# ON PRECISION AND PROMINENCE IN WTP RESPONSES

David K Whynes
University of Nottingham

Emma J Frew
University of Birmingham

Zoë N Philips
AstraZeneca

Judith Covey
University of Durham

Richard D Smith
University of East Anglia

*Paper presented at the Health Economists Study Group, University of Birmingham, 3-5 January 2007. Paper not for quotation or citation without authors' permission.*

**Correspondence to:**
Richard Smith
Reader in Health Economics
Health Economics Group
School of Medicine, Health Policy & Practice
University of East Anglia
Norwich
NR4 7TJ
Tel: +44 (0)1603 593617
Fax: +44 (0)1603 593604
Em: Richard.smith@uea.ac.uk

**ABSTRACT**

The presence of 'prominent numbers' reflects imprecision in willingness-to-pay (WTP) responses. This might arise because subjects find precision unnecessary, the valuation task intrinsically difficult, or that prominent numerical 'cues' are given in the study. This paper reviews work by the authors concerning evidence for the presence of prominent numbers within contingent valuation (CV) studies. Two papers (Whynes *et al*, 2005, Covey & Smith, 2006) present mixed evidence, but were based on secondary analysis of CV studies, and hence an indirect assessment of the presence of prominent numbers and/or reasons for imprecision. A third paper (Whynes *et al*, 2007), outlined in more detail here, undertook an experiment specifically to test a number of hypotheses associated with prominent numbers and the precision of WTP values. The evidence from this suggests that when people give their WTP values they do tend to use prominent numbers – the implication being that their responses are imprecise and CV practitioners should therefore be concerned that responses may be too crude to be of use. This result seems robust and even unaffected by the presence of non-prominent cues.

The evidence on prominence that we have collected is however limited to WTP surveys in health care contexts where goods tend to be quite low in value (up to only a couple of hundred pounds at most). We are therefore left with the possibility that the prominence effect is actually reflecting a lack of cognitive effort on the participants' part – when people are giving WTP amounts that represent a small proportion of their income the implications of being imprecise are not so great, so why bother striving for a precise value of £18 when £15 or £20 is near enough? The problem is that when these values are aggregated for policy purposes the imprecision is magnified. It might therefore be that prominence will be less of a problem for WTP surveys which elicit large WTP values where the implications to individuals of imprecision are much greater (because the WTP is a much bigger proportion of their income). Indeed there is related evidence that the size of the WTP response required might explain the difference in scope sensitivity effects demonstrated between health and environmental literature.

Results like these suggest that we can perhaps only be confident of WTP values where we expect those values to comprise a significant proportion of respondent income. Clearly, however, further research is required to test whether this conjecture is robust. This paper therefore reflects on the results to date, but also provides an embryonic research agenda for future work to test more directly how the magnitude of the WTP response required relates to the amount of cognitive effort that people put into their WTP responses – and whether this increased cognitive effort goes on to result in more precise and reliable WTP responses that are less affected by numerical cues. The intention is to use this paper to provoke some discussion by HESG members concerning the issues of prominence and precision in WTP estimation.

## 1. INTRODUCTION

The contingent valuation (CV) method of valuing (the benefits of) non-market goods has become increasingly used in health economics in recent years (Smith, 2000; Olsen & Smith, 2001; Smith, 2003). Briefly, the method presents individuals with a hypothetical scenario where they are asked to reveal their maximum willingness to pay (WTP) for a good or service. This is then taken as the consumer surplus, or value, of that good relative to other goods. Individual values are aggregated to calculate the population value.

An important element of the specification of the contingent market is the payment elicitation technique. Initially, CV was conducted using an 'open-ended' format, where respondents were simply asked to state their maximum WTP with no prompting or framing. Issues with the validity of this technique led to developments in 'bidding' (where respondents accept or reject an initial amount and are then bid up or down in increments until their maximum WTP is achieved) (Bennett & Tranter, 1998). Following the influential National Oceanic and Atmospheric Administration (1993) report, the 'discrete' format (also referred to as closed-ended, dichotomous-choice, referendum or binary format) became widespread, where respondents are presented with a single WTP value (or bid), which they either accept or reject (a form of bidding truncated at one bid). By varying the bid across sub-samples it is possible to calculate the aggregate demand curve for the commodity valued, with the area under the curve representing societal WTP. However, dichotomous-choice surveys are associated with yea-saying bias and highly statistically inefficient, as a very large number of observations is required to identify a distribution of values with any degree of accuracy (Smith, 2000; Yeung *et al*, 2006). An alternative, and the most popular in health economics over the last 20 years, is the 'payment scale' which presents respondents with an ordered list of numbers representing money values and asks respondents to indicate which value on this scale they would pay (and sometimes which they would not) (Smith, 2006a).

