



Contingent Valuation: Is Some Number better than No Number?

Peter A. Diamond, Jerry A. Hausman

The Journal of Economic Perspectives, Volume 8, Issue 4 (Autumn, 1994), 45-64.

Your use of the JSTOR database indicates your acceptance of JSTOR's Terms and Conditions of Use. A copy of JSTOR's Terms and Conditions of Use is available at <http://www.jstor.ac.uk/about/terms.html>, by contacting JSTOR at jstor-info@umich.edu, or by calling JSTOR at (888)388-3574, (734)998-9101 or (FAX) (734)998-9113. No part of a JSTOR transmission may be copied, downloaded, stored, further transmitted, transferred, distributed, altered, or otherwise used, in any form or by any means, except: (1) one stored electronic and one paper copy of any article solely for your personal, non-commercial use, or (2) with prior written permission of JSTOR and the publisher of the article or other text.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The Journal of Economic Perspectives is published by American Economic Association. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.ac.uk/journals/aea.html>.

The Journal of Economic Perspectives
©1994 American Economic Association

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2000 JSTOR

Contingent Valuation: Is Some Number Better than No Number?

Peter A. Diamond and Jerry A. Hausman

Most economic analyses aim at explaining market transactions. Data on transactions, or potentially collectible data on transactions, are the touchstone for recognizing interesting economic analyses. However loose the connection between a theoretical or empirical analysis and transactions, this connection is the basis of the methodology of judging the credibility and reliability of economic analyses. Generally, individuals do not purchase public goods directly. Lack of data on transactions implies that economists must find other methods to assess surveys asking for valuations of public goods.

To address this problem, we begin with a discussion of the methodology of evaluating contingent valuation surveys. While there is some experimental evidence about small payments for public goods, we work with the assumption that we do not have data on actual transactions for interesting environmental public goods to compare with survey responses of hypothetical willingness-to-pay. This situation creates the need for other standards for evaluating survey responses. Evaluation involves the credibility, bias (also referred to as reliability in the literature), and precision of responses. Credibility refers to whether survey respondents are answering the question the interviewer is trying to ask. If respondents are answering the right question, reliability refers to the size and direction of the biases that may be present in the answers. Precision refers to the variability in responses. Since precision can usually be increased by the simple expedient of increasing the sample size, we will not discuss precision further in this paper. Problems of credibility or of bias are not reduced by increases in sample size. Thus credibility and bias must be evaluated when

■ *Peter A. Diamond and Jerry A. Hausman are Professors of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts.*

considering the use of such surveys—in benefit-cost analyses, in the determination of damages after a finding of liability, or as general information to affect the legislative process.¹

We discuss how to judge the content in contingent valuation surveys together with evidence from surveys that have been done. Surveys designed to test for consistency between stated willingness-to-pay and economic theory have found that contingent valuation responses are not consistent with economic theory. The main contingent valuation anomaly that we discuss is called the “embedding effect,” and was first analyzed systematically by Kahneman and Knetsch (1992).² The embedding effect is the name given to the tendency of willingness-to-pay responses to be highly similar across different surveys, even where theory suggests (and sometimes requires) that the responses be very different.³ An example of embedding would be a willingness-to-pay to clean up one lake roughly equal to that for cleaning up five lakes, including the one asked about individually. The embedding effect is usually thought to arise from the nonexistence of individual preferences for the public good in question and from the failure of survey respondents, in the hypothetical circumstances of the survey, to consider the effect of their budget constraints. Because of these embedding effects, different surveys can obtain widely variable stated willingness-to-pay amounts for the same public good, with no straightforward way for selecting one particular method as the appropriate one.

In short, we think that the evidence supports the conclusion that to date, contingent valuation surveys do not measure the preferences they attempt to measure. Moreover, we present reasons for thinking that changes in survey methods are not likely to change this conclusion. Viewed alternatively as opinion polls on possible government actions, we think that these surveys do not have much information to contribute to informed policy-making. Thus, we conclude that reliance on contingent valuation surveys in either damage assessments or in government decision making is basically misguided.

¹With two estimates of an economic value, one can analyze directly whether one is a biased estimate of the other. With nonuse value, the lack of an alternative direct estimate of willingness-to-pay makes it relevant to consider credibility directly, as well as the differences between survey results and behavior in other contexts where transactions data are available.

²Another failure of contingent valuation surveys to be consistent with economic preferences is that stated willingness-to-pay is usually found to be much less than stated willingness-to-accept. From economic theory, willingness-to-pay differs from willingness-to-accept only by an income effect. Thus, their values should be extremely close in typical contingent valuation circumstances, where the stated willingness-to-pay is a small share of the consumer's overall budget, and willingness-to-pay amounts show a small income elasticity. For further discussion of this problem with contingent valuation surveys and other problems, see Diamond and Hausman (1993) and Milgrom (1993).

³The term embedding came from the research approach of “embedding” a particular good in a more inclusive good, and contrasting the stated willingness-to-pay for the good with that obtained by allocating the willingness-to-pay for the more inclusive good among its components (Kahneman, personal communication).

Judging Surveys of Willingness-to-Pay for Public Goods

A number of bases exist for forming judgments about whether particular respondents are answering the right question and whether the response is roughly correct. One widely accepted basis is by reaching the conclusion that a particular response is simply not credible as an answer to the question the interviewer is trying to ask. It is standard practice in the contingent valuation literature to eliminate some responses as being unreasonably large to be the true willingness-to-pay. Thus trimming responses that are more than, say, 5 percent of income for an environmental public good that contains only nonuse value may be criticized for having an arbitrary cutoff, but not for omitting answers that are believed to be credible. Similarly, it is standard practice to eliminate some responses of zero on the basis that these are "protest zeros," that answers to other questions in the survey indicate that individuals do put a positive value on changes in the level of the public good, and thus zero is not a credible answer.

A widely accepted incredibility test indicates that it is not automatic that the response given is an answer to the question that the interviewer wants answered. But we need to go further in considering how to form a judgment on the survey responses; it is not adequate to assume that any response that is not obviously wrong is an accurate response to the question the survey designer had in mind.

A number of additional bases have been used by people arguing that responses are or are not acceptable. The methods we shall discuss include verbal protocol analysis, the patterns of willingness-to-pay responses across individuals, and across surveys.

