}274$ ($P = 0{\cdot}000$)	$H_0: v(P)^1_{DA} = v(c)^2_{DA}$	$t = 6{\cdot}67$ ($P = 0{\cdot}000$)	$Z = -3{\cdot}827$ ($P = 0{\cdot}000$)
Test of component insensitivity[c]					
$H_0: v(P)^1_{AS} = v(a)^1_{BU}$	$t = 1{\cdot}98$ ($P = 0{\cdot}048$)	$Z = -2{\cdot}163$ ($P = 0{\cdot}031$)	$H_0: v(P)^1_{DA} = v(a)^1_{BU}$	$t = 1{\cdot}25$ ($P = 0{\cdot}213$)	$Z = -0{\cdot}885$ ($P = 0{\cdot}376$)
$H_0: v(P)^1_{AS} = v(b)^1_{BU}$	$t = 2{\cdot}76$ ($P = 0{\cdot}006$)	$Z = -0{\cdot}862$ ($P = 0{\cdot}389$)	$H_0: v(P)^1_{DA} = v(b)^1_{BU}$	$t = 2{\cdot}19$ ($P = 0{\cdot}029$)	$Z = -0{\cdot}1389$ ($P = 0{\cdot}903$)
$H_0: v(P)^1_{AS} = v(c)^1_{BU}$	$t = 1{\cdot}26$ ($P = 0{\cdot}207$)	$Z = -0{\cdot}078$ ($P = 0{\cdot}938$)	$H_0: v(P)^1_{DA} = v(c)^1_{BU}$	$t = 0{\cdot}43$ ($P = 0{\cdot}671$)	$Z = -0{\cdot}926$ ($P = 0{\cdot}354$)

[a] The statistical tests used in the Embedding effect test were a parametric two-tailed independent samples t-test and a non-parametric Mann-Whitney U test.
[b] The statistical tests used in the Nested sequence test were a parametric one-tailed paired t-test and a non-parametric Wilcoxon Signed Rank test.
[c] The statistical tests used in the test of component insensitivity were a parametric one-tailed independent samples t-test and a non-parametric Mann-Whitney U test.

$v(c)$. This finding supports the validity of both treatments since they confirm that respondents were internally consistent in their responses to the nested sequence of WTP questions (Carson and Mitchell, 1995).

Carson and Mitchell's test of component sensitivity (Equation (7)) examines the inequality of WTP for the policy package and its component programmes when both elements are valued in the same position within the elicitation sequence. The null hypotheses (component insensitivity) was only convincingly rejected in the $v(P)_{AS} = v(a)_{BU}$ comparison (Table 7). Thus, five of the six treatment comparisons demonstrated insensitivity to scope effects. Although, Kahneman and Knetsch (1992) argue that scope insensitivity invalidates the findings of a CV study, Carson and Mitchell (1995) take the opposite view and argue that such effects may alternatively be the result of weaknesses in the design of the CV instrument. It is unfortunate that the exact cause of scope insensitivity could not be determined from the data collected. However, the lack of statistical power resulting from the small sample size used in the BU treatment would have contributed towards the insensitivity (Carson and Mitchell, 1995).

Discussion

The data from the empirical experiment have indicated a number of inconsistencies between the WTP estimates obtained using the different elicitation treatments. These inconsistencies infer that at least one of the elicitation treatments results in unreliable WTP estimates. We now re-examine the empirical findings to assess which elicitation method provides the most robust estimation of WTP for both the policy package and its component programmes.

The validity of policy package evaluations

The value of the policy package $v(P)$ was estimated both directly (AS and DA treatments) and indirectly through the aggregation of the component programme values (BU and AS treatments). Intuition would suggest that direct valuation approaches are more likely to provide valid estimates of WTP than indirect approaches since direct approaches specifically address the composite package. In our experiment, equality was found between the direct estimates of $v(P)$ established using the AS and DA treatments. Hoehn states that a valid valuation design is one where the '*valuation of a policy change is unique*' (Hoehn, 1991:298). It is therefore argued that these two direct approaches appear to produce valid WTP estimations of $v(P)$.

