Misleading Quantification

The Contingent Valuation of Environmental Quality

Robert K. Niewijk

Regulators in Southern California have proposed an ambitious new plan designed to bring the area into compliance with federal air quality standards. The plan's costs to California's economy are estimated at $10 billion to $13 billion per year. Californians ought to ask if the cleaner air would be worth the cost.

That, however, is a difficult question to answer. To weigh costs against benefits, one must measure the benefits. People would likely pay *something* for the chance to live in Los Angeles without interminable sore throats and stinging eyes. But how *much* would they pay?

Another hypothetical: a company discharges toxic waste or spilled oil onto publicly owned land or waterways. The law should require the company to pay the full amount of the harm it causes (to give firms proper incentives to take precautions against other spills and to compensate the public for its loss). Once again, how does one place a dollar value on the harm to a natural resource?

These types of valuation problems crop up constantly in the environmental field. Society often must decide questions such as the extent of polluters' legal liability, whether to invest in greater regulation, and how much to spend on environmental cleanup efforts, public land purchases, and the like. Yet no market exists for much of the value people derive from natural resources such as waterways, beaches, parks, or the air itself. Without a market, there is no market price, and therefore no direct way to measure value.

Researchers can estimate the "use values" people derive from recreational activities such as camping or fishing based on the money people spend on the activities or on travel to the recreational sites. For several decades, however, economists have also recognized "nonuse values" that people enjoy even if they never use a particular piece of the environment, such as the option to use it in the future, the opportunity to preserve it for posterity, and the mere knowledge of its existence. By definition, no activity exists that can serve as a base for estimates of nonuse values. Some uses—breathing clean air, for example—are so passive that they pose the same difficulty.

In response to this problem, researchers have turned to an innovative technique to measure people's nonuse values: ask them. Called contingent valuation (or CV), the technique uses surveys that provide a detailed description of the resource, its current condition, a hypothetical improvement in its condition or decrease in the chances of its degradation, and a way in which the person would pay for the improvement (such as an increase in taxes or higher prices). The survey asks how much

Robert K. Niewijk is an attorney practicing antitrust and commercial litigation at Katten Muchin & Zavis in Chicago.

a person would be willing to pay for the specified improvement or decreased chance of harm. (The technique is called "contingent valuation" because the person's willingness to pay is contingent on the hypothetical change to the resource.) Researchers then combine all the survey responses to calculate an average willingness to pay (WTP) and multiply that average by the total population to come up with society's value for the resource.

Although academics have been conducting and writing about CV surveys for some time, CV has gained practical prominence only in the past few years. The Department of the Interior began including CV as a potential technique for measuring natural resource damages under the Superfund laws in 1986. The National Oceanic and Atmospheric Administration (NOAA) proposed the use of CV in similar damage assessment regulations covering oil spills in January 1994. The Environmental Protection Agency (EPA) has begun to use CV studies to estimate the economic benefits of its regulations when conducting cost-benefit regulatory impact analyses. And the California clean air hypothetical at the beginning of this article is not hypothetical; at least four studies have used CV surveys to value the health benefits of the 1989 air quality plan adopted by California's South Coast Air Quality Management District.

The academic literature discussing CV has long noted its potential for biased estimates—but only in the context of fine-tuning CV surveys in an attempt to eliminate that bias. Along with the heightened importance attached to CV's use has come a heightened scrutiny of its reliability. Under the glare of that scrutiny, CV looks dreadful: experimental evidence suggests that, at best, CV grossly overestimates the actual values that people hold for a resource, and at worst, it does not measure those values at all. In light of these results, CV should not be used in either cost/benefit public policy analyses or natural resource damage assessments.

Sources of Bias in Contingent Valuations

CV's shortcomings can be seen simply by describing how it's done: a survey offers a description of a complex natural resource that a person may have never heard of, along with a hypothetical event that would harm or improve the resource, an explanation of the complex impact that event would have on the resource, and a hypothetical payment mechanism. A person is then asked what amount, hypothetically, she would be willing to pay to prevent the harm or cause the improvement.

Researchers have long known that people's answers to survey questions about their behavior often bear little resemblance to what they actually do. Talk is cheap: because there is no cost to being wrong, survey respondents have little incentive to undertake the mental effort needed to be accurate.

CV surveys of natural resource values are far more susceptible to bias than the average survey. Respondents are often unfamiliar with both the resource being valued and the effects that some form of pollution will have on it. For example, if a person is asked her willingness to pay to prevent a toxic waste discharge in a lake, she must know the size, location, and current condition of the lake, whether other lakes (and how many) could serve as substitutes for the lake in question, the details of the ecosystem the lake supports, the damage the chemicals would do to humans and the ecosystem, and the amount of time it would take for the chemicals to break down.

