Journal of Economic Perspectives—Volume 26, Number 4—Fall 2012—Pages 43–56

Contingent Valuation: From Dubious to Hopeless

Jerry Hausman

Approximately 20 years ago, Peter Diamond and I wrote an article for this journal analyzing contingent valuation methods (Diamond and Hausman 1994). At that time Peter's view was that contingent valuation was hopeless, while I was dubious but somewhat more optimistic. But 20 years later, after millions of dollars of largely government-funded research, I have concluded that Peter's earlier position was correct and that contingent valuation is hopeless.

In this paper, I selectively review the contingent valuation literature, focusing on empirical findings. I find that three long-standing problems continue to exist: 1) hypothetical response bias that leads contingent valuation to overstatements of value; 2) large differences between willingness to pay and willingness to accept; and 3) the embedding problem which encompasses scope problems. In their overview essay in this journal, Kling, Phaneuf, and Zhao discuss all three of these issues. On the first two points, I do not find their conclusions differ too much from mine. But I think they underestimate the problems of embedding and scope, which are likely to be the most intractable of the problems. Indeed, I believe that respondents to contingent valuation surveys are often not responding out of stable or well-defined preferences, but are essentially inventing their answers on the fly, in a way which makes the resulting data useless for serious analysis. In this comment, I first discuss these issues. I then offer a case study of a prominent contingent valuation study done by recognized experts in this approach, a study that should be only minimally affected by these concerns but in which the answers of respondents to the survey are implausible and inconsistent.

■ *Jerry Hausman is the John and Jennie S. MacDonald Professor of Economics at the Massachusetts Institute of Technology, Cambridge, Massachusetts. His email address is jhausman@mit.edu.*

http://dx.doi.org/10.1257/jep.26.4.43. doi=10.1257/jep.26.4.43

I am often asked what should be done given my view that contingent valuation should not be used. Should nonuse value be ignored? My view is that expert government agencies and Congress should make informed decisions and enact regulations that attempt to improve the economic allocation process (see also Diamond and Hausman 1994). To the extent that contingent valuation is interpreted as an opinion poll about the environment in general, rather than a measure of preferences about a specific project, public officials and regulators should recognize this concern.[1] However, public policy will do better if expert opinion is used to evaluate specific projects, including nonuse value, and to set appropriate financial incentives to reduce the risk of accidents such as the *Exxon Valdez* and BP disasters.

Responses to contingent valuation surveys for a single environmental issue are typically based on little information, given the limited time involved for each survey respondent. Thus, the results of such surveys are unlikely to be accurate predictors of informed opinion. Contingent valuation about specific projects does not improve the inputs to the analysis, so it should not be included in the policy analysis. Contingent valuation does not provide a good basis for either informed policymaking or accurate damage assessments in judicial proceedings.

Have the Empirical Problems of Contingent Valuation Been Addressed?

Possible problems of contingent valuation have been discussed in the literature for at least the past 30 years. While I focus on three of those problems, my chosen focus should not be taken to imply that other problems do not exist for individual studies or for the method as a whole.

Hypothetical Bias and Upward-Biased Results

The nature of a survey is that it asks a hypothetical question. Hypothetical bias is the bias that arises in answering a hypothetical question with which the respondent has no market experience; put simply, what people say is different from what they do. When hypothetical questions are asked about willingness to pay, the results tend to be upward-biased. This fact is well-known. For example, Jamieson and Bass (1989) studied people's stated intentions to purchase new products, and found that such measures were overstated. Other studies have affirmed this finding, like

[1] Referenda are similar to opinion polls except the results are often binding. The claim is sometimes made that contingent valuation studies, to the extent they can forecast how voters would respond in a binding referendum, should be used to design public policy. But even for referenda, the necessity of calibration to individual preferences to do welfare analysis remains, an issue which I discuss subsequently. Further, the use of actual referenda to obtain economic values is highly questionable. For example, because no immediate obvious budget constraint exists for voters in referenda, the evaluation of individual preferences from referenda is highly problematic. The importance of a budget constraint is fundamental to economic choices and its absence has an important distorting effect in contingent valuation studies, as we discuss in Diamond and Hausman (1994).

Hsiao et al. (2002) and Morwitz et al. (2007). (The latter paper finds that familiarity with the new product leads to more successful forecasts of whether people will buy, but familiarity with the product will not be present in most contingent valuation studies.) The NOAA panel (Arrow, Solow, Portney, Leamer, Radner, and Schuman 1993) found upward bias to be present in willingness-to-pay responses for both private goods and public goods, which they determined would extend to contingent valuation studies. In their overview paper, Kling, Phaneuf, and Zhao find that significant biases exist and their "net impact varies with the characteristics of participants and the commodity, and the type of script used."

