When is some number really better than no number? On the optimal choice between non-market valuation methods

Munro, Alistair

National Graduate Institute for Policy Studies, Japan

20 September 2007

When is some number really better than no number? On the optimal choice between non-market valuation methods.

Alistair Munro
National Graduate Institute for Policy Studies, Tokyo, Japan.

Abstract

Decision-makers have a wide variety of competing and complementary methods for non-market valuation, but there is little formal advice on the choice of method. I offer a formal approach, using a loss function (the mean square error) to compare contingent valuation, Citizens' Jury and methods where by intention only a portion of total value is estimated, when a) preferences vary across the population and b) methods are more or less susceptible to framing effects. Illustrative simulations suggest conditions under which the Citizens' Jury may dominate contingent valuation when framing effects are significant.

1 Introduction.

There are many different means by which non-market decisions can be formalised. Some of the approaches regularly employed such as contingent valuation and the travel cost method share a common underlying philosophy (e.g. welfarism), but clearly some do not. For instance, advocates of Citizens' Juries (e.g. Crosby, 1991) or consensus conferences, attach as much value to the process by which decisions are made as to whether the conclusions are reflective of preferences. On the other hand, typical cost-benefit techniques such as hedonic pricing or contingent valuation are almost exclusively concerned with the outcome (preference satisfaction) rather than the process.

Although there is some discussion of the issue, there has been relatively little consensus on which methods are most appropriate or why one might choose a method that viewed from the underlying normative principles appears dominated by another method.[1] For instance, within non-market valuation it is generally only stated preference techniques such as contingent valuation which can produce estimates for total values, including the values of non-users. In theory therefore, revealed preference methods, such as the travel cost approach

[1] For instance, the standard guidance for cost-benefit analysis in the UK is provided by the Treasury's 'Green Book' which states only that, 'The technique chosen will depend on the individual circumstances, and should be judged on a case-by-case basis.'

or dose response methods are inferior because they cannot estimate benefits to non-users. Nevertheless these techniques are alive and well and frequently chosen in preference to contingent valuation, raising the question of why one might choose such methods, or more generally what is the optimal valuation technique.

The basic idea of the paper starts from the observation that most decision processes place some weight on individual preferences. The problem is that expressions of preference are potentially susceptible to framing effects, where by a framing effect, I mean that elicited values may vary according to aspects of the choices faced by consumers that, in standard consumer theory, should have no effect.

Typically the reliability of data on preference becomes weaker the further decisions are removed from the realm of actual and repeated choice (List 2002)[2]. In choosing a process for producing information on preferences, the decision-maker therefore faces choices between methods which are more prone to framing effects and methods, which might be less prone to anomalies, but which capture only part of the value placed by the population on a potential benefit. For instance, regular users of a threatened wilderness may have much more robust and precise preferences towards its conservation, compared to non-users, but nevertheless non-users may still place some value on the wilderness. One way to conceptualise the dilemma is in the form of a trade-off between the variance of an estimate of value and its bias: for instance regular wilderness users may provide lower variance estimates of value, compared to values obtained from all citizens, but by design an estimate based only one the values of one section of the population will be biased. It is this trade-off that is analysed here.

The trade-offs between closely related methods may be relatively straightforward to analyse because of the shared normative principles, but as I have already remarked many alternative decision-making frameworks differ quite fundamentally in their underlying rationale. One way to deal with conflicting views on the appropriate decision process is to try to judge processes according to a common set of criteria, even if the original normative motives for the processes differ. If, for instance, the strong sustainability criterion leads to higher preference satisfaction than say contingent valuation (which rests on a goal of preference satisfaction), then on purely instrumental grounds one can judge strong sustainability as superior to contingent valuation without having to debate the relative merits of the underlying philosophies. Thus we may be able to produce at least an incomplete ordering of decision processes even in the absence of agreement about deeper philosophies about how decisions should be made.[3]

[2] "An obvious hypothesis, for which there is some support, is that the more ambiguity in one's preferences... the more one's expressed preferences will be subject to procedural and descriptive influences". (Schkade and Payne, 1994, p. 105).

[3] In this paper I take it as a given that the decision-maker wishes to know something about a parameter that is related to preference. There are various responses one can make to the evidence on framing and anomalies (Sugden 2005). One option is to dismiss the value of preference elicitation exercises (Diamond and Hausman 1994) for public policy decisions. The fundamental perspective in this paper is more pragmatic: preference information may be noisy and unreliable, but unless it is infinitely susceptible to framing effects, it may provide useful

Figure 1 summarises many of the alternative means by which decisions can be made in an environmental context. As can be seen, methods vary to the extent that they are base themselves on consumer preferences and on the extent to which an attempt is made to estimate the a total figure for valuation. Strong sustainability for instance, gives little weight to preference. Meanwhile, contingent valuation does attempt to estimate a total value, but many other methods such as the travel cost technique, dose response and hedonic pricing are partial valuation methods, in that by design they omit some aspects of valuation, such as non-use value. In what follows I compare different processes according to the degree that they elicit a reliable figure for the Hicksian consumer surplus. Other criteria could be used, but given consumer surplus is at the heart of the cost-benefit approach and its associated controversies, it makes sense to use it as the vehicle for comparing decision processes. In order to motivate the comparison further, I also use three prototypical processes that have widespread use and advocacy.

1. Contingent valuation
2. Partial valuation.
3. Citizen's Jury, an example of deliberative methods.

Contingent valuation is perhaps the most widely used non-market valuation method and possibly the most controversial. (e.g. Diamond and Hausman, 1994). Meanwhile, partial valuation methods (see Figure 1) are processes that by intent do not seek to capture all of consumer surplus. They might do this in one of three ways:

1. Restrictions on subjects. The researcher may for instance only gather preference data on actual users of a resource. This is typically the case with techniques such as the travel cost method.
2. Restrictions on aspects of valuation. The user/non-user distinction is fairly obvious, but a process might also seek to identify only some part of use value and ignore other contributions to welfare. For instance, the damage done by pollutants might be estimated purely from data on mortality, ignoring morbidity, the damage to plant life animals and the built environment, as well as possibly harder to measure costs such as the reduced amenity value from poor visibility.
3. Restrictions on preference variability. Some valuation methods, such as the QALYS used in health evaluation can only be formally reconciled with preference theory if preferences fit a particular pattern. One argument for such an approach is that, by imposing this restriction on elicitation procedures, researchers simplify the cognitive requirements for respondents.

guidance to decision-makers.

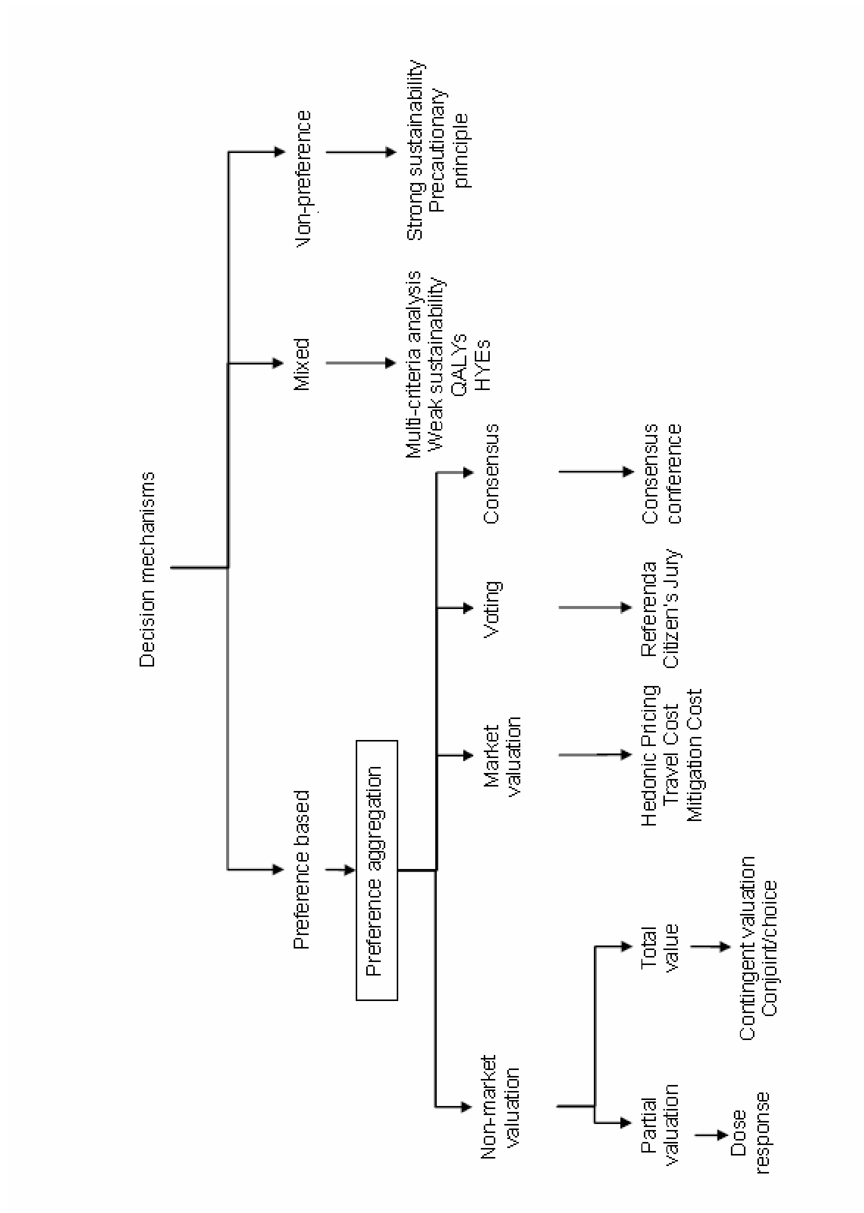

Figure 1: Decision Methods.