Experimental evidence suggests that, at best, contingent valuation grossly overestimates the actual values that people hold for a resource, and at worst, it does not measure those values at all.

The CV survey itself must provide all of this information. And because that information may (indeed, should) be the sole basis for the person's valuation, slight variations in how it is presented can have a huge impact on the value a person reports.

Even if a survey provides sufficiently detailed information in an unbiased manner, respondents still must assimilate that information and accept it as true. Experiments suggest that the few minutes a respondent spends answering a survey are too short for this process to occur; when respondents in one group are asked their WTP at the time they are given the survey and respondents in a second group are given a day to think before giving their answers, the second group's WTP is significantly lower. And apart from time constraints, a person

may not believe the information she is given; she may instead rely on preconceived notions of chicken-little environmentalists or land-raping corporations to discount or inflate the extent of the harm or length of the recovery period stated by the survey.

Added to those technical concerns is a more fundamental conceptual problem with the information a CV survey provides. When trying to assess damages, a CV survey is presumed to measure people's preexisting values for a resource, which are lost when the resource is harmed. But a person cannot hold a value for something she does not know exists, and the average person is unaware of the specifics of most environmental goods. By educating the person about a resource, the CV survey *creates* the very nonuse value it purports to measure.

Before a person concludes that he would spend $5 per year to prevent logging in a particular tract of forest, he must keep in mind the existence of thousands of acres of comparable forest in the same areas that he might also pay to protect.

People's unfamiliarity with the very concept of expressing a personally held value for a natural resource further compounds the problem. A much-cited study's finding that people give more-or-less accurate answers about their WTP for strawberries is unsurprising; people are familiar with strawberries, they are familiar with paying for them, and they have a good idea of how much strawberries are worth to them.

In contrast, most people are not used to "purchasing" environmental quality; they have never considered how much they would pay for a slight increase in visibility in the Grand Canyon, for a specified decrease in the chances of an oil spill off the coast of Washington and British Columbia, or to prevent the extinction of the whooping crane. (All were the subjects of actual CV surveys.)

Inexperience in valuing natural resources produces one problem in particular that appears to dwarf all others: when determining their WTP, people do not keep in mind the substitutes for the resource in question, the other environmental goods they might pay to protect or improve, and all the other goods and services they might purchase with their limited household income. For example, before a person concludes that he would spend $5 per year to prevent logging in a particular 1,000-acre tract of forest, he must keep in mind the existence of thousands of acres of comparable forest in the same area that he might also pay to protect, as well as any equivalent desire to protect forests in other areas, wetlands in Louisiana, coastline in Alaska, water quality in Minnesota lakes, air quality in Colorado, and all other environmental goods—all while still paying his taxes and purchasing food, clothes, housing, and entertainment. The average WTPs reported by most CV surveys suggest that respondents do not construct their WTP with a realistic view of substitute resources, their alternative purchases, and their budget constraints.

Once again, to reach an accurate WTP, respondents must sort through a thicket of complex, technical data about a resource and an environmental effect with which they have little or no experience. For what is likely the first time in their lives, they must state a value for a resource that they may have never used or even seen, while keeping in mind possible substitutes and their budget constraints. They must do so in the context of a survey that lasts perhaps 10 or 20 minutes. Arriving at an accurate answer would take a great deal of effort, and respondents have little reason to invest such effort when answering hypothetical questions in a survey.

Finally, all of the above assumes that a respondent will attempt to answer the question truthfully. In fact, she may have an incentive to lie. A person may purposefully overstate her WTP because she views the survey as a costless way to send a message that society should spend more money on the resource or on the environment generally. Or she may overstate her answer because she views a higher number as a more socially acceptable answer. (For similar reasons, survey estimates of charitable giving consistently overstate actual charitable giving by a substantial margin.)

Recent Assessments of Contingent Valuation's Validity

Most of CV's biases have been discussed in CV literature for years. That literature, however, was written by researchers specializing in CV studies. Understandably, their approach was to

list the sources of bias and suggest techniques to minimize that bias rather than to question whether even the most accurate CV study is accurate enough to be useful. Any assessments of CV as a methodology relied on preexisting CV studies that had been performed to determine a value for a resource rather than determine the validity of the valuation method; almost no attempt was made to test the reliability of CV measurements of nonuse values directly.

The lack of direct research into CV's validity was understandable for practical as well as motivational reasons. One can test the reliability of nonmarket measurements of *use* values by comparing the study results to people's actual behavior or to the results from other methodologies. But CV is a potentially valuable tool precisely because no behavior and no other indirect methods exist to measure nonuse values—there is no yardstick against which one can measure its accuracy. Even so, CV's reliability can be tested indirectly by examining the internal consistency of CV results, comparing different CV studies' measurements of the same good, and comparing CV results to people's behavior in equivalent situations.

