Sources of inadequate scope sensitivity in WTP for risk reductions: An experimental approach

Katharine Bolt, Ian Bateman, Brett Day, and Graham Loomes
University of East Anglia, Norwich, UK

Abstract:

The issue of WTP estimates exhibiting inadequate scope sensitivity has been the focus of much debate in the non-market valuation literature. This problem is particularly acute in the context of WTP for risk reductions. This study examines the determinants of this scope insensitivity asking whether unfamiliarity with the valuation scenario and/or the risk denominator pose an excessive cognitive burden on respondents resulting in the magnitude of the risk reduction being overlooked. Using experimental techniques we show that as familiarity with the valuation scenario declines so too does sensitivity to the magnitude of the risk reduction. We also find that increasing the risk denominator reduces sensitivity to scope. These results suggest limitations to the application of stated preference methods in unfamiliar, low risk scenarios such as mortality risk-reduction studies.

JEL Classifications: D61, D81, I10, J17, Q51

1 Overview

Stated preference methods are a long established method for eliciting individuals' willingness to pay (WTP) for non-market benefits. However, the reliability of these WTP estimates has been the focus of much debate in the literature (see for example, Jones-Lee, 1989; Hammitt and Graham, 1999; Beattie *et al*, 1998). Reliable estimates should correspond to expectations of standard neo-classical theory of consumer behaviour, in which the basic test is that 'more is better'. Therefore, an individual should be willing to sacrifice more money for more of a good and, under certain reasonable assumptions, WTP should be roughly proportional to the level of good.

However, the literature has often reported that WTP estimates are not adequately sensitive to the quantity of the good offered (see Hammitt and Graham, 1999; Beattie et al, 1998 for example). In these cases it appears that the quantity of the good was overlooked by the respondent, or not given adequate consideration. An explanation of why this may be is suggested by the psychology literature which reports a collection of models that describe an individual's decision-making as a dual-process. The first process assumes that people approach decision making rationally (otherwise referred to as analytically or deliberatively). The second process assumes that people use intuition (otherwise referred to as experience or automatic) to make decisions. The rational process is the system that standard economic theory relies upon people being dominated by.

The observed insensitivity to the quantity of the good offered may be explained by the relative influence of these two systems. For certain valuation scenarios respondents may rely more heavily on their rational system and therefore comply with assumptions of behaviour in standard economic theory. However, for others respondents may switch to be more heavily influenced by their intuitive system. It is conjectured that for complex valuation scenarios respondents may be unable to process all the available information in a rational deliberative manner and switch to an intuitive process of decision making.

This calls into question the applicability of stated preference methods to scenarios which pose an excessive cognitive burden on the respondent. Scenarios that the respondent is

likely to find difficult to value are those goods for which he or she does not have pre-existing fully defined preferences. This would be the case for goods with which the respondent has little experience. If respondents are not familiar with trading the good for other goods then it is likely they will find the valuation task more complex. Estimates of WTP for everyday goods would therefore be more likely to be reliable than unfamiliar goods.

Another category of goods that may pose an excessive cognitive burden on respondents are those where the outcome is uncertain. It has been found in the literature that individuals have difficulty understanding probabilities. If this is the case, then this would cause difficulty for the respondent to have a complete understanding of the valuation task.

This study focuses on the reliability of WTP estimates for these two categories of goods. The question of the validity of WTP estimates for unfamiliar and/or risky goods is an important issue for many other non-market goods for which stated preference may be used to value. For example, changes in the risk of certain health impacts are a common outcome of many environmental programmes. The risks involved are often small and therefore cognitively challenging as part of a valuation scenarios. Another good often valued using stated preference is biodiversity or ecosystems. These are complex goods with which the respondent is unlikely to be familiar. Another example is that of valuing the impacts from climate change as these are uncertain and therefore would require respondents to value risk. To varying degrees the question of familiarity and/or risk affects many non-market goods which have been or could be valued using stated preference methods.

Using techniques of experimental economics, this study investigates whether the reliability of WTP estimates is affected by 1) the degree of familiarity with the good and/or 2) the level of risk. The study then goes on to consider whether WTP estimates for risk reductions for goods of varying familiarity are subject to other biases such as framing bias, part-whole effects, and hypothetical bias. It also considers whether different types of respondents are more successful at giving valid WTP estimates than others. In particular

it is tested whether respondents who are more numerate give more valid WTP estimates than other less numerate respondents.

Section 2 discusses the theoretical expectations of WTP sensitivity to the magnitude of a risk reduction. The experimental design is described in section 3. Section 4 presents the results and section 5 offers concluding observations.

2 Theoretical expectations and scope sensitivity

The prominence of the scope test as a criterion for reliable non-market valuation can be traced back to some of the earliest empirical applications. However, the test rose to particular prominence with the endorsement of the so called National Oceanic and Atmospheric Administration (NOAA) 'Blue Ribbon' panel report on the reliability of contingent valuation for estimating passive use damages (Arrow *et al.*, 1993). This report noted that WTP estimates should be 'adequately' responsive to scope and concluded that the observation of insensitivity to scope was 'perhaps the most important internal argument against the reliability of the CV approaches' (*ibid*, p4607).

The interpretation of an 'adequate' response to change in scope depends on assumptions made about the respondent's underlying preferences. Without assumptions about the respondent's utility function, utility theory merely suggests that respondents can order bundles of goods in terms of preference (Powe and Bateman, 2005; p261). So, where *A* is a larger good than *a:*

- $WTP_A \geq WTP_a$ if weak monotonicity with respect to magnitude is assumed
- $WTP_A > WTP_a$ if strong monotonicity with respect to magnitude is assumed

The standard assumptions are that, under conditions of non-satiation (which seems a reasonable assumption for most health risk reductions), the respondent's utility function is increasing with respect to consumption, strictly concave, and bounded from below. Under these assumptions it is possible to show that for certain risks when WTP is

concentrated on small changes in risks, WTP should not only (a) increase as the risk reduction increases but also (b) increase roughly proportionally (Hammitt and Graham, 1999; p34). Taking the case of mortality risk reductions, if a reduction in annual mortality risk from 20/100,000 to 18/100,000 is valued at $20, then a larger reduction from 20/100,000 to 16/100,000 should be valued at about $40 (Hammitt and Graham, 1999; p35). Note that, while the assumptions required for this expectation are reasonable for health risk changes, they may be less plausible for many other goods, such as non-use environmental resources. This makes the expectations of the proportional scope test less clear a-priori and the focus upon health of particular interest as if scope tests fail here then it may have wider consequences.

