New Work on the Benefits of Environmental Regulation
(Albert McGartland, USEPA, presiding)

The Issue of Scope in Contingent Valuation Studies

Richard T. Carson and Robert C. Mitchell

Natural resource damage assessment under various federal and state laws has given research in non-market valuation high visibility. This paper is about a consistency check suggested by the recent NOAA Blue Ribbon panel as one basis for evaluating contingent valuation (CV) surveys. This check relates to the responsiveness of valuation estimates to changes in the scope or size in an environmental commodity. Critics of the CV method have recently argued, on the basis of several empirical studies, that the CV method is incapable of demonstrating scope and is therefore too unreliable to obtain useful information for natural resource damage assessments.[1] In this paper we question their conclusion on the basis of some evidence from the larger CV literature and a reanalysis of the studies conducted by the critics. While the issues of scope and sequence are often confounded under the ill-defined rubric of embedding, due to space considerations, we confine our remarks in this paper solely to the issue of scope.[2]

Lack of Sensitivity to Scope

The first empirically based claim that CV surveys were likely to be insensitive to the scope of the good being valued was put forward by Kahneman. Kahneman displayed a graph with results from a telephone survey involving three subsamples, two which were asked about cleaning up lakes in specific regions of Ontario (i.e., Muskoka and Haliburton) and one which was asked about cleaning up all lakes in Ontario.[3] From this graph, Kahneman argued that respondents were willing to pay (WTP) only slightly more for cleaning up the lakes in one part of Ontario than for cleaning up all the lakes in Ontario. Kahneman contends that this lack of sensitivity to scope is almost inevitable because respondents are expressing "ideological values," "receiving a warm glow," or "purchasing moral satisfaction," terms which are used interchangeably. Other critics of contingent valuation have used terms such as a "good cause dump" or "lack of crystallized preferences."

In contrast to Kahneman's conclusion stands the result of Carson and Mitchell (1993a) who compared an estimate of WTP for a change from boatable to fishable quality water on the Monongahela River from a sample taken from that area (Smith and Desvousges) to WTP from a national sample for the same improvement in national water quality using a very similar survey instrument (Mitchell and Carson 1986a). After correcting for differences in sample characteristics (which narrowed the difference between the two estimates), we obtained an estimate for Pittsburgh area residents of $26 for improvement in Monongahela River quality and an estimate of $68 for the same improvement in national water quality. The difference is large in percentage terms (over 160%) and highly significant ($t = 4.88$, $p < .01$). A benefits-transfer exercise shows that a plausible set of assumptions maps one estimate to the other.

Richard T. Carson is an associate professor in the Department of Economics, University of California, San Diego; Robert Cameron Mitchell is a professor in the Graduate School of Geography, Clark University.

[1] See, for instance, the volume from the Exxon sponsored symposium on contingent valuation edited by Hausman and testimony on behalf of several oil and chemical companies by Peter Diamond, William Desvousges, Jerry Hausman, Daniel Kahneman, Daniel McFadden, Steven Shavell and others before the NOAA Blue Ribbon Panel on Contingent Valuation co-chaired by Kenneth Arrow and Robert Solow.

[2] Carson and Mitchell (1993b) formally look at the structure of experiments designed to look at sequence and scope issues and the testable hypotheses that arise from them. Scope issues involve comparisons of values for goods which differ in a quantitative or qualitative fashion. Sequence issues involve how the value of a good changes with the order in which it is valued and the nature of the goods offered before it.

[3] There have been a reasonably large number of contingent valuation surveys which demonstrate that respondents are "internally" consistent. By that we mean within a single survey respondents value larger goods more than smaller ones. Kahneman's experiment differed from these studies in that it was an "external" comparison of whether different subsamples of respondents valued different goods differently.

Amer. J. Agr. Econ. 75 (December 1993): 1263–1267

If one accepts the line of argument advanced by Kahneman and other CV critics, CV studies would be unable to get such a result. Kahneman and Knetsch are inclined to dismiss the result on the grounds that "Given the uncertainties of comparisons across studies and sampling areas, this evidence against embedding is not persuasive." This ignores the fact that our comparison is based on the results of two surveys which used large samples, experienced in-person interviewers, and well tested survey instruments which were explicit about the nature of the good being valued and its method of provision and payment. In contrast, the results reported by Kahneman, Kahneman and Knetsch, Diamond et al., and Desvousges et al. which are the typically cited evidence for respondents being insensitive to the scope of the good being valued, are based on short telephone surveys or self-administered mall-intercept surveys. Further, the survey scenarios used in those experiments are generally unclear about the nature of the good being valued, the manner of the good's provision, or the payment obligation.