A standard response to this problem has long been to apply a "fudge factor"—that is, to deflate the stated willingness to pay by some amount. How much? NOAA proposed dividing contingent valuation results by two (*Federal Register* 1994). But of course, there is no reason that the degree of overstatement should be the same across all survey methods, commodities, and types of survey respondents. I do not see how past studies provide a basis for an appropriate estimate of the needed adjustment. Suppose that a new econometric estimator was proposed that had a mean bias of 300 percent, but the 95 percent confidence interval varied from 150 to 800 percent. Such an estimator would not be used in serious policy formation.

Why the bias and large variation in answers to hypothetical questions? One issue, common in the opinion survey literature, is that those being interviewed often seek to please the interviewer (Tourangeau, Rips, and Rasinski 2000). Other issues may arise because of the specific nature of contingent valuation surveys. As Horowitz (2000) points out, the standard form of a contingent valuation question asks about one's willingness to pay a certain amount for a certain outcome, but doesn't say explicitly how the answer will be used, nor offer a range of options, nor offer a chance for discussion and interaction with others (as does a public voting process). Supporters of such surveys spend considerable time and energy on the precise wording of their questions and they test different wording choices in focus groups in an effort to make respondents feel that their answer really matters and that they should take the task of answering seriously. But as Harrison (2007) writes: "The literature on non-market valuation in environmental economics is littered with assertions that one can somehow trick people into believing something that is not true. . . . The claims tend to take the form, 'if we frame the hypothetical task the same way as some real-world task that is incentive compatible, people will view it as incentive compatible.'"

But despite such efforts, for whatever reason, hypothetical bias persists. For example, Murphy and Stevens (2004) note that the literature shows hypothetical bias across a wide variety of contingent valuation approaches. Johnston (2006, p. 469) concurs: "Most research finds significant divergence between stated and actual behaviors." Kling, Phaneuf, and Zhao agree as well.

Sometimes supporters of contingent valuation surveys compare them to polls about public referenda—and thus seek to give them a presumption of legitimacy because polls are useful at predicting the outcome of the democratic process. But of course, polls predicting the outcome of referenda are sometimes accurate, sometimes

not. Often, the polls about a referenda change considerably over the period of time leading up to an election, suggesting that preferences about the choice were not especially stable at the beginning of the process. Vossler, Kerkvliet, Polasky, and Gainutdinova (2003) look at survey responses and an actual referendum vote on a proposal to protect open space in Corvallis, Oregon. They find that it is necessary to treat the "undecided" vote as "no" if the survey is to avoid hypothetical bias and to reflect the actual outcome of the vote. Of course, one should be cautious about extrapolating from surveys about a vote on a local public good to contingent valuation surveys on other subjects. The empirical basis for weighting undecided or other responses to a contingent valuation survey in some way that is intended to reduce hypothetical bias is scanty at best.

Difference between Willingness to Pay and Willingness to Accept

Contingent valuation questions can be phrased in two broad ways: the willingness-to-pay approach seeks to discern what the respondent would pay to avoid a negative outcome (or to achieve a positive outcome), while the willingness-to-accept approach seeks to discern how large a payment the respondent would need to receive in order to accept the negative outcome (or not to receive a positive outcome). Basic economic theory suggests that these two approaches should give (approximately) the same answer, but both supporters and skeptics of contingent value methods recognize that large and persistent disparities commonly arise in answers to contingent valuation surveys.

Broadly speaking, there have been two approaches to rationalizing the large gaps between willingness to pay and willingness to accept: use a theoretical background rooted in behavioral economics, or relax enough assumptions in the neoclassical model. The difficulty with either approach is that if benefit–cost analysis is to be logically coherent, it requires a theoretical framework. The Hicksian foundations of standard welfare analysis are based on compensated demand curves and potential Pareto improvements, and no substitute foundational framework has received wide acceptance to replace the Hicksian approach. (For a discussion of the Hicksian basis for welfare analysis, see Hausman 1981, and Hausman and Newey 1995.) Rationalizations of the gap between willingness to pay and willingness to accept come at the expense of introducing assumptions that render standard benefit–cost analysis invalid.