## 1.1. Prominence and precision

Recent analysis of previously-collected CV data using payment scale, iterative bidding and open-ended techniques led the authors to conjecture that particular numerical forms – termed prominent numbers – might be significantly over-represented amongst WTP valuations elicited from general populations (Covey & Smith, 2006; Whynes *et al*, 2005).

Prominent numbers are those which belong to the sequence $N \times 10^X$, where $N$ takes the value 1, 2 or 5 and $X$ is an integer, 0, 1, 2, and so on. This produces the series: 1, 2 and 5, 10, 20 and 50, 100, 200 and 500, 1000 and so forth. According to prominence theory, the prominent number sequence is the ‘‘least-exact’’ or ‘‘coarsest-grained’’ series of the counting numbers. As with the sequence of round numbers, it is decimal-based, although prominent numbers are also defined by ratios. Each number in the series is approximately twice the preceding one, halving and doubling being fundamental to the way in which humans make quantitative comparisons (Albers, 1998).

When individuals are asked to make numerical estimates, the precision of their responses reflects their perceptions of the difficulty in, and the consequences of, making the assessments (Hertwig *et al*, 1999). Individuals are more likely to provide vague, inexact or imprecise numerical estimates when they feel their knowledge of the subject matter is sufficiently poor so as to preclude exactness without disproportionate effort and/or when precision in response is deemed unimportant. In contrast, exact responses are more probable when individuals feel well informed and/or when they believe that imprecision in response could prove costly to themselves. Providing a less coarse-grained estimate than that offered by the prominent number sequence would entail recourse to the series of ‘‘spontaneous numbers’’, which are approximate half-steps between the prominent numbers, i.e., 15, 30 or 40, 70 or 80, 150, 300 or 400, and so on. A finer-grained series would interpolate still further, and so forth. Albers’ relative exactness ratio offers a more formal measure of precision, in terms of the extent to which interpolation occurs (Albers, 2001). Under Albers’ formulation, prominent numbers, such as 100 and 200, always have a ratio of 1, and the ratio falls as we move progressively through intermediate steps towards higher degrees of exactness. The numbers 150, 180 and 196, for example, have exactness ratios of 0.33, 0.11 and 0.01, respectively.

In the previous studies (Covey & Smith, 2006; Whynes *et al*, 2005), the over-representation of WTP responses configured as prominent numbers was attributed to subjects being either willing to accept inexactness (in that they believed the valuation tasks were unimportant) or unable to offer precise estimates (owing to perceived task difficulty). The possibility therefore arises that deliberate changes to the character of a CV task might change the subjects' exactness in response. As the earlier data had been collected for evaluation rather than for methodological purposes, they were not suitable as a basis for testing this possibility scientifically. A bespoke study was necessary, the results of which are formally presented in a forthcoming paper by the authors (Whynes *et al*, 2007). In this paper we briefly summarise the methods and results of this study, using it as a platform to suggest what the implications of this phenomenon might be and propose future areas for research.

## 2. METHOD

### 2.1. Hypotheses tested

The study undertaken was designed to test three specific hypotheses.

***Hypothesis 1: the perceived difficulty of the CV task decreases as the exactness of the numerical cues provided increases.***

It has been suggested that consumers interpret a retail price characterised by a high degree of exactness, using our present terminology, as being more legitimate or authoritative than a less-exact price-number. Consumers tend to conclude that suppliers who select prices from the finer-grained series of numbers must be employing well reasoned and rational price-setting processes (Schindler, 1991). Accordingly, we hypothesise, first, that CV subjects who are offered exact numerical cues will view them as authoritative. They will be more willing to be guided by such cues, and will accordingly find the CV task easier. Second, and as the least-precise cue is that which is not provided at all, subjects provided with numerical cues will find the valuation task easier than those receiving no such cues, by virtue of having been supplied with a menu of alternative number-values which can be interpreted as legitimate possibilities.

***Hypothesis 2: the subject's WTP valuation will be characterised by a higher degree of exactness if s/he has been cued by numbers characterised by higher degrees of exactness.***

It is now well established that preferences and values are not simply revealed by the elicitation tasks, but are intrinsically influenced, even constructed, by them (Gold & List, 2004; Slovic, 1995). The earlier conjecture about the over-representation of prominent numbers in response had been based on the analysis of valuations obtained from payment scales which had themselves contained a sizeable proportion of prominent numbers. Therefore, we cannot exclude the possibility that the high proportion of prominent-number WTP values actually observed resulted from either (i) a purely-coincidental equivalence between most individuals' subjective valuations and prominent numbers, or (ii) the structure of the numerical cues which the subjects had received. With respect to the first possibility, the strength of our conjecture would be proved by attempting to replicate the results using a different sample faced with a task involving different pay-off parameters. If the second possibility were to be the case, any subject confronted with numerical cues of high exactness, and where prominent numbers were accordingly absent, should be significantly less inclined to select a prominent number.