In considering the relevance of this evidence for the question of whether survey responses are accurate measures of true preferences, it is useful to have in mind some possible alternative hypotheses of how people respond to such surveys, since the responses are not simply random numbers. Several hypotheses have been put forward as alternatives to the hypothesis that the responses are measures of true economic preferences. Individuals may be expressing an attitude toward a public good (or class of public goods), expressed in a dollar scale because they are asked to express it in a dollar scale (Kahneman and Ritov, 1993). Individuals may receive a "warm glow" from expressing support for good causes (Andreoni, 1989).⁴ Individuals may be describing what they think is good for the country, in a sort of casual benefit-cost analysis (Diamond and Hausman, 1993). Individuals may be expressing a reaction to actions that have been taken (for example, allowing an oil spill) rather than evaluating the state of a resource.

⁴This approach was developed for actual charitable contributions, not survey responses. Kahneman and Knetsch (1992) call it the purchase of moral satisfaction.

Under all of these alternative hypotheses, responses are not an attempt by an individual to evaluate his or her own preference for a public good. For example, people doing casual benefit-cost analyses may be reflecting how much they think people generally care about the issue. We think that different hypotheses are likely to be appropriate for different people. Thus the question is not whether the hypothesis of an accurate measurement of preferences is the single best hypothesis, but whether the fraction of the population for whom the hypothesis of accuracy is reasonable is sufficiently large to make the survey as a whole useful for policy purposes.

All of these alternatives are based on what individuals are trying to do; there are further questions of standard survey biases (such as interviewer bias, framing bias, hypothetical bias) and whether people have enough information to express a preference with any accuracy, even if they are attempting to express a preference. Insofar as this understanding is faulty, expressed preferences are not an expression of true economic preferences.

Verbal Protocol Analysis

For verbal protocol analysis, individuals are asked to “think aloud” as they respond to a questionnaire, reporting everything that goes through their minds. Everything the subjects say is recorded on audio tapes that are transcribed and coded for the types of considerations being mentioned. Schkade and Payne (1993) have done such an analysis using a contingent valuation survey that asks for willingness-to-pay to protect migratory waterfowl from drowning in uncovered waste water holding ponds from oil and gas operations.

The transcripts show the inherent difficulty in selecting a willingness-to-pay response and the extent to which people refer to elements that ought to be irrelevant to evaluating their own preferences. If people are trying to report a preference, we would expect them to consider inputs into the forming of their preferences, such as how much they care about birds, how important the number of killed birds are relative to the numbers in the species. Conversely, we would not expect them to report a willingness-to-pay just equal to what they think the program will cost. Respondents verbalized many diverse considerations. Perhaps the most common strategy involved first acknowledging that something should be done and then trying to figure out an appropriate amount. About one-fourth of the sample mentioned the idea that if everyone did his part then each household would not have to give all that much. About one-sixth of the sample made comparisons with donations to charities. About one-fifth of the sample said they just made up a number or guessed an answer. Many respondents seemed to wish to signal concern for a larger environmental issue. This pattern may reflect the unfamiliarity of the task the respondents faced.

These findings strongly suggest that people are not easily in touch with underlying preferences about the type of commodity asked about. The findings do not lend support to the hypothesis that responses are an attempt to measure and express personal preferences. To the extent that individuals consider costs

to everyone, the analysis supports the hypothesis of casual benefit-cost analysis. To the extent that individuals look to their own charitable contributions for a guide, the analysis is consistent with hypotheses that explain actual contributions, such as the warm glow hypothesis.

Variation in Willingness-To-Pay Across Individuals

If stated willingness-to-pay is a reflection of true preferences, then we would expect certain patterns of answers across different individuals (other things equal). We would expect self-described environmentalists to have larger willingnesses-to-pay. We would expect individuals with higher incomes to have larger willingnesses-to-pay. Both results do occur. However, such results do not distinguish among the various hypotheses that were spelled out above since we would expect roughly similar results from any of them. Thus this potential basis for evaluation does not have much bite.⁵ We do observe that the income effects that are measured in typical surveys are lower than we would expect if true preferences are measured, lower for example than measured income elasticities for charitable giving.⁶

Variation in Willingness-To-Pay Across Surveys

Another approach to forming a judgment is to compare willingness-to-pay responses to different questions, whether in the same or in different surveys.

Multiple Questions. If a survey question reveals a true valuation, it should not matter whether the question is asked by itself or with other questions, nor if asked with any other questions, what the order of questioning is. However, when Tolley et al. (1983) asked for willingness-to-pay to preserve visibility at the Grand Canyon, the response was five times higher when this was the only question, as compared to its being the third such question. Attempts to claim this result to be consistent with preferences have relied on income effects and substitution effects. Neither of these rationalizations for the anomalous results is compelling, as we explain in a moment.

The importance of question order was also shown in a study by Samples and Hollyer (1990) asking for the values of preserving seals and whales. Some respondents were asked for willingness-to-pay to preserve seals first, followed by a question about whales. Others were asked for willingness-to-pay in the reverse order. Seal value tended to be lower when asked after whale value, while whale value was not affected by the sequence of questions.⁷ Thus the sum

⁵The importance of the lack of bite of such considerations comes, in part, from the fact that the contingent valuation study of the Exxon Valdez spill that was done for the state of Alaska (Carson et al., 1992) included such analyses, but none of the more powerful split-sample consistency tests that we discuss below.

⁶The empirical finding of low income elasticities is also inconsistent with the typical finding of a large divergence between willingness-to-pay and willingness-to-accept, discussed in footnote 2.

⁷Samples and Hollyer used dichotomous choice surveys. They estimated that whales were valued at \$125 when asked about first, and \$142 when second. Seals were valued at \$103 when asked about first and \$62 when second. When they asked about both (together) in a single question, the estimated values were \$131 and \$146 in two surveys.

of willingness-to-pay depended on the sequence of the questions asked. The authors offer an explanation (p. 189) "based on debriefing sessions held with the interviewer."

Apparently, when respondents valued seals first, they used their behavior in this market situation to guide their responses to whale valuation questions. Since whales are generally more popular than seals, respondents were reluctant to behave more benevolently toward seals compared with humpback whales. Consequently, whale values were inflated in the S-W questionnaire version to maintain a relatively higher value for the humpbacks. This behavioral anchoring effect did not exist in the W-S version, where whales were valued first.