Careful examination of the conclusions made by these researchers shows several of them to be questionable or wrong. For example, taking the data revealed in the graph in Kahneman, we find that his respondents had a median WTP that is approximately 50% higher when queried about all Ontario lakes than when queried about the Muskoka lakes. For any reasonable sample size, we estimate the difference between these two WTP distributions is highly significant. While it is possible to argue about what the magnitude of the difference should be, the difference is clearly not "only slightly higher."

An examination of the Diamond et al. scope result also serves to illustrate this point. In that paper, the authors claim that the stated WTP of three separate subsamples for three different wilderness areas varying in size from 700,000 to 1,300,000 acres is the same. They report a p-value of 0.68 for their test of no difference in the WTP amounts. However, an examination of the mean WTP amounts for the three areas ($49.85, Selway Bitterroot [1,300,000 acres]; $36.57, Bob Marshall [1,000,000 acres]; $29.90, Washakie [700,000 acres]), suggests that the amounts line up as one would expect them to if the respondents were sensitive to the size of the wilderness areas being valued. Indeed, comparing the smallest and largest of the two areas we see an 86% difference in size and a 67% difference in the mean WTP. Carson and Flores (1992) show that a simple t-statistic for this pair of wilderness areas yields a t-value of 1.22, which for a one-tailed test has a p-value of 0.11. For the cleaned data set reported by Diamond et al., which a CV researcher would likely use, the t-value is 1.96 ($p = 0.025$). How did Diamond et al. get a test statistic which suggested the opposite conclusion? They used a non-parametric test which had little power with respect to the relevant hypothesis.

The Desvousges et al.'s (1992) migratory waterfowl experiment appears to be the most credible demonstration that CV respondents are insensitive to scope. Desvousges et al. claim that their three subsamples of respondents who filled out self-administered surveys in Atlanta shopping malls, gave close to the same values for preventing the deaths of 2,000, 20,000, and 200,000 waterfowl from oil ponds in the Rocky Mountain flyway. In contrast to the Diamond et al. paper, inspection of the information available in Desvousges et al. suggests reasonable support for a claim that there are not large differences in the WTP amounts for the three treatments. However, there is strong reason to question whether the differences in the quantities being offered are really the two orders of magnitude apart they appear to be. The respondents in the 2,000 bird treatment were told that they would be saving much less than 1% of the waterfowl using the flyway, in the 20,000 bird treatment that they would be saving less than 1% of the waterfowl, and the 200,000 bird treatment that they would be saving about 2% of the waterfowl.

If respondents based their answers on the percentage changes, which is likely, given that the birds at issue were quite distant from Atlanta, one might not expect to see large differences in WTP for the programs. The strongest signs of problems with the study, however, come from two other sources. The first is internal, Desvousges et al. have little success in explaining the variance in the amounts given by respondents.[4] The second is from Schkade and Payne who used the Desvousges et al. survey instrument in a shopping mall intercept setting, but asked respondents to say what they were thinking about when filling out the questionnaire. Under these conditions, respondents were forced to pay more attention to the questions. Instead

[4] Desvousges et al. performed a second shopping mall experiment involving prevention of oil spills. This experiment also shows multiple problems such as the failure of respondents to take the payment obligation seriously and an inability to explain the variation in the responses.

of showing no difference between the median WTP amounts for 2,000 and 200,000 bird treatments, Schkade and Payne show a difference of 100%.[5]

The Kahneman and Knetsch embedding experiment is often cited as the principal evidence that CV respondents are insensitive to the scope of the good being valued. That experiment has a different if all too common problem—the goods offered to respondents are never clearly defined. Due to space constraints, we refer the reader to Smith (1992) for more insight into the particular problems with that paper.[6]

Sensitivity to Scope

In addition to the study already discussed, there are a number of other CV studies that also have shown sensitivity to scope. We will briefly summarize some of this evidence, concentrating on those studies in which we were involved. These studies span a fairly wide range of environmental amenities: (*a*) different levels of low-level risk from a drinking water contaminant, (*b*) different levels of water reliability, (*c*) air quality changes with different characteristics, and (*d*) prevention of different mining impacts in the Kakadu Conservation Zone.