***Hypothesis 3: the subject's WTP valuation is lower, absolutely, if the perceived difficulty of the CV task is higher.***

Evidence suggests that numerical framing and cueing go further than influencing the numerical form of the chosen WTP amount; they influence the magnitude also (Bonini *et al*, 2002). Previous studies have reported that, in comparison with payment scales, the open-ended elicitation format yields lower values (Brown *et al*, 1996; Donaldson *et al*, 1997; Frew *et al*, 2001), although the reason for this remains unclear. One specific conjecture is that, because the open-ended format has been believed to provide more cognitive difficulty than formats employing numerical cues, task difficulty and estimates of WTP amount are associated (Brown *et al*, 1996). The data necessary to test our first two hypotheses will enable us to explore this third hypothesis also.

### 2.2. Data collection

Data were obtained from questionnaires completed by students at the universities of Durham, East Anglia and Nottingham. Each subject received one of three variants of the questionnaire instrument at random. Common to all variants was an initial explanation of the CV method as a means of valuing and prioritizing government services. Thereafter, the specific valuation scenario was presented, which entailed the subject being the victim of a hypothetical road traffic accident, based on a broadly-similar scenario used successfully in the past (Carthy *et al*, 1999). Each subject was informed that the usual treatment for the injuries sustained in the accident would require a 2–3 day stay in hospital whilst enduring slight-to-moderate pain, up to 4 weeks of pain and discomfort after discharge, and 3–4 months of restricted activity. Thereafter, s/he would return to his/her pre-accident health level with no permanent or further disabilities. The subject was then told that an alternative treatment path was available although, unlike the usual therapy funded by the National Health Service or by private health insurance, this treatment would have to be paid for directly. The alternative offered the putative advantage of collapsing the entire treatment and recovery process into just 3–4 days.

Subjects were asked whether they would be "willing to pay something, however large or small, for the alternative treatment" on the presumption that, if they were not, they would be willing to accept the consequences of undergoing the usual treatment. Those indicating they would be willing to pay now moved on to the valuation task proper, which was introduced as follows: "Since you would consider paying something for the alternative treatment, we now want to know the most you would be willing to pay for it. When answering this question, please bear in mind that (i) whatever you pay or borrow to get better more quickly will reduce the amount of money you have available to spend on other things, (ii) we want to know what the treatment is worth to you, not what you think it might cost".

## 2.3. Payment scales

Each subject was instructed: “write your best estimate of the most you would pay for treatment in the box below”, and each proceeded to this task via one of three different elicitation routes. In the first, no prompts or cues were provided and the valuation task was presented immediately (the open-ended (OE) elicitation format). In the second variant, subjects were provided with a payment scale, comprising the following amounts (£) listed vertically: 10, 20, 50, 80, 100, 120, 150, 180, 200, 250, 300, 350, 400, 450, 500, 800, 1000, 1500, 2000, 2500, 3000, 4000, 5000. Subjects were invited to indicate on the scale the amounts they were sure they would be willing to pay (by ticking) and the amounts they were sure they would not (by crossing). Thereafter, each subject entered his or her estimate into the box, as with the OE subjects. This payment scale was constructed under the requirements that all scale points should be round numbers and that all the prominent numbers within the range should appear (more than one-third of the scale points are prominent numbers). This second elicitation variant we term the prominent payment scale (PPS), and it possesses a mean Albers exactness ratio of 0.56.

The third variant also required subjects to consider a payment scale prior to entering their maximum WTP although, in this case, the scale was the following: 9, 18, 45, 73, 94, 116, 148, 177, 196, 244, 291, 347, 392, 441, 503, 819, 1032, 1476, 2013, 2613, 2992, 4011, 4993. This second scale had both differences and similarities in comparison with the first. Specifically, both scales had 23 points, although the second contained no round numbers, no prominent numbers and each of the nine possible trailing digits was represented at least twice (only ‘5’ appears once). To avoid distortions of magnitude, points on the second scale were made as close as possible to those on the first, given the limits of feasible construction from the restricted range of digits. This entailed an unavoidable minimum difference of 10% between the initial points (10 vs. 9), although the discrepancy was smaller for higher points. Because payment scales can inform valuations based on relative as opposed to absolute amounts, we constructed the second scale to mimic the first with respect to scale point ratios, again constrained by the digits available. Thus, the absolute value of the ratio of the (n + 1)th scale point to the nth point for the second scale is, on average, 1.6% of the equivalent ratio for the first, and is never more than 4.7%. By way of illustration, the ratio of points 6 and 7 on the second scale is 148/116 = 1.28. The equivalent ratio for the first is 150/120 = 1.25, implying a difference of 0.03%, or 2.1%. This second scale, and third elicitation variant, we term the non-prominent payment scale (NPS), and it possesses a mean Albers relative exactness ratio of 0.02.