To have the value of preserving both seals and whales depend on the sequence in which the questions are asked is not consistent with the hypothesis that stated willingness-to-pay accurately measures preferences. These results can be interpreted in two ways. One interpretation is that contingent valuation studies that ask two questions rather than one are unreliable. The other interpretation is that the warm glow hypothesis is supported, since having expressed support for the environment in the first question permits a sharp fall in the second response. This effect is not present, however, when such a response would seem illogical to the respondent. More generally, one needs to decide whether a given pattern of responses is a result of survey design issues or a result of the underlying bases of response. This distinction is especially important when the pattern of results appears anomalous with or contradictory to the hypothesis that preferences are accurately measured.

Single Questions and the Embedding Effect. Alternatively, one can ask a single willingness-to-pay question each to different samples. For example, assume that one group is asked to evaluate public good X ; a second is asked to evaluate Y ; and a third is asked to evaluate X and Y . What interpretations could we make if the willingness-to-pay for X and Y (together) is considerably less than the sum of the willingness-to-pay for X and the willingness-to-pay for Y ?⁸ One interpretation is that we are seeing an income effect at work. That is, having "spent" for X , one has less income left to purchase Y . Given that the stated willingness-to-pay amounts are very small relative to income and that measured income elasticities are very small, the attempted income effect argument does not explain the differences found.

A second interpretation is to assume that individual preferences have a large substitution effect between X and Y . In some settings the assumption on preferences needed to justify the results is implausible. For example, Diamond et al. (1993) asked for willingness-to-pay to prevent logging in one, two, and

⁸This approach is similar to the work that was initiated by Kahneman (1986) and done recently by Kahneman and Knetsch (1992), Kemp and Maxwell (1993), Desvousges et al. (1993), Diamoud et al. (1993), McFadden and Leonard (1993), Loomis, Hoehn and Hanemann (1990).

three particular wilderness areas. Stated willingness-to-pay to preserve two (and three) areas was less than the sum of willingnesses-to-pay to preserve each of them separately.

At first look, this result appears to be an appropriate substitution effect, since protecting one area results in being less willing to protect another. However, preferences should be defined over wilderness remaining, not over proposals for development that are defeated. If preferences are concave over the amount of wilderness available (or, more generally, if different wilderness areas are substitutes), then willingness-to-pay is larger the smaller the quantity of wilderness remaining. This implies that the willingness-to-pay to preserve two threatened areas should be larger than the sum of willingness-to-pay to preserve each as the lone area threatened with development.⁹ Instead, stated willingness-to-pay was roughly the same for preserving one, two or three threatened areas, making the amount for several areas together significantly less than the sum of the amounts for the areas separately. Note that these surveys vary both the number of areas threatened and the number to be preserved. Neither the income effect nor the substitution effect can plausibly explain the embedding effect in this experiment. The hypothesis that this survey is eliciting individual preferences is not consistent with individuals having reasonably behaved preferences. However, from the point of view of the warm glow hypothesis, this pattern makes sense. That is, the warm glow hypothesis is that individuals are primarily reporting an expression of support for the environment, an expression that does not vary much with small changes in the precise environmental change being described.

A similar variation in responses across surveys appears in the study of Desvousges et al. (1993). They described a problem killing 2000, 20,000 and 200,000 birds. The willingness-to-pay to solve this problem was roughly the same in all three cases. Since the number of surviving birds is smaller the larger the problem, concave preferences over surviving birds should have resulted in more than a 100-fold variation in willingness-to-pay across this range.¹⁰ Thus

⁹For derivation of the convexity of willingness-to-pay when preferences are concave and the scenario is varied in this way, see Diamond (1993). That paper also contains a number of other implications of preferences for willingness-to-pay that can be used for internal consistency tests.

¹⁰Proponents of contingent valuation have made several critiques of this study. One critique is that it was a mall stop survey. But similar results followed when the questionnaire was used for the verbal protocol study cited above, which involved subjects coming to be interviewed. Another criticism is that in addition to the absolute numbers, the survey questions described the number of birds at risk as "much less than 1%" of the population, "less than 1%," and "about 2%." Thus, one can wonder whether respondents were paying attention to the absolute numbers which varied 100-fold or the percentages which varied from "much less than 1%" to "about 2%." Interpreting "much less than" as less than half, about 2% is at least a four-fold increase over less than half of 1%. If some people were paying attention to the percentages and some to the absolute numbers, the range should have been between four-fold and 100-fold. If, as Hanemann suggests, respondents did not perceive any real difference between "much less than 1%" and "about 2%," it is noteworthy that they perceived a large difference between zero and "much less than 1%." Moreover, these percentages were selected by the authors since they were the percentages in three actual oil spills: Arthur Kill, Nestucca, *Exxon Valdez*. This pattern of results is consistent with the responses being dominated by a "warm glow."

this study shows a contradiction between stated willingness-to-pay and the usual economic assumptions on preferences. Again, the study is consistent with the hypothesis that the responses are primarily warm glow, and so need not vary noticeably over moderate differences in the resource.

Adding-up Test. One difficulty in the approach described above is that the plausibility of the willingness-to-pay patterns depends on assumptions on the plausible (concave) structure of preferences. Another approach to tests of consistency that does not rely on an assumption of concave preferences is to attempt to measure the same preference in two different ways. This test can be constructed by varying the background scenario as well as varying the commodity to be purchased. For example, assume that one group is asked to evaluate public good X ; a second group is told that X will be provided and is asked to evaluate also having Y ; and a third is asked to evaluate X and Y (together). Now the willingness-to-pay for X and Y (together) should be the same as the sum of the willingness-to-pay for X and the willingness-to-pay for Y , having been given X (the same up to an income effect that can be measured in the survey and that empirically is small).¹¹ Thus, Diamond et al. (1993) varied the number of wilderness areas being developed as well as the number that could be protected. In this way the sum of two areas separately evaluated (with different degrees of development) should be the same as the value of preserving two areas (apart from a very small income effect). Again, the results of the survey are inconsistent with the responses being a measure of preferences.¹²

Embedding still infects even very recent work done by experienced contingent valuation analysts who were well aware of the problem. Schulze et al. (1993) asked for willingness-to-pay for partial and complete cleanup of contamination of the Clark Fork National Priorities List sites in Montana. After

¹¹Willingness-to-pay is a function of the two vectors giving alternative levels of public goods and the level of income. Thus the willingness-to-pay to improve the environment from z to z'' of someone with income I can be written $WTP(z, z'', I)$. The change from z to z'' can be broken into two pieces, a change from z to z' and a change from z' to z'' . From the definition of willingness-to-pay, one has $WTP(z, z'', I) = WTP(z, z', I) + WTP(z', z'', I - WTP(z, z', I))$.

