

Report of the NOAA Panel on Contingent Valuation

January 11, 1993

Kenneth Arrow
Robert Solow
Paul R. Portney
Edward E. Leamer
Roy Radner
Howard Schuman

Report of the NOAA Panel on Contingent Valuation

Date: May 9, 2001

I. INTRODUCTION

Under the Oil Pollution Act of 1990, the President--acting through the Under Secretary of Commerce for Oceans and Atmosphere--is required to issue regulations establishing procedures for assessing damages to or destruction of natural resources resulting from a discharge of oil covered by the Act. These procedures are to ensure the recovery of restoration costs as well as the diminution in value of the affected resources and any reasonable costs of conducting the damage assessment.

At least some of the values that might be diminished by such a discharge are relatively straightforward to measure through information revealed in market transactions. For instance, if the discharge kills fish and thereby reduces the incomes of commercial fishermen, their losses can reasonably be calculated by the reduced catch multiplied by the market price(s) of the fish (less, of course, any costs they would have incurred). Similarly, if the discharge of oil discourages tourist travel to an area, the lost incomes of those owning and/or operating motels, cottages, or other facilities can be reasonably represented by the difference in revenues between the affected period and a "normal" season. Even the losses to recreational fishermen, boaters, swimmers, hikers, and others who make active use of the areas affected by the discharge can be included in the estimate of diminished value, although these losses will

generally be somewhat more difficult to value than the more obvious out-of-pocket losses.

The losses described above have come to be known as lost "use values" because they are experienced by those who, in a variety of different ways, make active use of the resources adversely affected by the discharge. But for at least the last twenty-five years, economists have recognized the possibility that individuals who make no active use of a particular beach, river, bay, or other such natural resource might, nevertheless, derive satisfaction from its mere existence, even if they never intend to make active use of it.

This concept has come to be known as "existence value" and it is the major element of what are now referred to as "non-use" or "passive-use" values (the latter term is employed in the balance of this report). In regulations promulgated by the Department of the Interior in 1986 under the Comprehensive Environmental Response, Compensation, and Liability Act -- regulations that also pertained to natural resource damage assessments -- passive-use values were included among the losses for which trustees could recover. The inclusion of passive-use values was recently upheld by the D. C. Court of Appeals (State of Ohio v. Department of the Interior, 880 F.2d 432 (D.C. Cir. 1989)), as long as they could be reliably measured.

This begs an interesting and important question, however. If passive-use values are to be included among the compensable losses for which trustees can make recovery under the Oil

Pollution Act, how will they be estimated? Unlike losses to commercial fishermen or recreational property owners, there are no direct market transactions that can be observed to provide information on which estimates can be based. Unlike losses to boaters, swimmers, recreational fishermen and others, there exist no indirect methods through which market data can provide at least some clues as to lost values. In other words, there appear to be neither obvious nor even subtle behavioral trails that can provide information about lost passive-use values.

Some experts believe that there exists an approach that can provide useful information about the economic significance of the lost passive-use values individuals may suffer when oil discharges damage natural resources. Known as the contingent valuation (or CV) technique, this approach is based on the direct elicitation of these values from individuals through the use of carefully designed and administered sample surveys. Its appeal lies in its potential to inform damage assessment in an area (lost passive-use values) where there appear to be no behavioral trails to be followed.

Typically, CV studies provide respondents with information about a hypothetical government program that would reduce the likelihood of a future adverse environmental event such as an oil spill, chemical accident, or the like. Respondents are usually given some specific information about the exact nature of the damages that the program in question would prevent. And they are also confronted in the study with a question or questions that

provide information about the economic sacrifice they would have to make to support the environmental program. This may take the form of an open-ended question asking what is the maximum amount they would be willing to pay for the program in question; it may involve a series of questions confronting them with different prices for the program depending on their previous answers; or it may take the form of a hypothetical referendum (like a school bond issue) in which respondents are told how much each would have to pay if the measure passed and are then asked to cast a simple "yes" or "no" vote. (The conceptually correct measure of lost passive-use value for environmental damage that has already occurred is the minimum amount of compensation that each affected individual would be willing to accept. Nevertheless, because of concern that respondents would give unrealistically high answers to such questions, virtually all previous CV studies have described scenarios in which respondents are asked to pay to prevent future occurrences of similar accidents. This is the conservative choice because willingness to accept compensation should exceed willingness to pay, if only trivially; we say more about other biases below.)

The CV technique has been used for twenty years or so to estimate passive-use values. In the last five years, however, there has been a dramatic increase in the number of academic papers and presentations related to the CV technique. This is due in part to the availability of comprehensive reference texts on the subject (Mitchell and Carson (1989), for instance), and to

the growing interest both nationally and internationally in environmental problems and policies. But it is also attributable to the growing use of the CV technique in estimating lost passive-use values in litigation arising from state and federal statutes designed to protect natural resources. Since Ohio v. Department of the Interior admitted the concept of passive-use values in damage assessments, this can only give added impetus to the use of CV in such litigation.

The CV technique is the subject of great controversy. Its detractors argue that respondents give answers that are inconsistent with the tenets of rational choice, that these respondents do not understand what it is they are being asked to value (and, thus, that stated values reflect more than that which they are being asked to value), that respondents fail to take CV questions seriously because the results of the surveys are not binding, and raise other objections as well. Proponents of the CV technique acknowledge that its early (and even some current) applications suffered from many of the problems critics have noted, but believe that more recent and comprehensive studies have already or soon will be able to deal with these objections.