For example, suppose that consumers do not have neoclassical preferences, but instead are subject to "loss aversion," and thus they will weight prospective losses more heavily than equivalent gains. Such "behavioral" preferences will indeed drive a wedge between willingness to accept and willingness to pay. But it becomes unclear how to do welfare analysis of gains to some groups and losses to others with these assumptions; necessary compensation for aggrieved losers from any policy may well outstrip gains to the winners.

Various efforts have been made to extend the neoclassical framework in a way that rationalizes the gap. Proponents of contingent valuation have attempted rationalizations of these differences, but have not overcome the findings of Diamond and Hausman (1994) or the results of Milgrom (1993). Both papers demonstrate

that the attempts to rationalize the well-recognized and persistent disparity between willingness to pay and willingness to accept fail as a matter of economic theory and observed empirical outcomes.

Of course, one can claim that consumers do not have neoclassical preferences. But standard cost–benefit analysis, and the underlying logic of being able to sum the willingness to pay of many individuals, requires that individual preferences are being measured (Diamond and Hausman 1994, pp. 55–58). Essentially, use in policy analysis or damage analysis depends on willingness to pay being a measure of the compensating variation for avoiding a negative outcome. If the neoclassical assumptions are relaxed so that willingness to pay includes, say, a component for altruism or for sympathy, then willingness to pay will diverge from willingness to accept—but then addition across the willingness to pay of individuals is no longer appropriate. In aggregation, the neoclassical model requires that preferences be over states of the world and not over acts: for example, preferences must be over the choice between two different states of a wilderness area, not over whether the respondent receives a warm glow from the idea of saving a wilderness area, nor about a general attitude about providing public goods in general. Again, this assumption is required for consistent economic policy choices (see also the NOAA Report, Arrow et al. 1993).

Of course, there are a number of other ways to attempt to rationalize the large gaps between willingness to pay and willingness to accept in contingent valuation surveys. It's possible to create a theory that is consistent with (almost) any given set of facts. But the task of building the foundations of a benefit–cost analysis on top of those alternative theories has not been done. And as I demonstrate throughout this essay, the gaps are likely due to the reality that answers to contingent valuation surveys do not actually reflect stable or well-defined preferences but instead are opinions invented on the fly.

Scope and Embedding

The most fundamental challenge to the contingent value method, and the strongest evidence that the answers to such surveys are invented in response to the questions, comes from concerns that are referred to as "scope" and "embedding." Kahneman and Knetsch (1992) were the first to explore the "embedding effect," which demonstrates the nonexistence of preferences in a contingent valuation setting. As they wrote, "perhaps the most serious shortcoming of CVM [contingent value methods]" is that "the assessed value of a public good is demonstrably arbitrary, because willingness to pay for the same good can vary over a wide range depending on whether the good is assessed on its own or embedded as part of a more inclusive package." In Diamond and Hausman (1994), we provide an example of the embedding effect, where willingness to pay to clean one lake is approximately equal to stated willingness to pay to clean up five lakes—including the one asked about individually. Embedding is related to the scope effect, which is the broader proposition that respondents to contingent valuation surveys should be more willing to pay for a large effect than for a subset of that effect. Proponents of contingent valuation would like to demonstrate a scope affect, but the scope

effects typically found are not nearly large enough to make contingent valuation results credible.

In the earlier version of their overview paper, Kling, Phaneuf, and Zhao acknowledged that "scope effects are typically present and positive, if not always large." But I find this result to be similar to a finding in econometric estimation that demand curves slope downward—a very weak test with almost no power. We do not know how large scope effects should be. Indeed, since contingent valuation surveys are typically pretested, the survey design can be manipulated to ensure that at least minimal scope effects are present.

Thus, in Diamond and Hausman (1994), we proposed a more stringent version of a scope test called an "adding-up test." The test works this way: a first group of respondents is asked their willingness to pay for a public good X; a second group is asked their willingness to pay for public good Y; and a third groups is asked their willingness to pay for X and Y together. The total value of the entire project minus the value of the first project should approximately equal the value of the incremental projects.[2] A specification test then permits one to statistically determine whether embedding is present. If Diamond–Hausman tests are done correctly, they first establish what the willingness to pay is for a given project. They then establish the willingness to pay for a larger project in which the first project is present, which establishes the *incremental amount* of willingness to pay for the additional projects. In this way, the adding-up test overcomes the problem of a more general scope test.