This adding-up test makes no use of an assumption on the magnitude or sign of income or substitution effects. One could do an adding-up test without the adjustment of income shown in the equation by comparing $WTP(z, z'', I)$ with $WTP(z, z', I) + WTP(z, z'', I)$. This comparison would involve a deviation from exact adding-up because of the income effect. With a willingness-to-pay on the order of \$30 and a household income level of \$30,000, even an income elasticity of one—higher than the elasticity typically measured in contingent valuation surveys—would lead to a \$.03 deviation from exact adding-up. For a formal derivation, see the revised version of Diamond (1993).

¹²In brief response to Hanemann's criticisms of our analysis, we note that he does not address this adding-up test and seems comfortable accepting the idea that the less wilderness preserved, the less people care about any particular area of wilderness. These two tests do not rely on any assumption of different wilderness areas being interchangeable, as indicated by the vector interpretation of z in the previous footnote. In terms of Hanemann's test mentioned in his note 25 of whether willingness-to-pay to protect each of the areas is the same, we note that he did not do the statistical test correctly. Moreover, this reference is an example of Hanemann's trait of ignoring the central criticism while attacking a side issue. In Diamond et al., the focus is on the adding-up test, not a scope test. The adding-up test was clearly rejected.

removing protest zeros and high responses, the mean stated willingness-to-pay for complete cleanup was \$72.46 (standard error of \$4.71) while the mean response for a considerably smaller partial cleanup was \$72.02 (s.e. \$5.10). As part of the survey, respondents were asked whether their responses were just for this cleanup or partly to cleanup other sites or basically as a contribution for all environmental or other causes (or other). Only 16.9 percent reported their answers as just for this cleanup; that is, a vast majority of respondents recognized an embedding effect in their own responses. These respondents were asked what percentage of their previous answer was for this cleanup, and the willingness-to-pay responses were adjusted by these percentages. After this adjustment, the mean stated willingness-to-pay for complete cleanup was \$40.00 (s.e. \$2.62) while the mean response for partial cleanup was \$37.15 (s.e. \$2.71).

These numbers (and the large fraction of people recognizing that they are embedding) support the hypothesis that the responses are dominated by a warm glow. No reason is offered by the authors for the conclusion that the adjustment they do removes the dominance of warm glow. Neither do they perform an adding-up test such as that described above. This adding-up test could have been done by asking a third sample for willingness-to-pay to extend a "planned" partial cleanup to a complete cleanup. In short, the embedding problem does not appear to be one that contingent valuation practitioners know how to solve.

With a pattern of results that are inconsistent with the usual economic assumptions, two interpretations are always possible: the surveys were defective or the contingent valuation method as currently practiced does not measure with accuracy. One should consider all the surveys that attempt to test for consistency in order to judge which interpretation is likely to be correct. The studies we have described have been criticized as not done well enough to be an adequate test.¹³ However, they are the only quantitative tests we are aware of. No comparable comparison tests have been done by proponents of the accuracy of contingent valuation, although the embedding effect has long been recognized.

Differing Payment Vehicles. It is interesting to note what two contingent valuation proponents, Mitchell and Carson (1989), have written about the question that respondents are trying to answer. In discussing the sensitivity of responses to the payment vehicle (the way in which the hypothetical payment is to be collected), they write (pp. 123-24):

It was earlier assumed that only the nature and amount of the amenity being valued should influence the WTP [willingness-to-pay] amounts; all other scenario components, such as the payment vehicle and method of provision, should be neutral in effect... More recently, Arrow (1986),

¹³One can ask whether the patterns of thought reflected in the responses to the questions in any particular survey also occur in other survey settings. Cognitive psychology has found a number of such patterns that are robust. We think that the patterns reflected in these surveys are similarly robust.

Kahneman (1986), and Randall (1986) have argued against that view, holding that important conditions of a scenario, such as the payment vehicle, should be expected to affect the WTP amounts. In their view, which we accept, respondents in a CV [contingent valuation] study are not valuing levels of provision of an amenity in the abstract; they are valuing a policy which includes the conditions under which the amenity will be provided, and the way the public is likely to be asked to pay for it.

In other words, Mitchell and Carson appear to accept the idea (consistent with the findings about some respondents by Schkade and Payne, 1993) that individuals' responses arise from casual benefit-cost analyses, not solely from an examination of their own preferences over resources. For welfare analysis and damage measurement, benefit-cost studies may be different from preferences. We will return to this issue.

Evaluation of Bias: Calibration

Surveys about behavior often have systematic biases relative to the behavior they ask about. Thus, it is common to "calibrate" the responses—that is, adjust for the biases—as part of using them for predictive purposes. In particular, when using surveys to estimate demand for new products, it is standard practice to use a calibration factor to adjust survey responses in order to produce an estimate of actual demand (Urban, Katz, Hatch, and Silk, 1983). As Mitchell and Carson (1989, p. 178) have written: "Such 'calibration' is common in marketing designed to predict purchases. If a systematic divergence between actual and CV [contingent valuation] survey behavior existed and could be quantified, calibration of CV results could be undertaken."

As some evidence on the need for calibration, comparisons of hypothetical surveys and actual offers often find large and significant differences. These comparisons have been done for private goods (Bishop and Heberlein, 1979; Dickie, Fisher and Gerking, 1987; Neill et al., 1993).¹⁴ Comparisons have also been done for charitable donations (Duffield and Patterson, 1992; Seip and Strand, 1992). These studies find a need to calibrate, with calibration factors involving dividing stated willingness-to-pay by a number ranging from 1.5 to 10.

How this calibration should be extended to the public good context is unclear, since the public good context includes both unfamiliar commodities and unfamiliar transactions. But the lack of study of appropriate calibration factors is not a basis for concluding that the best calibration is one-for-one.¹⁵

¹⁴On the Dickie, Fisher and Gerking (1987) study, see also the critique by Hausman and Leonard (1992).

¹⁵In its proposed rules for damage assessment, the National Oceanic and Atmospheric Administration (1994) has proposed a default calibration of dividing by two, in the absence of direct arguments by trustees of natural resources for a different calibration factor.