This (sometimes acrimonious) debate has put the National Oceanic and Atmospheric Administration (NOAA) in a very difficult spot. NOAA must decide in promulgating the regulations under the Oil Pollution Act whether the CV technique is capable of providing reliable information about lost existence or other passive-use values. Toward this end, NOAA appointed the

Contingent Valuation Panel to consider this question and make recommendations to it.

This report is the product of the Panel's deliberations and is organized in the following way. Following this introduction, the drawbacks to the CV technique are discussed in Section II. Section III discusses several key issues concerning the design of CV surveys, including use of the referendum format to elicit individual values, ways of addressing the so-called "embedding" problem, and the evaluation of damages that last for some period but not forever. Section IV presents guidelines to which the Panel believes any CV study should adhere if the study is to produce information useful in natural resource damage assessment. (These are elaborated upon in an Appendix.) In Section V a research agenda is described; it is the Panel's belief that future applications of the CV technique may be less time-consuming and contentious if the research described in the agenda is carried out. Section VI presents the Panel's conclusions.

II. CRITICISMS OF THE CONTINGENT VALUATION METHOD

The contingent valuation method has been criticized for many reasons and the Panel believes that a number of these criticisms are particularly compelling. Before identifying and discussing these problems, however, it is worth pointing out that they all take on added importance in light of the impossibility of validating externally the results of CV studies. It should be

noticed, however, that this same disadvantage must inhere in any method of assessing damages from deprivation of passive-use. It is not special to the CV approach although, as suggested in Section I, there are currently no other methods capable of providing information on these values.

One way to evade this difficulty, at least partially, is to construct experiments in which an artificial opportunity is created to pay for environmental goods. The goods in question can perfectly well involve passive use. Then the results of a CV estimate of willingness to pay can be compared with the "real" results when the opportunity is made available to the same sample or an analogous sample.

A few such experiments have been attempted. The most recent, due to Seip and Strand (1992), used CV to estimate willingness to pay for membership in a Norwegian organization devoted to environmental affairs, and compared this estimate with actual responses when a number of the same respondents were presented with an opportunity actually to contribute. The finding was that self-reported willingness to pay was significantly greater than "actual" willingness to pay. A recent study by Duffield and Patterson (1991) took as the environmental amenity in question the maintenance of stream flow in two Montana rivers. The rivers in question provided spawning grounds for two rare species of fish; passive use was believed to be the main motivation for respondents. One of two parallel samples was asked about hypothetical willingness to contribute to the Montana

Nature Conservancy which would then maintain stream flow; the other was offered an opportunity actually to contribute to the same organization for the same purpose. It was found that response rates and expressed willingness to contribute were significantly higher when the contribution was hypothetical than when "expressed willingness" meant an immediate cash contribution. On the other hand, the size of contributions, hypothetical in one case and actual in the other, was not much different as between those who said they would contribute and those who did so.

These studies suggest that the CV technique is likely to overstate "real" willingness to pay. Duffield and Patterson, however, hold out hope that the differences are small enough and predictable enough that CV estimates could be discounted for possible overstatement and then used as a conservative estimate of willingness to pay. Clearly more such experiments would be useful.

A less direct test of the "reality" of CV estimates of lost passive use values is to use the technique to estimate willingness to pay for ordinary market goods and then to compare the results with actual purchases. This has been tried by Dickie, Fisher, and Gerking (1987) using the demand for strawberries. When the data were re-analyzed by Diamond, Hausman, Leonard, and Denning (1992), it was found that the CV approach tended systematically to overestimate quantity demanded at each price, sometimes by as much as 50 percent. This result

has to be qualified in two ways. First, the original CV study seems to have been fairly casual by the standards now proposed by practitioners; pre-testing and improvement of the survey instrument might (perhaps) have narrowed the gap. And second, it seems to go too far to conclude from systematic over-estimation that the CV study, even as conducted, provides no information about the demand for strawberries. Much of the same could be said about a study submitted to the Panel by Cummings and Harrison (1992) comparing hypothetical and demonstrated willingness to pay for small household goods. (See also Bishop and Heberlein (1979).)

External validation of the CV method remains an important issue. A critically important contribution could come from experiments in which state-of-the-art CV studies are employed in contexts where they can in fact be compared with "real" behavioral willingness to pay for goods that can actually be bought and sold.

Of the other problems arising in CV studies, the following are of most concern to the Panel: (i) the contingent valuation method can produce results that appear to be inconsistent with assumptions of rational choice; (ii) responses to CV surveys sometimes seem implausibly large in view of the many programs for which individuals might be asked to contribute and the existence of both public and private goods that might be substitutes for the resource(s) in question; (iii) relatively few previous applications of the CV method have reminded respondents

forcefully of the budget constraints under which all must operate; (iv) it is difficult in CV surveys to provide adequate information to respondents about the policy or program for which values are being elicited and to be sure they have absorbed and accepted this information as the basis for their responses; (v) in generating aggregate estimates using the CV technique, it is sometimes difficult determining the "extent of the market;" and (vi) respondents in CV surveys may actually be expressing feelings about public spiritedness or the "warm glow" of giving, rather than actual willingness to pay for the program in question. We discuss each of these briefly.