Exxon, which faced multi-billion-dollar damage claims in the *Exxon Valdez* litigation that were largely based on CV studies, was the first party with the incentive to undertake an empirical analysis of CV's validity. Even though Exxon settled all claims by the federal and Alaskan governments, the potential for similar claims from future spills was clear, and federal agencies were proposing damage assessment rules for oil spills that would give government officials the discretion to base their claims on CV estimates. Faced with these prospects, Exxon decided to fund direct studies of CV's validity, with the obvious hope that the results would be critical of CV.

Exxon did not skimp on funding. The participants were among the nation's most prominent economists and econometricians: professors from schools like MIT, Harvard, Stanford, and Berkeley, members of leading economics consulting firms, recipients of the John Bates Clark Medal, and at least one expected recipient of a future Nobel Prize. They presented the results of their studies at a conference attended by federal agency officials and established CV practitioners. The editor of the collected papers (Hausman, 1993) ended the preface with a thinly veiled estimation of the comparative quality of the new research: "CV has been studied almost solely by economists who specialize in environmental economics. However, the research presented here is mainly from economists with specialties in economic theory, econometrics, and public finance, rather than from the more narrowly focused research of environmental economists."

The conclusions of the conference papers were indeed "quite critical." One paper reviewed recent studies by CV practitioners and made the simplest observation about CV's validity: CV-produced estimates are just not credible, on the individual or aggregate level. Surveys regularly produce individual WTP responses of $1,000 or more for a commodity with an average WTP of only around $20. No one defends the $1,000 answers—but even the averages are not credible in the context of a household's budget constraints. A household WTP of $20 (which is lower than most average WTPs) implies a similar WTP for each of the thousands of other similar environmental goods. Summing the household's WTPs across all these resources produces a total many times the household's entire income.

A study estimated a mean individual WTP of up to $149 per year to prevent the extinction of the whooping crane, which equals $27.7 billion per year for all U.S. adults. Because only 168 whooping cranes exist, this total represents $165 million per bird, per year.

When average WTPs are multiplied by the number of households to produce a total value for a resource, those totals are even less credible, in light of what society actually spends on environmental quality and other public goods. Consider:

- A study estimated a mean individual WTP of up to $149 per year to prevent the extinction of the whooping crane, which equals $27.7 billion per year for all U.S. adults. Because only 168 whooping cranes exist, this total represents $165 million per bird, per year.
- A study estimated a median WTP of $124 to $143 (Australian dollars) and a mean WTP of

$1,299 to $1,525 per year (10 percent of the average household's income) to prevent mining for 10 years on a 20-square-mile tract of national park in northern Australia. When summed across Australia's 12.3 million adult population, the total value is up to $1.8 billion per year using the median and $18.7 billion using the mean.

- Based on a mean WTP of $50 to $144 per year and a "best estimate" household WTP of $86 per year to protect the old-growth habitat of the spotted owl in the Pacific Northwest, a study found a nationwide WTP of $8.3 billion per year. Using a 4 percent discount rate, the present value of a 30-year commitment is $215 billion.
- A study of the *Nestucca* oil spill estimated that the people of Washington and British Columbia alone would be willing to pay as much as $11,950 for each seabird saved from oil spills off their coasts, even though the seabird populations would fully recover in five to 10 years.

With so many bids of zero, one would expect an equally large number of very small bids, such as 25¢, 50¢, or $1. But in fact, respondents almost never give positive responses of less than $5.

Even apart from the intuitive implausibility of the size of average stated WTPs, an examination of individual WTP responses within a single CV survey reveals several inconsistencies. Along with a large number of $1000+ answers, CV surveys generate a large proportion of $0 responses—30 to 40 percent in many studies. Follow-up questions show that some of these answers are not true statements of WTP but rather "protest zeros": a refusal to state a positive number because a person has an ethical objection to the concept of trading money for environmental harm, because he thinks someone else should pay for the improvement, or because he did not feel he was given enough information to make a good estimate.

Many of the $0 answers, however, are actual statements that the person is willing to pay nothing for the improvement in the resource. With so many bids of zero, one would expect an equally large number of very small bids, such as 25¢, 50¢, or $1. But in fact, respondents almost never give positive responses of less than $5.

This unusual distribution of many $0 and extremely large answers makes it difficult to calculate a mean WTP with any degree of confidence. A tiny fraction of individuals with extremely large stated WTPs can have a huge effect on the final estimate. If those individuals are overstating their true WTP either intentionally or unintentionally, they can inflate the estimated value by several orders of magnitude. Even if their answers are accurate, the dispersal of large and small answers causes enormous variations in the calculated mean depending on which statistical model is used to perform the calculation. For example, different specifications of a single model used in the northern Australia study produced means ranging from $123 to $2,616,726 *per person per year*. The economist reviewing this study concluded that the model used to produce the high estimate was just as statistically valid as the other models used.