Desvousges, Mathews, and Train (2012a) review 109 contingent valuation studies on environmental goods since 1994 that apply a scope test. They find that most studies do not provide sufficient information to determine whether the difference in survey response to variations in scope is "adequate" (p. 4). Those who conduct contingent valuation surveys have typically not collected their data in way that makes an adding-up test possible. They find only one study that permits a test of adding-up: a study by Chapman et al. (2009). This study passes a scope test, but fails the more stringent adding-up test. Desvousges, Mathews, and Train (2012b) expand the Chapman et al. (2009) survey to measure the value of each increment directly with contingent valuation, and they find that the sum of the estimated values of the incremental parts is three times greater than the estimated value of the whole. They conclude—as do Heberlein, Wilson, Bishop, and Schaeffer (2005) and Bateman (2011)—that "standard scope tests are uninformative."

My view is that until contingent valuation surveys can reliably pass the Diamond–Hausman adding-up test (or a similar test) to demonstrate that embedding is not present, the results do not indicate stable or coherent individual preferences.

[2] This statement of the test leaves out income effects, but income effects are typically quite small for projects considered by contingent valuation. In a paper we are still working on (Hausman and Newey, no date), my coauthor and I develop bounds that take account of the share of income spent on a good and its income derivative. Our results demonstrate that for the size of contingent valuation projects, typically less than $100, the upper and lower bounds are almost identical for consumer surplus.

Recent empirical evidence demonstrates that some problems that exist in contingent valuation studies also exist in actual market situations. For example, framing of questions can lead to very different results in contingent valuation studies, as I discuss below in my case study. Evidence demonstrates that framing can also affect consumer choices in the market. Thus, some of the problems in contingent valuation also exist in revealed preference outcomes. However, I expect that consumers do better (even though they still make mistakes) for important decisions and for repeated decisions. And consumers have a budget constraint which has a large effect on their decisions. For public policy purposes, expert analysis, as I discuss above, will hopefully avoid the "mistakes" that would arise with the use of contingent valuation and come to better allocative outcomes than if we were to depend on results from contingent valuation surveys that are not consistent with fundamental economic preferences on which we base economic welfare analysis.

A Case Study: Contingent Valuation and Australian Cable Television

To provide a more concrete illustration of these issues, I will consider a particular contingent valuation study.[3] This particular study is chosen for several reasons. It was implemented by Richard Carson, a participant in this symposium and someone widely recognized as a top expert in contingent valuation studies.[4] The design and implementation of the study was not constrained in any meaningful way by lack of a budget. It had large sample sizes. Unlike some contingent value surveys that deal with issues far-removed from the daily experience of the respondents—like the value to be placed on cleaning up Prince William Sound where the *Exxon Valdez* ran aground—this survey dealt with a product well-known to those being surveyed: cable television. Yet despite these advantages, I will demonstrate that the results from this contingent value survey are unreliable. The results demonstrate that although people responded to the survey, their answers cannot correctly be treated as a meaningful measure of preferences.

Of course, one study does not discredit contingent valuation methodology. And the decisionmakers in this case, on the Australian Copyright Tribunal, do not have the last word on economic methodology. Yet the case study is potentially useful for thinking about the issues of contingent valuation. The contingent valuation study did not solve the well-recognized problems of contingent valuation, even though it

[3] As an econometrician, I typically do not rely on case studies in my academic research. However, contingent valuation studies that include both a critique and then a decision from an outside party are limited to "high stakes" proceedings. I am not aware of many such cases. Thus, I cannot do a meta-analysis of many contingent valuation studies, because the cost of analyzing them is typically quite high. The Australian case study seems especially useful since it was a "high stakes" proceeding with significant analysis and a decision by an informed tribunal.

[4] All the participants in this study, including myself, involved in the Australian case study were paid consultants. The contingent valuation study was paid for by Screenrights (the copyright holder) and my report was paid for by the cable TV providers.

had experienced academic experts in contingent valuation who were not subject to a tight budget constraint to finance their analysis. And the Australian contingent valuation study failed the Diamond–Hausman test as I describe below.

The background situation is that cable TV companies in Australia retransmit the free-to-air TV channels' broadcasts, as do cable TV companies in most of the world. In 2001, a change in Australian law defined retransmission as an infringement of copyright, requiring the cable TV companies to pay "equitable remuneration" to the copyright owners via their declared collecting society, Screenrights. The cable TV companies and Screenrights were unable to agree upon what constituted "equitable remuneration," and the matter was brought before the Copyright Tribunal, which is administered by the Federal Court of Australia. Screenrights' primary evidence before the Tribunal was a contingent valuation study designed by Jeff Borland and Richard T. Carson, and conducted by a leading Australian research company. The primary evidence opposing this approach was a study provided by Tim Bock and myself.[5] Thus, I confess that yet another reason for choosing this study is that it was the most recent contingent valuation study that I have analyzed in detail. The 2006 Copyright Tribunal decision has a useful extended overview of all the issues of the case, with several sections focused on contingent valuation, available online at http://www.austlii.edu.au/au/cases/cth/ACopyT/2006/2.html.