Inconsistency with Rational Choice

Some of the empirical results produced by CV studies have been alleged to be inconsistent with the assumptions of rational choice. This raises two questions: What requirements are imposed by rationality? Why are they relevant to the evaluation of the reliability of the CV method?

Rationality in its weakest form requires certain kinds of consistency among choices made by individuals. For instance, if an individual chooses some purchases at a given set of prices and income, then if some prices fall and there are no other changes, the goods that the individual would now buy would make him or her better off. Similarly, we would expect an individual's preferences over public goods (i.e., bridges, highways, air quality) to reflect the same kind of consistency.

Common notions of rationality impose other requirements which are relevant in different contexts. Usually, though not always, it is reasonable to suppose that more of something regarded as good is better so long as an individual is not satiated. This is in general translated into a willingness to pay somewhat more for more of a good, as judged by the individual. Also, if marginal or incremental willingness to pay for additional amounts does decline with the amount already available, it is usually not reasonable to assume that it declines very abruptly.

This point assumes importance in view of some empirical evidence from CV studies that willingness to pay does not increase with the good. In one study, Kahneman (1986) found that willingness to pay for the cleanup of all lakes in Ontario was only slightly more than willingness to pay for cleaning up lakes in just one region. Evidence of this kind has multiplied (see Kahneman and Knetch (1992), Desvousges, et al. (1992), and Diamond et al. (1992)). Desvousges' result is very striking; the average willingness to pay to take measures to prevent 2,000 migratory birds (not endangered species) from dying in oil-filled ponds was as great as that for preventing 20,000 or 200,000 birds from dying. Diminishing marginal willingness to pay for additional protection could be expected to result in some drop. But a drop to zero, especially when the willingness to pay for the first 2,000 birds is certainly not trivial, is hard to explain as the expression of a consistent, rational set of

choices.

It has been argued on a more technical level that the studies finding such apparent inconsistencies are defective, that the choices are not presented clearly to the respondents. In the study referred to immediately above, for instance, respondents were told that 2,000 birds was "...much less than 1%" of the total migratory bird population while 200,000 birds was "...about 2%" of the total. This may have led respondents to evaluate the programs as being essentially the same. But on the face of it, the evidence certainly raises some serious questions about the rationality of the responses.

It could be asked whether rationality is indeed needed. Why not take the values found as given? There are two answers. One is that we do not know yet how to reason about values without some assumption of rationality, if indeed it is possible at all.

Rationality requirements impose a constraint on the possible values, without which damage judgments would be arbitrary. A second answer is that, as discussed above, it is difficult to find objective counterparts to verify the values obtained in response to questionnaires. Therefore, some form of internal consistency is the least we would need to feel some confidence that the verbal answers corresponded to some reality.

Implausibility of Responses

The CV method is generally used to elicit values for a specific program to prevent environmental damage, whether it be

dead animals, spoilage of a pristine wilderness area, or loss of visibility in some very unusually clear area. Though in each case, individuals often express zero willingness to pay, average willingness to pay over the whole sample is often at least a few dollars and frequently \$20 to \$50. With 100,000,000 households in the United States, these responses result in very large totals, frequently over \$1 billion. Some have argued that these large sums are in themselves incredible and cast doubt on the CV method. The Panel is not convinced by this argument, since it is hard to have an intuition as to a reasonable total.

But there is a different problem with these answers. one can envision many possible types of environmental damage -- oil spills or groundwater contamination in many different locations, visibility impairment in a variety of places, and so on. Would the average individual or household really be willing to pay \$50 or even \$5 to prevent each one? This seems very unlikely, since the total resulting willingness to pay for all such programs could easily become a very large fraction of one's income or perhaps even exceed it.

In other words, even if the willingness to pay responses to individual environmental insults are correct if only one program is to be considered, they may give overestimates when there are expected to be a large number of environmental problems. Similarly, if individuals fail to consider seriously the public or private goods that might be substitutes for the resources in question, their responses to questions in a CV survey may be

unrealistically large.

Absence of a Meaningful Budget Constraint

Even if respondents in CV surveys take seriously the hypothetical referendum (or other type of) questions being asked them, they may respond without thinking carefully about how much disposable income they have available to allocate to all causes, public and private (see Kemp and Maxwell (1992), for instance). Specifically, respondents might reveal a willingness to pay of, say, \$100 for a project that would reduce the risk of an oil spill; but if asked what current or planned expenditures they would forgo to pay for the program, they might instead re-evaluate their responses and revise them downward. This is similar to the problem identified immediately above where individuals fail to think of the possible multiplicity of environmental projects or policies they might be asked to support. To date, relatively few CV surveys have reminded respondents convincingly of the very real economic constraints within which spending decisions must be made.

Information Provision and Acceptance

If CV surveys are to elicit useful information about willingness to pay, respondents must understand exactly what it is they are being asked to value (or vote upon) and must accept the scenario in formulating their responses. Frequently, CV surveys have provided only sketchy details about the project(s)

being valued and this calls into question the estimates derived therefrom.

Consider the following example. Suppose information is desired about individuals' willingness to pay to prevent a chemical leak into a river. Presumably, their responses would depend importantly on how long it would take for the chemical to degrade naturally in the river (if it would at all), what ecological and human health damage the chemical would do until it had degraded, and so on. Absent information about such matters, it is unreasonable to expect even very bright and well-informed respondents to place meaningful values on a program to prevent leaks.