Similarly, different specifications of a single model used in the whooping crane study produced a range from $21 to $149; the study's authors described their selection of one model producing a $130 annual mean WTP as simply "a judgment call by researchers."

Individual responses within a single study contain other inconsistencies as well. Most CV surveys include follow-up questions about the respondent's demographics. The answers to those questions reveal that a person's WTP varies only slightly with her income, a curious result given that environmental quality is essentially a luxury good. WTP responses also do not vary when people are asked about their individual WTP instead of the WTP for their entire household and do not vary with the size of their household; researchers do not know whether people are ignoring the question and consistently giving only their own WTP or ignoring the question and consistently giving a WTP for their household. Because the mean WTP is summed across either all households or all individuals depending on the question asked, this effect can create an immediate bias of 2.3 times the total estimate (the size of the average household) in either direction.

Inconsistencies Among Different CV Surveys Measuring the Same Resource

The most telling lessons from the Exxon confer-

ence papers come from the results of experiments that compared the results of studies measuring the same resource using slightly different surveys or measuring different resources using identical surveys. Researchers had long known that changes in the presentation of questions can produce large changes in people's answers. For example, when people are asked their willingness to accept some amount of money in exchange for harm to a resource (WTA) rather than their willingness to pay to prevent that identical harm, their answers are substantially higher. No valid theoretical justification for more than a tiny discrepancy exists. (Most commentators agree on the actual explanation: by offering money in exchange for pollution, WTA questions connote a bribe to look the other way; some respondents become indignant and give unrealistically high answers.)

The ordering of questions is also significant. One well-known study found that Chicago residents had a mean WTP of $90 to improve visibility in the Grand Canyon. A study the following year asked Chicago residents the same question after first asking their WTP for visibility improvements in Chicago and in the Eastern United States; the mean WTP for the Grand Canyon was only $16.

CV practitioners try to explain this anomaly as the result of income effects and substitution effects: after a person has "spent" hypothetical amounts on goods valued in earlier questions, she has less income to "spend" on a good valued in a later question, and the hypothetical improvements to the earlier goods substitute for and devalue a hypothetical improvement to the later good. But both effects would need to be enormous to produce the anomaly created by changing the question sequence. Direct measurements show that income and substitution effects are actually quite small.

The Exxon conference studies also were among the first to measure the effect that the *form* of the question has on the resulting WTP. A survey may ask for a person's WTP by simply asking an open-ended question ("What is the most that your household would agree to pay . . . ?"), it may provide a number and then ask if the person would pay that amount (called "referendum format" or "dichotomous choice"), or it may provide a list of dollar values and ask the person to check the appropriate amount (called a "payment card").

One conference study gave different groups of people surveys that were identical except for the question format: one set used open-ended questions, while the other used a referendum format. The average WTP from the referendum format was significantly higher than that from the open-ended-question format, and the distributions of answers were quite different. Most significantly, over 30 percent of referendum-format respondents stated that they would pay $1,000, while only 3 percent of the open-ended responses were $1,000 or more. As with the disparity between WTP and WTA responses, no theoretical justification for these discrepancies exists.

Another study used two different questioning techniques to explore the extent to which people overlook other possible resources on which they might spend their limited incomes. The "single focus" survey used the standard CV survey technique of simply asking people their WTP—in this case, their WTP to prevent oil spills in Alaska. The second, "top-down disaggregation" survey asked people their total WTP for a broad list of social programs. It then asked people to disaggregate that amount, step by step, into the proportion they would devote to environmental programs; the proportion of *that* amount they would devote to wilderness areas; to human-caused problems in wilderness areas; to marine oil spills; and finally, to marine oil spills in Alaska. The average WTP obtained by the single-focus survey was *290 times larger* than that from the top-down survey.

When people are asked their willingness to accept some amount of money in exchange for harm to a resource rather than their willingness to pay to prevent that identical harm, their answers are substantially higher.

Inconsistencies Among CV Surveys Measuring Different Quantities of a Resource

The most damning tests of CV's validity used several different surveys that were identical in every respect except for the extent of the resource being valued. If people's stated WTP represents their true economic preferences (as

they must if CV surveys are measuring the value of a resource), those stated values must be continuous and additive: a person must prefer more of a resource rather than less, and the sum of values she places on A and B separately should equal the value she holds for A + B.

Casual observations of CV results had long suggested that people's stated WTP did not meet those conditions: most CV estimates tend to fall within a similar range, regardless of the type or extent of the resource being valued. But the studies funded by Exxon were among the first to test this proposition directly. One study asked people their WTP for wire-net covers on waste-oil holding ponds that would protect migratory waterfowl against landing in the oil and dying—a simple procedure with concrete, easy-to-understand, and quantifiable benefits. Each respondent received one of three versions of the survey, indicating that either 2,000, 20,000, or 200,000 birds would be saved annually. Despite the variation in resource size by a factor of 100, the average WTPs from all three versions were virtually identical.