After the valuation task, subjects for all three elicitation variants were asked to indicate how easy or difficult they had found the exercise, by placing a mark on a visual analogue scale. Five numbered and equally-spaced points were provided on a 13 cm. linear scale, from 1 (“Easy”), on the left, to 5 (“Difficult”) on the right. For the purpose of subsequent analysis positions were measured and coded on a continuous decimal scale. Subjects completing either of the two payment scale variants were then asked to indicate (i) whether they felt that the scale provided had offered a “reasonable and realistic” range of values and (ii) if the scale had helped them in reaching their eventual WTP valuation.

Finally, all three variants of the questionnaire requested socio-demographic data, including age, ethnicity, health-status and smoking behaviour. We felt it appropriate to enquire whether subjects had experience of, or familiarity with, road traffic accidents on the part of themselves, a family member or a close friend, given that such experience might plausibly influence their valuation decisions. Although income has been shown to be a major influence on WTP values, we chose not to measure it directly, owing to the expectation that income would exhibit little variability in a student population. Instead, we asked subjects to categorise themselves by selecting one from a list of four statements which described their “situation with regard to money”.

For the Nottingham and Durham samples, copies of the questionnaires were distributed during formal classes and collected at the end of each session, time having been allocated for their completion. For the East Anglia sample, the questionnaire was posted on an intranet and was completed electronically at the respondents’ convenience.

## 3. RESULTS

Of the 468 questionnaires distributed to and completed by subjects, 407 (87.0%) indicated a willingness to pay for the intervention, and these subjects proceeded to the valuation task. Representation of the three questionnaire variants amongst those offering WTP values was approximately equal: 140 responses for OE, 131 for PPS and 136 for NPS. For the sample offering valuations, the mean WTP was £982 (SD 1811), although the distribution of values was significantly skewed (coefficient of skew 4.43, SE 0.12). The median WTP was £400 (IQR 200–1000).

The mean reported degree of difficulty in completing the WTP task, as marked on the visual analogue scale, was 3.4 (SD 1.0, median 3.5), with 17.0% of the sample reporting a difficulty level of 2 or below and 19.9% reporting a level of 4 or above. For the two payment scale samples, the proportions perceiving the scale as offering a reasonable and realistic set of values was 88.5% and 80.1% for PPS and NPS, respectively (v2 = 3.5, p = 0.04). As the ranges available on the payment scale variants were essentially similar, we might tentatively ascribe the difference in perceptions of realism to the arcane nature of the NPS numbers. This difference notwithstanding, the proportions indicating that they had found their particular scale helpful in deciding on their maximum WTP did not vary between scales (68.9% and 62.8% for PPS and NPS, respectively, v2 = 1.1, p = 0.17).

With respect to gender, 52.7% of those offering a value were male and 72.9% described themselves as ethnically “white”. Sizeable ethnic minorities included Chinese (7.6%) and Indian (7.1%). An overwhelming majority (93.1%) declared their health to be excellent or good and most subjects (79.3%) had never smoked. More than half (54.1%) had experience of, or familiarity with, a minor injury from a road traffic accident, whilst a smaller proportion (22.6%) were familiar with a serious injury. The majority of the sample (66.0%) categorised their financial situation as: “I have enough money, so long as I plan my spending carefully”, although 13.4% responded: “I normally have enough money for anything I want”. For the purpose of analysis, we classified the latter group as “affluent”. The remaining respondents indicated either that they had money sufficient only for “basic things” (17.9%) or that even basic things were hard for them to afford (2.7%). These last two groups we combined and classified as “impecunious”.

### 3.1. Hypothesis 1

To investigate our first hypothesis, we fitted regression models with the reported degree of difficulty as the dependent variable. All other variables noted above were entered into a backward stepwise procedure, some re-coded as dummy variables where appropriate. Two models were necessary, as data associated with the usefulness of the payment scales was not collected for the OE variant. The final models appear in Table 1. Evidently, females found the task more difficult than did males, whereas the affluent found it less difficult. Not surprisingly, those finding the scale useful in deciding on their valuation found the task easier. Having hypothesised that the difficulty of the CV task decreases with the precision of the numerical cues provide to subjects, we would predict that, amongst the three variants, reported difficulty would be highest for subjects under the OE format (where cues are non-existent and therefore axiomatically imprecise) and lowest for those under NPS. The coefficients associated with the dummy variables for the elicitation variants are significant and consistent with hypothesis H1. Evaluated at the variable means, the mean degrees of difficulty estimated from Table 1 models are 3.9, 3.4 and 3.1 for the OE, PPS and NPS samples, respectively.

### 3.2. Hypothesis 2

With respect to hypothesis 2, the relevant results appear in Tables 2 and 3. Table 2 lists, for each questionnaire variant, the 12 most commonly chosen WTP values. Eight of these 12 values are common to all three lists – 100, 200, 250, 300, 400, 500, 1000, 1500 – and three values appear in two lists – 150, 2000, 5000. All values, excepting 250, are either prominent or spontaneous numbers. Six values – 20, 50, 120, 503, 750, 2500 – appear in one list only. For the PPS list, all of the ‘top 12’ WTP values were provided as numerical cues, i.e., as scale values available for selection. For NPS, however, only 503 was a scale value.