Finally, a Citizens' Jury (Crosby 1991) is an example of a deliberative method (Bohman and Rehg 1997), a process where citizens are actively and publicly involved in decision-making in a manner that encourages reason and argument. The jury method involves placing a small group of ordinary people together over a period of hours or days and asking them to reflect upon a policy issue. Typically, the jurists receive written and oral information from experts and get the chance to interrogate witnesses about the basis of their evidence.

Citizens' Juries originate in the USA (Crosby 1991), but have since spread to many parts of the globe (Coote and Lenaghan, 1997) and used in environmental decision-making (Aldred and Jacobs 2000). Taken-up by public agencies as well as NGOs much of the justification for the use of juries arises from the notion that in a functioning democracy a good decision is defined by the process by which the decision is made as much as by the match between means and ends created by the final choice. However, at least one potentially important justification of Citizens' Juries is that they enable information to be transmitted and understood in such a way that judgement biases are eliminated. Blamey et al, 2002, for instance argue that deliberative methods are less likely to provoke yea-saying, strategic behaviour, protest voting and biases created by social desirability. Meanwhile Kenyon et al, 2003, in a discussion of two juries on environmental risks in Scotland state that 'citizens' jury process helps the participants to construct their preferences in a rational and transparent manner' p. 223. It follows therefore that Citizens' Juries are at least potential alternatives to methods such as contingent valuation even when viewed through the lens of welfarist principles (Wilson and Howarth 2002).

The plan for the remainder of the paper is as follows. In the next section I outline the basic methodology of comparison and illustrate the problem of comparing decision processes using the example of elicitation effects in contingent valuation. Following that some simple formulae are derived for comparing the performance of different processes. These formulae are then used to carry out some illustrative calculations. In the final section, a summary is offered, along with some thoughts on the implications of the paper.

2 A Formal Approach to Optimal Decisions.

Consider a decision-maker who must estimate an unknown parameter, μ, using some estimator $\widehat{\mu}$. Acting on $\widehat{\mu}$ rather than μ produces a potential loss, L, for the decision maker in the sense that a less than optimal decision may be made. The loss is zero if $\mu = \widehat{\mu}$, but otherwise positive and increasing in the difference between $\widehat{\mu}$ and μ. A conventional model for the loss function is that the costs of a mistaken decision are proportional to the square of the difference between the true value and the estimate. That is,

$$Loss = aE\left[(\widehat{\mu} - \mu)^2\right] = a.var\left(\widehat{\mu}\right) + a\left(bias\left(\widehat{\mu}\right)\right)^2 = aMSE\left(\widehat{\mu}\right) \qquad (1)$$

Where $var(.)$ is the variance, $bias(\widehat{\mu}) = E\left[(\widehat{\mu} - \mu)\right]$, $a > 0$ and MSE is the mean squared error. In practice loss functions may not be symmetric, convex or

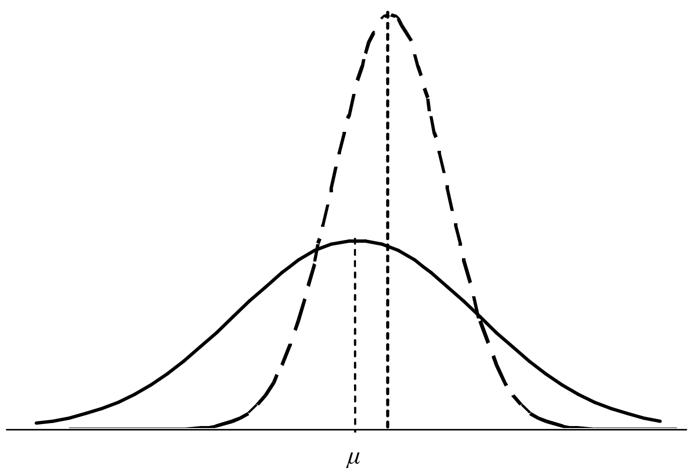

Figure 2: Estimators

continuous in the manner implied by the quadratic formulation, but if costs do have this form, it follows that if the optimal estimator minimizes the expected loss from the decision, then it also minimizes the mean square error. Consequently, the optimal estimator is not necessarily unbiased. Figure 2 illustrates the issue, showing two probability distributions for estimators: one estimator (shown by the solid line) is unbiased, but because it has a larger variance, the probability that is closer to its true value is higher with the alternative estimator (shown by the broken line), despite the fact that the latter is biased.

When the parameter estimate is an aggregate measure of preferences, estimated from data obtained from a sample of the population affected by a project, variance in can arise in two ways: first, through a combination of sampling and variability of the underlying preference parameter in the population. Secondly, variance can arise through *framing variation*.

As defined here, we can think of frame variation arising from two distinct sources. On the one hand the experimenter typically does not have the time or the resources to examine all aspects of his or her elicitation method (Smith, Desvousges and Fisher 1986). Many features of the design will be pre-tested through piloting and focus groups and the past experience of other researchers provides a stock of guidance, but there may also be aspects of the design (or the interaction of the design and the environment in which it is used) which can elude the researcher. The variables which determine values are almost countless: particular payment vehicles may be more or less acceptable; data and questions early in the survey might provide anchors for later valuation questions; the attitude and dress of interviewers can also influence subjects. Thus even in the best of surveys, there will be some residual uncertainty about the exact frame elicited by the researcher.

Secondly, recall that the researcher is attempting to identify a true valuation of the object. In order to do so, there must be a model of the relationship between the frame and the true valuation of the resource. Running a revealed preference study in parallel with the hypothetical exercise can give insight into the correct frame, but the nature of non-market valuation is that this is impractical for many goods. Without revealed preference behaviour therefore, the best frame represents something of a guess (albeit calculated) on the part of the researcher and there may be conflicting arguments what frame should be used. For instance, suppose that a proposed project involves environmental degradation of some kind, so that viewed from a property rights perspective willingness to accept compensation (WTA) is the appropriate measure of lost value. But as is widely noted in the contingent valuation literature (Mitchell and Carson 1989), WTA appears more prone to producing unreliable estimates of value, compared to willingness to pay (WTP) and in cases where through repetition and learning the WTA-WTP gap is closed, it appears that most of the change occurs in the WTA estimate (Shogren, Shin, Hayes and Kliebenstein 1994). This suggests that the figures obtained from a one-shot WTP measure may be a better estimate of true WTA than a one-shot WTA figure. With arguments on both sides the researcher may consider both WTA and WTP acceptable frames for eliciting preferences over the environmental degradation.

2.1 Elicitation effects.

Since the notion of frame variance is at the heart of this paper it is worthwhile illustrating the problem with an example. Elicitation effects occur when responses gathered from subjects are sensitive to the method of elicitation in a manner inconsistent with standard, Hicksian consumer theory. Two widely reported elicitation effects are starting point effects and yea-saying. Starting point effects (SPE) occur when reported valuations are correlated with some initial valuation cue, such as the bid value in dichotomous choice (DC) questions. Yea-saying describes the phenomenon of a subject agreeing to a proposal in the form of a direct question that she or he would reject under other conditions. For instance, a subject may agree to a bid price in a dichotomous choice format but then provide a lower stated valuation in a subsequent valuation exercise. A key difference between the two elicitation effects is that yea-saying is a unidirectional phenomenon, i.e., it raises willingness to pay or reduces willingness to accept whereas starting point bias can work in either direction depending on the value of the cue.

Starting point effects could be caused by anchoring, which, in the context of valuation, occurs when an individual's reported or revealed valuation is correlated with some prior numerical cue. Since its preliminary identification by Slovic and Lictenstein, 1971 manifestations of anchoring have been identified in numerous and diverse settings including the guessing of answers to multiplication problems and estimating the number of African countries in the United Nations (Kahneman, Slovic and Tversky 1982). A particularly stark example of anchoring can be found in the recent work of Ariely et al, 2003, who asked subjects for the final two digits of their US social security number and found that it was closely correlated with individuals' subsequent valuations of a variety of goods. Anchoring is therefore one possible reason for SPE, with the anchor provided by the initial value of x offered to subjects. The initial bid value might act as a clue or hint towards the good's value, especially when respondents are confused or unfamiliar with the good concerned (Bishop et al, 1983, McFadden, 1999, Brown et al, 1996). Since the domain of CVM often involves the valuation of unfamiliar, non-marketed goods, this starting point problem has become recognised as a potentially serious flaw inherent in iterative bidding techniques. (Boyle et al, 1996, p.193).