Welfare Analysis

If an accurate measure of willingness-to-pay for the pure public good of the existence of an environmental amenity were available, the measured willingness-to-pay would belong in benefit-cost analysis, just like a pure public good based on resource use. Similarly, the measure should be included in the incentives government creates (through fines and damage payments) to avoid damaging an environmental amenity. As we know from the pure theory of public goods, we would simply add individual willingness-to-pay across the population.¹⁶ In this section, we consider the welfare implications of using stated willingness-to-pay as if it were an accurate measure of preferences in the case that the responses are generated by the alternative hypotheses given above.

One set of problems arises even if willingness-to-pay is being measured accurately, if measured willingness-to-pay contains an altruism component. That is, individuals may be willing to pay to preserve an environmental amenity because of their concerns for others (who may be users or also nonusers). Consider what happens if society adds up everyone's willingness-to-pay and compares the sum with the cost of some action. As a matter of social welfare evaluation we might conclude that such altruistic externalities are double counting, since a utility benefit shows up in the willingness-to-pay of both the person enjoying the public good and the people who care about that person. For example, consider the income distribution problem in a three-person economy. If two of the people start to care about each other, is this change in preferences a reason for a government to increase the level of incomes allocated to the two of them? Similarly, we can ask if the government should devote more taxes to cleaning up lakes where neighbors are friendly with each other than to lakes where neighbors do not know (or care about) each other.

Moreover, if altruistic externalities are thought to be appropriately included in the analysis, it is necessary to include all such externalities for accurate evaluation. In particular, if people care about each other's utilities, they care about the costs borne by others as well as the benefits received by others. An adjustment for altruism must include external costs as well as external benefits if we are to avoid the possibility of a Pareto worsening from an action based on a calculation that appears to be a Pareto improvement (Milgrom, 1993).

A second general problem arises when stated willingness-to-pay may be a poor guess, even though it may be the best guess individuals have of their true willingness-to-pay. Individuals often face the problem of trying to form judgments about the gains from a purchase in settings where the link between the

¹⁶For the correct use of a benefit-cost calculation, we need to be considering the marginal project for finding the optimum. With many projects under consideration, and a nonoptimal starting point, one does not get the right answer by asking about many projects independently and carrying out all that pass the test (Hoehn and Randall, 1989).

commodity and utility is hard to evaluate. One example is the grade of gasoline to buy, assuming that one wants to minimize cost per mile. In the case of environmental amenities, individuals may have a derived demand based on their beliefs about the relationship between the amenity and variables they really care about. For example, they may care about the survival of a species and not know about the range of natural variation in population size, about the probability of survival as a function of population size, nor about the effect of environmental damage on population size. Such derived preferences may be a poor guide to policy; it may be more informative to have expert evaluation of the consequences of an environmental change than to consult the public directly about environmental damage.

The issues just discussed were based on the hypothesis that stated willingness-to-pay is a measure of an individual preference over an outcome. Under the hypothesis that responses reflect casual benefit-cost evaluations rather than preferences, it would be inappropriate to add any other benefits to those coming from a contingent valuation survey since such benefits are presumably included by the respondents, however imperfectly, in their benefit-cost analyses. But if contingent valuation is just a survey of benefit-cost estimates, rather than preferences, it might be better to have a more careful analysis done by people knowing more about environmental issues and about the principles of benefit-cost analysis. Moreover, if responses are benefit-cost estimates rather than preferences, they do not measure a compensable loss in damage suits.

The embedding effect is supportive of the hypothesis that responses are primarily determined by warm glow. If respondents get pleasure from thinking of themselves as supportive of the environment, the willingness-to-pay for this warm glow is not part of the gain from a *particular* environmental project—unless there are no cheaper ways of generating the warm glow. That is, if an individual wants to see the government do at least one environmental project (or n projects) a year in order to feel “environmentally supportive,” the person would support one project, but not any particular project. Moreover, if different samples are asked about different projects, the responses will appear to support many projects, even though the warm glow comes from the desire to support a single project.

An illustration of this view comes from the fact that when individuals are asked simultaneously about many projects, stated willingness-to-pay is far below the sum of stated willingnesses-to-pay from asking about the projects separately. For example, Kemp and Maxwell (1993) asked one group for willingness-to-pay to minimize the risk of oil spills off the coast of Alaska, and found a mean stated willingness-to-pay of \$85 (with a 95 percent confidence interval of \pm \$44). Then they asked a different sample for willingness-to-pay for a broad group of government programs, followed by asking these people to divide and subdivide their willingness-to-pay among the separate programs. By the time they reached minimizing the risk of oil spills off the coast of Alaska,

they found a mean of \$0.29 (with a 95 percent confidence interval of \pm \$0.21).

These findings make little sense if responses are measures of preferences, and considerable sense if the response is primarily a warm glow effect from a desire to express support for protecting the environment. In the latter circumstance, we would expect little warm glow for any single project in a context where respondents are asked about many government projects affecting the environment. Therefore warm glow may need to be purged from stated willingness-to-pay even if (as witnessed by charitable contributions) people really are willing to pay for some warm glow.¹⁷

A different complication arises if people do not really care about the resource, but care about the activity that might harm a resource. For example, the stated willingness-to-pay to clean up a natural oil seepage might be zero while the stated willingness-to-pay to clean up a man-made oil spill is positive. This outcome is the flip side of the "protest zero," where people state no willingness-to-pay to repair environmental damage that they feel is someone else's responsibility. As noted earlier, it is standard practice to consider this zero not to be an accurate measure of preferences, on the assumption that people care about the resource.

Survey results suggest that many answers are heavily influenced by concern about actions, not resources. For example, Desvousges et al. (1993) find a large stated willingness-to-pay to save small numbers of common birds. The finding seems much more likely to reflect a feeling that it is a shame that people do things that kill birds rather than a preference over the number of birds. Concern over the actions of others is different from concern about the state of the environment. Concern about actions is conventionally part of the basis of punitive damages, but not compensatory damages. That is, deliberately or recklessly destroying the property of others opens one up to liability for compensatory damages for the value of the property destroyed and also punitive damages. On the other hand, the legal system does not compensate people who are upset that others engage in actions such as reading *Lady Chatterley's Lover*. When and how such concerns should affect public policy is a complex issue, one not explored here.