The WTP for all 57 areas was only three times larger than the WTP for any one area—not nearly large enough given the huge change in the resource being valued.

The same researchers conducted a different study that asked one group their WTP to build a nationwide network of local oil-spill response centers that would reduce the effects only of small oil spills (less than 50,000 gallons). A second group of respondents were asked their WTP to build not only the local response centers, but also larger regional centers that would reduce the effects of all oil spills. Even though the second survey measured all of the benefits of the first survey plus some additional benefits, it produced an average WTP that was 50 percent *lower*.

Finally, a different group of researchers asked different groups of respondents their WTP to preserve three different federally-protected wilderness areas out of the 57 such areas located in the western United States. Three groups were asked their WTP to protect a specified one of the three areas given that seven other areas would be developed; two other group were asked their WTP to preserve one area given that eight and nine others would be developed, respectively; another was asked its WTP to preserve two of the areas together; another was asked its WTP to preserve all three areas together; and a final group was asked its WTP to preserve all 57 areas. The researchers ran statistical tests on the data using a variety of models. The results were almost surprisingly consistent from each model:

- The WTP for each of the three areas was the same.
- The WTP was the same whether seven, eight, or nine other areas are to be developed (i.e., substitution effects were so small as to be unobservable).
- The sum of the WTPs for two (or three) areas individually was far greater than the WTP for the same two (or three) areas jointly.
- Indeed, statistical tests could not reject the possibility that the WTP to protect all three areas together was the same as the WTP to protect only one area.
- The WTP for all 57 areas was only three times larger than the WTP for any one area—not nearly large enough given the huge change in the resource being valued.

Does Contingent Valuation Measure Real Values at All?

Most of the anomalies in CV data are well-known to practitioners. They attempt to account for them by adjusting the data to create more "conservative" estimates. Surveys ask for people's willingness to pay for an improvement rather than their willingness to accept more pollution, even though many practitioners argue that WTA is the theoretically correct measurement for damage assessments. Researchers eliminate "outlier" responses that exceed some specified cutoff, such as $1,000 or 1 percent of a household's income, even though such culling is essentially an admission that at least some respondents answer CV surveys using a mental model other than the one assumed by CV.

Some studies (for example, the northern Australia study detailed above) have used the median WTP rather than the mean, even though the mean is the correct variable for calculating the total value for a population. And researchers reject certain statistical models used to calculate a mean WTP when they produce results they feel are unrealistic.

All the anomalies, however, raise some fairly

basic questions. When a person's answer is so sensitive to minor and theoretically irrelevant changes in questioning format, is CV measuring a value that he actually holds? When a person's answer is *not* sensitive to the factors it should be, is he expressing a value that he holds for the resource being measured, or for something else? When CV produces results that far exceed any credible number, are those results merely inaccurate, or are they completely divorced from the reality they purport to measure?

The experiments reported in the conference papers appear to answer those questions. If CV results strongly violate fundamental principles of the economic theory on which they are based, CV does not measure people's economic preferences for natural resources, and no amount of survey refinement or data adjustment will produce an accurate number. Such adjustments (and the implicitly admitted need for them) are curious in the first place: no theoretical justification exists for any of them. The process is akin to generating a random number with a Rube Goldberg device and then turning a few dials and levers to produce a lower number if the first one "seems too high." The second number is no better an estimate of anything real than the first.

What Does Contingent Valuation Actually Measure?

If CV surveys do not measure the values people place on improvements in specific resources, what *do* they measure? The constancy of WTP responses across widely different quantities of a resource provides part of the answer: people are expressing their support for preserving the environment in general, and their value for the specific resource being measured is embedded in that larger amount.

Recent articles have gone further and suggested that people are expressing support not just for the environment, but for *whatever* good cause a survey covers. People view the hypothetical amount they "pay" as an imaginary gift to charity, and that gift creates the warm glow associated with altruism—the very act of answering the survey has value for the respondent. Thus, the numbers people report in CV surveys reflect a "purchase" of not only the resource being measured, but also moral satisfaction.

This logic helps explain several characteristics of CV results: people give very few answers that are above $0 but less than $5 because they first decide whether the resource in question is a cause they would support financially and then pick a round number to "donate" to that cause. As one commentator noted, charities that solicit individuals directly "see the same pattern of gifts—many zeros and some sizeable checks." The moral satisfaction component also accounts for the similar size of average WTPs for different quantities of a resource: contributing to some subset of a cause may be as morally satisfying as contributing to the cause in its entirety.