The contingent valuation study actually involved two parts. In the first part, 2,622 subscribers participated in a 10-minute personal interview by surveyors who knocked on their doors at home. The survey asked general questions about household structure and behavior, and then respondents were both read and asked to read descriptions about the benefits of retransmission and available substitutes. They were then asked, using a formal script, if they had the choice of 1) paying $X extra per month to continue receiving the regular TV channels through cable TV, or 2) paying the same as before and losing these channels from cable TV (but perhaps getting them through a TV aerial), what would they choose? The respondents were randomly allocated to one of five monthly fees: $1.00, $2.50, $5.00, $7.50, and $10.00. In an unexpected twist, it was discovered that the first part of the study had not correctly implemented procedures for recontacting those households not at home ("call-backs"), therefore the study was repeated with a new sample of 2,369 households, and some minor wording changes were made in the second study.

Specific responses to the first and the second survey appear in Table 1. One oddity jumps out immediately: In both studies, the quantity demanded at $10 is higher than demanded at $7.50. Although it is highly unusual in real markets to find that a 33.3 percent increase in price does not cause an outright decline in quantity demanded, at least the increase in quantity demanded here is not statistically significant. Other questions also come to mind: for example, the Copyright Tribunal questioned why the lowest value surveyed was set at $1, rather than some

[5] More explanation and detail are provided in Hausman and Bock (2007).

Table 1
Binary Choice Data from the Australian Cable Television Contingent Valuation Survey

	Monthly subscription fee				
	$1.00	*$2.50*	*$5.00*	*$7.50*	*$10.00*
Study 1					
Pay	299	224	176	139	140
Not pay/Don't know	224	294	333	403	390
Study 2					
Pay	292	207	152	85	90
Not pay/Don't know	199	288	321	375	360

lower number. But I will focus on two fundamental issues and then mention some other points.

First, preferences in this contingent valuation study appear to be irrationally unstable, in the sense that minor differences in wording—that is, the framing of the questions—led to large differences in response. Study 1 finds a 32 percent higher share of respondents who state they are willing to pay $10 per month for retransmission compared to Study 2. Similarly, 39 percent more respondents in Study 1 said they would pay $7.50 than in Study 2. Yet the questions in the two studies are essentially identical, with only a small amount of additional information in Study 2. Borland and Carson agreed that the discrepancies in the demand curves were attributable to changes in question wording. In my view, the only significant change was that respondents were shown both a monthly and annual fee in Study 2, while in Study 1 only the monthly fee was shown. If relatively minor changes in wording lead to significant differences in results, I would refer to this situation as "irrational preference instability." Such results support a conclusion that consumers did not reveal true preferences in the stated preference questionnaire and are instead, to some unknown extent, "making-up" or "inventing" their answers to a hypothetical situation with which they are unfamiliar. I conclude that consumers do not have well-formed preferences, which is why their responses to the main contingent valuation question were significantly influenced by the survey wording.

Second, the study results fail a Diamond–Hausman (1994) adding-up test, discussed earlier. Specifically, this test checks to see if average willingness to pay (WTP) for divisible good X is equivalent to the sum of average WTP for kX and average WTP for $(1 - k)X$ *conditional* upon kX already having been supplied, where $0 < k < 1$. For example, if a consumer is willing to pay $50 for two items together, such as local telephone service and a broadband Internet connection, the consumer should be willing to pay (approximately) this same amount if the consumer first purchases the local telephone service and then buys the broadband service (after purchasing the local telephone service).

The Diamond–Hausman test was administered by another fieldwork company using the protocols employed in Study 1. Some minor wording changes were made to the contingent question to improve its intelligibility, but these changes were not contentious between the parties and their experts involved in the proceeding. Three independent samples ($N = 200$ in each; 600 in total) received different versions of the questionnaire, as follows:

Version 1: Retransmission of only ABC, Channel 9 and SBS (that is, kX).

Version 2: Retransmission of Channel 7 and Channel 10, given that ABC, Channel 9 and SBS are already being retransmitted to the household (that is, $(1 - k)X$ given kX).

Version 3: Retransmission of all the free-to-air channels (that is, X).