One complication from the perspective of benefit-cost analysis is that preferences over acts (as opposed to states of the world) do not provide the consistency that is necessary for consistent economic policy. For example, if people are willing to pay to offset an act, then proposing and not doing an act appears to generate a welfare gain. For example, consider the warm glow from blocking development of a wilderness area. If one proposes two projects and

¹⁷In the context of the bird study by Desvousges et al. (1993), Kahneman (personal communication) has proposed to purge the warm glow by extrapolating willingness-to-pay as a function of birds saved back to zero and then subtracting this amount from the estimate of willingness-to-pay at any particular level of birds. This approach involves a curve-fitting extrapolation and the assumption that warm glow is totally insensitive to the magnitude of the problem, an assumption that is probably not completely correct.

has one blocked, are people better off (from the warm glow) than if one project is proposed and happens? Does this imply that the government would do good by proposing projects that it does not mind seeing blocked? More generally, the relationship of benefit-cost analysis and Pareto optimality has been developed and is understood in a setting where preferences are defined over resources.

We note that under the hypothesis of Kahneman and Ritov (1993), responses to contingent valuation surveys are expressions of attitudes toward public goods that the respondents are required to state in dollar terms. Responses are then not measures of willingness-to-pay and provide no quantitative basis for estimates of environmental damages, although like polls generally, they do alert the government about concerns of the public.

The “Some Number is Better than No Number” Fallacy

We began this essay by arguing that stated willingnesses-to-pay from contingent valuation surveys are not measures of nonuse preferences over environmental amenities. We then considered some of the welfare implications of treating the responses as if they were a measure of nonuse preferences when they were generated by different considerations. We concluded that such welfare analysis would not be a guide to good policy. Our conclusion is often challenged by the common Washington fallacy that even if stated willingness-to-pay is inaccurate, it should be used because no alternative estimate exists for public policy purposes. Put more crudely, one hears the argument that “some number is better than no number.”¹⁸ This argument leads to the claim that it is better to do benefit-cost studies with stated willingness-to-pay numbers, despite inaccuracy and bias, rather than use zero in the benefit-cost analysis and adjust for this omission somewhere else in the decision-making process.

To evaluate this argument, one needs a model of the determination of government policy.¹⁹ Ideally, one would like to carry out a number of government decisions twice: once using zero in the benefit-cost study, and a second time using stated willingness-to-pay, with associated adjustments of the decision process in recognition of the inclusion or omission of a contingent valuation number. Such a comparison would recognize that much more input goes into government decisions than just the benefit-cost study. That is, the comparison is not between relying on contingent valuation and relying on Congress, but between relying on Congress after doing a contingent valuation study and

¹⁸The history of economic policy awaits an investigation, similar to the famous study of the sociologist R. K. Merton on the history of Newton’s “on the shoulders of giants” remark, to trace the lineage of the “some number is better than no number” fallacy.

¹⁹One can also consider how a social welfare maximizing planner might use the information in contingent valuation surveys. There is useful information if people are expressing preferences that are not otherwise accessible to the planner. However, if the other hypotheses are the correct description of the bases of willingness-to-pay responses, then the planner would not be receiving useful information. Treating the responses as measures of what they do not measure would mislead such a planner.

relying on Congress without doing a contingent valuation study.²⁰ Thus one is asking whether inclusion of such survey results tends to improve the allocation process, even if the numbers are not reliable estimates of the preferences called for by the theory. Similarly, one can ask whether the combination of fines and damage payments will result in more efficient decisions to avoid accidents with or without a contingent valuation estimate of nonuse value.

Judge Stephen Breyer (1993) has recently reviewed government responses to public perceptions of risk. Since he feels that public perceptions of risk are inaccurate and that Congress is responsive to these public perceptions, Breyer wants to increase the role of administrative expertise in designing public policy to deal with risks. A similar situation seems to exist with respect to contingent valuations of nonuse value. If we conclude that contingent valuation is really an opinion poll on concern about the environment in general, rather than a measure of preferences about specific projects, public policy is likely to do better if the concern is noted but expert opinion is used to evaluate specific projects and to set financial incentives to avoid accidents. One could hope for a more consistent relative treatment of alternative natural resources in this way.

In both economic logic and politics, we expect that using contingent valuation in decision making about the environment would soon be extended to other policy arenas where existence values are equally plausible. We do not expect that policy would be improved by using contingent valuation to affect the levels and patterns of spending for elementary school education, foreign aid, Medicaid, Medicare, AFDC, construction of safer highways, medical research, airline safety, or police and fire services. Yet people have concerns for others in all of these areas that parallel their concern for the environment.

Concern for other people naturally includes concern about their jobs. Thus, in considering rules that limit economic activity to protect the environment, it is as appropriate to include a contingent valuation of existence value for destroyed jobs as the one for protection of the environment. The fact that jobs may be created elsewhere in the economy does not rule out concern about job destruction *per se*. These possible extensions of the use of contingent valuation increase the importance of considering the "some number is better than no number" fallacy.

Referenda

We have heard the argument that if referenda are legitimate, so too is contingent valuation. That is, one can consider a contingent valuation survey to be a forecast of how voters would respond to a binding referendum. This

²⁰The results of a contingent valuation survey are not binding. Thus a respondent who was behaving strategically would select a response that reflected his or her belief in how the results of the survey would affect actual outcomes. Thus we do not understand how the NOAA Panel could conclude that with a dichotomous choice question there is no strategic reason for the respondent to do otherwise than answer truthfully.

perspective raises the same issues considered above. How should we decide how to interpret the bases of how people vote in referenda? Since different bases imply different appropriate uses of the responses, how should voting responses be used for economic analysis? Moreover, the necessity of calibration remains, since no obvious reason exists for people necessarily to vote the same in binding and nonbinding referenda. And, as in the previous section, we can ask whether we think we get better policies with or without such surveys.

It is interesting to consider issues raised by polls about actual referenda, as well as by the referenda themselves. Sometimes polls are accurate predictors of voting outcomes; sometimes, they are not, even when they are taken close to election day. Sometimes, repeated polls about the same referendum find very large changes in expressed intentions as a referendum campaign proceeds.