One additional conference study provided some of the only direct evidence of those motivations. The researchers focused not on the numbers produced by survey respondents, but rather on the thought processes people used in reaching their answers. The authors suggested that people have well-defined values only for very familiar goods; when questioned about unfamiliar natural resources that they have never bought or sold, people must *construct* a value rather than simply retrieve a previously formed value. People may use any number of different methods to construct their answer that have little or nothing to do with the maximization of their expected utility.

People view the hypothetical amount they "pay" as an imaginary gift to charity, and that gift creates the warm glow associated with altruism—the very act of answering the survey has value for the respondent.

To examine these methods, researchers asked people to vocalize their thoughts while they answered a CV survey and then recorded, coded, and analyzed the vocalizations. The results were surprising even to skeptics:

- 23 percent of respondents began by accepting the cost as inevitable and then estimated how much of the burden would fall on them personally. Notably, people's calculations in estimating their personal share were often highly inaccurate.
- 23 percent suggested a desire to show concern for preserving the environment generally. For example: "I think $500 is not very much to

spend each year in taking care of our world."
• 20 percent said that they just made up a number or guessed: "Um, I have no idea. I guess $500 sounds like a nice round number." "There was no67 thought really put into it. I think the $100 figure just popped into my head and that's why I put it down, really."
• 17 percent explicitly viewed their WTP amount as a gift to charity. "I, uh, was just thinking about how I make a donation to, like, maybe the Fraternal Order of Police or to, uh, MADD or different types of, of things that are for the good of our society." "To charities I usually give no more than $20, so that's how I arrived at my $20 figure."

If contingent valuation does not measure people's economic preferences at all, it should not be used in either cost-benefit policy analyses or damage assessments.

When asked a follow-up question of whether they would be willing to support other important issues with a similar dollar amount, 26 percent said no, and those who said yes were on average willing to support only 2.9 other issues. Just as with true charitable giving, respondents would "support" only a small number of causes even though they knew other causes exist that are worthy of support. Indeed, this follow-up question prompted many respondents to realize the far-reaching implications for their household budget of the WTP response they had given; they indicated that their answer was really too large or that it should go for all similar issues.

Implications for the Use of Contingent Valuation

The conference researchers drew the natural conclusion from their studies: if CV does not measure people's economic preferences at all, it should not be used in either cost-benefit policy analyses or damage assessments. Since CV is currently the *only* way to measure nonuse values, mothballing it means giving up any precise estimate of such values. But a precise number that is wildly inaccurate is worse than no number at all: including estimates that are as highly biased as CV results appear to be could result in massive misallocations of resources. The summing of an average WTP across millions of people can produce enormous estimates for even small environmental changes. With approximately 100 million U.S. households, a mean WTP of only $10 per household (well below the results of most CV studies) produces a total estimate of $1 billion for the nation. An error of only $1 per household creates $100 million in phantom value. (Recall that actual CV studies often estimate a mean WTP as being *somewhere* within a range that can be as wide as $21 to $149, for the whooping crane study.)

With the potential for such large estimates and with such uncertainty over the outcome of any particular CV study, the risk of paying CV-based damages for a relatively small environmental accident becomes a "bet the company" proposition for even the largest corporations. Companies would be forced to take safety precautions whose costs far exceed their true benefits or forgo the activity altogether. They must include the cost of the precautions (and of potential liability) in the price that consumers pay for their products. As an example, if the average household WTP for a dolphin is only 1 cent and large-net tuna fishing kills one dolphin per 10,000 cans of tuna, including the CV-determined cost of dolphin deaths could add $100 to the price of a can of tuna.

Using CV as a public-policy tool would have similar effects. Environmental regulations are extremely costly (in terms of spending on direct compliance and enforcement, indirect effects on the economy, and the personal inconvenience of forced car-pooling and the like). So are cleanup efforts. Using CV to estimate the benefits of a proposed policy can counsel for the policy's implementation when its true benefits are dwarfed by the costs.

Balanced against those effects are the costs of excluding a measurement of nonuse values from damage calculations or policy analyses. In the damage assessment context, however, nonuse values are likely to be significant only when there is long-term harm to a unique and well-known natural resource. If a resource recovers quickly, both knowledge of its existence and the option to use it remain; if it is not unique, the presence of substitutes should diminish a person's nonuse value to near-zero; if it is unknown to most people, they cannot attach a value to it. In the policy context, society can and does account for nonuse values when making important decisions by simply acknowledging their

existence, without attempting to place a number on them. We should continue doing so.