Table 3 results indicate that the proportion of final values which were prominent numbers was highest for the OE sample, whilst the proportions for the PPS and NPS samples were not significantly different from one another. The mean Albers exactness ratios tell a similar story, with the ratio for the OE sample being significantly higher than that for either of the payment scales. At first sight, therefore, it appears that hypothesis 2 cannot be confirmed, but we note that the exactness ratio distributions, like the WTP values, are themselves skewed. In contrast to the mean ratios, the median exactness ratios presented in Table 3 do indeed suggest that subjects who were offered the more exact cues did provide, on average, the more exact responses. This result must be considered in the light of an additional response characteristic, namely, the propensity of a proportion of the payment scale samples to translate numerical cues directly into maximum WTP. The OE and the PPS samples each used a total of 32 different value-numbers to express their WTPs. Of these, 5 were unique to OE and 11 were unique to PPS. The NPS sample, however, made use of 49 different numbers, of which 23 were unrepresented in both the OE and PPS samples. Of these, 17 values were offered by a single subject only in each case, and none of the remaining 6 were offered by more than 3 subjects in each case. Twelve of the 23 were scale points specific to NPS which possessed, by design, very low exactness ratios. The lower median for the NPS Albers ratio, therefore, should be attributed to the existence of a far longer tail of more-exact responses in the distribution of values.

Table 3 shows that PPS subjects were more likely to select a scale point as the final WTP, although the relationship between selected scale points and prominence was not straightforward. For 34.1% of the PPS sample, the highest scale point ticked and the final WTP value offered were both prominent numbers. The proportion for which neither scale point nor WTP value was a prominent number was 38.1%. For 18.9%, the highest point ticked was prominent but the final value was not whilst, for the remaining 8.3%, the reverse was the case (v2 = 29.2, $p < 0.01$). Non-roundness in WTP numbers selected was significantly more prevalent amongst NPS subjects, although most of these cases were accounted for by individuals using (non-round) scale points as final valuations.

### 3.3. Hypothesis 3

Table 4 displays the regression models which, following a backward stepwise procedure, best explain the WTP values finally selected by the subjects. The independent variables initially entered were all the socio-demographic variables, as before, and the data pertaining to task difficulty and completion. The significant negative coefficient associated with finding the range of the payment scale reasonable implies that those finding it reasonable offered lower maximum WTP values. It is therefore possible that those finding the scale unreasonable did so because of its omission of sufficiently high values from which to choose. The coefficient for degree of difficulty is significant in both regressions and its sign is consistent with hypothesis 3, namely that those finding the task more difficult tend to offer lower WTP values. Calculating from the equations at the variables' mean values, Table 4 model predicts a mean WTP of £1024 for the degree of difficulty experienced by the PPS subjects. The average WTP for subjects with the NPS degree of difficulty is 7.3% higher, whilst, in the OE case, it is 22.0% lower.

## 4. DISCUSSION

### 4.1. Summary of findings

This study, together with our previous analyses, seems to support the view that prominence in response remains prevalent. The majority of subjects appeared to be either unwilling or unable to respond to valuation questions with high degrees of exactness, even when supplied with particularly exact cues. For instance, the number 250 figures amongst the values in Table 2 for each of the three elicitation variants, and it is the one value amongst these numbers which is neither spontaneous nor prominent. Albers argues that the use of the $25x10^X$ series is a statement about precision in decision which, if anything, implies even more inexactness than does the use of the prominent number sequence. The $25x10^X$ series is used by individuals who stop their decision processes at the crude level of exactness characterised by indecision (Albers, 2001). Thus, the number 250 will be used by someone who believes that the proper value lies somewhere between zero and 500, but is unable able to specify where. The presence of 250 as a common value in our samples thus strengthens our belief that inexact valuations are being offered.

An interesting finding was that the new results are similar to those observed in other accident CV studies (Covey & Smith, 2006) rather than those obtained from an earlier CV study of cancer screening (Whynes *et al*, 2005) as the new payment scale evidence reveals both a far lower dominance of specific prominent numbers and a flatter distribution across both prominent and spontaneous numbers. One possible explanation is a magnitude effect. The payment scale values used in the cancer screening studies were an order of magnitude lower than those for the accident scenario. It seems plausible that subjects might view the payment scale's numerical cues not only in relative terms but in absolute ones also. For example, whilst a scale move from £10 to £20 and one from £100 to £200 each represents a doubling of valuation, the latter clearly signals the net addition of 10 times the amount (£100 vs. £10). As valuation magnitude increases in a sequence of given graininess, so too does the absolute size of each step in the sequence, and individuals sensitive to absolute magnitude effects may accordingly be inclined to move to numbers drawn from a finer-grained series. That is, the larger the WTP the more likely it is to be 'precise', supporting evidence elsewhere reported (Yeung *et al*, 2003; Smith, 2005). This explanation would account for why the strong concentration of particular numerical values was less apparent in the PPS sample, relative to the OE sample. The absence of numerical cues may conceal the absolute effect from those using OE elicitation, thus giving rise to similar distributions dictated by relativities alone.