The phenomenon of yea-saying, well documented in the psychology literature (Arndt and Crane, 1975 and Couch and Keniston, 1960), has also been proposed in the CVM context (Kristrom, 1993, Kanninen, 1995, Brown et al, 1996) as a possible influence on DC responses. Brown et al (1996) argued that the simplicity of the take-it-or-leave-it choice might generate a conflicting objective in respondents. Torn between answering truthfully and showing a positive preference, if a DC bid is above her/his maximum WTP, a subject may still respond positively because s/he would like to demonstrate a positive preference for the good in question. In addition to this, we might include the notion of the good respondent (Orne 1962). Orne described how subjects, when faced with officialdom, might respond positively to questions, only because they wrongly believe that such a response is exactly what the interviewer (in a position of perceived authority) wishes to hear.

One point to emphasise is that starting point effects are not necessarily confined to hypothetical choices concerning unfamiliar goods. Frykblom and Shogren, 2000, for instance, use real choices and a split-sample design to value an environmental economics text using 108 Swedish university students. One treatment undergoes a Vickrey auction while the others face dichotomous choice (DC) questions set at a variety of bid levels. The authors argue that both yea-saying and anchoring will increase the acceptance of the proposal at high bid levels, while the two effects work in opposite directions for low bid levels. Hence it is possible to test between the impact of these two effects by comparing the distribution of values derived from the auction with the upper and lower parts of the distribution derived from the DC exercise. On the basis of their results, they conclude that yea-saying is present, but not anchoring.

Alternatively, one could design an experiment in which for some subjects, after the DC questions there is an incentive compatible open-ended (OE) valuation question. If only anchoring occurs the values derived from open-ended questions should be consistent with the values from the DC questions, but if only yea-saying is present then the distribution of values derived from the OE questions should be independent of the bid level in the DC question and equal to the distribution obtained from subjects who face an open-ended question

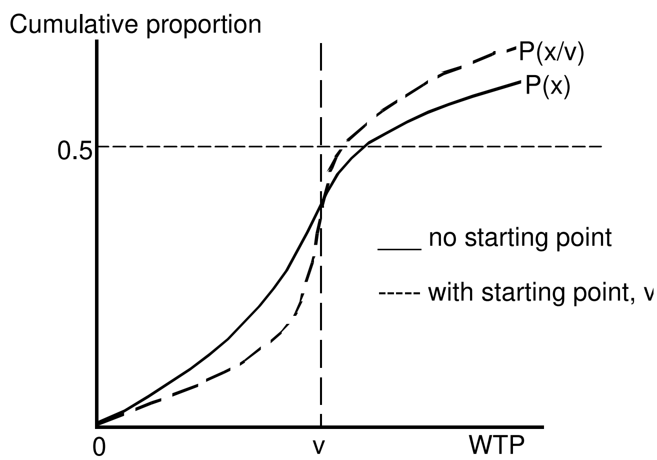

Figure 3: The Effect of a Staring Point on Valuations.

without a prior DC question. This provides a clear cut means of distinguishing between anchoring and yea-saying. Bateman et al, 2006, use this method, with students as subjects and teabags and wine as the commodities As the Figure 4 suggests, they conclude in favour of yea-saying as the cause of the results. This diagram shows none of the properties of the stylised anchoring effect shown in Figure 3 . Instead, the cumulative percentages bidding at each level (in UK pounds) appear highly similar for all three treatments. Such results are at variance with the work of Ariely et al, 2003, who admittedly use less familiar goods. On the other hand the Bateman et al results are consistent with Frykblom and Shogren, 2000 and with the recent evidence on US consumption presented by van Soest and Hurd, 2003. Partly the difference may lie in the goods involved. In Bateman et al 2006 experiment and in the last two studies cited, familiar goods were the objects of valuation, whereas anchoring effects seemed to have been found most clearly when subjects were facing novel or unfamiliar valuation tasks, which is often the case in environmental valuation.

The deeper conclusion from this literature is that even in an incentive compatible environment with well-trained and experienced subjects, with real goods and simple choices, different frames produce different valuations. Thus the researcher cannot be entirely certain about which method of elicitation is superior.

2.2 Gold standard frames.

There are a number of possible responses to framing effects of the kind illustrated above. When Diamond and Hausman, 1994, asked 'is some number better

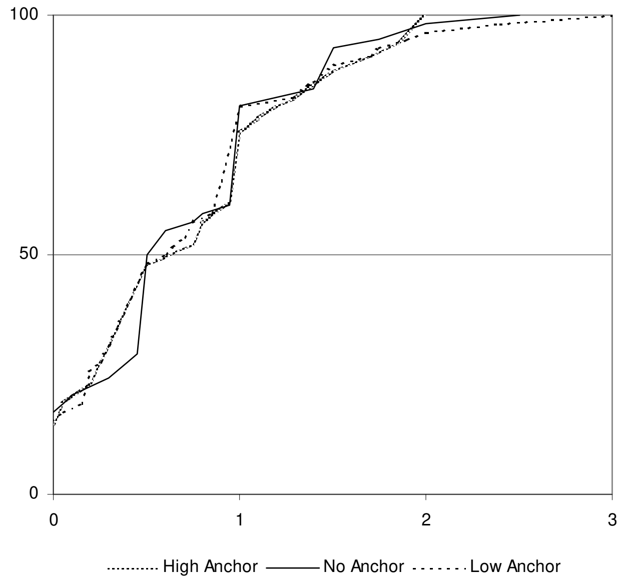

Figure 4: WTP for Teabags

than no number' they responded that very often with stated preference it was better to have no number. But possibly a more common response has been to advocate a golden standard for preference elicitation and by doing so to eliminate the frame variance. A gold standard or optimal frame is therefore one where the manner in which preference information is elicited is specified precisely and which passes all tests of validity. Methods which meet the gold standard produce acceptable information, but methods which do not meet the standard require the rejection of information or at least a significant reduction in its value.

For economists, especially those involved in environmental valuation, guidance on what constitutes an optimal or gold standard frame is available in large quantities. Within the contingent valuation approach, Mitchell and Carson's 1989 comprehensive book is probably still the dominant source of framing advice, supplemented by the guidelines produced by the 1992 NOAA panel under the chairmanship of Kenneth Arrow (Arrow et al, 1993) which produced what it called "a fairly complete set of guidelines compliance with which would define an ideal CV survey". For instance there is a general requirement that,

> in the absence of a set of reliable reference surveys, the burden of proof shall remain on the researchers. A survey should be judged unreliable to the extent it exhibited the following defects:a high non-response rate, inadequate responsiveness to scope, lack of

understanding of the task by respondents, lack of credibility of the restoration scenario". (Arrow et al, 1993, p. 37)

More recent advice has been put forward by Carson, 1998, 2000, amongst others and there are a number of environmental valuation manuals being produced or available for government economists.(e.g. Bateman et al, 2002).

Richard Carson, 2000 is a clear and thoughtful summary of some general principles in the wake of the NOAA panel. For him, fundamentally a well-conducted CV exercise must have three attributes: the good to be valued must be well-defined; there must be plausible means of provision and there must be a plausible means of making trade-offs. A good that is vaguely defined by the surveyors may be understood in widely differing ways by subjects in the survey. To aid the credibility of the exercise with regard to the payment vehicle, Carson recommends coercion where this is credible. Ideally, the vehicle should also appear reasonably fair, as well as plausible and understandable to the respondents in the survey. In addition to these features of the survey frame, the sampling process should be carefully designed so that population estimates of value can be derived; statistical methods should allow for the possibility of zero willingness to pay values and the whole procedure should be transparent.

Transparency is also a feature advocated by the practitioners who argue that preferences are not always in place ready for the survey analyst, but must be constructed by the individual respondent in the course of the valuation exercise. In other ways the recommended methodology is fundamentally different. Gregory and Slovic, 1997, for instance argue that

> the analyst therefore functions as an architect, helping respondents build their values from simpler pieces rather than (following the economics model) as an archaeologist whose task is to uncover values presumed to exist '1997:177.

Typically, options differ along many dimensions. For some dimensions, such as environmental damage or risks to health, subjects may be unused to expressing values in terms of a single metric such as money. They may also be cognitively overloaded if faced with choices between two complex options which differ in many important ways. Consequently, they argue that subjects should be first asked to compare alternatives along single dimensions, using valuation scales natural to the dimension. Only later should subjects be asked to make trade-offs between different dimensions and invited to attach weights. This they argue, limits the cognitive overload that might otherwise hinder subject's ability to choose between alternatives.