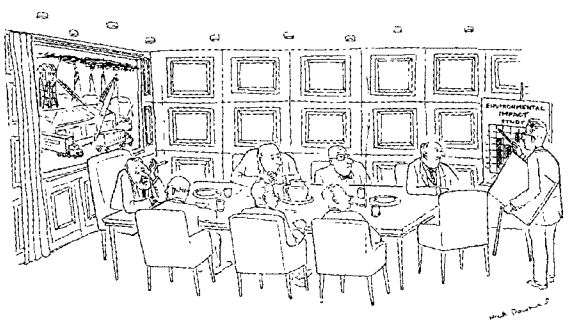
Reactions to the Exxon Conference Studies

The CV practitioners who attended the conference (and whose careers and livelihoods depend in part on conducting CV studies) were less than sanguine about those conclusions. The book collecting the papers (Hausman, 1993) includes transcripts of the discussions that occurred after each paper was presented, and the comments from CV practitioners were quite acrimonious.

One of the first comments was the most obvious and, potentially, powerful one: the funding for the research and the conference came from Exxon, which had just lived through the *Valdez* disaster, will inevitably be the source of at least small oil spills in the future, and thus has a strong interest in seeing CV discredited. The researchers were top-level, but anyone familiar with expert witnesses in litigation knows that even academics at the top of their field can have remarkably malleable opinions. Were the conference papers the best research on CV's validity to date, or merely the best opinions money could buy?

The answer lies in the research itself: it was well-documented, and if either the results or the conclusions to be drawn from them were biased, that bias should be evident. It is not. CV practitioners pointed to no real flaws at the conference, and they have not done so to date. Most of the criticisms at the conference concerned details of the survey design or implementation, and because the studies largely involved comparisons of surveys with identical design and implementation, those details should not have affected the comparisons.

Another criticism was notable for its broader implications: the survey valuing the deaths of 2,000, 20,000, or 200,000 birds told respondents that those bird losses represented "much less than 1 percent," "less than 1 percent," and "about 2 percent" of the relevant waterfowl population, and some argued that this information made the number of birds seem the same in each of the surveys. The three quantities of birds, however, were chosen to track the numbers actually killed by three different oil spills (the Arthur Kill pipeline spill, the *Nestucca* spill, and the *Exxon Valdez* spill). Losses as a percentage of total populations really are less than 1 or 2 percent in such spills, yet the effects of a large spill dwarf the effects of the small one; it is precisely this level of distinction that survey respondents must make if CV studies are to have any meaning.



"Did he say someday we'd be sorry, or Sunday we'd be sorry?"

Reaction of Government Agencies

More important to the future of CV are the reactions of government agencies that are either regulators or potential users of CV studies. Those responses have varied widely. Over nine months before the conference, the Department of the Interior (the agency that started the controversy) had closed off the comment period for its proposed second set of Superfund damage assessment regulations allowing for CV. After the conference, Interior reopened its comment period for additional comments on CV.

In the policy context, society can and does account for nonuse values when making important decisions by simply acknowledging their existence, without attempting to place a number on them.

Interior is being especially cautious—as one might expect after being told by a federal court of appeals that its first set of regulations strictly limiting CV was based on a gross misreading of the Superfund laws and being told by commentators that its reaction to that court ruling would allow CV-based damage assessments that were more fantasy than reality. Its official notice informing the public of the new comment peri-

od states that "the Department remains concerned about the reliability of CVM [contingent valuation method] in calculating nonuse values," but it goes on to assert that "a well-designed CVM survey can satisfy many of the reference operating conditions cited by commentators." Notably lacking is any assertion that CV actually measures nonuse values; a CV survey could satisfy many "reference operating conditions" of design and implementation and still measure something other than people's true economic preferences. Interior is now expected to propose regulations concerning CV that mirror those just released by NOAA.

The EPA has taken a different approach: it has continued to use CV studies in the cost-benefit regulatory impact analyses that are required for all major regulations, without mentioning the new evidence. One recent EPA analysis that used CV to estimate the nonuse values of cleaning up contaminated ground water was rejected by the environmental economics committee of the EPA's own Science Advisory Board. Another CV study that the EPA used to measure benefits from proposed Great Lakes water quality standards notes "an inherent degree of skepticism among some economists regarding the accuracy of results derived from [CV]," but then continues with the same fallacy CV proponents have committed for a decade: "As in any economic research technique, the credibility, accuracy, and robustness of CVM-derived results depend *entirely* on the research protocol applied by the practitioners in designing and implementing the CVM survey instruments. . . . [C]areful design and implementation of surveys allow researchers to test for (and account for) potential biases and embedded values."

The EPA has continued to use contingent valuation studies in the cost-benefit regulatory impact analyses that are required for all major regulations.

Such wishful thinking comports with that of the state environmental departments, who stand to claim and collect higher amounts from hazardous waste and oil spill damage assessments that use CV estimates. Their comments on Interior's proposed regulations included such bald assertions as "CVM is an acceptable methodology period."