### 4.2. Some implications

Further experimentation is clearly required to establish the extent to which the conjectured prominent number effect is robust. Nevertheless, if the evidence continues to support the existence of the phenomenon then we must consider the implications for the conduct of CV surveys and the use of their findings in cost-benefit analysis. It seems to us that the high frequency of prominent number responses in this and earlier studies implies that valuation estimates derived from CV tasks must be less-than-precise, due to either subjects' difficulties in estimating values or their lack of concern over exactness. Given this conclusion, a range of implications may follow. For example:

1. Is there now an argument for rejecting point estimates of value and accepting, even eliciting explicitly, WTP values as ranges? It is, after all, an established scientific principle that analytical results should not be presented at levels of precision higher than those set by the constituent data.
2. Does the pursuit of precision require that future CV value elicitation take place iteratively? Beginning with the elicitation of the coarse-grained response, subjects would be taken along an 'ever-fine-graining' exercise, until they arrive at as low an Albers exactness ratio as is possible.
3. To what extent does the prominence effect damage the process of estimating perceived benefit, in the form of consumer surplus, from the activities being evaluated, given that the shadow demand curve derived from WTP values is as precise only as the constituent data?

### 4.3. Future research possibilities

One specific implication of this research concerns the assessment of the construct validity of WTP estimates. One common test for construct validity is to assess the sensitivity of WTP to the size or scope of the good being valued. In this respect it has been suggested that studies report far better results within the environmental and transport literature than health care literature (Smith, 2005). One hypothesis for this concerns the type of good being valued. It is common practise within transport and environmental economics for individuals to be asked to value public goods that will have a significant and widespread impact, such as a large pollution disaster or the development of a motorway. The context is different within health care as normally what is being valued is a private good which is not regarded as being as significant or having such a widespread impact. It is inevitable therefore that when asked to consider a good that has such large consequences (as in environmental/transport economics) that individuals will devote more cognitive effort to revealing a WTP value as the consequences of getting it wrong are greater.

From our previous work on WTP and sensitivity and the prominence phenomenon outlined in this paper, we suggest that when individuals are asked to consider interventions that are potentially large in terms of cost and impact they will devote more cognitive effort to revealing a WTP value which will in turn lead to both a more precise and a more reliable value (Smith, 2004; Smith, 2005; Smith, 2006b). If this hypothesis is substantiated, then it may suggest a revision to the manner with which WTP values are elicited, reported and used. For instance, there may have to be exploration of a CV equivalent of the 'experienced utility' concept that Kahneman *et al* are moving towards, or the expression of WTP as a range with an associated confidence distribution for use in analysis, or a focus on the 'marginal' approach advocated by Donaldson *et al*.

We therefore propose to undertake further quantitative and qualitative investigation of how individuals use numbers to reflect value. More specifically, to assess and explain the implications of the degree of cognitive effort devoted to answering WTP questions and to identify approaches to strengthen the CV design taking this into consideration. Under this broad aim, we propose a basic experimental design to compare WTP values that participants give to two different scenarios – whether the WTP either represents a small or large proportion of their income. We could then examine how long it takes participants to give their WTP response as a measure of the amount of cognitive effort that they put into it (time = effort). The hypothesis being that respondents will take longer to give a WTP response when the value represents a large proportion of their income. Asking respondents to repeat the exercise at a later point in time would allow a test-retest of reliability of responses over time, with the hypothesis that reliability will be greater when WTP represents a large proportion of their income, as this value is more precise. Finally, we would examine whether responses are precise or not (prominence), with the hypothesis that WTP responses will be more precise (less prominent values) when the value represents a large proportion of their income. It might also be possible to test the hypothesis that test-retest reliability will be worse when the numerical cues presented to participants are different in the retest phase than they were in the test phase. However, the detrimental effect of changing the cue structure will have more of an effect on test-retest reliability when the value represents a small proportion of income.

We would very much welcome the views of HESG members on the work presented here, and that proposed for the future.