The drinking water study was conducted for the U.S. Environmental Protection Agency (Mitchell and Carson 1986b) and involved trihalomethane, a low level carcinogen which is formed when organic materials in surface water react with chlorine during the process of disinfecting drinking water from biological agents. The risk from not chlorinating water is much greater than the risk from the trihalomethanes. The risk from the trihalomethanes can be reduced, although not eliminated, by a carbon filtration system. Because changes in low-level risks are difficult to communicate to people, a considerable effort was put into developing a way to explain these risk reductions in terms that could be understood by the average household. This effort included an experiment to assess the success of our effort. A random sample of respondents in a small town in southern Illinois were interviewed in-person and randomly assigned to one of two treatments. In the first treatment, respondents were asked to value a risk reduction of 0.04 annual deaths per 100,000. In the second treatment respondents were asked to value a risk reduction of 2.43 annual deaths per thousand. Mean WTP of the 117 respondents in treatment A was \$3.78 while the mean WTP of the 110 respondents in treatment B was \$15.23. A *t*-test of the difference between these amounts is -4.16 ($p < .01$). Thus the hypothesis of lack of sensitivity of scope is clearly rejected in spite of a fairly small sample.[7] Further evidence can be seen in the significant difference ($p < 0.01$) in the percentage of zeros received for the two treatments: 87% for treatment A and 58% for treatment B.

The next study is a water reliability study Carson and Mitchell conducted for the Metropolitan Water District of Southern California. This study estimated household WTP to avoid different possible water shortage scenarios. Two thousand respondents were interviewed by telephone and randomly assigned to one of two versions of the questionnaire. In version I of the questionnaire, respondents were asked about their annual WTP to avoid a 10–15% water shortage; the median estimate was \$83. In version II of the questionnaire respondents were asked about their WTP to avoid a 30–35% water shortage once every five years; the median estimate was \$113. The t-statistic on the difference from the estimated likelihood function is -4.64 ($p < 0.01$).

Version I respondents were subsequently asked about their WTP to avoid two 10–15% water shortages in a five year period; the median WTP estimate was \$151. Version II respondents were asked about their WTP to avoid one 30–35% water shortage and one 10–15% water shortage in a five year period. Again, the difference between the WTP for the two water shortage programs was highly significant ($t = -4.92$; $p < 0.01$).

We now focus on a study (Mitchell, Carson, and Ruud) involving air quality improvements conducted for the Electric Power Research Institute. In that study 130 Cincinnati residents were

[5] Schkade and Payne report a *p*-value of .184 for a Kruskal-Wallis test on the median differences using a sample of only 105 observations split between the three treatments. They report a *p*-value of .113 after dropping the zeros in the sample.

[6] It is worth noting that the anomalies claimed by Kahneman and Knetsch for their major experiment disappear if one drops the group I column containing the ill-defined good "significant improvements in environmental services" and its subsequent decomposition.

[7] The survey went on to ask respondents in each treatment about two additional higher levels of risk. Comparison of mean willingness to pay across treatments for these two additional risk reductions shows that both are also significant at the 0.01 level.

interviewed in-person and randomly assigned one of two treatments. In treatment A, respondents were asked about their WTP for 9 different visibility improvements. In treatment B, respondents were asked about their WTP for 9 different air quality improvements some of which involved both visibility improvements and health improvements. We found identical means in the two programs which were the same in the two treatments (5 visibility days; 0 health days) and found significant differences where ratio of health days to visibility days was large. For instance, in the comparison of (3 visibility days, 0 health days) in the first treatment to (3 visibility days, 3 health days) in the second treatment we found a mean WTP of \$8 versus \$29 in the second treatment ($t = -3.60$; $p < 0.01$). Likewise, we found significant differences when comparing (6 visibility days, 0 health days) to (6 visibility days, 3 health days) ($t = -3.70$; $p < .01$) and (7 visibility days, 0 health days) to (7 visibility days, 2 health days) ($t = -2.50$, $p < .01$). Equally important, we did not find significant differences when we compared programs with a very low ratio of health to visibility days to programs in the treatment with just visibility days.