Unlike the typical CV instrument which seeks to gain values from a representative sample of the population, Gregory and Slovic's constructive approach uses only small numbers of key stakeholders. They recommend a five-step approach to valuation in which subjects are given constructive support in their decision-making. This approach differs in a number of important ways from the standard cv method, but whichever way it is approached, there are two fundamental problems with the golden frame perspective on valuation. First the guidance may not be exact enough, in the sense that the researcher however well-funded is not in a position to control exactly the perception of the survey and its purpose in all respondents. The impact of changes in the way information is presented visually, or in elicitation formats, or in the order in which questions are asked may all be tested by the researcher, but some aspect of the survey design or its implementation is bound to escape the attention of the researcher or may interact with the target sample's experience in an unanticipated manner. In other words there is always some residual variance in the frame elicited in the mind of the subject and this is one source of the frame variance discussed above.

Secondly, in drawing up the definition of the gold frame, there is often no definitive reason for favouring a frame with aspect x rather than aspect y. We see this in the anchoring and yea-saying experiments discussed above, where even with real goods and real choices changes in the way in which preference information is elicited can affect values. In the case of contingent valuation, an experienced practitioner claims that "Each of the three main response formats has strengths and weaknesses" (Boyle, 2003 p. 137,) referring to the choice between open-ended, dichotomous choice and payment cards in a recent guide to the contingent valuation method. The dichotomous choice elicitation method is often recommended (e.g. by the NOAA panel,), largely on the grounds of its perceived incentive compatibility. However, the bid levels provided to subjects in DC may provoke starting point effects and therefore responses at variance to individuals' true preferences (if they have them). Dichotomous choice is also inefficient in that it produces very little information per subject. In response to the anchoring problem, one might use payment cards - which is where a subject sees a card or screen with a set of bid levels and is asked for his or her preferred option. In response to the inefficiency of the DC method, there has been a recent trend towards conjoint methods, in which subjects face a sequence of choices, often between more than two options. Both the alternative elicitation formats lack the clear incentives for truth-telling provided by the DC method, but they clearly have other advantages and it is not obvious where the balance lies. This then is the second source of frame variance: the set of acceptable frames contains more than one element.

3 Formal comparison.

Recall, that processes are compared according to their mean squared error, that is the sum of the variance and the square of the bias.. Suppose that there are F acceptable frames (with F>1) from within which one frame is selected at random.[4] The estimator is a random variable, $\widehat{\mu}$, which in frame j, we label as

[4] Although this is not an issue pursued here, the astute reader will note from the following derivations that MSE may be lowered by splitting the sample across frames. Such a strategy may be feasible when it is the researcher who has full control over the frame. However when there is residual uncertainty over how subjects will interpret the valuation process, then it

$\widehat{\mu}_j,\ \ j = 1, ..., F$,with an expected value of $\overline{\overline{\mu}}_j$. within frame j and an expected value across frames of $\overline{\overline{\mu}}$. In other words,

$$\overline{\overline{\mu}} = \frac{1}{F}\sum_{j=1}^{j=F}\overline{\overline{\mu}}_j = \frac{1}{F}\sum_{j=1}^{j=F}E\left[\widehat{\mu}_j\right] \tag{2}$$

The *within frame* variance of $\widehat{\mu}_j$ is $\sigma_j^2 = E\left[\left(\widehat{\mu}_j - \overline{\overline{\mu}}_j\right)^2\right]$, with the expected value of it across frames defined as $\sigma^2 = \frac{1}{F}\sum_{j=1}^{j=F}\sigma_j^2$. . Meanwhile, the *frame variance,* denoted by σ_f^2 is the variance across the frames of the mean estimates. That is,

$$\sigma_f^2 = \frac{1}{F}\sum_{j=1}^{j=F}\left(\overline{\overline{\mu}}_j - \overline{\overline{\mu}}\right)^2 \tag{3}$$

Let the sample size used in the process be n, with individual observations indexed by i = 1,...,n. Each random variable within this sample is assumed to be i.i.d. We let $\widehat{\mu}_{ji}$ be the ith element of the sample when the frame is j. The total variance of the process, $var\left(\widehat{\mu}\right)$, is then

$$\begin{aligned}
var\left(\widehat{\mu}\right) &= E\left[\left(\sum_i\left(\frac{\widehat{\mu}_i}{n}\right) - \overline{\overline{\mu}}\right)^2\right] = E\left[\left(\sum_i\left(\frac{\widehat{\mu}_i - \overline{\overline{\mu}}}{n}\right)\right)^2\right] \\
&= E\left[\frac{1}{F}\sum_{j=1}^{j=F}\left(\sum_{i=1}^{i=n}\left(\frac{\widehat{\mu}_{ji} - \overline{\overline{\mu}}}{n}\right)\right)^2\right] \\
&= \frac{1}{F}\sum_{j=1}^{j=F}E\left[\left(\sum_i\left(\frac{\widehat{\mu}_{ji} - \overline{\overline{\mu}}_j + \overline{\overline{\mu}}_j - \overline{\overline{\mu}}}{n}\right)\right)^2\right] \\
&= \frac{1}{F}\sum_{j=1}^{j=F}\left(\frac{\sigma_j^2}{n} + \left(\overline{\overline{\mu}}_j - \overline{\overline{\mu}}\right)^2\right) \\
&= \frac{1}{F}\sum_{j=1}^{j=F}\left(\frac{\sigma_j^2}{n} + \left(\overline{\overline{\mu}}_j - \overline{\overline{\mu}}\right)^2\right) \\
&= \frac{\sigma^2}{n} + \sigma_f^2
\end{aligned} \tag{4}$$

Note that the penultimate line of this expression arises from the i.i.d. nature of the sample. Using this expression, the formula for MSE becomes:

$$MSE = \frac{\sigma^2}{n} + \sigma_f^2 + \left(\overline{\overline{\mu}} - \mu\right)^2 \tag{5}$$

may not be in the power of the researcher to choose the exact frame.

Comparing processes is then a matter of comparing the MSE for different decision-making approaches.

Now, each of the processes selected for the comparison has its strengths and weaknesses which we need to formalise. Most jury exercises begin by recruiting a 'representative' sample of individuals. To be manageable, Citizens' Jury have to be small, so there is a potential problem of high variance due to the small sample of individuals involved. For instance, the Jefferson Institute (which trade-marked the term Citizens' Jury in the USA, but not elsewhere) argue in favour of a jury of 18 people (Crosby 1991). Unless preference variability across the population is extremely small, sampling variability with deliberative processes is likely to be large. In the guidelines for the original Jefferson Centre exercise, one of the essential elements was random sampling for members of the jury. [5] Let n be the size of the Citizens' Jury. When the jury convenes, n individuals arrive at a collective value for the mean value. Suppose that, by conversation, they arrive at the correct frame. I shall call this an *ideal* Citizens' Jury (ICJ). It has already been stressed that Citizens' Juries are most often justified in terms other than preference satisfaction, so the term 'ideal' here simply refers to the lack of bias in the outcome, without prejudging any other features of the process. Further, suppose that the mean value produced by the jury is the mean of their individual values.[6] If the n members are chosen randomly, then the expected value obtained from this process is also μ.

With contingent valuation on the other hand judgement biases are not necessarily eliminated within any given frame, but samples are potentially large enough to limit sampling variability and the set of acceptable frames is ideally chosen so that there is no ex ante bias.

For the PV method the benefits can be decomposed. For the purposes of the exercise it does not matter what these components are, only that they can be conceptually separated in what follows. Let $\mu = \mu_e + \mu_n$ where the labels of the components are chosen to indicate 'estimated' and 'non-estimated' values respectively. For the partial valuation method only μ_e is estimated. Ideally, a partial valuation method may be chosen because the part of the benefits estimated are not subject to significant framing effects[7] - in other words the frame variance is zero, but by definition it will omit the values of some users or uses or force data on preferences into a very specific functional form. As a result we can expect systematic bias. However, as with contingent valuation, the

[5] In many actual CJ exercises, the process of juror recruitment has been rather less formal or different criteria have been used . In some cases volunteers are recruited by word of mouth or advertisement. A typical worry is that this slants the jury and therefore its verdict. We might suppose that individuals at one extreme or other of the distribution of values would be attracted to serve. A symmetric distribution of values might then conceivably still generate an unbiased estimate of the population mean value, with the extremes 'cancelling' one another out. With skewed distributions such a comforting hypothesis seems less reasonable.

[6] In an original paper, Howarth and Wilson, 2006, put forward a formal argument for why the willingness to pay of a group engaged in deliberative valuation should be less than the average willingness to pay of individual members. In other words, deliberative valuation induces its own bias in valuation.

[7] This is the underlying claim about the superior value of revealed preference methods.

sample may be chosen to be sufficiently large and random to all-but eliminate sampling variation.

Formally therefore, I define each of the methods to be compared as follows:

Definition 1 *Contingent valuation (CV):* $n = \infty$, $\overline{\overline{\mu}} - \mu = 0$.