Empirically, the evidence of scope sensitivity (or lack thereof) is mixed. Some studies have demonstrated scope sensitivity (e.g. Norinder *et al*, 2001; Carson, 1997; Smith and Osborne, 1996), others have not (e.g. Jones-Lee, 1989; Beattie *et al*, 1998), and still others show that it is possible to observe both sensitivity and insensitivity to scope in the same study (e.g. Veisten *et al*, 2004; Olsen *et al*, 2004).

Scope sensitivity appears to be a particularly salient problem in the case of valuing changes in risk. According to Carson (1997), there is not a general problem with scope insensitivity in valuing non-market benefits. However, Carson does acknowledge a scope insensitivity problem with respect to the valuation of small changes in health risk. A review of the literature by Hammitt and Graham (1999) finds that risk reduction studies typically pass tests of the statistical significance of scope sensitivity (i.e. the typical respondent does state a greater WTP for larger risk reductions), but that the proportionality assumption generally fails.

Some commentators have argued that empirical evidence of scope insensitivity provides support for non-standard preference theories. These critiques range from technical questions regarding the specification of the respondent's utility function, to wide ranging reformulations of consumer behaviour theory such as prospect or reference-dependent utility theory (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991). However,

even these models are locally linear[1] thus demanding WTP to be roughly proportional to the magnitude of the good offered.

This suggests that the question of whether or not standard theory holds may not be the most pertinent issue in understanding why scope insensitivity (or, more accurately, insufficient scope sensitivity) is such a prevalent problem. Another factor, arguably one which more readily satisfies Occam's Razor (Blaug, 1997), concerns unfamiliarity and consequent cognitive load of many non-market valuation scenarios. If valuation questions are too far removed from survey respondent's experience of assessment tasks then they may begin to lack 'evaluability' (Hsee 1996; 1998; Slovic *et al*., 2004). In such situations those dimensions of the valuation task which appear objectively pertinent to the analyst (such as the scope of the good in question), may be overlooked by the respondent. Inadequate scope sensitivity seems a likely outcome of such situations.

The issue of whether unfamiliarity may limit respondents' ability or preparedness to engage in non-market valuation exercises is perhaps surprisingly under researched. When the outcome of a scenario is unfamiliar it raises the possibility that the underlying model of well-formed preferences may not be sound. In such situations, rather than being able to refer to some pre-existing set of preferences (Carson *et al.*, 1999), or even being able to quickly 'discover' preferences of the type envisaged under standard economic theory (Plott, 1996), the danger is that respondents may 'construct' preferences (Slovic, 1995). Under the constructed preference hypothesis, respondents do not refer to some underlying, theoretically compatible preference set, but rather use the question frame to provide heuristic cues (or 'rules of thumb') to form responses (Tversky and Kahneman, 1973; 1974). Both the description of good and that of the contingent market may be a source of such cues (Braga and Starmer, 2005). In such situations, as the question frame alters so do the heuristic cues and consequently, so do the implied preferences (Kahneman, *et al*, 1999; Ariely, *et al.*, 2003). Without the touchstone of true underlying preferences to guide responses they become malleable, fixed within a given question frame rather than transferable across such frames. In such situations changes in the objective dimensions of

[1] Except possibly at zero risk level.

the good become overshadowed by changes in the question frame and problems such as scope insensitivity become almost inevitable.

If respondents are unfamiliar with the good then it would follow that it would be harder for them to value a change in it. This may result in respondents focusing on particular characteristics of the good rather than assessing all the information in the valuation scenario. The respondent may then overlook the magnitude of the good resulting in WTP being less sensitive to the size of the good. Veisten *et al* (2004) found that less familiar goods exhibited less scope than familiar goods when they looked at the sensitivity of WTP for an unfamiliar sub-group of endangered species (insects, cryptogams, and birds) compared to 'all endangered species'. According to the authors, the fact that respondents knew relatively little about the unfamiliar species and their role in the ecosystem may have been the origin of the problems observed here.

Alongside unfamiliarity with the good, respondents may fail to adequately understand the change in the provision level hypothesised in the valuation exercise. In particular, small risk reductions pose a cognitive burden on respondents who are not used to dealing with probabilities. If risks are not communicated in a meaningful way, respondents may fail to distinguish between risk reductions of different sizes, or be confused about absolute versus relative risk reductions (Baron, 1997).

This study empirically tests the following hypotheses:

H_0^1: Willingness to pay (WTP) is equal for larger and smaller risk reductions

$$WTP^{b-n;j}(Z) = WTP^{b-m;j}(Z)$$

H_1^1: Willingness to pay (WTP) is higher for larger rather than smaller risk reductions

$$WTP^{b-n;j}(Z) > WTP^{b-m;j}(Z)$$

where, $WTP^{b-n;j}(Z)$ = willingness to pay for risk reduction from the baseline risk *b/j* to *n/j* and $WTP^{b-m;j}(Z)$ = willingness to pay for risk reduction from the baseline risk *b/j* to *m/j* where $n>m$.

H_0^2: Respondents WTP is proportional to the level of risk

$$WTP^{xb-xn;j}(Z) = x \cdot WTP^{b-n;j}(Z)$$

H_1^2: Respondents WTP is less than proportional to the level of risk

$$WTP^{xb-xn;j}(Z) < x \cdot WTP^{b-n;j}(Z)$$

where $x > 0$

H_0^3: As familiarity with the scenario decreases scope sensitivity of WTP is not affected.

$$SS^{\frac{b-n}{c-d};j}(y) = SS^{\frac{b-n}{c-d};j}(Z)$$

H_1^3: As familiarity with the scenario decreases scope sensitivity of WTP decreases.

$$SS^{\frac{b-n}{c-d};j}(y) > SS^{\frac{b-n}{c-d};j}(Z)$$

where $SS^{\frac{b-n}{c-d};j}(y)$ is the scope sensitivity of WTP for reductions in the risk of good *y* measured as the ratio of WTP for the risk reduction from *b/j* to *n/j* and WTP for the risk reduction from *c/j* to *d/j*. $SS^{\frac{b-n}{c-d};j}(Z)$ is the scope sensitivity of WTP for reduction in risk for good *Z*. Good *Z* is less familiar than good *y*.

H_0^4: As the risk denominator increases scope sensitivity of WTP is not affected

$$SS^{\frac{b-n}{c-d}=X;j}(Z) = SS^{\frac{e-f}{g-h}=X;k}(Z)$$

H_1^4: As the risk denominator increases scope sensitivity of WTP decreases

$$SS^{\frac{b-n}{c-d}=X;j}(Z) > SS^{\frac{e-f}{g-h}=X;k}(Z)$$

where $SS^{\frac{b-n}{c-d}=X;j}(Z)$ is the ratio of WTP for a risk reduction where the ratio of the risk reduction *b/j* to *n/j* and *c/j* to *d/j* is *X*. $SS^{\frac{e-f}{g-h}=X;k}(Z)$ is the ratio of WTP for a risk reduction where the ratio of the risk reduction *b/j* to *n/j* and *c/j* to *d/j* is the same as the previous ratio of risk reductions, *X*. The risk denominator $j<k$.