Even if detailed information were supplied, there are limits on the ability of respondents to internalize and thus accept and proceed from the information given. It is one thing to tell respondents matter-of-factly that complete recovery will occur in, say, two years. It is another thing for them to accept this information completely and then incorporate it in their answers to difficult questions.

To return to the example above, respondents who take a pessimistic view of the probable consequences of a chemical leak are likely to report relatively high willingness to pay to prevent the contamination -- too high, in fact, if in actuality such an event had less serious effects. On the other hand, respondents with an exaggerated sense of the river's assimilative capacity or regenerative power could be expected to report a

willingness to pay that understates their "true" valuation if provided with a more complete description of likely consequences.

To repeat, even when CV surveys provide detailed and accurate information about the effects of the program being valued, respondents must accept that information in making their (hypothetical) choices. If, instead, respondents rely on a set of heuristics ("these environmental accidents are seldom as bad as we're led to believe," or "authorities almost always put too good a face on these things"), in effect they will be answering a different question from that being asked; thus, the resulting values that are elicited will not reliably measure willingness to pay.

Extent of the Market

Suits for environmental damages are brought by trustees on behalf of a legally definable group. This group limits the population that is appropriate for determining damages even though individuals outside of this group may suffer loss of passive and active use. Undersampling and even zero sampling of a subgroup of the relevant population may be appropriate if the subgroup has a predictably low valuation of the resource. For example, the authors of the CV study conducted in connection with the Nestucca oil spill limited their sample to households in Washington and British Columbia possibly because the individuals living elsewhere were presumed to have values too low to justify examination (or possibly because the sponsors of the study were

agencies of the State of Washington and the province of British Columbia and so defined the legally appropriate population) (Rowe, Shaw, and Schulze, 1992).

"Warm Glow" Effects

Some critics of the CV technique (e.g., Diamond and Hausman (1992)) have observed that the distribution of responses to open-ended questions about willingness to pay often is characterized by a significant proportion of "zeros" -- people who would pay nothing for the program -- and also a number of sizable reports. This bi-modal distribution also characterizes individual giving: most of us give nothing to most charities, but give non-trivial amounts to the ones we do support (at least \$10 or \$20, say). This has led these critics to conclude that individuals' responses to CV questions serve the same function as charitable contributions -- not only to support the organization in question, but also to feel the "warm glow" that attends donating to worthy causes (see Andreoni (1989)). If this is so, CV responses should not be taken as reliable estimates of true willingness to pay, but rather as indicative of approval for the environmental program in question.

III. KEY ISSUES IN THE DESIGN OF CONTINGENT VALUATION INSTRUMENTS

In the course of its deliberations, the Panel discussed many issues surrounding the design of CV surveys. Here we provide our

views on several issues that are especially important. In Section IV and in an Appendix to this report, we provide much greater detail on the characteristics of a valid application of the CV method.

The Referendum Format

Considered as a survey, a CV instrument is descriptive rather than explanatory. Description may be as simple as reporting univariate averages of one kind or another, such as the percentages of those employed, seeking work, and not seeking work in the United States, the mean number of rooms occupied by American households, or the proportion of "likely" voters favoring one or another candidate in an upcoming election. A CV study seeks to find the average willingness to pay for a specific environmental improvement. Nevertheless, as will be seen later, it is often desirable to ask respondents to specify the reasons for their reported choices.

Univariate descriptive results are meaningful mainly when the alternative responses to a question are simple and can be well specified and there is a high consensus among both respondents and investigators about the precise meaning of the questions and answers. In some cases where consensus would initially not be adequate, simple definitions can be added to a questionnaire to attain satisfactory agreement -- e.g., in asking people how many rooms they have in their homes, one states whether bathrooms, basements, etc. are to be included in the

count; most respondents will conform to this specification.

With questions about subjective phenomena, such as attitudes and values, treating answers as simply descriptive is seldom meaningful. Too much depends on how questions are worded, and there is neither sufficient social consensus about precise meaning, nor an external reference to facilitate such consensus.

There are many examples in the survey literature of how changes in wording or context will affect results based on questions about subjective phenomena (see Schuman and Presser (1981)). For example, in national surveys close to a quarter of the population will choose the "don't know" response to most attitude questions if it is explicitly offered; yet these same people will select a substantive alternative if "don't know" is not specifically provided, even though accepted when asserted spontaneously. More puzzlingly, a question about "forbidding" a particular action tends to elicit less agreement than a question about "not allowing" the same action, although the two questions are logically equivalent. Beyond these examples, most attitude objects are simply too complex to be summarized by a single survey question, e.g., attitudes toward abortion are too dependent on the reasons for abortion and the time in pregnancy to be adequately captured by a single question; attitudes toward "gun control" vary enormously depending on the exact framing of the issue (e.g., handguns vs. all guns, registration vs. banning, and other concrete policy distinctions).

Contingent valuation studies seek descriptive information,

yet call for a response similar to those elicited by questions about subjective phenomena. Thus they risk many of the same response effects and other wording difficulties that turn up regularly in attitude surveys. Minimizing these effects presents a considerable challenge to anyone wishing to elicit reliable CV estimates. The simplest way to approach the problem is to consider a CV survey as essentially a self-contained referendum in which respondents vote on whether to tax themselves or not for a particular purpose. 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.