Definition 2 *Ideal Citizens' Jury (ICJ):* $\overline{\overline{\mu}} - \mu = 0$; $\sigma_f^2 = 0$.

Definition 3 *Partial valuation (PV):* $n = \infty$, $\overline{\overline{\mu}} - \mu = -\overline{\mu}_n < 0$; $\sigma_f^2 = 0$.

Given these formalisations we therefore have:

$$\begin{aligned} MSE_{CV} &= \sigma_f^2 \\ MSE_{ICJ} &= \frac{\sigma^2}{n} \\ MSE_{PV} &= \left(\overline{\overline{\mu}} - \mu\right)^2 = \bar{\mu}_n^2. \end{aligned} \tag{6}$$

The optimal decision process is then the one with the lowest MSE. [8]

Define $\alpha = \sigma_f/\mu$ as the *coefficient of frame variation*, let $\beta = \sigma/\sqrt{n}\mu$ be the *coefficient of sample variation* with $B = \sigma/\mu$ as the *coefficient of population variation* and let s $= \bar{\mu}_n/\mu$ be the share of total benefits omitted by the partial valuation method. Then we can easily manipulate 6 to show that CV is as least as good as PV when,

$$\alpha \leq s \tag{7}$$

Meanwhile CV is at least as good as ICJ provided,

$$\alpha \leq \beta \tag{8}$$

Finally, PV is as least as good as ICJ, when

$$s \leq \beta \tag{9}$$

Figure 5 summarises the circumstances under which each decision rule dominates the other two. We have three regions of dominance. As the proportion

[8] The simplicity of the expression for MSE$_{PV}$ suggests a relatively straightforward algorithm for choosing which components of valuation to estimate when there are more than two alternatives: add a component if the square of its mean value is greater than the variance or equivalently, add if its coefficient of variation is less than 1. However, suppose for instance there are N components of total value, j are included in the set to be estimated and we wish to consider a j+1th component. In the absence of correlation between the components, the change in the loss function from adding the j+1th component is $\overline{\mu}_{j+1}[-\overline{\mu}_{j+1} - 2(\overline{\mu} - \overline{\mu}_{J+1})] + \sigma_{j+1}^2$ where $\overline{\mu}_{j+1}$ is the mean value of the j+1th component, $\overline{\mu}$ is the mean of total value, $\overline{\mu}_{J+1}$ is the sum of the mean values of components 1 to J+1 and σ_{j+1}^2 is the variance of the j+1th component. It follows that, compared to the simple expression, this more general expression gives more weight to reducing the bias compared to raising the variance. Thus a coefficient of variation less than one is a sufficient condition for a component to be added, but not necessary (except when it is the final component to be considered).

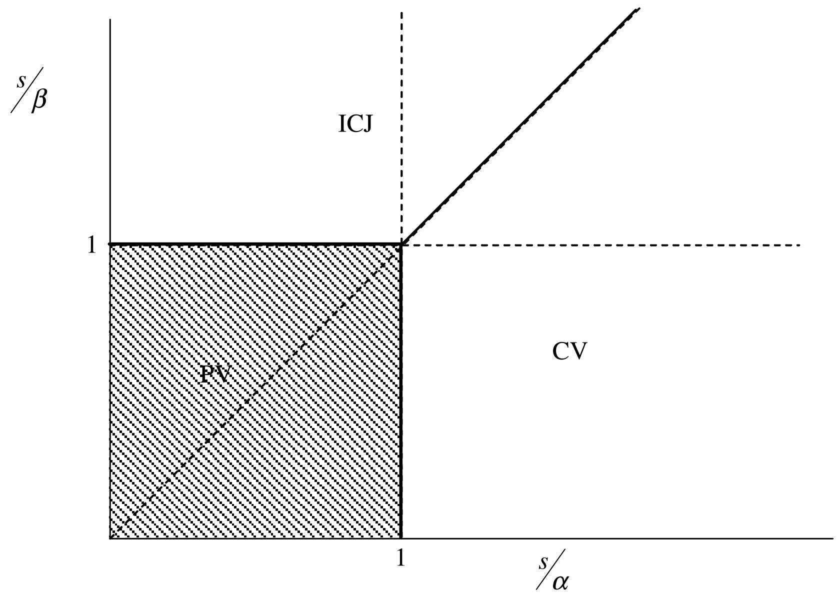

Figure 5: A Comparison of Three Methods

of total benefits omitted by the PV estimation technique becomes smaller, the PV method raises its advantage compared to the other two rules. Once the coefficient of frame variation is larger than s, then PV produces a lower MSE than contingent valuation. Similarly if the coefficient of sample variation is large compared to s, then PV is superior to the ideal Citizens' Jury. For the comparison between ICJ and CV, s is irrelevant – what matters is the relative size of the coefficients of variation. If the frame coefficient is larger then ICJ is superior to CV, but if the sample coefficient is larger then CV is the superior elicitation method. Thus for the comparison between CV and ICJ three things matter: the size of the Citizens' Jury sample; the variance of tastes within the population and the sensitivity of expressions of preferences to changes in the frame.

Recall that the particularly simple formulae and diagram represent the result of some strong simplifying assumptions, particularly that ICJ and partial valuation methods do not suffer from framing variation. Adding in these elements would raise the relative advantage of the contingent valuation process. On the other hand, allowing for sampling variation in contingent valuation and partial valuation would lower their advantages relative to the ICJ.

3.1 Example 1: WTA versus WTP.[9]

To illustrate the formulae we consider two examples. The first example supposes that there are only two possible frames, namely buying and selling and therefore two possible values: willingness to pay (WTP) and willingness to accept (WTA). There is uncertainty over the correct frame for valuation purposes, not because of uncertainty over property rights and the researcher believes that WTA and WTP are equally likely to be the correct way of eliciting valuation. Let $WTA = (1+\gamma)WTP$ then the mean value across frames is $(1+0.5\gamma)WTP$. The frame variance is $\gamma^2 WTP^2/4$ so the coefficient of frame variation is $\gamma/(2+\gamma)$.

We first calculate critical values of n (the size of the jury) when the CV approach and ICJ have equal MSE. In other words,

$$n = \frac{\sigma^2}{\sigma_f^2} = \frac{(2+\gamma)^2 B^2}{\gamma^2} \tag{10}$$

To determine the range of sensible values for the coefficients in this expression, recall the experimental evidence: the ratio of mean WTA to WTP typically exceeds 1 by a significant amount and ratios of four or more are not uncommon (Horowitz and McDonnell, 2002, or Sayman and Onculer, 2005).

There is no systematic evidence on the coefficient of population variation from the CV literature, but a sample of CV studies is assembled in Table 1. Though this sample is not entirely systematic and the studies do not claim to randomly sample members of the population, it suggests that a range for B from 0.5 to 2.5 is reasonable.

As can be seen in Table 2 the critical size for the Citizens' Jury is sensitive to the value of B. While the Jefferson Institute favour a jury size of 18, other sources suggest 16-25 members as a sensible range of number for an effective process, which would imply that despite the small numbers involved, the ICJ may outperform the CV exercise when it is not clear whether the optimal frame should present the policy changes as a loss or a gain and the variance of preferences across the population is small.

The final column of the table illustrates a different comparison: between contingent valuation and partial valuation for different values of the WTA/WTP ratio. In the column, s^* is the critical share of total benefits not estimated by the partial valuation method such that $\alpha = s^*$. In the example $s^* = \gamma/(\gamma+2)$. From the column we can see that when WTA is only 50% bigger than WTP then the partial valuation method must estimate 75% of total benefits to be of equal value to the contingent valuation method. However, by the time the ratio of WTA/WTP is equal to 4, the partial valuation method can be superior even if it omits 3/5 of the benefits.

Estimates of s are hard to come by (see Table 3). Typically when contingent valuation is compared to the travel cost method, for example, both procedures

[9] It is of course well-known that WTA will exceed WTP for normal goods and that in theory, the gap between the two values can be large in the absence of close substitutes for the good being valued. In this example we suppose that is is unknown to what extent the disparity is due to framing effects.

Table 1: Evidence on the coefficient of population variation, B.

Authors	Method	Measure	Good	B
Halvorsen and Sœlensminde, 1998	OE	WTP	Reducing premature deaths	0.9
Adamowicz and Bhardawaj, 1993	OE	WTA	Movie tickets	0.92
Bateman et al, 1995	OE	WTP	Flood defences	1.70
Johannson et al, 1997	OE	WTP	Chocolates	0.84
Whittington et al, 1990	OE	WTP	Water, public standpipe	0.54
Whittington et al, 1990	OE	WTP	Water, private supply	1.32
Boyle et al, 1996	OE	WTP	Oil spill clear up	0.34
Silberman et al, 1992	OE	WTP	Beach quality (users)	1.37
Silberman et al, 1992	OE	WTP	Beach quality (never will use)	1.83
Kontoleon and Swanson, 2003	OE	WTP	Panda reserves	1.05
Thayer, 1981	Bidding game	WTP	Landscape preservation	1.08
Amigues et al, 2002	OE	WTA	Riparian habitat conservation	2.80

Notes: OE = open ended. Sample standard deviations used.