**Table 1 – Explaining the degree of difficulty in valuation**

| | $\beta$ | SE | *t*-ratio | *p*= |
|---|---|---|---|---|
| Full sample | | | | |
| Constant | 3.05 | 0.10 | 30.11 | 0.00 |
| Gender (female=1) | 0.22 | 0.10 | 2.17 | 0.03 |
| Finance (affluent=1) | -0.31 | 0.15 | -2.12 | 0.03 |
| PPS format (=1) | 0.23 | 0.12 | 1.91 | 0.06 |
| OE format (=1) | 0.71 | 0.12 | 5.89 | 0.00 |
| $\check{R}^2=$ | 0.09 | | | |
| Payment scales only | | | | |
| Constant | 3.12 | 0.17 | 18.17 | 0.00 |
| Gender (female=1) | 0.28 | 0.13 | 2.20 | 0.03 |
| Ethnicity (white=1) | 0.31 | 0.14 | 2.17 | 0.03 |
| Finance (affluent=1) | -0.36 | 0.19 | -1.87 | 0.06 |
| PPS format (=1) | 0.26 | 0.13 | 2.08 | 0.04 |
| Scale help decide final WTP? (yes=1) | -0.50 | 0.13 | -3.74 | 0.00 |
| $\check{R}^2=$ | 0.09 | | | |

**Table 2 - 12 most frequently chosen values, in descending order, by elicitation format**

WTP = willingness to pay value chosen, £; AER = Albers' exactness ratio

| | Open-ended (OE) | | | Prominent scale (PPS) | | | Non-prominent scale (NPS) | |
|---|---|---|---|---|---|---|---|---|
| WTP | % | AER | WTP | % | AER | WTP | % | AER |
| 500 | 18.7 | 1.00 | 200 | 11.8 | 1.00 | 200 | 9.9 | 1.00 |
| 100 | 10.8 | 1.00 | 1,000 | 10.2 | 1.00 | 500 | 8.4 | 1.00 |
| 1,000 | 10.8 | 1.00 | 300 | 7.9 | 0.67 | 300 | 7.6 | 0.67 |
| 200 | 9.4 | 1.00 | 500 | 7.9 | 1.00 | 100 | 6.1 | 1.00 |
| 250 | 5.8 | 0.20 | 250 | 6.3 | 0.20 | 150 | 6.1 | 0.33 |
| 300 | 5.0 | 0.67 | 400 | 6.3 | 0.25 | 250 | 6.1 | 0.20 |
| 50 | 3.6 | 1.00 | 100 | 5.5 | 1.00 | 1,000 | 6.1 | 1.00 |
| 1,500 | 3.6 | 0.33 | 1,500 | 4.7 | 0.33 | 400 | 4.6 | 0.25 |
| 2,000 | 3.6 | 1.00 | 120 | 3.9 | 0.17 | 5,000 | 3.8 | 1.00 |
| 5,000 | 3.6 | 1.00 | 150 | 3.9 | 0.33 | 503 | 2.3 | 0.00 |
| 400 | 2.9 | 0.25 | 2,000 | 3.9 | 1.00 | 750 | 2.3 | 0.07 |
| 20 | 2.2 | 1.00 | 2,500 | 3.1 | 0.20 | 1,500 | 2.3 | 0.33 |
| Cumulative proportion, % | 79.9 | | | 75.6 | | | 65.6 | |

**Table 3 - Numerical characteristics of WTP values, by elicitation format**

| | OE | PPS | NPS | $\chi^2$/Z | *p*= |
|---|---|---|---|---|---|
| Proportion offering a prominent number as the maximum WTP, % | 63.3 | 42.4 | | 11.6 | 0.00 |
| | | 42.4 | 36.8 | 0.9 | 0.21 |
| | 63.3 | | 36.8 | 19.4 | 0.00 |
| Median Albers relative exactness ratio | 1.00 | 0.67 | | -3.1 | 0.00 |
| | | 0.67 | 0.33 | -2.1 | 0.04 |
| | 1.00 | | 0.33 | -4.7 | 0.00 |
| Proportion selecting any scale point as their maximum WTP, % | | 90.6 | 13.7 | 152.3 | 0.00 |
| Proportion selecting the highest scale point ticked as their maximum WTP, % | | 57.5 | 10.1 | 64.5 | 0.00 |
| Proportion for whom the maximum WTP is not a round number, % | 2.2 | 2.3 | 16.9 | 29.4 | 0.00 |

**Table 4 - Explaining maximum willingness to pay**

| | $\beta$ | SE | *t*-ratio | *p*= |
|---|---|---|---|---|
| Full sample | | | | |
| Constant | 481.87 | 645.87 | 0.75 | 0.46 |
| Age (years) | 55.78 | 20.14 | 2.77 | 0.01 |
| Ethnicity (white=1) | -613.55 | 202.04 | -3.04 | 0.00 |
| Health (excellent or good=1) | 812.55 | 359.04 | 2.26 | 0.02 |
| Smoking (current smoker=1) | 675.75 | 291.25 | 2.32 | 0.02 |
| Finance (affluent=1) | 1054.37 | 261.99 | 4.02 | 0.00 |
| Finance (impecunious=1) | -416.66 | 217.90 | -1.91 | 0.06 |
| Degree of difficulty | -321.41 | 84.84 | -3.79 | 0.00 |
| $\check{R}^2$= | 0.14 | | | |
| Payment scales only | | | | |
| Constant | 1328.84 | 708.78 | 1.87 | 0.06 |
| Age (years) | 46.74 | 19.93 | 2.34 | 0.02 |
| Ethnicity (white=1) | -744.62 | 214.87 | -3.47 | 0.00 |
| Health (excellent or good=1) | 859.45 | 383.86 | 2.24 | 0.03 |
| Finance (affluent=1) | 964.79 | 284.03 | 3.40 | 0.00 |
| Degree of difficulty | -371.26 | 90.73 | -4.09 | 0.00 |
| Reasonable range on scale ? (yes=1) | -683.63 | 264.39 | -2.59 | 0.01 |
| $\check{R}^2$= | 0.18 | | | |