The last study is the Australian Kakadu study (Carson, Wilks, and Imber). Here 2,000 respondents from a national sample were randomly assigned to one of two treatments. In the first treatment, respondents were asked to value, in a discrete choice context, preventing a set of mining impacts on the Kakadu National Park roughly consistent with the mining industry's view. In the second treatment, respondents were asked to value a set of mining impacts roughly consistent with the environmental groups' view. The principal difference in the impacts centered on potential effects on the Park which surrounds the potential mine site. The median WTP was A\$60 for the minor impact scenario and A\$124 for the major impact scenario. This difference of over 100% ($t = -3.95$; $p < .01$) was highly significant.

The tests we report come either from studies where policy makers requested benefit estimates for different scenarios or from pilot study tests we undertook of respondent understanding when the good being valued was particularly difficult to convey clearly. While tests of scope are not common in the literature because of their expense, we are certainly not the only CV researchers to have demonstrated that CV estimates are sensitive to the scope of the offered good (e.g., Loomis, Lockwood, and DeLacy; Magnussen; Romer and Pommerehne).

Concluding Remarks

The notion that the lack of sensitivity by respondents to the characteristics of the good being valued might pose a fatal flaw to contingent valuation seemed remote to most CV researchers until the recent debate. That is because CV studies have produced a wide range of estimated values for goods and casual inspection suggests those values tend to vary with the nature of the good being offered respondents. To make an extreme comparison, Carson et al. (1992) found respondents were willing to pay on average less than \$1 to improve visibility in the Grand Canyon on ten poor weather days during the winter, while Randall and Kriesel found that respondents were willing to pay almost \$700 for substantial improvements in several national environmental programs. Some CV studies looking at human health issues (e.g., Thompson) have found estimates for substantial improvements in a patient's health state averaging around \$5,000 for a sample of patients with rheumatoid arthritis with patients in the highest income category willing to pay an order of magnitude more than those in the lowest income category. Taking a broader view, Walsh, Johnson, and McKean performed a metaanalysis of 129 CV estimates involving outdoor recreation made between 1968 and 1988 and found these estimates were sensitive to site quality, region of the country, and type of activity. Smith (forthcoming) finds similar results with respect to the visibility change measured in a number of contingent valuation air quality studies.

We wish to emphasize strongly that we are not saying that CV estimates are immune from issues related to the scope of the commodity. Mitchell and Carson (1989) devote substantial attention to the specific issues of part-whole and symbolic bias and to the larger issue of amenity misspecification.[8] From our perspective, the issue of sensitivity to scope has always been one of survey design and administration. If there is a central problem with contingent valuation it is that people will try to answer whatever question is put to them. The quality of the response is crucially dependent on the information provided

[8] The recent Fischhoff et al. experiments provide some excellent examples of the types of problems with respondent perceptions and understanding that can occur when using short telephone and self-administered questionnaires. In particular, it sheds some light on respondent discounting of the likelihood that large projects can be delivered and a corresponding reduction in respondent WTP for projects in such contexts.

to and perceived by the respondent, and the seriousness with which the respondent takes the survey interview. In this regard, we find it useful to draw a clear distinction between a WTP question and a CV response. That is, simply asking respondents for information about their WTP for a public good does not ensure that one gets a valid CV response. For that to happen, respondents must (*i*) clearly understand the characteristics of the good they are being asked to value; (*ii*) find the CV scenario elements related to the good's provision plausible; and (*iii*) answer the CV questions in a deliberate and meaningful manner.

References

Carson, R. T., M. B. Conaway, W. M. Hanemann, R. J. Kopp, K. Martin, R. C. Mitchell, S. Presser, and P. A. Ruud. "Valuing Winter Visibility Improvement in the Grand Canyon." Paper presented at Association of Environmental and Resource Economists Meeting. New Orleans, January, 1992.

Carson, R. T., and N. E. Flores. "A Second Look at 'Does Contingent Valuation Measure Preferences?: Experimental Evidence' How Evident is the Evidence?" Comment No. 44 submitted to the NOAA Blue Ribbon Panel on Contingent Valuation. October 1992.

Carson, R. T., and R. C. Mitchell. "The Value of Diamonds and Water." Paper presented at European Association of Environmental and Resource Economists Meeting. Stockholm, June 1991.

——. "The Value of Clean Water: The Public's Willingness to Pay for Boatable, Fishable, and Swimmable Quality Water." *Water Resources Research* 29(1993a):2445–54.

——. "Sequencing and Nesting in Contingent Valuation Surveys." Discussion Paper 93–36, Department of Economics, University of California, San Diego, August 1993b.