Table 2: Critical values of n for equivalence of CV and ICJ.

	Coefficient of population variation					Critical s^*
WTA/WTP ratio	0.5	1	1.5	2	2.5	
1.5	6	25	56	100	156	0.20
1.75	3	13	30	54	84	0.25
2	2	9	20	36	56	0.33
2.5	1	5	12	22	34	0.43
3	1	4	9	16	25	0.50
4	1	3	6	11	17	0.60

are partial valuations, since it is only visitors to the site who are sampled (Carson et al, 1996), not all potential beneficiaries. The early study on hunting permits by Bishop and Heberlein, 1979, also included a travel cost exercise, where for a zero cost of time, the method produced estimates of \$159,000 for total benefits. This compares to \$880,000 obtained from the real purchase of permits off the hunters (i.e. a WTA figure), suggesting a crude figure of $s = 0.82$, if we ignore possible differences between WTA and WTP. When higher values for the cost of time were used, s fell to 0.28. Meanwhile, when the zero cost of time figures were compared to hypothetical WTP elicited from a comparison group of hunters, $s = 0.46$. In the case of non-use values, Greenley et al, 1981, estimate 19-27% of benefits for improvements to water quality in a south Colorado, USA, river system represent existence value.[10] If non-use benefits were the ones omitted from a partial valuation (i.e. $s = 0.19 - 0.27$), this would suggest that partial valuation outperforms a full valuation for standard figures for the WTA/WTP ratio.

In the context of healthcare, Clarke, 2002, compares travel cost[11] and contingent valuation estimates for rural pregnancy services in Australia using a mail-based survey. He finds that estimates from the travel cost method are approximately 56% of those obtained via contingent valuation (i.e. $s = 0.44$). Meanwhile, Kennedy 2002 compares revealed preference and contingent valuation measures for radon protection and obtains a figure for the former which is 67% of the latter (i.e. $s = 0.33$). Again, making the heroic assumption that this difference is purely due to benefits omitted from the revealed preference methods it would suggest that partial valuation outperforms contingent valuation for many realistic values of the WTA/WTP ratio.

Table 3: Some Estimates for omitted benefits, s.

Study	Good	Partial Valuation	s
Bishop and Heberlein, 1979	Hunting permits	Travel cost	0.28-0.82
Greenley et al, 1981	Water quality	Use values	0.19-0.27
Clarke, 2002	Maternity services	Travel cost	0.44
Huang et al, 1997	Recreational water quality	Travel cost	0.27
Kennedy, 2002	Radon protection	Hedonic pricing	0.33

[10] Cummings and Harrison, 1995 argue forcefully that it is not possible to separate the values held by one individual into existence or non-use and use values.

[11] It is well-known that travel cost methods produce estimates for the Marshallian consumer surplus. For normal goods this underestimates the Hicksian compensating surplus. In theory if income and price elasticities are identified, the Hicksian surplus can be estimated. The point is this figure would still omit the non-use value non-users and hence even the adjusted figure from travel cost studies will be a partial valuation measure.

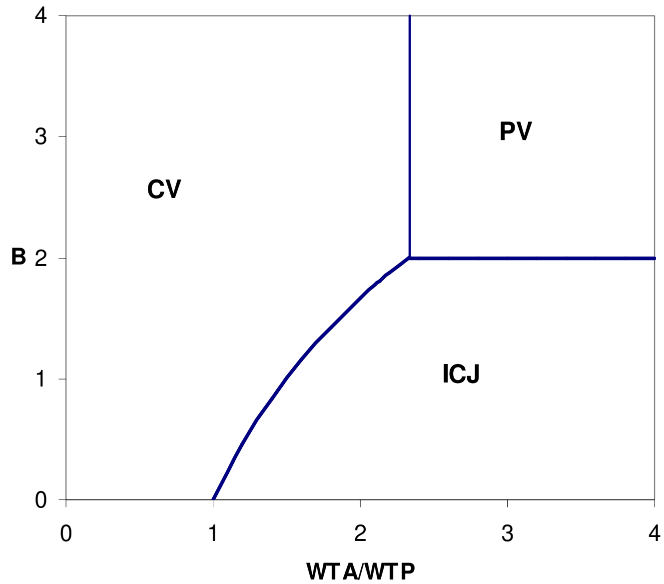

Figure 6: Optimal decisions for s=0.4 and n=25

Figure 6 offers a simultaneous comparison of all three approaches. For the purposes of this illustration, I assume that $s = 0.4$ and that a jury of up to 25 is possible - the latter is fairly generous to the ICJ method and this is reflected in the figure, where we can see that none of the approaches has a decisive advantage over the other methods for typical parameter combinations. Comparing all three approaches (CV, ICJ and PV) we can see that contingent valuation has its clearest advantage when the WTA/WTP ratio is low (i.e. close to 1), preference variation within the population is high and, at the same time partial valuation methods capture only a small percentage of the total value.[12]

It might reasonably be assumed that the sample size for any CV or PV exercise cannot be infinite. It is a straightforward matter to adjust the relevant equations for a finite sample of m:

$$\begin{aligned} MSE_{CV} &= \sigma_f^2 + \frac{\sigma^2}{m} \\ MSE_{PV} &= \bar{\mu}_n^2 + \frac{\sigma^2}{m}. \end{aligned} \tag{11}$$

[12] In the Bishop and Heberlein, 1979, article referred to above, the WTA/WTP ratio for hypothetical valuation was 4.8, giving a critical value of 66% for s. The intermediate cost-of time travel cost estimate would therefore dominate hypothetical valuation, but an ICJ would easily dominate both given a value of B $\approx$ 1 (see Li et al, 1996 for this last figure).

The borderline between PV and CV is unchanged[13]. Meanwhile the critical size of the ICJ becomes:

$$n = \frac{(2+\gamma)^2 B^2/\gamma^2}{1+(2+\gamma)^2 B^2/m\gamma^2} \tag{12}$$

The effect on the comparisons of this sections is small. For instance, if we take a representative CV sample of 400 individuals, a value of B of 4 and a value of 5 for the WTA/WTP ratio then the critical value of n is reduced by 4% compared to the case where m is infinite. This amounts to approximately 1 person for the typical critical values in Table 2.

Instead of there being sampling variation with CV, if there is framing variation of $\delta\sigma_f^2$ $(0 \leq \delta \leq 1)$ for the Citizens' Jury method, then the MSE becomes,

$$MSE_{ICJ} = \delta\sigma_f^2 + \frac{\sigma^2}{n} \tag{13}$$

The critical size of the ICJ is then,

$$n = \frac{\sigma^2}{(1-\delta)\,\sigma_f^2} = \frac{(2+\gamma)^2 B^2/\gamma^2}{1-\delta} \tag{14}$$

As a result, the critical values of n reported in Table 2 are inflated by $1/(1-\delta)$. We have no evidence on reasonable values for δ, but the key point here is that as long as δ is small, then ICJ will maintain its advantage over CV in the conditions illustrated by Table 2.

Finally, a more general expression for comparing PV with the other measures can be obtained if we suppose that the frame variance of the partial valuation method is some fraction θ of the CV frame variance. We then obtain $\alpha\sqrt{1-\theta} \leq s$ for the comparison between CV and partial valuation. To put this into some kind of perspective: if the frame variance for partial valuation is half the figure for CV, then the critical figures for s in the final column of Table 2 would be reduced by about 30%.

3.2 Example 2: Elicitation methods.

Table 3 provides an example where all the elicitation methods used produce a figure for WTP. In the well-known Bateman et al, 1995, paper a CV exercise is conducted on an aspect of the Norfolk Broads (a national park located near to Norwich, UK). Some respondents face a straightforward open ended (OE) question. Others face a dichotomous choice, with a follow-up open-ended question. In the table IB OE refers to this question; c^{***} refers to the estimates of WTP from the dichotomous choice answers based on the assumption that any values between zero and infinity are possible, while c^{**} uses the same data, but

[13] This is only true if the population preference variance is the same for partial and full valuation. If it is not then another parameter is required for the comparison of valuation methods. However, the thrust of this section is that for typicaly feasible sample sizes, sampling variation plays only a small part in the MSE for PV and CV methods.

Table 4: Mean WTP values and critical jury sizes for different acceptable sets.