To estimate a mean from the kind of data in Table 1, one standard approach is to use the Turnbull Lower Bound estimator. (The nonparametric Turnbull estimator begins by determining the fraction of refusals falling into each dollar interval, and a lower-bound estimate of the mean follows from these fractions.[6]) By this measure, the mean for Version 1 is $2.96, the mean for Version 2 is $1.64, and the sum of these two is $4.60. However, the mean is $2.81 for Version 3. Thus, the sum of Version 1 and 2 is 64 percent greater than Version 3.

One potential objection to this comparison is that something is amiss with the follow-up survey. However, the estimated willingness to pay for in Versions (2) and (3) of the adding-up test survey are quite close to the willingness to pay in Study 2 shown in Table 1. This outcome strongly suggests that respondents reacted in a similar manner to these two surveys. Another objection sometimes raised is that asking a respondent to "pretend" they have already obtained part of a good is problematic because it may be difficult to get respondents to take such an exercise seriously. But if this objection is true, it would invalidate both the potential problem and the original contingent valuation study. After all, contingent valuation questions—including the ones in this study— often set up scenarios that ask respondents to pretend.

As one might expect, there were a number of other points at issue in the discussion before the Copyright Tribunal, which are summarized in its report. For example, we also demonstrated that if the demand curves estimated from the contingent valuation survey were to be taken seriously, the cable television companies would increase

[6] The Turnbull estimator is computed using the Pooled Adjacent Violators Algorithm and treating "don't know" as "not pay" (as discussed in Bateman et al. 2002, p. 231). The algorithm for the Turnbull estimator is roughly as follows: For the lowest bid level, calculate the proportion of refusals. Then move to the next-highest bid level and again calculate the number of refusals. Continue this process up through the bid levels, and use this data to calculate a cumulative density function from which you can derive a probability density function. Multiply the probability density function by the bid defining the lower bound, and then sum over all bid levels. The other method often used for this calculation is a Weibull distribution, which is a two-parameter location-scale distribution often used in duration models. In this case, the approach is to estimate the two parameters that would characterize a Weibull distribution that fits the pattern of the data, and then to use the properties of that distribution to calculate the mean. The Weibull distribution will typically give a mean estimate greater than the Turnbull estimator.

their profits by charging an extra $10 per month (or more): that is, the extra revenue they would gain by charging more would more than offset the losses from those who decided not to subscribe to cable television at all. Instead, it turned out that one cable company, Foxtel, was charging the same for digital satellite and digital cable, but was including all the free-to-air stations in the first delivery mechanism but not the second. Apparently, providing the free-to-air stations was not a service for which the company thought it could charge more. But the primary argument that is relevant for thinking about contingent valuation methods as a whole is that the answers from such studies are unstable and inconsistent, invented for the moment of the survey, and cannot be treated as preferences in the sense that economist understand that term.

After reviewing all the arguments, Australia's Copyright Tribunal (2006, par. 510 and 512) chose to disregard completely the evidence from the contingent value survey. It quoted a 1965 case to the effect that "[A] person exercising quasi-judicial functions must . . . not spin a coin or consult an astrologer, but he may take into account any material which, as a matter of reason, has some probative value. . . . If it is capable of having any probative value, the weight attached to it is a matter for the person to whom Parliament has entrusted the responsibility of deciding the issue." Having expressed a willingness to give at least some weight to any evidence that might be relevant, the Tribunal wrote: "Courts and tribunals must proceed on the basis of probative evidence, not speculation. . . . We have such a level of doubt about the Survey that we attach no weight to it."

Although the Copyright Tribunal decided that in this situation—to paraphrase the title of Diamond and Hausman (1994)—no number was better than the contingent valuation number, it did rely on a range of other evidence to decide that the cable companies should pay 22.5 cents per subscriber per month in exchange for transmitting the free-to-air content. This amount was not even in the range of possibilities considered in the contingent valuation study. Thus the Copyright Tribunal chose to use its expert opinion to set the rate per subscriber, and it completely ignored the outcome of the contingent valuation study.

Of course, this particular case study addresses only one contingent value survey—but, again, it is presumably a "high quality" study by the standards of this literature. The study by Chapman et al. (2009) of the aesthetic and ecosystem value of certain water resources in Oklahoma is another "high quality" contingent study designed and implemented by proponents of such studies. It had large sample sizes, a budget constraint that did not bind very tightly, and claimed to meet best-practice guidelines. Yet as discussed earlier, Desvousges, Mathews, and Train (2012a, b) show that the results of this contingent valuation survey are unreliable for various reasons, including failing an adding-up test.