The Panel is of the opinion that open-ended CV questions -- e.g., "What is the smallest sum that would compensate you for environmental damage X?" or, "What is the largest amount you would be willing to pay to avoid (or repair) environmental damage X?" -- are unlikely to provide the most reliable valuations. There are at least two reasons for this conclusion. In the first place, the scenario lacks realism since respondents are rarely asked or required in the course of their everyday lives to place a dollar value on a particular public good. Their responses to such questions are therefore likely to be unduly sensitive to trivial characteristics of the scenario presented. In the second place, an open-ended request for willingness to pay or willingness to accept compensation invites strategic

overstatement. The more seriously the respondent takes the question, the more likely it is that he or she will see that reporting a large response is a costless way to make a point. Both experience and logic suggest that responses to open-ended questions will be erratic and biased.

However, the referendum format, especially when cast in the willingness to pay mode -- "Would you be willing to contribute (or be taxed) D dollars to cover the cost of avoiding or repairing environmental damage X?" -- has many advantages. It is realistic: referenda on the provision of public goods are not uncommon in real life. There is no strategic reason for the respondent to do other than answer truthfully, although a tendency to overestimate often appears even in connection with surveys concerning routine market goods. The fact that market surveys continue to be used routinely suggests that this tendency is not an insuperable obstacle. Of course, the respondent in a CV survey understands that the referendum is hypothetical; there is no implication that the tax will actually be levied and the damage actually repaired or avoided. This suggests that considerable efforts should be made to induce respondents to take the question seriously, and that the CV instrument should contain other questions designed to detect whether the respondent has done so. Although Carson, et al. (1992), included a useful question to determine whether respondents believed the survey was biased in any direction, they did not sufficiently test whether the completeness of, and time period for, restoration stated in

the survey were fully accepted by respondents. But, as far as strategic reasons go, a respondent who would not be willing to pay D dollars has no reason to answer "Yes," and a respondent who would be willing to pay D dollars has no reason to answer "No."

There are, however, several other reasons why one's response to a hypothetical referendum question might be the opposite of one's actual vote on a real ballot. On one hand, a respondent unwilling to pay D dollars in reality might feel pressure to give the "right" or "good" answer when responding to an in-person or telephone interviewer. This could happen if the respondent believes that the interviewer would herself favor a yes answer. On the other hand, a respondent actually willing to pay the stated amount might answer in the negative for several reasons: (i) belief that the proposed scenarios distributed the burden unfairly; (ii) doubt of either the feasibility of the proposed action, so that any contribution would be wasted, or the ability of the relevant agency to carry out the action efficiently; or (iii) refusal to accept the hypothetical choice problem, because of either a generalized aversion to taxes or a view that someone else -- the "oil industry", for example -- should pay for repair or avoidance as the responsible party. The same considerations suggest that a CV instrument should include questions designed to detect the presence of these sources of bias. This is in fact often done, but we do not know how successfully.

There are two further problems that could detract from the reliability of CV responses without producing any determinate

bias: (i) a feeling that one's vote will have no significant effect on the outcome of the hypothetical referendum, leading to no reply or an unconsidered one; and (ii) poor information about the damage being valued. Of course, either of these could occur in real referenda.

Here we must decide on the standard of knowledgeableability of the respondents that we want to impose on a CV study. It is clear that it should be at least as high as that which the average voter brings to a real referendum on the provision of a specific public good, but should it be higher? A "conservative" CV study, i.e., one that avoids overestimating true willingness to pay, will no doubt exceed the minimum standard of information and will also lean over backwards to avoid providing information in a way that might bias the response upwards. In particular, a conservative study will provide the respondent with some perspective concerning the overall frequency and magnitude of oil spills, the amount of money currently being spent on preventing and remedying them, the overall scale of their consequences, the peculiar features of the spill in question, and similar relevant information. Placing the choice problem in a broader context helps the respondent to arrive at a realistic or even conservative valuation.

Most of the provision of public goods in this country is decided by representatives and bureaucrats rather than by direct vote of the citizens. It is presumed that these agents are more "expert" or at least draw on more knowledge than the citizens

themselves. The agents' expertise, if it really exists, is about the means and cost of providing public goods, though elected officials may sometimes be presumed to "represent" judgments of ultimate value to the citizens. Nevertheless, to increase one's confidence that a CV study is conservatively reliable, one might want to compare its outcome with that provided by a panel of experts. This will help check whether respondents and those conducting the study or studies are reasonably well-informed and well-motivated. This comparison could be made on a sample of CV studies to give an idea of their reliability in general.

The above considerations suggest that a CV study based on the referendum scenario can produce more reliably conservative estimates of willingness to pay, and hence of compensation required in the aftermath of environmental impairment, provided that a concerted effort is made to motivate the respondents to take the study seriously, to inform them about the context and special circumstances of the spill or other accident, and to minimize any bias toward high or low answers originating from social pressure within the interview. This implies that, in the present state of the art, a reliably conservative CV study should be conducted with personal interviews of significant duration and will therefore be relatively costly. It follows therefore that, in order that the cost of the study not be disproportionately large compared to the amount of damages, the CV approach would likely be used only in relatively major spills, at least until further improvements in methodology can be developed and

accepted. (A suggestion for doing so is offered in Section V.)

The referendum format offers one further advantage for CV.

As we have argued, external validation of elicited lost passive-use values is usually impossible. There are however real-life referenda. Some of them, at least, are decisions to purchase specific public goods with defined payment mechanisms, e.g., an increase in property taxes. The analogy with willingness to pay for avoidance or repair of environmental damage is far from perfect but close enough that the ability of CV-like studies to predict the outcomes of real-world referenda would be useful evidence on the validity of the CV method in general.