	Acceptable Frames				
	All	Open ended	Un-truncated	All DC	OE after DC
OE	67	67	67	-	-
IB OE	74.9	74.9	-	-	74.9
C**	112	-	-	112	112
C***	144	-	144	144	144
Mean value across frames	99.5	71.0	105.5	128.0	109.5
Coefficient of frame variation	0.28	0.05	0.30	0.10	0.26
Critical n, B = 0.5	3	121	3	24	4
Critical n, B = 1	13	484	11	96	15
Critical n, B = 1.5	29	1089	25	216	34

assumes that no values beyond the highest bid level (£500) are credible. The mean figures from the four estimation methods form a pattern that might be anticipated: OE produces the lowest estimate, the dichotomous choice estimates are somewhat higher and the follow-up OE question produces an estimate which is lower than that from the dichotomous choice questions but possibly because of anchoring it is still higher than the basic OE estimate.

I consider different groups of acceptable frames. For the first one, the prior is that all frames appear equally credible; in the remaining four cases only two of the elicitation methods are possible candidates for being the elicitors of the true value. One group consists of just the open-ended WTP figures; one eliminates the datasets where there is some truncation by the researchers of the acceptable values; a fourth group consists of just the dichotomous choice variants, while the final group uses two sources of data (OE and dichotomous choice) drawn from the same individuals. In the table therefore five estimates of the critical n are produced for each possible B. I vary B over three values from 0.5 to 1.5.

As can be seen, the critical value of n varies enormously. When the choice is across OE or DC-based methods of elicitation, an ideal CJ consisting of 16-25 members might outperform CV for some low values of B. On the other hand, if the set of possible frames lies within OE variants or within DC variants then ICJ is inferior. In this context it is worth noting that in the actual study, the value for B from the raw open-ended WTP was 1.7 which would rule out ICJ for any set of acceptable frames. Meanwhile the highest value for frame variation is 0.28, meaning that any partial valuation would have to capture at least 72% of the total valuation in its estimates for it to dominate the mix of contingent valuation methods used in the research.[14]

[14] As above, setting a finite level for the CV sample size produces only a marginal reduction in the critical values of n in this table.

4 Summing up.

This paper has examined one means by which non-market valuation methods can be compared. In a sense it provides an answer to Hausman and Diamonds' when is no number better than any number question. Generally the answer is never, given a choice between decision processes. Nevertheless we have seen clear reasons why the figures obtained from a contingent valuation exercise may be less useful from those obtained from a method that appears a priori inferior.

In the examples, I stress the possibility that partial valuation methods and approaches such as the Citizens' Jury may, under specific circumstances be superior to contingent valuation, particularly when WTA is viewed as an acceptable frame, but it is worth reiterating the point made earlier, that whereas contingent valuation and its associated techniques have been exhaustively examined for anomalies and inadequacies many of the alternatives have not received similar scrutiny. It is particularly worth emphasising that the potential advantage of the partial valuation methods lies in the absence or reduction of framing variance when such alternatives are used. To the extent that a partial valuation method is subject to the same kinds of framing variance that plagues fully specified contingent valuation, then the partial valuation method loses its superiority. The same applies to Citizens' Juries and consensus methods which have not thus far been tested extensively for framing invariance.[15]

One interpretation of the examples is that the frame variance drives the results. In other words, without any restriction on the set of acceptable frames, partial valuation and ICJ methods dominate contingent valuation. To the extent that some frames are excludable from the set of acceptable frames, for instance by using the criteria set out by the NOAA panel, then viewed from the perspective of preference satisfaction, the advantage of alternative decision-processes over contingent valuation diminishes.

The results also suggest what meta-data we need to know in order to make a choice between decision processes. Estimates of the variation in tastes within the sample are often not reported from valuation exercises, but would provide a useful guide to the choice of technique. Similarly, a database comparing values from users to total values would be beneficial. More fundamentally, more exercises in which frame variance was compared amongst different groups such as users versus non-users would be a step towards the optimal choice of decision processes.

References

Adamowicz, W. L. and Bhardwaj: 1993, Experiments on the difference between willingness to pay and willingness to accept, *Land Economics* **69**, 416–427.

[15] Indeed the evidence from group decision-making (see for example, Kerr et al, 1996) is far from comforting on the issue. Wilson and Howarth, 2002, include a discussion of the limited evidence currently available on biases arising from group-based valuation.

Aldred, J.: 2005, Consumer valuation and citizen deliberation, *in* M. Getzner, S. Stagl and C. Spash (eds), *Developing Alternatives to Valuing Nature*, Edward Elgar, pp. 187–208.

Aldred, J. and Jacobs, M.: 2000, Citizens and wetlands evaluating the Ely citizens' jury, *Ecological Economics* **34**, 217–232.

Amigues, J.-P., (Broadhead), C. B., Desaigues, B., Gauthier, C. and Keith, J. E.: 2002, The benefits and costs of riparian analysis habitat preservation a willingness to accept/willingness to pay contingent valuation aroach, *Ecological Economics* **43**(1), 17–31.

Ariely, D., Loewenstein, G. and Prelec, D.: 2003, Coherent arbitrariness: Stable demand curves without stable preferences, *Quarterly Journal of Economics* **118**, 73–105.

Armour, A.: 1995, The Citizens' Jury model of public participation: a critical evaluation, *in* O. Renn, T. Webler and P. Widemann (eds), *Fairness and Competence in Citizen Participation*, Kluwer Academic Publishers, Dordrecht.

Arndt, J. and Crane, E.: 1975, Response bias, yea-saying and the double negative, *Journal of Market Research* **12**, 218–220.

Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R. and Schuman, H.: 1993, Report of the NOAA panel on contingent valuation, *Federal Register* **58**(10), 4602–4614.

Barnes, M.: 1999, *Building a Deliberative Democracy An evaluation of two Citizens' Juries*, Institute for Public Policy Research, London.

Bateman, I. J., Carson, R., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozdemiroglu, E., Pearce, D. W., Sugden, R. and Swanson, T.: 2002, *Economic valuation with stated preference techniques: a manual*, Edward Elgar.

Bateman, I. J., Langford, I. H., Turner, R. K., Willis, K. G. and Garrod, G. D.: 1995, Elicitation and truncation effects in contingent valuation studies, *Ecological Economics* **12**(2), 161–179.

Bateman, I., Munro, A., Rhodes, B., Starmer, C. and Sugden, R.: 2006, Anchoring and yea-saying with private goods: an experiment, *in* J. List (ed.), *Using Experimental Methods in Environmental and Resource Economics*, Edward A. Elgar, Cheltenham.

Bishop, R. C. and Heberlein, T. A.: 1979, Measuring values of extramarket goods: Are indirect measures biased?, *American Journal of Agricultural Economics* **61**(5), 926–930.

Bishop, R. C. and Heberlein, T. A.: 1986, Does contingent valuation work, *in* R. G. Cummings, D. S. Brookshire and W. Schulze (eds), *Valuing Environmental Goods. An Assessment of the Contingent Valuation Method*, Totowa, pp. 123–147.

Bishop, R., Heberlein, T. and Kealy, M.: 1983, Hypothetical bias in contingent valuation: Results from a simulated market, *Natural Resources Journal* **23**(3), 619–633.

Blamey, R., Smith, R. J. R. and Niemeyer, S.: 2002, *Citizens' Juries and Environmental Value Assessment*, Research School of Social Sciences.

Bohman, J. and Rehg, W. (eds): 1997, *Deliberative Democracy: Essays on Reason and Politics*, MIT Press.

Bowker, J. and Stoll, J.: 1988, Use of dichotomous choice nonmarket methods to value the whooping crane resource, *American Journal of Agricultural Economics* **70**, 372–381.

Boyce, R. R., Brown, T. C., McClelland, G. H., Peterson, G. L. and Schulze, W. D.: 1992, An experimental examination of intrinsic values as a source of the wta-wtp disparity, *The American Economic Review* **82**(5), 1366–1373.

Boyle, K. J.: 2003, Contingent valuation in practice, *in* P. A. Champ, K. J. Boyle and T. C. Brown (eds), *A Primer on Non-Market Valuation*, Kluwer Academic Press.

Boyle, K. J. and Bishop, R. C.: 1988a, Valuing wildlife in benefit-cost analysis a case study involving endangered species, *Water Resources Research* **23**, 943–950.

Boyle, K. J. and Bishop, R. C.: 1988b, Welfare measurements using contingent valuation: A comparison of techniques, *American Journal of Agricultural Economics* **70**(1), 20–28.

Boyle, K. J., Johnson, F. R., McCollum, D. W., Desvouges, W. H., Dunford, R. W. and Hudson, S. P.: 1996, Valuing public goods: Discrete versus continuous contingent-valuation responses, *Land Economics* **72**(3), 381–396.

Brookshire, D. S. and Coursey, D. L.: 1987, Measuring the value of a public good: An empirical comparison of elicitation procedures, *American Economic Review* (77), 554–566.

Brookshire, D. S., Randall, A. and Stoll, J. R.: 1980, Valuing increments and decrements in natural resource service flows, *American Journal of Agricultural Economics* **62**, 478–488.

Brookshire, D., Schulze, W. and D'Arge, R.: 1982, 'valuing public goods a comparison of survey and hedonic aroaches', *American Economic Review* **72**, 165–177.