Scope sensitivity will be measured by comparing 'within respondent' estimates of WTP and 'between respondent' estimates of WTP. The 'within respondent' or 'internal' test compares WTP responses for the larger and smaller risk reductions from the same respondent whereas the 'between respondent' or 'external' test compares responses between different respondents. The internal test is arguably less demanding, as respondents may well consistently base their answers to WTP questions about one risk reduction on their responses to previous questions about larger or smaller risk reductions. Even if respondents are extremely uncertain about their WTP for risk reduction, anchoring on their previous responses may enforce some degree of apparent consistency. In an external test, different respondents are asked about their WTP for different risk reductions and there is no possibility of coordinating their responses.

The following section describes the experimental design undertaken to test these four hypotheses.

3 Experimental design

Various stages of the experiment design were pre-tested using one-to-one interviews on 55 respondents. Data was collected by a computerised experiment administered at the University of East Anglia (UEA) which took approximately 35 minutes to complete. Respondents were recruited from the student population at the UEA by email invitation. Emails were sent to nearly 2,000, mostly undergraduate, students from a cross-section of

academic disciplines inviting them to participate in the experiment. Of this, 154 students replied and 99 participated in the experiment. Recruited students were mostly undergraduates from a wide range of academic disciplines.

Testing the impact of familiarity on scope sensitivity

In order to ascertain whether respondents' WTP scope sensitivity was affected by either (or both) the familiarity of the good, and their experience with the specified market, respondents were asked to value risk changes for three very different scenarios: (1) Avoiding losses of money (high familiarity good); (2) Avoiding stomach illness (medium familiarity good); (3) Avoiding risks of temporary blindness (low familiarity). A brief description of each good is given in Box 1 and discussed here.

Scenario (1), concerning a risk of having money stolen, was intended to be the most familiar. The reasons for this are two fold. First, most people are familiar with losing money and even if it has not happened to them, such a scenario is readily imaginable. Second, it overcomes the potential hurdle of unfamiliarity with certain contingent markets, such as the unusual task of valuing non-market goods. As in this example the potential loss is in direct money terms, we have a simple money risk-money trade situation such that this hurdle is overcome.

The second scenario concerns the risk of getting a stomach bug. While the outcome of this is likely to be well known, respondents may not be familiar with putting a monetary value on a health-risk reduction. Therefore scenario 2 is assumed to be less familiar than scenario 1.

Scenario (3) presented respondents with a practically uniformly unfamiliar outcome; temporary blindness (a genuine but very rare condition). The aim of the scenario was to present respondents with something unfamiliar but not, long-term, overwhelmingly serious.

Box 1: Scenario descriptions

Scenario 1: Chance of money being stolen

Suppose that for some reason you must put £75 in a jar and leave it at a specified location overnight. There is a risk that a thief will steal your money overnight. If the thief steals the money you will lose £75. If he doesn't, you can collect your money from the jar the next day. However, there is a security device that you can buy which will reduce the chance that the thief will steal the money.

Scenario 2: Chance of getting a stomach bug

There is a risk that you will get a stomach bug. It affects everyone:

- *Young and old*
- *Fit and healthy and ill and weak*

If you get the stomach bug you will have the following symptoms for 3 days:

- *Vomiting every two hours*
- *Diarrhoea every two hours*
- *Continuous stomach cramps*
- *Continuous exhaustion*

There is a pill that will reduce your risk of getting the stomach bug. The pill is:

- *100% effective*
- *thoroughly tested and has absolutely no side effects*

Scenario 3: Chance of temporarily going blind

There is a virus that leads your to lose your sight completely for 3 days. For people who get the virus they are unable to see anything at all—not even light.

- *You are completely blind for 3 days*
- *But sight returns to normal on the 4^{th} day*

There is a pill available that will reduce your risk of getting the virus and losing your sight for 3 days. The pill is:

- *100% effective*
- *thoroughly tested and has absolutely no side effects*

Testing the impact of the risk denominator on scope sensitivity

To examine the impact upon scope sensitivity of changes in the denominator used to convey an objectively identical risk, respondents were asked to value a given risk reduction expressed both with a denominator of 10 and with a denominator of 1000 (e.g. contrasting WTP for a risk reduction from 1/10 to 0/10 as well as from 100/1,000 to 0/1,000). Within both denominator treatments, scope sensitivity tests examined WTP for different levels of risk reduction. So, for the denominator of 10, respondents were asked to value a risk reduction of 5/10 to 0/10 and 1/10 to 0/10. Similarly, for the risk

denominator of 1,000 respondents valued risk reductions 100/1,000 to 0/1,000 and 20/1,000 to 0/1,000. In both cases the larger risk reduction was five times greater than the smaller risk reduction. Therefore, if the proportionality condition holds, WTP for the larger risk reduction should be roughly five times greater than WTP for the smaller risk reduction. By comparing scope shown in the WTP estimates when the risk denominator is 10 and when it is 1,000 it should be possible to ascertain whether scope sensitivity alters with changes in the risk denominator.

The experimental design allowed both 'internal' (within sample) and 'external' (between sample) tests of scope sensitivity to the magnitude in risk. The less demanding internal test compares WTP responses for the larger and smaller risk reductions from the same respondent whereas the external test compares responses between different respondents.

To reduce the possibility of respondents anchoring and constructing consistent WTP responses, a more demanding internal test was designed. This was done by giving half the respondents the related WTP questions (i.e. 5/10 to 0/10; and 1/10 to 0/10 for the same scenario) apart in the experiment and including distraction tasks between them. Half the respondents were posed the related WTP questions at least three questions apart (usually 6 to 7 questions apart) and with at least one distraction task in between. The other respondents were posed the related WTP questions adjacent to one another. We call this test the '**i**nternal test **c**ontrolling for **c**onstruction of **p**references' (ICCP) test'. For the ICCP test if respondents exhibit more sensitivity to the magnitude of the risk when the relevant questions are adjacent rather than far apart then this may be due to respondents merely constructing consistent answers rather than having well-defined preferences.