The most detailed (and reasoned) response to date has come from the NOAA, which is charged with creating damage assessment regulations for oil spills. Initially scheduled to propose its regulations in August 1992, NOAA extended its comment period five times to receive more evidence on CV and just released its proposed regulations this past January. Concurrent with the comment period, NOAA appointed a Contingent Valuation Panel of economic experts who were more-or-less unattached to either side of the CV debate, co-chaired by two Nobel laureates.

Both the Panel's report (NOAA, 1993) and the proposed rules (NOAA, 1994) are overly judicious. The Panel found that CV's many biases and anomalies, described at length in its report, are "particularly compelling," "are not adequately addressed by even the best CV studies," and "will need to be convincingly dealt with in order to assure the reliability of estimates." However, it rejected as "extreme" the suggestion that "there can be no useful information content to CV surveys" and instead asserted that "CV studies can produce estimates reliable enough to be the starting point" of an estimate. The report provides a set of "stringent" guidelines for CV studies used in federal damage assessments and states that "[m]any departures from the guidelines or even a single serious deviation would suggest unreliability."

Although NOAA's proposed rules do impose many of the Panel's guidelines as requirements that will lower the amount of bias in CV estimates, they take a step backwards from the Panel's conclusions. The rules' preamble downplays commentators' concerns with CV's reliability (shared by the Panel), and the obligatory "response to comments" section answers many of the concerns with a terse "NOAA finds no evidence" or "NOAA is not convinced."

Truly astounding, though, is a requirement buried in the proposed rules with almost no discussion or mention anywhere in the 130-page document: "Because of the various possible biases, a discount factor is included in the proposed rule to apply to estimated WTP. The proposed rule gives a default factor of 50 percent for the purposes of soliciting comment." In short, spend millions to generate a precise num-

ber; then divide by two. This section of the rules expressly notes the disparity between "hypothetical WTP" and "actual WTP." Although it contemplates a method of calibrating the two that is more sophisticated than a naked 50-percent discount, it gives no indication of how this might be done.

CV Surveys as Surrogate Referenda

Less astounding but just as noteworthy is the focus of both the report and rules on the use of a referendum format as the most important way to reduce bias. The Panel noted that, because people commonly vote in real referenda, a referendum-format survey should increase accuracy by providing a more familiar and realistic context to respondents. Although the Panel conceded that many biases would remain, it stated that these same biases "could occur in real referenda"; "since real referenda are exposed to most of the response effects that occur with attitude surveys, and since we take the result of referenda as telling us something about 'true' preferences, it is not necessary to claim they can be eliminated completely in a CV study."

That conclusion confuses the different purposes served by referenda and CV studies. Referenda offer citizens an opportunity to make policy decisions directly rather than through their representatives. They are most useful (and usually succeed) only when the majority's position is clear and their legislators' individual incentives prevent that position from becoming law through normal legislative channels. (California's tax-slashing Proposition 13 and the current wave of referenda on legislative term limits are perfect examples.) In those situations, the potential for inaccuracy is far less important than the opportunity for pure democracy to override the imperfections of a representative system of government.

Referenda are far less reliable or useful when they address issues that are complex, technical, or closely balanced among the electorate. In those situations, all of the biases that "could occur in real referenda" can overwhelm the process, and many people have little faith that the outcome reflects either well-informed or well-reasoned opinions. Indeed, commentators often give near-total credit for a referendum's defeat or passage to the ability of propaganda-like advertising campaigns to sway an uninformed or careless electorate.

CV studies are not referenda that produce a law; they are essentially opinion polls that produce a number, which will then be used either by bureaucrats who make regulatory decisions about the environment or by courts who order private parties to pay damages for environmental harm. CV studies offer none of the benefits of a direct vote by citizens but present all of the pitfalls of the worst referenda—and they require far greater reliability. Both regulators and courts would be better served by foregoing the use of numbers that cloak gross overestimation and inaccuracy in the illusion of precision.

Selected Readings

Cummings, R.G., Brookshire, D.S., and Schulze, W.D. *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method.* Towata, N.J.: Rowan & Allanheld, 1986.

Hausman, J.A. *Contingent Valuation: A Critical Assessment.* Amsterdam: North-Holland, 1993.

Mitchell, R.C. and Carson, R. *Using Surveys to Value Public Goods: The Contingent Valuation Method.* Washington, D.C.: Resources for the Future, 1989.

Niewijk, R.K. "'Ask a Silly Question . . . ': Contingent Valuation of Natural Resource Damages." Harvard Law Review, Vol. 105, No. 8 (1992).

National Oceanic and Atmospheric Administration. "Natural Resource Damage Assessments Under the Oil Pollution Act of 1990." Federal Register, Vol. 58, No. 10 (Jan. 15, 1993); Vol. 59, No. 5 (Jan. 7, 1994).