Conclusion

The controversy over contingent valuation studies often follows a predictable pattern. A contingent value study is designed and carried out, with much talk about

how methodology has strengthened over time. When the results are announced, critics point out potentially severe problems, like hypothetical bias and overstatement, disagreements between willingness to pay and willingness to accept, and problems of scope or embedding. Supporters then respond that perhaps this particular study wasn't well-designed, and that there are ways to make adjustments, and that it would be wrong to conclude from one study that the enterprise of contingent valuation is fundamentally flawed. Then the next study arrives and is criticized and defended in the same way. For those of us who have criticized a number of contingent valuation studies, it feels as if proponents of contingent valuation retreat to the position that all studies shown to be inaccurate are examples of poor practice rather than any inherent flaw. But despite all the positive-sounding talk about how great progress has been made in contingent valuation methods, recent studies by top experts continue to fail basic tests of plausibility.

I expect that if contingent value respondents had been asked about Prince William Sound (where the *Exxon Valdez* ran aground) and another group was asked about Prince Andrew Sound (fictitious) after being told that Price William Sound had been saved, and a third group was asked about Price William Sound and Price Andrew Sound together, the combined response would not be much different than the individual responses, so that the sum of the individual responses would be significantly greater than the combined response. When contingent studies can routinely pass Diamond–Hausman adding-up tests I am willing to reconsider my conclusion of little or no progress over the past 20 years in solving the most important problems with contingent valuation. But even if that event occurs, contingent valuation would still face problems like how to address the upward bias in responses and how to build a framework for cost–benefit analysis in a setting where the data show a gulf between willingness to pay and willingness to accept.

I do not expect these problems to be resolved, so in my view "no number" is still better than a contingent valuation estimate. Moreover, as the discussion of Australian Copyright Tribunal (2006) showed, other pieces of evidence can be brought to bear on goods that are not directly valued in the market. For example, in environmental damage situations, the method of "habitat equivalency analysis" relies on a group of trustees appointed through government or the courts to analyze what expenditures are needed to restore the environment (Damage Assessment and Restoration Program 2006). The political process can also provide outcomes. As Diamond and I wrote in our 1994 essay in this journal (pp. 58–59), "the choice is between relying on Congress after doing a contingent valuation study and relying on Congress without doing such a contingent valuation study." My theme is that unless or until contingent value studies resolve their long-standing problems, they should have zero weight in public decision-making.

I do not expect that proponents and opponents of contingent valuation will ever agree. Some bad ideas in economics and econometrics maintain a surprising viability. Numerous branches of the federal government continue to fund contingent valuation research in the hope that it will support their favored policies subject to cost–benefit analyses. In turn, the proposed regulations lead to push-back from

those who would bear the costs. In cases like the *Exxon Valdez* spill or the BP *Deepwater Horizon* spill, vast amounts of money are at stake. I do not find my view that such debates will persist to be at all cynical; rather, it is the expected outcome given the incentives that all parties face.

■ *I am not involved in any ongoing paid research or litigation involving contingent valuation. I previously served as a paid consultant on the* Exxon Valdez *matter and the Australian Copyright matter discussed in this paper. I have also testified before Congress on contingent valuation, but I do not accept payment for Congressional testimony. I thank the editors for help in revising the paper.*

References

Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, and Howard Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." National Oceanic and Atmospheric Administration (U.S. Department of Commerce).

Australian Copyright Tribunal (see Copyright Tribunal of Australia)

Bateman, Ian J. 2011. "Valid Value Estimates and Value Estimate Validation: Better Methods and Better Testing for Stated Preference Research." Chap. 16 in *The International Handbook on Non-Market Environmental Valuation,* Jeff Bennett, editor, Cheltenham UK: Edward Elgar Publishing.

Bateman, Ian J., et al. 2002. *Economic Valuation with Stated Preference Techniques: A Manual.* Cheltenham: Edward Elgar.

Chapman, David J., Richard C. Bishop, W. Michael Hanemann, Barbara J. Kanninen, Jon A. Krosnick, Edward R. Morey, and Roger Tourangeau. 2009. "Natural Resource Damages Associated with Aesthetic and Ecosystem Injuries to Oklahoma's Illinois River System and Tenkiller Lake." Expert Report for State of Oklahoma. Case No. 05-CV-0329-GKF-SAJ. *State of Oklahoma v. Tyson Foods, et al.* In the United States District Court for the Northern District of Oklahoma. Volume I.