In order to carry out the external scope sensitivity test the experimental design ensured that half of the respondents were posed the larger risk reduction in 10 first while the remainder of the sample initially faced the question concerning the smaller risk reduction in 10. The same split treatment was applied to the risk in 1,000. Our external scope sensitivity test then compares WTP responses for identical risk reductions across ordering treatments.

Risk communication

In order to minimise the possibility that any observed inadequacy in scope sensitivity was merely due to risk changes being poorly communicated to respondents, in all treatments the first half of the experiment was dedicated to introducing risk concepts and the representation of risk. Following recent psychological research in risk communication (Hibbard and Peters, 2003; Lipkus and Hollands, 1999; Slovic *et al*., 2002; Peters *et al*., 2005a,b; Hibbard *et al.*, 2002), risks were represented both numerically and visually throughout the experiment.

The experiment began by presenting some 'warm up' exercises which asked respondents to estimate the chance of something happening. The risks were familiar and the questions were phrased so as to get respondents thinking about risks in 10. For example respondents were asked to estimate on how many days they thought it would rain during the first 10 days of August in Norwich?

After these 'warm-up' exercises, respondents were then introduced to risks being conveyed visually on a grid. Corso *et al* (2001) found a version of this approach to be the most effective means of communicating risk. Respondents were presented with the current risk on a grid of 10 squares. A risk of 3 in 10 for example, was represented by a grid of 10 squares, 3 of which were coloured red and 7 white. The advantage of running the experiment on the computer was that respondents could then click to change the grid to show the new level of risk being offered. If the new risk level being offered was 1 in 10 then the grid would change to show 1 red square and 9 white ones. Clicking between the two grids gave the respondent a visual sense of the change in risk. After familiarising themselves with risks with a denominator of 10, respondents were introduced to risks with a denominator of 1,000.

Before presenting respondents with the WTP questions central to the experiment, respondents were given some further practice exercises, in this case introducing them to the idea of trading off risk and other goods as well as the mechanism for stating their

WTP estimates. The WTP elicitation format utilised an interactive payment card in which the respondent was asked to use a slider to determine their maximum WTP. An example of this is shown in Figure 1. As the respondent dragged the pointer along the slider their WTP estimate was shown in the bid box. The amount increased exponentially from £0.00, increasing faster as the pointer reached towards the end of the slider with the highest possible amount of £1,000.00. Respondents were encouraged to use the slider to get roughly to the amount they were WTP. The 'Up' and 'Down' buttons were then used to increase or decrease their bid amounts by increments of 10 pence. The idea behind this was to reduce the amount of rounding error as respondents often state WTP estimates rounded to at least the nearest one pound. The 'Up' and 'Down' buttons encouraged respondents to think in terms of 10 pence increments. To the best of our knowledge this is first time that a stated preference method has used this interactive payment card approach.

Figure 1: Example of payment card

4 Results

The first objective of the experiment was to ascertain whether respondents' WTP passed the 'weak' test of scope sensitivity. If this were the case then respondents should be willing to pay more for larger risk reductions than smaller risk reductions. So, it would be expected that respondents were willing to pay more for a risk reduction from 5/10 to 0/10

than for a risk reduction from 1/10 to 0/10. Similarly, individuals should also be willing to pay more for a risk reduction from 100/1,000 to 0/1,000 than for a risk reduction from 20/1,000 to 0/1,000.

We consider 'internal' (within sample), 'ICCP' (within sample but with controls to reduce the ability of respondents 'constructing' scope sensitivity), and 'external' (between sample) tests of sensitivity to the magnitude in risk.

The internal test of sensitivity to the magnitude in risk shows that respondents' mean WTP is indeed higher for larger risk reductions than for smaller risk reductions. The results are shown in Table 1. The Wilcoxon signed-rank test shows that there is a statistically significant increase in WTP for larger over smaller risk reductions. As the research hypothesis is that WTP for the larger risk reduction is greater than WTP for the smaller risk reduction the 1-tailed probability is calculated. Table 1 shows that this is significant at the 99 percent level for all goods and for both denominator levels. So, for the internal test the null hypothesis is rejected giving support to the research hypothesis H_1^1 that respondents are WTP more for a larger rather than smaller risk reduction.

Table 1: Internal scope sensitivity tests

	Mean WTP (£)		**Wilcoxon signed ranks test[a]**	
	Larger risk reduction	**Smaller risk reduction**	**Z**	**Sig. (1-tailed)**
Risk in 10	*5/10 to 0/10*	*1/10 to 0/10*		
Money stolen	21.46	6.98	-8.236(a)	.000
Stomach bug	25.02	9.76	-8.174(a)	.000
Blindness	54.65	26.49	-7.482(a)	.000
Risk in 1,000	*100/1,000 to 0/1,000*	*20/1,000 to 0/1,000*		
Money stolen	11.71	7.45	-4.985(a)	.000
Stomach bug	13.26	7.11	-6.521(a)	.000
Blindness	32.16	19.46	-5.844(a)	.000

Note: a Based on positive ranks.

The 'ICCP' test of sensitivity to the magnitude of risk finds that both respondents who were presented with consecutively ordered and non-consecutively ordered risk questions exhibited scope sensitivity (see Table 2 and Table 3 respectively).

Table 2: ICCP Test of Scope Sensitivity: Questions Adjacent

	Mean WTP (£)		**Wilcoxon signed ranks test[a]**	
	Larger risk reduction	**Smaller risk reduction**	**Z**	**Sig. (1-tailed)**
Risk in 10	*5/10 to 0/10*	*1/10 to 0/10*		
Money stolen	21.30	7.33	-5.907(a)	.000
Stomach bug	30.57	12.25	-5.910(a)	.000
Blindness	53.76	23.97	-5.580(a)	.000
Risk in 1,000	*100/1,000 to 0/1,000*	*20/1,000 to 0/1,000*		
Money stolen	11.33	8.09	-3.300(a)	.000
Stomach bug	13.10	7.63	-4.229(a)	.000
Blindness	35.03	22.13	-4.308(a)	.000

Note: a Based on positive ranks.

Table 3: ICCP Test of Scope Sensitivity: Questions Apart

	Mean WTP (£)		**Wilcoxon signed ranks test[a]**	
	Larger risk reduction	**Smaller risk reduction**	**Z**	**Sig. (1-tailed)**
Risk in 10	*5/10 to 0/10*	*1/10 to 0/10*		
Money stolen	21.62	6.64	-5.766(a)	.000
Stomach bug	19.58	7.33	-5.656(a)	.000
Blindness	55.51	28.96	-4.915(a)	.000
Risk in 1,000	*100/1,000 to 0/1,000*	*20/1,000 to 0/1,000*		
Money stolen	12.09	6.82	-3.879(a)	.000
Stomach bug	13.42	6.61	-4.955(a)	.000
Blindness	29.34	16.84	-3.960(a)	.000

Note: a. Based on positive ranks.