The test we envision is not an election poll of the usual type. Instead, using the referendum format and providing the usual information to the respondents, a study should ask whether they are willing to pay the average amount implied by the actual referendum. The outcome of the CV-like study should be compared with that of the actual referendum. The Panel thinks that studies of this kind should be pursued as a method of validating and perhaps even calibrating applications of the CV method (see Magleby, 1984).

Addressing the Embedding Problem

Perhaps the most important internal argument against the reliability of the CV approach (as against general criticisms about vagueness, lack of information, or unreality of the scenario) is the observation of the "embedding" phenomenon (see

the discussion in Section II). Different but similar samples of respondents are asked about their willingness to pay for prevention of environmental damage scenarios that are identical except for their scale: different numbers of seabirds saved, different numbers of forest tracts preserved from logging, etc. It is reported that average willingness to pay is often substantial for the smallest scenario presented but is then substantially independent of the size of the damage averted, rising slightly if at all for large changes in size.

The usual interpretation proposed by critics of the CV method is that the responses are not measuring the equivalent dollar value of the utility of the environmental assets preserved, because that would certainly be measurably larger for substantially larger programs of preservation. Instead, the fixed sum offered is the value of a feeling of having done something praiseworthy; a "warm glow" is the phrase often used.

This is potentially a very damaging criticism of the method. CV studies almost always seek to measure willingness to pay to avoid a particular incident rather than compensation that would be required for damage that has already occurred. This is because respondents are more likely to exaggerate the compensation they would require than their willingness to pay, and because the latter is expected to be less than the former and so is conservative. If reported willingness to pay accurately reflected actual willingness to pay, then, under the "warm glow" interpretation, willingness to pay might well exceed compensation

required because the former contains an element of self-approbation. It might be real but not properly compensable.

Defenders of the CV approach reply to this criticism in various ways. Sometimes it is argued that the evidence used to support "embedding" simply indicates diminishing marginal utility of the asset in question. In many cases, however, the constancy or near-constancy of willingness to pay does not appear consistent with the large reported amounts for the first small increment of environmental preservation.

A second defense of CV against the embedding phenomenon is that CV questions have to be posed carefully and in context. It is argued that carelessly formulated CV instruments leave respondents with the impression that they are being asked, "Would you pay \$X to avert a certain small environmental harm?" In a very large population of birds, the death of 1,000 is not seen as noticeably different from the death of 100,000 -- and may not actually be very different -- so that respondents simply answer the question just asked.

This second response leads to the obvious question: how should a CV instrument be framed to elicit an answer that responds to the precise scenario and not to a generalized "warm glow" effect? We must reject one possible approach, that of asking each respondent to express willingness to pay to avert incidents of varying sizes; the danger is that embedding will be forcibly avoided, still without realism. This issue is best considered as part of the broader question: How much context

about the incident itself and about the respondent's circumstances and choices should be included in the CV instrument?

We are recommending a high standard of richness in context to achieve a realistic background. Our proposed guidelines regarding this issue are embodied in Section IV below.

Time Dimension of Passive Use Losses

Typically, environmental damages from oil spills or similar accidents are severe for some period of time -- weeks, months, or sometimes a few years -- and gradually are reduced by natural forces and human efforts to a low or possibly even zero steady state level. In some circumstances, passive-use losses derive only or mostly from the steady state conditions; thus, if passive use value derives from species diversity, even a considerable loss of birds or mammals which does not endanger any species will give rise to no loss. If, on the contrary, considerable passive-use value is attached to the interim state of the natural resource, then respondents have to do a very difficult present value calculation properly to compute their current willingness to pay for the difference between the fully restored state of the resource and the actual state as the level of restoration varies over time. CV surveys accordingly have to be carefully designed to allow respondents to differentiate interim from steady state passive-use loss, and, if there is interim passive-use loss, to report its present value correctly.

It is reasonable to assume that interim passive-use values are additive over time. Hence, we need a calculation of present values of the interim losses. The discounting and the estimation of the rate of recovery of the resource should be done by technical experts and not by the respondents, who are unlikely to handle these tasks adequately. Respondents should be asked only their willingness to pay to eliminate the difference between some partially restored level of the resource and the pristine state for a specific period of time, say a year, on the assumption that after that time full restoration is assured. Technical experts would estimate how the state of the resource will vary from year to year as the restoration takes place. The technical information about the state of the resource, together with the respondent's assessments of the flow valuation of the resource, can be used to construct a time series of passive-use losses which can be discounted to the present at an appropriate rate of interest to determine the present value of the damages.

IV. SURVEY GUIDELINES

In this section we try to lay down a fairly complete set of guidelines compliance with which would define an ideal CV survey.

A CV survey does not have to meet each of these guidelines fully in order to qualify as a source of reliable information to a damage assessment process. Many departures from the guidelines or even a single serious deviation would, however, suggest unreliability *prima facie*. To preserve continuity, we give only

a bald list of guidelines here. They are repeated together with further explanatory comments in the Appendix to this Report.

GENERAL GUIDELINES

- ☐ Sample Type and Size: Probability sampling is essential for a survey used for damage assessment.¹ The choice of sample specific design and size is a difficult, technical question that requires the guidance of a professional sampling statistician.