## References


Albers W (1998). *Foundations of a theory of prominence in the decimal system. Parts I–VI*. Bielefeld: Institute of Mathematical Economics, University of Bielefeld.

Albers W (2001). *Prominence theory as a tool to model boundedly rational decisions*. In: Gigerenzer G, Selten R (Eds.), Bounded rationality: The adaptive toolbox (pp. 297–317). Cambridge, MA: MIT Press.

Bennett R, Tranter R (1998). The dilemma concerning choice of contingent valuation willingness-to-pay elicitation format. *Journal of Environmental Planning and Management*; 41: 253–257.

Bonini N, Biel A, Garling T, Karlsson N (2002). Influencing what the money is perceived to be worth: Framing and priming in contingent valuation studies. *Journal of Economic Psychology*; 23: 655–663.

Brown TC, Champ PA, Bishop RC, McCollum DW (1996). Which response format reveals the truth about donations to a public good? *Land Economics*; 72: 156–166.

Carthy T, Chilton S, Covey J, Hopkins L, Jones-Lee M, Loomes G (1999). On the contingent valuation of safety and the safety of contingent valuation: Part 2 – the CV/SG chained approach. *Journal of Risk and Uncertainty*; 17: 187–213.

Covey J, Smith RD (2006). How common is the prominence effect? Additional evidence to Whynes et al. *Health Economics*; 15: 205–210.

Donaldson C, Thomas R, Torgerson DJ (1997). Validity of open-ended and payment scale approaches to eliciting willingness to pay. *Applied Economics*; 29: 79–84.

Frew EJ, Wolstenholme JL, Whynes DK (2001). Willingness-to-pay for colorectal cancer screening. *European Journal of Cancer*; 37: 1746–1751.

Gold N, List C (2004). Framing as path dependence. *Economics and Philosophy*; 20: 253–277.

Hertwig R, Hoffrage U, Martignon L (1999). *Quick estimation: Letting the environment do the work*. In: Schindler G (1991). Symbolic meanings of a price ending. Advances in Consumer Research, 18(1), 794–801.

National Oceanic and Atmospheric Administration (1993). Report of the NOAA panel on contingent valuation. *Federal Register*; 58: 4602-4614

Olsen JA, Smith RD (2001). Theory versus practice: A review of 'willingness-to-pay' in health and health care. *Health Economics*; 10: 39-52.

Schindler RM (1991). Symbolic meanings of a price ending. *Advances in Consumer Research*; 18: 794–801.

Slovic P (1995). The construction of preference. *American Psychologist*; 50: 364–371.

Smith RD (2000). The discrete-choice willingness to pay question format in health economics: Should we adopt environmental guidelines? *Medical Decision Making*; 20: 194–206.

Smith RD (2003). Construction of the contingent valuation market in health care: A critical assessment. *Health Economics*; 12: 609–628.

Smith RD (2004). The reliability of willingness to pay for changes in health status. *Applied Health Economics and Health Policy*; 3(1): 35-38.

Smith RD (2005). Sensitivity to scale in contingent valuation: the importance of the budget constraint. *Journal of Health Economics*; 24: 515-529.

Smith RD (2006a). Its not just what you do, its the way that you do it: the effect of different payment card formats and survey administration on willingness to pay for health gain. *Health Economics*; 15: 281-293.

Smith RD (2006b). The reliability of willingness to pay values: a qualitative insight. *Health Economics Letters*; 10: 21-26.

Whynes DK, Philips Z, Frew E (2005). Think of a number. . .any number? *Health Economics*; 14: 1191–1195.

Whynes DK, Frew EJ, Philips ZN, Covey J, Smith RD (2007). On the numerical forms of contingent valuation responses. *Journal of Economic Psychology (forthcoming)*.

Yeung RYT, Smith RD, McGhee SM (2003). Willingness to pay and size of health benefit: an integrated model to test for 'sensitivity to scale'. *Health Economics*; 12: 791-796.

Yeung RYT, Smith RD, Ho LM, Johnston J, Leung G (2006). Empirical implications of response acquiescence in discrete-choice contingent valuation. *Health Economics*; 15: 1077-1089.