The result that even when the relevant questions are posed far apart, mean WTP for the larger risk reduction is still larger than for the smaller risk reduction, gives stronger evidence that respondents do exhibit 'weak' scope sensitivity. This is because the ICCP

test reduces the ability of respondents to merely base their answers to WTP questions about one risk reduction on their responses to previous questions about larger or smaller risk reductions.

The most demanding test of sensitivity to the magnitude of risk is the 'external' (across respondent) test. The non-parametric Mann-Whitney test looks at the difference in the ranked positions of scores in different groups. The results are shown in Table 4.

Table 4: External Test of Scope Sensitivity

	Mean WTP (£)		**Mann Whitney Test[a]**		
	Larger risk reduction	**Smaller risk reduction**	**Mann-Whitney U**	**Z**	**Sig. (1-tailed)**
Risk in 10	*5/10 to 0/10*	*1/10 to 0/10*			
Money stolen	18.35	8.47	477.00	-5.05	.000
Stomach bug	22.88	13.03	708.50	-3.37	.000
Blindness	57.36	34.84	871.50	-2.09	.014
Risk in 1,000	*100/1,000 to 0/1,000*	*20/1,000 to 0/1,000*			
Money stolen	8.92	7.83	832.00	-2.49	.006
Stomach bug	10.67	8.53	1094.00	-.59	.278
Blindness	24.44	25.64	1049.50	-.87	.180

Note: a Grouping Variable: Riskred

The 'external' test shows that when changes in risk are presented using a denominator of 10 (i.e. from 5/10 to 0/10) WTP is significantly higher for larger than smaller risk reductions. This is significant at the 99 percent level for the two more familiar scenarios (money stolen and stomach bug) and at the 95 percent level for the least familiar scenario (going blind). Therefore, for valuation scenarios with a risk denominator of 10, the null hypothesis H_0^1 is rejected. However, when changes in risk are presented using a denominator of 1,000 (i.e. 100/1,000 to 0/1,000) the question of whether WTP is higher at a statistically significant level for larger rather than smaller risk reduction depends on how familiar the scenario is. For the most familiar scenario (that of having money stolen) the test shows that WTP is higher for the larger risk reduction and that this is significant

at the 99 percent level. However, for the stomach bug scenario the surplus of WTP for the larger risk reduction over that for the smaller risk reduction is not statistically significant. Furthermore, for the least familiar good, not only does the scope sensitivity test fail but even the mean WTP values are incorrectly ordered. Therefore, it is not possible to reject the null hypothesis H_0^1 for valuation scenarios when the risk denominator is 1,000 and the scenario is unfamiliar.

In summary, while respondents' exhibit an ability to maintain internal consistency in their responses, external testing reveals a less satisfactory picture. While external scope sensitivity tests are satisfied for risk reductions with a denominator of 10 affecting familiar risks, as risks move to denominators of 1,000 affecting less familiar goods so scope sensitivity tests fail. Furthermore, as discussed previously, even when tests of statistical scope significance are passed, this does not imply that the degree of scope sensitivity is adequate and consistent with reasonable expectations. To consider this issue I examine the second hypothesis of the experiment: To ascertain whether WTP responses satisfy the proportionality condition.

As respondents were asked to value certain risk changes that were five times greater than the smaller risk changes, then if the proportionality condition holds WTP should also be five times higher for the larger than the smaller risk reduction. To test this second hypothesis, a new variable was calculated for both risk denominators by multiplying the WTP responses for the small risk reduction by five. If WTP were in fact proportional in risk then the mean WTP responses for the large risk reduction should not differ significantly from this calculated amount.

The Wilcoxon signed ranks test estimates whether the mean WTP for the larger risk reduction is equal to the mean WTP of five times the smaller risk reduction. Both the 'internal' and 'ICCP' test shows that WTP is less than proportional in risk. This is statistically significant at the 99 percent level as shown in Table 5 and Table 6 and therefore the null hypothesis (H_0^2) is rejected. As the Wilcoxon test is based on negative

ranks it shows that the mean level of WTP is higher for five times the smaller risk reduction than for the larger risk reduction. Given that the less demanding 'internal' tests failed the proportionality condition, it is unsurprisingly that the more demanding 'external' test also rejected the proportionality condition for all three scenarios and for both risk denominators (see Table 7).

Table 5: 'Internal' Wilcoxon test of proportionality condition

	£1to0x5 - £5to0	stom1to0x5 - stom5to0	blind1to0x5 - blind5to0	£20x5 - £100to0	stom20x5 - stom100to0	blind20x5 - blind100to0
Z	-3.984(a)	-5.790(a)	-6.294(a)	-6.999(a)	-7.434(a)	-7.496(a)
Asymp. Sig. (2-tailed)	.000	.000	.000	.000	.000	.000

a Based on negative ranks.

Table 6: 'ICCP' Wilcoxon test of proportionality condition

Questions posed adjacent						
	£1to0x5 - £5to0	stom1to0x5 - stom5to0	blind1to0x5 - blind5to0	£20x5 - £100to0	stom20x5 - stom100to0	blind20x5 - blind100to0
Z	-2.874(a)	-4.321(a)	-4.329(a)	-5.629(a)	-5.561(a)	-5.369(a)
Asymp. Sig. (2-tailed)	.004	.000	.000	.000	.000	.000
Questions posed far apart						
	£1to0x5 - £5to0	stom1to0x5 - stom5to0	blind1to0x5 - blind5to0	£20x5 - £100to0	stom20x5 - stom100to0	blind20x5 - blind100to0
Z	-2.886(a)	-3.784(a)	-4.680(a)	-4.420(a)	-4.973(a)	-5.063(a)
Asymp. Sig. (2-tailed)	.004	.000	.000	.000	.000	.000

a Based on negative ranks.

Table 7: 'External' test of proportionality condition

	£5v5x1	stom5v5x1	blind5v5x1	£100v5x20	stom100v5 x20	blind100v5 x20
Mann-Whitney U	812.500	777.500	770.000	659.000	454.000	662.000
Wilcoxon W	2087.500	2052.500	1995.000	1934.000	1729.000	1887.000
Z	-2.623	-2.875	-2.935	-3.734	-5.223	-3.716
Asymp. Sig. (2-tailed)	.009	.004	.003	.000	.000	.000

a Grouping Variable: Riskred

While respondents are generally found to be prepared to pay more for larger risk reductions, resultant WTP is not consistently proportional to risk reductions. The results are largely in agreement with findings in other studies. For example, in their review of the pertinent risk literature, while Hammitt and Graham (1999) find general evidence of statistically significant scope sensitivity within internal tests, evidence from external testing is mixed and, for both tests, the proportionality assumption generally fails.