- ☐ Minimize Nonresponses: High nonresponse rates would make the survey results unreliable.

- ☐ Personal Interview: The Panel believes it unlikely that reliable estimates of values could be elicited with mail surveys. Face-to-face interviews are usually preferable, although telephone interviews have some advantages in terms of cost and centralized supervision.

- ☐ Pretesting for Interviewer Effects: An important respect in

¹ This need not preclude use of less adequate samples, including quota or even convenience samples, for preliminary testing of specific experimental variations, so long as order of magnitude differences rather than univariate results are the focus. Even then, obvious sources of bias should be avoided (e.g., college students are probably too different in age and education from the heterogeneous adult population to provide a trustworthy basis for wider generalization).

which CV surveys differ from actual referenda is the presence of an interviewer (except in the case of mail surveys). It is possible that interviewers contribute to "social desirability" bias, since preserving the environment is widely viewed as something positive. In order to test this possibility, major CV studies should incorporate experiments that assess interviewer effects.

☐ Reporting: Every report of a CV study should make clear the definition of the population sampled, the sampling frame used, the sample size, the overall sample non-response rate and its components (e.g., refusals), and item non-response on all important questions. The report should also reproduce the exact wording and sequence of the questionnaire and of other communications to respondents (e.g., advance letters). All data from the study should be archived and made available to interested parties (see Carson et al. (1992), for an example of good practice in inclusion of questionnaire and related details; as of this date, however, the report has not been available publicly and the data have not been archived for open use by other scholars).

☐ Careful Pretesting of a CV Questionnaire: Respondents in a CV survey are ordinarily presented with a good deal of new and often technical information, well beyond what is typical

in most surveys. This requires very careful pilot work and pretesting, plus evidence from the final survey that respondents understood and accepted the main description and questioning reasonably well.

GUIDELINES FOR VALUE ELICITATION SURVEYS

The following guidelines are met by the best CV surveys and need to be present in order to assure reliability and usefulness of the information that is obtained.

- ☐ Conservative Design: Generally, when aspects of the survey design and the analysis of the responses are ambiguous, the option that tends to underestimate willingness to pay is preferred. A conservative design increases the reliability of the estimate by eliminating extreme responses that can enlarge estimated values wildly and implausibly.
- ☐ Elicitation Format: The willingness to pay format should be used instead of the compensation required because the former is the conservative choice.
- ☐ Referendum Format: The valuation question should be posed as a vote on a referendum.
- ☐ Accurate Description of the Program or Policy: Adequate

information must be provided to respondents about the environmental program that is offered. It must be defined in a way that is relevant to damage assessment.

- ☐ Pretesting of Photographs: The effects of photographs on subjects must be carefully explored.

- ☐ Reminder of Undamaged Substitute Commodities: Respondents must be reminded of substitute commodities, such as other comparable natural resources or the future state of the same natural resource. This reminder should be introduced forcefully and directly prior to the main valuation question to assure that respondents have the alternatives clearly in mind.

- ☐ Adequate Time Lapse from the Accident: The survey must be conducted at a time sufficiently distant from the date of the environmental insult that respondents regard the scenario of complete restoration as plausible. Questions should be included to determine the state of subjects' beliefs regarding restoration probabilities.

- ☐ Temporal Averaging: Time dependent measurement noise should be reduced by averaging across independently drawn samples taken at different points in time. A clear and substantial

time trend in the responses would cast doubt on the "reliability" of the finding.

- ☐ "No-answer" Option: A "no-answer" option should be explicitly allowed in addition to the "yes" and "no" vote options on the main valuation (referendum) question. Respondents who choose the "no-answer" option should be asked nondirectively to explain their choice. Answers should be carefully coded to show the types of responses, for example: (i) rough indifference between a yes and a no vote; (ii) inability to make a decision without more time or more information; (iii) preference for some other mechanism for making this decision; and (iv) bored by this survey and anxious to end it as quickly as possible.

- ☐ Yes/no Follow-ups: Yes and no responses should be followed up by the open-ended question: "Why did you vote yes/no?" Answers should be carefully coded to show the types of responses, for example: (i) It is (or isn't) worth it; (ii) Don't know; or (iii) The oil companies should pay.

- ☐ Cross-tabulations: The survey should include a variety of other questions that help to interpret the responses to the primary valuation question. The final report should include summaries of willingness to pay broken down by these categories. Among the items that would be helpful in

interpreting the responses are:

Income
 Prior Knowledge of the Site
 Prior Interest in the Site (Visitation Rates)
 Attitudes Toward the Environment
 Attitudes Toward Big Business
 Distance to the Site
 Understanding of the Task
 Belief in the Scenarios
 Ability/Willingness to Perform the Task

- ☐ Checks on Understanding and Acceptance: The above guidelines must be satisfied without making the instrument so complex that it poses tasks that are beyond the ability or interest level of many participants.

GOALS FOR VALUE ELICITATION SURVEYS

The following items are not adequately addressed by even the best CV surveys. In the opinion of the Panel, these issues will need to be convincingly dealt with in order to assure the reliability of the estimates.

- ☐ Alternative Expenditure Possibilities: Respondents must be reminded that their willingness to pay for the environmental program in question would reduce their expenditures for private goods or other public goods. This reminder should

be more than perfunctory, but less than overwhelming. The goal is to induce respondents to keep in mind other likely expenditures, including those on other environmental goods, when evaluating the main scenario.