The indirect estimates of the policy package, established in the BU and AS treatments by aggregating the value of component programmes, were found to be significantly higher than the direct valuation of $v(P)$ (Table 5). Although the indirect valuation approaches appear to produce over-estimations of the actual WTP for the policy package $v(P)$, Hanemann (1994) argues that this over-estimation is expected and is due to sub-additivity effects. Hoehn also argues that this over-estimation is consistent with economic theory if the goods valued are substitutes. It is clear that the programmes investigated in this study, i.e. path repair, path creation and provision of countryside facilities, are likely substitutes of each other. Therefore it is argued that although the aggregation approaches appear to overestimate the assumed 'true' value of $v(P)$, these overestimations do not provide legitimate grounds to invalidate these elicitation treatments. Based on these empirical results, it is concluded that the direct valuation approaches (AS and DA treatments) produce most reliable estimates of WTP for the composite policy package, while the indirect aggregation approaches appear to over-estimate WTP.

The validity of policy programme evaluations

Estimates of WTP for individual policy programmes were established using all three elicitation treatments investigated. Of these treatments, it was only in the BU treatment that respondents' attention was focused entirely on a single policy programme; respondents of the other two treatments were also asked to evaluate the composite policy package. Since the WTP bids made

using the BU treatment were not influenced by information on other programmes, it is argued that the BU approach is likely to provide the most reliable estimates of WTP for the individual policy programmes. Based on this assumption, the validity of WTP estimates measured using the other two treatments was tested through comparisons with those elicited using the BU treatment. These tests (Table 4) demonstrate consistency between the WTP estimates for the policy programmes elicited using the AS and BU treatments, inferring that the AS treatment can also produce valid and reliable estimations of WTP for the policy programmes. This conclusion is supported by Loomis *et al*. (1993) who similarly found consistency in WTP for the preservation of 6000 ha of forest in SE Australia when evaluated by itself (i.e. BU approach) or as a proportion of a larger area of forest (AS approach).

Significant differences were found between the value of the policy programmes estimated using the DA and BU treatments (Table 4). This inequality indicates the existence of embedding effects (Equation (5)) which, according to Kahneman and Knetsch (1992), invalidates the WTP estimates obtained using the DA treatments. Although controversy exists regarding the implications of embedding on the validity of a CV instrument, it is clear from this experiment that the WTP estimates for the policy programme established using the DA treatment were not consistent with the assumed true values established using the BU approach. It is therefore concluded that the DA treatment provides a less reliable method of establishing the WTP for nested policy programmes than the AS approach.

Conclusions

The purpose of this paper was to assess the validity of three alternative CV elicitation treatments for the evaluation of multi-dimensional policies. It is concluded above that only the AS treatment provided valid and reliable estimates of both the policy package and its component programmes. However, it was also noted that the AS treatment produced WTP estimates that were internally inconsistent in that the direct value of the policy package $v(P)$ was found to be significantly lower than the aggregate value of the three programmes $v(a)+v(b)+v(c)$. It is clear that this inconsistency needs further discussion before the AS elicitation treatment can be recommended for use.

Hanemann (1994) argues that the internal inconsistencies found in the AS treatment are the result of sub-additivity effects. He further argues that sub-additivity effects are to be expected and can be explained in terms of substitution. Brown and Duffield (1995), in their examination of the relationships among multi-dimensional goods, conclude that the precise relationship found in the AS results reflect the situation where the component programmes are 'imperfect substitutes'. They state that imperfect substitution occurs in situations where, within an overall budget constraint, survey respondents '*allocate enough to cover the separate value of* [the programme under investigation]' (Brown and Duffield, 1995:2343). In other words, the AS treatment allows respondents to incorporate substitution effects into their valuation of policy programmes. The incorporation of substitution effects is clearly desirable in evaluation studies since it ensures that the resultant valuations reflect the total impact of an individual programme. It is therefore concluded that the internal inconsistencies associated with the AS treatment are not a cause of concern, although it should be recognised that an appreciation of what causes these inconsistencies is desirable for the interpretation of CV studies.

The other two elicitation treatments were found to be less reliable. Although the BU treatment produces reliable estimates of WTP for individual policy programmes, the aggregation procedure required by this technique resulted in over-estimations of the value of the policy package. The DA treatment, on the other hand, was shown to provide valid estimations of WTP for the policy package, however, the restrictive nature of the disaggregation procedure resulted in underestimations of WTP for the component programmes. In its defence, the relative allocations of WTP bids between the component programmes were found to be consistent with the other techniques. This, and the fact that the DA approach only requires a single sample to be used, make the DA approach a useful and relatively inexpensive method for the elicitation of benefit information at a level

of reliability adequate for comparative purposes. However further empirical evidence is required to ensure that the relative consistencies found in this experiment is a repeatable finding before this treatment can be recommended.