Given these findings is there any evidence to suggest that familiarity with the scenario and/or the risk denominator affects the sensitivity of WTP estimates to the magnitude of the risk change? The third research hypothesis of this experiment asks whether lack of familiarity with the valuation scenario adversely affects the ability of respondents to give WTP estimates that are sensitive to the magnitude of the risk reduction. If this is the case then it could be expected that scope sensitivity will decrease as familiarity with the scenario decreases.

Figure 2 and Figure 3 respectively show the mean ratios between the WTP responses for the large risk reduction and five times the WTP for the smaller risk reduction for the pooled sample (N=97). As stated before, if WTP were proportional to the level of risk reduction then the ratio between the two should be equal to one. However, as can be seen in Figure 2 and Figure 3 the ratio for all scenarios and risk levels consistently falls short of this level, indicating that WTP is less than proportional to the level of risk.

Figure 2: Ratios of WTP for larger and five times smaller risk reduction for risk in 10

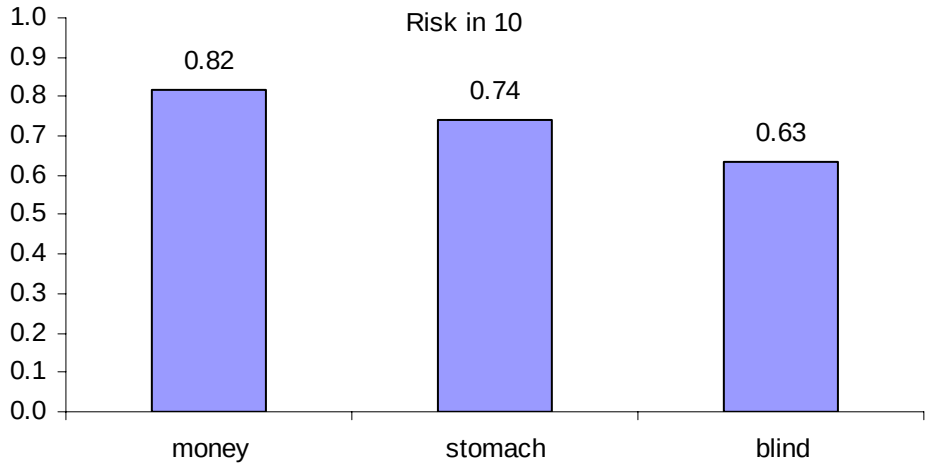
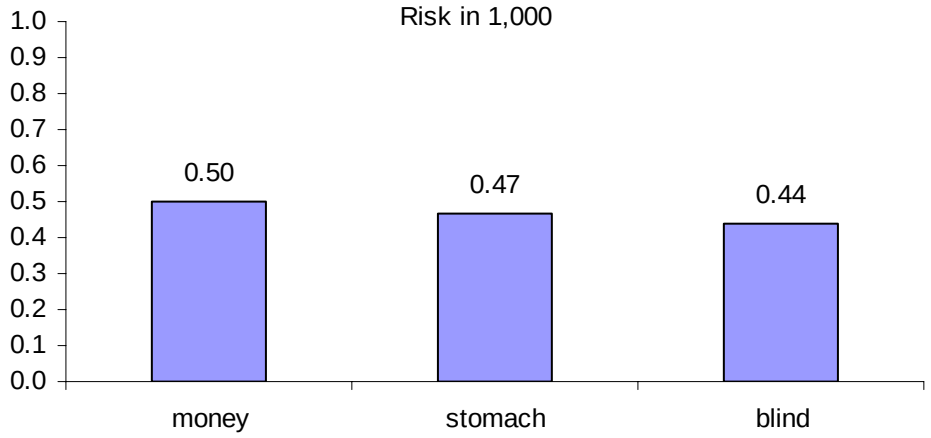


Figure 3: Ratios of WTP for larger and five times smaller risk reduction for risk in 1,000



In both Figure 2 and Figure 3 the mean ratio decreases as familiarity with the scenario decreases. This indicates that the proportionality problem increases as the level of familiarity with the scenario decreases. Table 7 details the Wilcoxon signed-rank tests examining whether the ratios for each scenario are significantly different from each other. The difference in ratios is not statistically significant for scenarios with one degree of difference in familiarity (i.e. money stolen and stomach bug or stomach bug and blind). However, the ratios between the most familiar (money stolen) and least familiar (going

blind) scenario are different at the 99 percent level when the risk denominator is 10 and at the 90 percent level when the risk denominator is 1,000. This supports the research hypothesis H_1^3 that as familiarity with the valuation scenario decreases, the sensitivity of WTP to the magnitude of the risk reduction decreases.

Table 8: Wilcoxon signed rank test of difference in ratio between scenarios

	stom5div5x1to0 - £5div5x1to0	blind5div5x1to0 - stom5div5x1to0	blind5div5x1to0 - £5div5x1to0	stom100div5x20 - £100div5x20	blind100div5x20 - stom100div5x20	blind100div5x20 - £100div5x20
Z	-1.396(a)	-.911(a)	-2.676(a)	-.376(a)	-.642(a)	-1.766(a)
Asymp. Sig. (2-tailed)	.163	.362	.007	.707	.521	.077

a Based on positive ranks.

The fourth research hypothesis H_1^4 of this experiment is that WTP responses are less sensitive to the size of the risk reduction being offered when the risk denominator is larger. The null hypothesis H_0^4 is that the risk denominator does not affect the scope sensitivity of WTP. Comparing Figure 2 and Figure 3 shows that the mean level of scope is lower when the risk denominator is 1,000 than when it is 10. For example, the mean level of scope for the money stolen scenario when the risk denominator is 10 is 0.82, whereas when the denominator is 1,000 the ratio is 0.50.

Table 9 shows that this difference is significant at the 99 percent level for the money stolen and stomach bug scenario and at the 95 percent level for the blind scenario. Therefore the null hypothesis that the risk denominator does not affect scope sensitivity of WTP is rejected. As the test is based on positive ranks it shows that the ratio is statistically larger for the risk in 10 than in 1,000. This result implies that respondents have more difficulty giving valid WTP responses to questions when the risk denominator is larger.

Table 9: Wilcoxon signed rank test of difference in scope sensitivity between risk denominators.

	£100div5x20 - £5div5x1to0	stom100div5x20 - stom5div5x1to0	blind100div5x20 - blind5div5x1to0
Z	-4.244(a)	-3.167(a)	-2.021(a)
Asymp. Sig. (2-tailed)	.000	.002	.043

Note: a Based on positive ranks.