Copyright Tribunal of Australia. 2006. *Audio-Visual Copyright Society Ltd v Foxtel Management Pty Ltd (No 4) [2006] ACopyT 2* (Ruling), Date of Judgment: May 3, 2006. http://www.austlii.edu.au/au/cases/cth/ACopyT/2006/2.html.

Damage Assessment and Restoration Program, National Oceanic and Atmospheric Administration, U.S. Department of Commerce. 2006. "Habitat Equivalency Analysis: An Overview." May 23. (Originally published March 21, 1995.) http://www.darrp.noaa.gov/library/pdf/heaoverv.pdf.

Desvousges, William, Kristy Mathews, and Kenneth Train. 2012a. "Adequate Responsiveness to Scope in Contingent Valuation." http://elsa.berkeley.edu/~train/scope.pdf.

Desvousges, William, Kristy Mathews, and Kenneth Train. 2012b. "Rationality Reduc: The Incremental Adding up Test for Contingent Valuation." Unpublished paper, Department of Economics, University of California, Berkeley.

Diamond, Peter A., and Jerry A. Hausman. 1994. "Contingent Valuation: Is Some Number Better Than No Number?" *Journal of Economic Perspectives* 8(4): 45–64.

***Federal Register*.** 1994. May 14, 59(85): 1183.

Harrison, Glenn W. 2007. "Making Choice Studies Incentive Compatible." Chap. 4 in *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice,* edited by Barbara Kanninen. Dordecht, The Netherlands: Springer.

Hausman, Jerry. 1981. "Exact Consumer Surplus and Deadweight Loss." *American Economic Review* 71(4): 662–76.

Hausman, Jerry A., and Timothy Bock. 2007. "The Falsification of Contingent Valuation: A Case Study." Unpublished paper.

Hausman, Jerry A., and Whitney K. Newey. 1995. "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss." *Econometrica* 63(6): 1445–76.

Hausman, Jerry A., and Whitney K. Newey. No date. "Non-separable Estimation of Exact Consumers Surplus and Deadweight Loss with Arbitrary Heterogeneity." Paper not completed.

Heberlein, Thomas A., Matthew A. Wilson, Richard C. Bishop, and Nora Cate Schaeffer. 2005. "Rethinking the Scope Test as a Criterion for Validity in Contingent Valuation." *Journal of Environmental Economics and Management* 50(1): 1–22.

Horowitz, John. 2000. Policy Discussion section of "Stated Preference: What Do We Know? Where Do We Go?" Proceedings of Session One of "Theory and Design of Stated Preference Methods," A Workshop Sponsored by the USEPA's National Center for Environmental Economics and National Center for Environmental Research. October 12–13, Doubletree Hotel, Washington, DC. http://epa.gov/ncer/publications/workshop/pdf/EE-0436-01.pdf.

Hsiao, Cheng, Baohong Sun, Vicki G. Morwitz. 2002. "The Role of Stated Intentions in New Product Purchase Forecasting." In *Advances in Econometrics,* vol. 16, pp. 11–28. Emerald Group Publishing.

Jamieson, Linda F., and Frank M. Bass. 1989. "Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods." *Journal of Marketing Research* 26(3): 336–45.

Johnston, Robert J. 2006. "Is Hypothetical Bias Universal? Validating Contingent Valuation Responses Using a Binding Public Referendum." *Journal of Environmental Economics and Management* 52(1): 469–81.

Kahneman, Daniel, and Jack Knetsch. 1992. "Valuing Public Goods: The Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management* 22(1): 57–70.

Milgrom, Paul. 1993. "Is Sympathy an Economic Value?" In *Contingent Valuation: A Critical Assessment,* edited by J. Hausman, 417–42. Amsterdam: North Holland Press.

Morwitz, Vicki, Joel H. Steckel, and Alok Gupta. 2007. "When Do Purchase Intentions Predict Sales?" *International Journal of Forecasting* 23(3): 347–64.

Murphy, James J., and Thomass H. Stevens. 2004. "Contingent Valuation, Hypothetical Bias, and Experimental Economics." *Agricultural Resource Economics Review* 33(2): 182–192.

Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski. 2000. *The Psychology of Survey Response.* Cambridge: Cambridge University Press.

Vossler, Christian A., Joe Kerkvliet, Stephen Polasky, and Olesya Gainutdinova. 2003. "Externally Validating Contingent Valuation: An Open-Space Survey and Referendum in Corvallis, Oregon." *Journal of Economic Behavior and Organization* 51(2): 261–77.