Magleby (1984) has analyzed statewide polls in California and Massachusetts for which at least three separate surveys were done. In some cases, the polls show roughly the same margin over time. Magleby calls these "standing opinions" and believes that this stability comes from the deep attachment to their opinions that voters hold on some controversial issues such as the death penalty and the equal rights amendment. In some cases, the polls show significant changes in the margin of preferences, but no change in the side that is ahead. Magleby calls these "uncertain opinions." Examples of such votes involve handgun registration and homosexual teachers. In some cases, significant changes in voting intentions occur as the campaign proceeds, with victory in the actual election going to the side that had at one time been far behind. Magleby calls these outcomes "opinion reversals." For example, in a referendum for flat rate electricity, a February poll showed 71 percent in favor, 17 percent opposed, and 12 percent undecided. The actual vote was 23 percent in favor, 69 percent opposed and 7 percent skipping this question. Other examples of such votes are a state lottery and a tax reduction measure. In his analysis of 36 propositions in California, Magleby found that on 28 percent of the issues, voters held standing opinions, on 19 percent voters had uncertain opinions, and on 53 percent he found opinion reversals. That is, in a majority of cases, early opinion polls were not good predictors of election outcomes. Moreover, they were not even good predictors of later opinion polls, after the campaign had run for some time.

It seems to us that responses to contingent valuation questionnaires for a single environmental issue are likely to be based on little information, since there is limited time for presentation and digestion of information during a contingent valuation survey. This conclusion suggests that the results of such surveys are unlikely to be accurate predictors of informed opinions on the same issues if respondents had more information and further time for reflection, including learning of the opinions of others. Such surveys are therefore unlikely to be a good basis for either informed policy-making or accurate damage assessment.

Even if a contingent valuation survey were a good predictor of an actual referendum, one can also question the use of actual referenda to obtain

economic values. Considerable skepticism exists about the extent to which voting on a referendum represents informed decision making (see, for example, Magleby, 1984). In the functioning of a democracy, it may be more important to place some powers directly with the voters, rather than with their elected representatives, than to worry about the quality of decision making by voters.²¹ However, incorporating contingent valuation survey responses in benefit-cost analyses or judicial proceedings does not seem to have a special role in enhancing democracy. In the looser context of legislative debate, such opinion polls may have a role to play, although the net value of that role is unclear.

NOAA Panel Evaluation of Contingent Valuation

In light of the controversy and the stakes involved, the National Oceanic and Atmospheric Administration recently appointed a prestigious panel to consider the reliability of contingent valuation studies of nonuse values in damage suits.²² The panel's Report (NOAA, 1993) begins with criticisms of contingent valuation. In discussing the alleged inconsistency of some results with rational choice, the Report states (p. 4604) that: "some form of internal consistency is the least we would need to feel some confidence that the verbal answers correspond to some reality." The Report also addresses the need for rationality (p. 4604).

It could be asked whether rationality is indeed needed. Why not take the values found as given? There are two answers. One is that we do not know yet how to reason about values without some assumption of rationality, if indeed it is possible at all. Rationality requirements impose a constraint on the possible values, without which damage judgments would be arbitrary. A second answer is that, as discussed above, it is difficult to find objective counterparts to verify the values obtained in the response to questionnaires.

In discussing "warm glow" effects, the Report recognizes the claim that contingent valuation responses include a warm glow. They write (p. 4605): "If this is so, CV [contingent valuation] responses should not be taken as reliable estimates of true willingness to pay."

The Report states that the burden of proof of reliability must rest on the survey designers. It states (p. 4609) that a survey would be unreliable if there were "[i]nadequate responsiveness to the scope of the environmental insult," as

²¹The allocation of a decision directly to the voters, rather than indirectly through the choice of elected representatives, and the form in which referenda are put to voters are both methods of agenda control. In many settings, design of the agenda has large effects on voting outcomes.

²²Kenneth Arrow (co-chair), Robert Solow (co-chair), Edward Leamer, Paul Portney, Roy Radner, and Howard Schuman.

occurred in the embedding examples we have discussed. Unfortunately, the Panel did not elaborate on how to test for reliability.²³ We interpret the view they express to call for testing of the internal consistency of responses to the same survey instrument with different levels of environmental problem and policy successes. The Report cites no existing study that has passed such internal consistency tests.

The Report presents a set of guidelines which would define an “ideal” contingent valuation survey (and are summarized in Portney’s paper in this issue). The Report asserts (p. 4610) that studies meeting such guidelines can produce estimates “reliable enough to be the starting point” of a judicial process of damage assessment. The Report offers no reason for reaching this conclusion, although the finding that surveys that do not meet their guidelines may be biased is not a basis for concluding that surveys that do meet their guidelines are not biased. In particular, they state no reason for reaching the conclusion that following their guidelines implies that responses are not dominated by a “warm glow.” The Panel does not explicitly call for testing whether a survey done according to their guidelines is reliable. In particular, they do not mention a need to check the internal consistency of responses. Nor do they explain their conclusion that the inconsistencies between stated willingness-to-pay and economic theory come from survey design issues and would go away if the survey had followed their guidelines.

Conclusion

We believe that contingent valuation is a deeply flawed methodology for measuring nonuse values, one that does not estimate what its proponents claim to be estimating. The absence of direct market parallels affects both the ability to judge the quality of contingent valuation responses and the ability to calibrate responses to have usable numbers. It is precisely the lack of experience both in markets for environmental commodities and in the consequences of such decision that makes contingent valuation questions so hard to answer and the responses so suspect.

We have argued that internal consistency tests (particularly adding-up tests) are required to assess the reliability and validity of such surveys. When these tests have been done, contingent valuation has come up short. Contingent valuation proponents typically claim that the surveys used for these tests were not done well enough. Yet they have not subjected their own surveys to such tests. (We note that Hanemann does not address the question of which split-sample internal consistency tests, if any, he thinks a contingent valuation survey needs to pass.) There is a history of anomalous results in contingent valuation surveys that seems closely tied to the embedding problem. Although

²³ Nor, we add, do Portney or Hanemann in this symposium.

this problem has been recognized in the literature for over a decade, it has not been solved. Thus, we conclude that current contingent valuation methods should not be used for damage assessment or for benefit cost analysis.

It is impossible to conclude definitely that surveys with new methods (or the latest survey that has been done) will not pass internal consistency tests. Yet, we do not see much hope for such success. This skepticism comes from the belief that the internal consistency problems come from an absence of preferences, not a flaw in survey methodology. That is, we do not think that people generally hold views about individual environmental sites (many of which they have never heard of); or that, within the confines of the time available for survey instruments, people will focus successfully on the identification of preferences, to the exclusion of other bases for answering survey questions. This absence of preferences shows up as inconsistency in responses across surveys and implies that the survey responses are not satisfactory bases for policy.