It should be noted that there are two aspects of change between the ratio of the risk reduction in 10 and in 1,000. First, the denominator changes from 10 to 1,000. Secondly, while the ratio of the risk reduction remains the same, the size of the risk changes. For the risk in 10 the smallest risk reduction valued is 1/10 to 0/10. This is mathematically equivalent to the largest risk reduction in 1,000 (100/1,000 to 0/1,000). Therefore the observed reduction in scope sensitivity cannot solely be attributed to either the change in denominator or change in the size of risk.

To test the impact of either of these attributes would have involved presenting respondents with four identical valuation questions or splitting the sample further. Pre-testing revealed that those respondents who recognised that the risk reduction of 1/10 to 0/10 and 100/1,000 to 0/1,000 were mathematically the same began to fear they were being 'tricked'. Therefore it was felt that presenting respondents with further mathematically identical questions would increase the likelihood that the survey was viewed as a test. In addition, as the key question is whether respondents can give reliable WTP estimates for small risks this would necessarily involve changing both the denominator and the absolute size of the risk reduction.

Conclusions

The aim of this survey was to ascertain whether stated preference could provide a reliable means of valuing uncertain and/or unfamiliar goods. The results of this study are not very encouraging and indeed are of concern to valuing a wide range of non-market goods that may be unfamiliar to the respondent and/or where the outcome is uncertain.

The reliability of WTP estimates was measured by the sensitivity of WTP to the quantity of the risk reduction offered. Inline with many of the studies in the empirical literature (e.g. Hammitt and Graham, 1999), this experiment finds that while certain (but not all) of the stated WTP sums satisfy 'weak' tests of statistically significant scope sensitivity, the same stated preferences do not satisfy the 'strong' scope test of theoretically 'adequate' sensitive to the magnitude of the risk reduction offered.

Considering the weak tests of scope significance first; while internal (within sample) tests showed that respondents were willing to pay significantly more for a larger risk reduction, external (cross sample) tests revealed a less satisfactory picture. Although the latter tests are satisfied for risk reductions for familiar risks where those risk reductions are expressed with a denominator of 10, as we move to consider risks with denominators of 1,000 affecting less familiar goods so the 'weak' external scope sensitivity tests fail. This leads to the first causes of concern that the reliability of WTP estimates declines with the degree of familiarity and as the level of the risk denominator increases.

Indeed the degree of scope sensitivity was found to decline for goods that were less familiar. This indicates that the reliability of WTP estimates from stated preference methods is lower for goods which are less familiar. In addition, it was also found that the reliability of WTP estimates decreased as the size of the risk decreased. The difference in the level of scope sensitivity between risk reductions when presented in 10 and in 1,000 was statistically significant at the 95 percent for goods of all levels of familiarity. For the most familiar good, that of having money stolen, the ratio of WTP for the different levels of risk reduction when the risk was presented in 10 was 0.82 compared to 0.50 when the risk was presented in 1,000.

These results give weight to the argument that the scope insensitivity often observed in the literature may, at least in part, be due to unfamiliarity with the valuation scenario

and/or the level of the risk denominator[2] used to express that risk. If the valuation task is too far removed from the respondent's experience, or if the risk denominator is too large, then the cognitive burden of the task may be considerable. If this burden is too great, then respondents may be unable or unwilling to rely on underlying theoretically compatible preferences to estimate their WTP. In such situations, psychological research argues that respondents will increasingly rely upon heuristic cues (or 'rules of thumb') to construct preferences and consequent WTP responses. This may lead respondents to overlook the objective dimensions of the good being offered thus resulting in WTP being insensitive to the magnitude of the good being offered.

Widespread evidence of non-standard preference construction would call into question the use of stated preference methods for valuing non-market goods where the valuation scenario is unfamiliar and/or involves large risk denominators. The results from this study where the risks presented where in 10 and 1,000 do not bode well for the reliability of estimates when the denominator is larger. This is of particular concern in the case of valuing mortality risk reductions where the risk denominator is often 100,000 or more. Clearly if respondents are challenged by valuing changes in risks in 1,000 the possibility of eliciting reliable estimates for risk in 100,000 or more are small.

Our results suggest that before stated preference methods can be confidently used to elicit WTP estimates for unfamiliar scenarios and/or where the scenario involves small risks, particular attention should be directed to reducing the cognitive burden of these scenarios. This could arise from either improving communication of the risk and/or the valuation scenario. Psychologists and risk analysts will not be surprised by the cognitive challenges of comprehending changes in magnitude as a substantial prior literature exists. However, more research in this area with respect to stated preference methods would be welcomed. Another potentially fruitful direction of research is the exploration of communicating scenarios with the aid of new technologies. As respondents become increasingly comfortable with the use of computers this opens the possibility of interactive means of

[2] Indeed, given that many studies in the area of health economics concern risk levels expressed with denominators far greater than 1,000 then this effect may be a particularly strong factor in explaining scope insensitivity in such studies.

communication as a means of incorporating recent research on the use of visual and other stimuli as a strategy for reducing reliance upon heuristics and thereby directly addressing a major source of preference anomalies.

Acknowledgments:

Funding for this study was obtained from the Economic and Social Research Council, the CEC Sixth Framework project Valuation of Environment-Related Health Impacts: Accounting for Differences Across Age, Latency and Risk Categories with a Particular Focus on Children (VERHI) and the Economics for the Environment Consultancy (EFTEC).

References

Acton, J., 1973. "Evaluating Public Progress to Save Lives: The Case of Heart Attacks," RAND Research Report R-73-02.Santa Monica: RAND Corporation, 1973.

Ariely, D., Loewenstein, G., and Prelec, Dražen (2003) "Coherent arbitrariness: Stable demand curves without stable preferences", Quarterly Journal of Economics, 118(1): 73-105.

Arrow, K. J. (1982). "Risk Perception in Psychology and Economics." Economic Inquiry 20: 1-9.

Arrow, K., R. Solow, P. Portney, E. Leamer, R. Radner and H. Schuman. (1993). "Report of the NOAA Panel on Contingent Valuation," Federal Register, 58(10) (January 15), pp. 4601-4614.

Baron, J. (1997). "Confusion of relative and absolute risk in valuation." Journal of Risk and Uncertainty **14**(3): 301-309.

Bateman, I. J., M. Cole, P. Cooper, S. Georgiou, D. Hadley, and G. Poe. (2004). "On visible choice sets and scope sensitivity." Journal of Environmental Economics and Management **47**(1): 71-93.