This paper also attempted to clarify, and examine further, the influence of embedding and related effects on the validity of the CV technique. In the comparison between the BU and AS treatment, the survey results demonstrate that the position of the WTP question in an elicitation sequence did not affect WTP responses for the policy programmes. According to Kahneman and Knetsch (1992) the equality found here indicates that embedding bias did not affect WTP responses and therefore provides evidence supporting the validity of both treatments for the evaluation of policy programmes. Survey responses were also examined to determine whether respondents were sensitive to the scope of the policies investigated. The internal nested sequence tests demonstrated that survey respondents were able to distinguish between the policy package and its component programmes; however, the external test of component sensitivity failed to demonstrate such sensitivity. Rather than inferring that the CV technique is insensitive to scope effects *per se*, it was hypothesised that the lack of sensitivity to scope found in the external test may have been induced as a result of having an inadequate size of sample in the BU treatment. Although it is argued that the evidence from this study does not support Kahneman and Knetsch's (1992) criticisms of the CV technique, the result can neither be used to dismiss their concerns. It is clear that more research needs to be undertaken to improve understanding of this issue. Perhaps what this paper has nevertheless demonstrated is that the tests proposed by Carson and Mitchell (1995) allow clearer definition of embedding issues, which in turn allows a systematic assessment of the influence of these issues on the validity of CV studies.

Finally, the research reported here has investigated three alternative elicitation treatments that can be used to value multidimensional policies. Clearly, there are a number of other potential approaches for this type of evaluation. For example, it may be possible to adopt an alternative bottom-up approach that aims to elicit the value of component programmes before establishing the value of the policy package. With such an approach, it would be interesting to compare the relationships found among the policy programmes and the composite packages, with those found in the experiment reported in this paper. The examination of these alternative approaches may help to further clarify the confusion associated with embedding and related effects, as well as potentially providing another valid elicitation treatment that enables CV researchers to successfully and efficiently assess the economic value of multidimensional policy.

References

Arrow, K., Solow, R., Portney, P. R., Learner, E. E., Radner, R. and Schuman, H. (1993). Report of the NOAA panel on contingent valuation. *Federal Register* **58**, 4601–4614.

Boyle, K. J., Reilings, S. D. and Philips, M. L. (1990). Species substitution and question sequencing in contingent valuation surveys evaluating the hunting of several types of wildlife. *Leisure Studies* **12**, 103–118.

Brown, T. C. and Duffield, J. W. (1995). Testing part-whole valuation effects in contingent valuation of instream flow protection. *Water Resources Research* **31**, 2341–2351.

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

Christie, M. (1998). *An Analysis of the Provision of Recreation Improvements in the Grampian Countryside*. PhD Thesis. University of Aberdeen.

Christie, M. (1999). An assessment of the economic effectiveness of recreation policy using contingent valuation. *Journal of Environmental Planning and Management* **42**, 547–564.

Desvousges, W. H., Johnson, F. R., Dunford, R. W., Boyle, K. J., Hudson, S. P. and Wilson, K. N. (1993). Measuring natural resource damages with contingent valuation: tests of validity and reliability. In *Contingent Valuation: A Critical assessment* (J. A. Hausman, ed.), pp. 91–164. Amsterdam: North-Holland.

Diamond, P. A., Hausman, J. A., Leonard, G. K. and Denning, M. A. (1993). Does contingent valuation measure preferences?: experimental evidence. In *Contingent Valuation: A Critical Assessment* (J. A. Hausman, ed.), pp. 14–89. Amsterdam: North-Holland.

Dillman, D. A. (1978). *Mail and Telephone Surveys: The Total Design Method*. Chichester: Wiley-Interscience Publications.

Frykblom, P. (1997). Hypothetical question modes and real willingness to pay. *Journal of Environmental Economics and Management* **34**, 275–287.

Hanemann, W. M. (1994). Valuing the environment through contingent valuation. *Journal of Economic Perspectives* **8**, 19–43.

Hausman, J. A. and Diamond, P. A. (1994). Contingent valuation: is some number better than no number? *Journal of Economic Perspectives* **8**, 45–64.

Hoehn, J. P. (1991). Valuing the multidimensional impacts of environmental policy: theory and methods. *American Journal of Agricultural Economics* **73**, 289–299.

Hoehn, J. P. and Loomis, J. B. (1993). Substitution effects in the valuation of multiple environmental programs. *Journal of Environmental Economics and Management* **25**, 56–75.

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

Loomis, J. Lockwood, M. and DeLacy, T. (1993). Some Empirical evidence on embedding effects in contingent valuation of forest protection. *Journal of Environmental Economics and Management* **23**, 45–55.

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