Brown, T. C., Champ, P. ., Bishop, R. and McCollum, D. W.: 1996, Which response format reveals the truth about donations to a public goods, *Land Economics* **72**, 152–166.

Carson, R.: 2000, Contingent valuation: a user's guide, *Environmental Science & Technology* **34**(8), 1413–1418.

Carson, R. T.: 1998, Valuation of tropical rainforests: philosophical and practical issues in the use of contingent valuation, *Ecological Economics* **24**, 15–29.

Carson, R. T., Flores, N. E., Martin, K. M. and Wright, J. L.: 1996, Contingent valuation and revealed preference methodologies: Comparing the estimates for quasi-public goods, *Land Economics* **72**(1), 80–99.

Clarke, P.: 2002, Testing the convergent validity of the contingent valuation and travel cost methods in valuing the benefits of health care, *Health Econ* **11**, 117–127.

Coote, A. and Lenaghan, J.: 1997, *Citizens' Juries From Theory to Practice*, IPPR, London.

Crosby, N.: 1991, *Citizens' Juries as a basic Democratic Reform*, Jefferson Centre, Minneapolis.

Crouch, A. and Keniston, K.: 1960, Yea-sayer and nay-sayers agreeing response set as a personal variable, *Journal of Abnormal and Social Psychology* **60**, 151–174.

Cummings, R. G. and Harrison, G. W.: 1995, The measurement and decomposition of nonuse values: A critical review, *Environmental and Resource Economics* **5**(3), 225–247.

Cummings, R. G., Harrison, G. W. and Rutstrom, E. E.: 1995, Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive-compatible?, *The American Economic Review* **85**(1), 260–266.

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

Fischoff, B.: 1991, Value elicitation is there anything in there?, *American Psychologist* **468**, 835–847.

Frykblom, P. and Shogren, J. F.: 2000, An experimental testing of anchoring effects in discrete choice questions, *Environmental and Resource Economics* **16**(3), 329–341.

Greenley, D., Walsh, R. G. and Young, R. A.: 1981, Option value: Empirical evidence from a case study of recreation and water quality, *Quarterly Journal of Economics* **96**(4), 657–673.

Gregory, R. S.: 2000, Valuing environmental policy options: A case study comparison of multiattribute and contingent valuation survey methods,, *Land Economics and 76 151-173.* **76**.

Gregory, R. S. and Slovic, P.: 1997, Commentary a constructive approach to environmental valuation, *Ecological Economics* **21**, 175–181.

Gregory, R. and Wellman, K.: 2001, Bringing stakeholder values into environmental policy choices: a community-based estuary case study, *Ecological Economics* **39**(1), 37–52.

Halvorsen, B., and Soelensminde, K.: 1998, Differences between willingness-to-pay estimates from open-ended and discrete-choice contingent valuation methods the effects of heteroscedasticity, *Land Economics* **74**(2), 262–282.

Hanemann, W.: 1984, Welfare evaluations in contingent valuation experiments with discrete responses, *American Journal of Agricultural Economics* **66**(3), 332–341.

Herriges, J. A. and Shogren, J. F.: 1996, Starting point bias in dichotomous choice valuation with follow-up questioning, *Journal Of Environmental Economics And Management* **30**(1), 112–131.

Hoehn, J. and Randall, A.: 1987, A satisfactory benefit cost indicator from contingent valuation, *Journal of Environmental Economic Management* **14**, 226–247.

Holmes, T. and Kramer, R. A.: 1995, An independent sample test of yea-saying and starting point bias in dichotomous-choice contingent valuation, *Journal of Environmental Economic Management* **29**, 121–132.

Horowitz, J. K. and McConnell, K. E.: 2002, A review of WTA/WTP studies, *Journal of Environmental Economics and Management* **44**(3), 426–447.

Howarth, R. and Wilson, M.: 2006, A theoretical approach to deliberative valuation: Aggregation by mutual consent, *Land Economics* **20**(1), 1–16.

Huang, J., Haab, T. and Whitehead, J.: 1997, Willingness to pay for quality improvements: Should revealed and stated preference data be combined?, *Journal of Environmental Economics and Management* **34**(3), 240–255.

Johannesson, M., Liljas, J. B. and O'Conor, R. M.: 1997, Hypothetical versus real willingness to pay some experimental results, *Applied Economics Letters* **4**(3), 149–159.

Kahneman, D., Slovic, P. and Tversky, A.: 1982, *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press.

Kanninen, B. J.: 1995, Bias in discrete response contingent valuation, *Journal of Environmental Economics and Management* **28**, 114–125.

Kennedy, C. A.: 2002, Revealed preference valuation compared to contingent valuation: radon-induced lung cancer prevention, *Health Economics* **117**, 585 – 598.

Kenyon, W., Nevin, C. and Hanley, N.: 2003, Enhancing Environmental Decision-making Using Citizens Juries, *Local Environment* **8**(2), 221–232.

Kerr, N. L., MacCoun, R. J. and Kramer, G. P.: 1996, Bias in judgment: Comparing individuals and groups, *Psychological Review* **1034**, 687–719.

Kontoleon, A. and Swanson, T.: 2003, The willingness to pay for property rights for the giant panda: Can a charismatic species be an instrument for nature conservation?, *Land Economics* **79**(4), 483–499.

Kristrom, B.: 1993, Comparing continuous and discrete contingent valuation questions, *Environmental and Resource Economics* **31**, 63–71.

Lazo, J. K., McClelland, G. H. and Schulze, W. D.: 1997, Economic theory and psychology of non use values, *Land Economics* **73**(3), 358–371.

Li, C.-Z., Löfgren, K.-G. and Hanemann, W. M.: 1996, *Real versus Hypothetical Willingness to Accept The Bishop and Heberlein Model Revisited*, CUDARE Working Paper 793.

List, J. A.: 2002, Preference reversals of a different kind: The "more is less" phenomenon, *The American Economic Review* **92**(5), 1636–1643.

List, J. A.: 2003, Does market experience eliminate market anomalies?, *Quarterly Journal of Economics* **118**(1), 41–71.

McFadden, D.: 1999, Rationality for economists?, *Journal of Risk and Uncertainty* **19**(1), 73–105.

Mitchell, R. C. and Carson, R. T.: 1989, *Using Surveys to Value Public Goods: the contingent valuation method*, Resources for the Future.

Orne, M.: 1962, On the social psychology of the psychological experiment, *American Psychologist* **17**, 776–783.

Philip, L., Hanley, N., Macmillan, D. C. and Alvarez-Farizo, B.: 2002, Valuing the non-market benefits of wild goose conservation: a comparison of interview and group based approaches, *Ecological Economics* **43**(1), 49–59.

Sagoff, M.: 1998, Aggregation and deliberation in valuing environmental public goods: A look beyond contingent pricing, *Ecological Economics* **24**(2), 213–230.

Sayman, S. and Onculer, A.: 2005, Effects of study design characteristics on the WTA-WTP disparity: A meta- analytical framework, *Journal of Economic Psychology* **26**, 289–312.

Schkade, D. and Payne, J.: 1994, How people respond to contingent valuation questions: A verbal protocol analysis of willingness to pay for an environmental regulation, *Journal of Environmental Economics and Management* **26**(1), 88–109.

Shogren, J. F., Shin, S. Y., Hayes, D. J. and Kliebenstein, J. B.: 1994, Resolving differences in willingness to pay and willingness to accept, *The American Economic Review* **84**(1), 255–270.

Silberman, J., Gerlowski, D. A. and Williams, N. A.: 1992, Estimating existence value for users and nonusers of New Jersey beaches, *Land Economics* **68**(2), 225–236.

Slovic, P. and Lichtenstein, S.: 1971, Comparison of Bayesian and regression approaches to the study of information processing in judgement, *Organizational Behaviour and Human Performance* **6**, 649–744.

Smith, V., Desvousges, W. and Fisher, A.: 1986, A comparison of direct and indirect methods for estimating environmental benefits, *American Journal of Agricultural Economics* **68**(2), 280–290.

Sugden, R.: 2005, Anomalies and stated preference techniques: A framework for a discussion of coping strategies, *Environmental and Resource Economics* **32**(1), 1–12.

Thayer, M.: 1981, Contingent valuation techniques for assessing environmental impacts: Further evidence., *Journal of Environmental Economics and Management* **8**(1), 27–44.

Van Soest, A. and Hurd, M.: 2003, *A Test for Anchoring and Yea-Saying in Experimental Consumption Data*, RAND Corporation.

Whittington, D., Briscoe, J., Mu, X. and Barron, W.: 1990, Estimating the willingness to pay for water services in developing countries a case study of the use of contingent valuation surveys in Southern Haiti, *Economic Development and Cultural Change* **38**(2), 293–311.

Wilson, M. and Howarth, R.: 2002, Discourse-based valuation of ecosystem services: establishing fair outcomes through group deliberation, *Ecological Economics* **41**(3), 431–443.