Beattie, J., J. Covey, P. Dolan, L. Hopkins, M. Jones-Lee, G. Loomes, N. Pidgeon, A. Robinson, and A. Spencer. (1998). "On the contingent valuation of safety and the safety of contingent valuation: Part I - Caveat investigator." Journal of Risk and Uncertainty **17**(1): 5-25.

Blaug, M. (1997) *Economic Theory in Retrospect*, Cambridge University Press, Cambridge.

Boadway R., and N. Bruce, 1984, Welfare Economics, Basil Blackwell, Oxford.

Braga, J. and C. Starmer (2005). "Preference anomalies, preference elicitation and the discovered preference hypothesis." Environmental & Resource Economics 32: 55-89.

Carson, R. T. (1997). "Contingent valuation surveys and tests of insensitivity to scope." in R.J. Kopp, W. Pommerhene, and N. Schwartz, eds., *Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy Relevant Aspects of Contingent Valuation Methods*. Boston, Kluwer, pp. 127-163.

Carson, R., T. Groves and M. Machina, 1999, Incentive and informational properties of preference questions, Plenary Address, *Ninth Annual Conference of the European Association of Environmental and Resource Economists* (EAERE), Oslo, Norway, June 1999.

Corso, P. S., J. K. Hammitt, and J. D. Graham. (2001). "Valuing mortality-risk reduction: Using visual aids to improve the validity of contingent valuation." Journal of Risk and Uncertainty. 23(2): 165-184.

Diamond, P. and J. Hausman, (1994) "Contingent Valuation: Is Some Number better than No Number?" The Journal of Economic Perspectives, 8(4): 45-64.

Guria, J., J. Leung, M. Jones-Lee, and G. Loomes. (2005). "The willingness to accept value of statistical life relative to the willingness to pay value: Evidence and policy implications." Environmental & Resource Economics **32**(1): 113-127.

Hammitt, J. K. and J. D. Graham. (1999). "Willingness to pay for health protection: Inadequate sensitivity to probability?" Journal of Risk and Uncertainty **18**(1): 33-62.

Hammitt, J. K. (2003). "Evaluating Contingent-Valuation Estimates of the Value of Reducing Environmental Health Risks: The Proportionality Test," Chapter 15 in D.W. Pearce, C. Santoro, and C. Palmer (eds.), Epidemiology and Air Pollution, Edward Elgar Publishing Ltd., Cheltenham, UK.

Hanemann, M. (1999), 'The economic theory of WTP and WTA', in I Bateman and K Willis (eds), *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU and Developing Countries*, Oxford: Oxford University Press, pp. 42-96.

Hibbard J., P. Slovic, E. Peters, and M. Finucane. (2002) "Strategies for Reporting Health Plan Performance Information to Consumers: Evidence from Controlled Studies". Health Services Research, 37(2): 291-313.

Hibbard J. and E. Peters. (2003) “Supporting Informed Consumer Health Care Decisions: Data presentation approaches that facilitate the use of information in choice”. Annual Review of Public Health. Vol. 24: 413-433

Hsee, C. K. (1996). "The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives." Organizational Behavior and Human Decision Processes 67: 247-257.

Hsee, C. K. (1998). "Less is better: When low-value options are valued more highly than high-value options." Journal of Behavioral Decision Making 11: 107-121.

Jones-Lee, M. W., (1976). “The Value of Life: An Economic Analysis”. Martin Robertson, London.

Jones-Lee, M. W. (1989). The Empirical Estimation of Individual Valuation of Safety: Results of a National Sample Survey, *The Economics of Safety and Physical Risk*, Oxford.

Lipkus, I. and J. Hollands (1999) “The Visual Communication of Risk”, Journal of the National Cancer Institute Monographs, 25: 149-163

Kahneman, D., I. Ritov, and D. Schkade, (1999). “Economic preferences or attitude expressions? An analysis of dollar responses to public issues”. Journal of Risk and Uncertainty, 19, 220-242.

Kahneman, D. and A. Tversky (1979). "Prospect Theory - Analysis of Decision under Risk." Econometrica **47**(2): 263-291.

Liu, J. T., J. K. Hammitt, Wang, J. D., and M. Tsou. (2005). "Valuation of the risk of SARS in Taiwan." Health Economics **14**(1): 83-91.

Norinder, A., K. Hjalte, and U. Persson. (2001). "Scope and scale insensitivities in a contingent valuation study of risk reductions." Health Policy **57**(2): 141-153.

Olsen, J. A., C. Donaldson, and J. Pereira. (2004). "The insensitivity of 'willingness-to-pay' to the size of the good: New evidence for health care." Journal of Economic Psychology **25**(4): 445-460.

Peters, E., P. Slovic and J. Hibbard (2005a) “Bringing Meaning to Numbers: Evaluability and Affect in Choice”, University of Oregon, Eugene, Oregon.

Peters, E., N. Dieckmann, D. Västfjäll and C. Mertz (2005b) “When five out of four people have trouble with fractions and other numbers: Numeracy and mood in decisions”, University of Oregon, Eugene, Oregon.

Plott, C., 1996, “Rational Individual Behavior in Markets and Social Choice Processes: The Discovered Preference Hypothesis,” in *Rational Foundations of Economic Behavior*.

K. Arrow, E. Colombatto, M. Perleman, and C. Schmidt, eds. London: Macmillan and NY: St. Martin's, pp. 225–50.

Slovic, P. 1995, "The construction of preferences", American Psychologist, 50: 364-371.

Slovic, P., M. Finucane, E. Peters, and D. MacGregor (2002). The affect heuristic. In T. Gilovich & D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397-420). New York: Cambridge University Press

Slovic, P., M. Finucane, E. Peters, and D. MacGregor (2004) "Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality", Risk Analysis, 24(2): 311-322.

Smith, V. K. and L. Osborne (1996). "Do contingent valuation estimates pass a "scope" test? A meta-analysis." Journal of Environmental Economics and Management **31**(3): 287-301.

Tversky, A. and D. Kahneman, (1973) "Availability: a heuristic for judging frequency and probability", Cognitive Psychology, 5, 207-232.

Tversky, A. and D. Kahneman (1974) "Judgement under uncertainty: Heuristics and biases", Science, 185: 1124-1131.

Tversky, A. and D. Kahneman, (1991) "Loss Aversion in Riskless Choice: A Reference-Dependent Model", Quarterly Journal of Economics, 106(4): 1039-1061.

Veisten, K., H. F. Hoen, S. Navrud, and J. Strand. (2004). "Scope insensitivity in contingent valuation of complex environmental amenities." Journal of Environmental Management **73**(4): 317-331.