Carson, R. T., L. Wilks, and D. Imber. "Valuing the Preservation of Australia's Kakadu Conservation Zone." Unpublished paper, Department of Economics, University of California, San Diego, March 1993.

Desvousges, W. H., F. R. Johnson, R. W. Dunford, K. J. Boyle, S. P. Hudson, and K. N. Wilson. *Measuring Nonuse Damages Using Contingent Valuation: An Experimental Evaluation of Accuracy*. Report to the Exxon Corporation, Research Triangle Institute Monograph 92–1, 1992.

Diamond, P. A., J. A. Hausman, G. K. Leonard, and M. A. Denning. "Does Contingent Valuation Measure Preferences? Experimental Evidence." Paper presented at the Cambridge Economics, Inc. Symposium, Washington, DC, April, 1992.

Fischhoff, B., M. Jacobs Quadrel, M. Kamlet, G. Loewenstein, R. Dawes, P. Fischbeck, S. Klepper, J. Leland, and P. Stroh. "Embedding Effects: Stimulus Representation and Response Mode." *Journal of Risk and Uncertainty*, 6(1993):211–34.

Hausman, J. A., ed. *Contingent Valuation: A Critical Assessment*. Cambridge Economics, Inc., Symposium. Washington, DC, April, 1992.

Kahneman, D. "Comments." *Valuing Environmental Goods*, edited by R. G. Cummings, D. S. Brookshire and W. D. Schulze. Totowa, NJ: Rowman and Allanheld, 1986.

Kahneman, D., and J. Knetsch. "Valuing Public Goods: The Purchase of Moral Satisfaction." *J. Environ. Econ. and Manag.* 22(1992):57–70.

Loomis, J., M. Lockwood, and T. DeLacy, "Some Empirical Evidence on Embedding Effects in Contingent Valuation of Forest Protection." *J. Environ. Econ. and Manag.* 25(1993):45–55.

Magnussen, K. "Valuation of Reduced Water Pollution Using the Contingent Valuation Method: Testing for Amenity Misspecification." *Pricing the European Environment*, S. Navarud (ed.). New York: Oxford University Press, 1992.

Mitchell, R. C., and R. T. Carson. *The Use of Contingent Valuation Data for Benefit-Cost Analysis In Water Pollution Control*. Report to U.S.E.P.A., 1986a.

——. *Valuing Drinking Water Risk Reductions Using the Contingent Valuation Method: A Methodological Study of Risks from THM and Giardia*. Report to U.S.E.P.A., 1986b.

——. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington, DC: Resources for the Future, 1989.

Mitchell, R. C., R. T. Carson, and P. A. Ruud. *Cincinnati Visibility Valuation Study: Pilot Study Findings*. Report to the Electric Power Research Institute, 1989.

Romer, A. U. and W. W. Pommerehne. "Valuing Reductions of Public Risks: The Case of Hazardous Waste." *Models and Measurement of Welfare and Inequality*, P. Eichhorn (ed.). Heidelberg: Springer, forthcoming.

Randall, A., and W. Kriesel. "Evaluating National Policy Proposals by Contingent Valuation." *Economic Valuation of Natural Resources: Issues, Theory, and Applications*, by R. L. Johnson and G. V. Johnson (eds.). Boulder: Westview Press, 1990.

Schkade, D. A., and J. W. Payne. "Where Do the Numbers Come From?: How People Respond to Contingent Valuation Questions." Paper presented at the Cambridge Economics, Inc., Symposium, Washington, DC, April, 1992.

Smith, V. K. "Comment: Arbitrary Values, Good Causes, and Premature Verdicts." *J. Environ. Econ. and Manag.* 22(1992):71–89.

——. "Lightning Rods, Dart Boards, and Contingent Valuation." *Natural Resources J.* (forthcoming).

Smith, V. K., and W. H. Desvousges. *Measuring Water Quality Benefits*. Boston: Kluwer-Nijhoff Publishing, 1986.

Thompson, M. S. "Willingness to Pay and Accept Risks to Cure Chronic Disease." *Amer. J. Public Health* 76(1986):392–96.

Walsh, R. G., D. M. Johnson, and J. R. McKean. "Benefits Transfer of Outdoor Recreation Demand Studies: 1968–1988." *Water Resources Research* 28(1992):707–13.