■ *The authors want to thank Bernard Saffran and four editors for helpful comments.*

References

- Andreoni, James**, "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *Journal of Political Economy*, December 1989, 97, 1447-58.
- Bishop, R. C., and T. A. Heberlein**, "Measuring Values of Extramarket Goods: Are Indirect Measures Biased?," *American Journal of Agricultural Economics*, December 1979, 61, 926-30.
- Breyer, Stephen**, *Breaking the Vicious Circle: Toward Effective Risk Regulation*. Cambridge: Harvard University Press, 1993.
- Carson, Richard T., et al.**, "A Contingent Valuation Study of Lost Passive Use Values Resulting From the Exxon Valdez Oil Spill," A Report to the Attorney General of the State of Alaska, 1992.
- Desvousges, W. H., et al.**, "Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability." In Hausman, J., ed., *Contingent Valuation: A Critical Assessment*. Amsterdam: North Holland Press, 1993, 91-164.
- Diamond, P. A.**, "Testing the Internal Consistency of Contingent Valuation Surveys," working paper, MIT, 1993.
- Diamond, P. A., and J. A. Hausman**, "On Contingent Valuation Measurement of Non-use Values" In Hausman, J., Ed., *Contingent Valuation: A Critical Assessment*. Amsterdam: North Holland Press, 1993, 3-38.
- Diamond, P. A., J. A. Hausman, G. K. Leonard, and M. A. Denning**, "Does Contingent Valuation Measure Preferences? Experimental Evidence." In Hausman, J., ed., *Contingent Valuation: A Critical Assessment*. Amsterdam: North Holland Press, 1993.
- Dickie, Mark, Ann Fisher, and Shelby Gerking**, "Market Transactions and Hypothetical Demand Data: A Comparative Study," *Journal of the American Statistical Association*, March 1987, 82, 69-75.
- Duffield, John W., and David A. Patterson**, "Field Testing Existence Values: An Instream Flow Trust Fund for Montana Rivers," mimeo, University of Montana, 1992.
- Hausman, J. A.**, *Contingent Valuation: A Critical Assessment*. Amsterdam: North Holland Press, 1993.
- Hausman, J. A., and G. Leonard**, *Contingent Valuation and the Value of Marketed Commodities*. Cambridge: Cambridge Economics, 1982.
- Hoehn, John, and Alan Randall**, "Too Many Proposals Pass the Benefit Cost Test," *American Economic Review*, June 1989, 79, 544-51.
- Kahneman, Daniel**, "Comments on the Contingent Valuation Method." In Cummings, Ronald G., David S. Brookshire, and William D. Schulze, eds., *Valuing Environmental Goods: A State of the Arts Assessment of the Contingent Valuation Method*. Totowa: Rowman and Allanheld, 1986, 185-94.

Kahneman, Daniel, and Jack L. Knetsch, "Valuing Public Goods: The Purchase of Moral Satisfaction," *Journal of Environmental Economics and Management*, January 1992, 22, 57-70.

Kahneman, Daniel and Ilana Ritov, "Determinants of Stated Willingness to Pay for Public Goods: A Study in the Headline Method," mimeo, Department of Psychology, University of California, Berkeley, 1993.

Kemp, M. A. and Maxwell, "Exploring a Budget Context for Contingent Valuation Estimates," In Hausman, J., ed., *Contingent Valuation: A Critical Assessment*, Amsterdam: North Holland Press, 1993, 217-70.

Loomis, John, John Hoehn, and Michael Hanemann, "Testing the Fallacy of Independent Valuation and Summation in Multi-part Policies: An Empirical Test of Whether 'Too Many Proposals Pass the Benefit Cost Test,'" mimeo, University of California, Davis, 1990.

Magleby, David B., *Direct Legislation, Voting on Ballot Propositions in the United States*. Baltimore and London: The Johns Hopkins University Press, 1984.

McFadden, Daniel, and Gregory K. Leonard, "Issues in the Contingent Valuation of Environmental Goods: Methodologies for Data Collection and Analysis." In Hausman, J., ed., *Contingent Valuation: A Critical Assessment*. Amsterdam: North Holland Press, 1993.

Milgrom, P., "Is Sympathy an Economic Value?," In Hausman, J., ed., *Contingent Valuation: A Critical Assessment*. Amsterdam: North Holland Press, 1993, 417-42.

Mitchell, Robert Cameron and Richard T. Carson, *Using Surveys to Value Public Goods*. Washington D. C.: Resources for the Future, 1989.

National Oceanic and Atmospheric Administration, 1993, "Report of the NOAA Panel on Contingent Valuation," *Federal Register*, 1993, 58, 10, 4602-14.

National Oceanic and Atmospheric Administration, "Natural Resource Damage Assessments; Proposed Rules," *Federal Register*, 1994, 59, 5, 1062-191.

Neill, Helen, R., et al., "Hypothetical Surveys and Real Economic Commitments," Economics Working Paper B-93-01, Department of Economics, College of Business Administration, University of South Carolina, 1993.

Samples, Karl C., and James R. Hollyer, "Contingent Valuation of Wildlife Resources in the Presence of Substitutes and Complements." In Johnson, Rebecca L., and Gary V. Johnson, eds., *Economic Valuation of Natural Resources: Issues, Theory and Applications*. Boulder: Westview Press, 1990, 177-192.

Schkade, D. A., and J. W. Payne, "Where do the numbers come from? How people respond to Contingent Valuation Questions." In Hausman, J., ed., *Contingent Valuation: A Critical Assessment*. Amsterdam: North Holland Press, 1993, 271-304.

Schulze, William, D., et al., "Contingent Valuation of Natural Resource Damages Due to Injuries to the Upper Clark Fork River Basin," State of Montana, Natural Resource Damage Program, 1993.

Seip, Kalle, and Jon Strand, "Willingness to Pay For Environmental Goods in Norway: A Contingent Valuation Study With Real Payment," *Environmental and Resource Economics*, 1992, 2, 91-106.

Tolley, George S., et al., "Establishing and Valuing the Effects of Improved Visibility in the Eastern United States," Report to the U.S. Environmental Protection Agency, Washington, D.C., 1983.

Urban, Glen L., Gerald M. Katz, Thomas E. Hatch, and Alvin J. Silk, "The ASSESSOR Pre-Test Market Evaluation System," *Interfaces*, 1983 13, 38-59.