- ☐ Deflection of Transaction Value: The survey should be designed to deflect the general "warm-glow" of giving or the dislike of "big business" away from the specific environmental program that is being evaluated. It is possible that the referendum format limits the "warm glow" effect, but until this is clear the survey design should explicitly address this problem.

- ☐ Steady State or Interim Losses: It should be made apparent that respondents can distinguish interim from steady-state losses.

- ☐ Present Value Calculations of Interim Losses: It should be demonstrated that, in revealing values, respondents are adequately sensitive to the timing of the restoration process.

- ☐ Advance Approval: Since the design of the CV survey can have a substantial effect on the responses, it is desirable that -- if possible -- critical features be preapproved by

both sides in a legal action, with arbitration and/or experiments used when disagreements cannot be resolved by the parties themselves.

☐ Burden of Proof: Until such time as there is a set of reliable reference surveys, the burden of proof of reliability must rest on the survey designers. They must show through pretesting or other experiments that their survey does not suffer from the problems that these guidelines are intended to avoid. Specifically, if a CV survey suffered from any of the following maladies, we would judge its findings "unreliable":

- A high nonresponse rate to the entire survey instrument or to the valuation question.
- Inadequate responsiveness to the scope of the environmental insult.
- Lack of understanding of the task by the respondents.
- Lack of belief in the full restoration scenario.
- "Yes" or "no" votes on the hypothetical referendum that are not followed up or explained by making

reference to the cost and/or the value of the program.

- ☐ Reliable Reference Surveys: In order to alleviate this heavy burden of proof, we strongly urge the government to undertake the task of creating a set of reliable reference surveys that can be used to interpret the guidelines and also to calibrate surveys that do not fully meet the conditions.

V. RECOMMENDATIONS FOR FUTURE RESEARCH

The Panel's major research recommendation goes toward a drastic reform of the CV procedure, extending beyond the guidelines suggestion in Section IV.

The problem of estimating the demand for highly innovative commercial products, including some that have not yet actually been produced, is much like the problem faced in CV research. It is the problem of estimating willingness to pay for a necessarily unfamiliar product. The field of market research has developed methods -- "conjoint analysis," for example -- that are very similar to the CV approach. (One important difference is that a new product may eventually reach the market, and projections of expected sales can be checked. Survey responses are usually found to be moderate overestimates of actual willingness to pay.)

Practitioners have found that survey methods are better at estimating relative demand than absolute demand. There is an

anchoring problem, even with private goods -- that is, absolute willingness to pay is hard to pin down. This leads to the following suggestion.

The federal government should produce standard damage assessments for a few specific reference oil spills, either hypothetical or actual, ranging from small to large. These standard valuations could be generated by any method. One possibility would be through a jury of experts. Such a jury of experts might wish to conduct a series of CV studies, satisfying the guidelines laid out above. These CV studies would be inputs into the jury process, to be combined with other information and expert judgment. Once these benchmarks were available, they could serve as reference points for later CV studies. When a damage assessment is required, surveys could be used to elicit answers to questions like: "Would you pay (much more, more, about the same, less, much less) to prevent this spill than you would to prevent Standard Spill A?" "Would you pay an amount to avoid this spill that is between the amounts you would pay to avoid Standard Spill B and Standard Spill C? If so, is the amount much closer to B than C, closer to B than C, halfway between B and C, closer to C than B, much closer to C than B?" These questions presumably would not be asked so schematically. Responses to such a study could then serve as one reliable source of information in the damage assessment.

We recognize that this technique would require that respondents be made familiar with the reference spills as well as

the particular spill whose damage is being assessed. We expect that the additional effort would be more than offset by the greater simplicity and reliability in estimating relative willingness to pay.

This possibility suggests a slightly more radical extension of the CV method. Respondents could be asked to compare their willingness to pay to avoid a specific case of environmental damage to their willingness to pay for a range of fairly familiar private goods. It would no doubt be best if the private goods were to bear some similarity to the environmental good in question, but that is not necessary. The anchoring purpose would be served if respondents could measure their willingness to pay in units of articles of clothing or small household appliances forgone.

This latter is a suggestion for research in the CV method, not necessarily a recommendation for current practical use.

The guidelines proposed in Section IV themselves suggest areas for further research, this time within the contingent valuation community. In particular, we emphasize the urgency of studying the sensitivity of willingness to pay responses to the number and extent of budgetary substitutes mentioned in survey instruments (that is, reminders of other things on which respondents could spend their money). In such research it would be helpful if parallel studies were conducted on the sensitivity of stated intentions to buy ordinary market goods -- both familiar and unfamiliar -- to reminders of alternative uses of

those resources. The point is to discover the extent to which the valuation of environmental public goods is intrinsically more difficult than similar exercises with respect to market goods.

A closely-related line of research is the sensitivity of responses in CV surveys to the number and extent of undamaged substitute commodities mentioned explicitly in the survey instrument (miles of nearby shoreline, miles of shoreline elsewhere, similarity for animal or bird life, alternative recreation possibilities and so on). This could be extended to variations in the way in which the budget constraint is presented to respondents. Here again, comparisons with market goods would be useful.

Finally, having urged that the availability of a no-vote option is an important component of the ability of the CV technique to mimic an actual referendum, we recommend further research into alternative ways of presenting and interpreting the no-vote option. In this respect, too, comparative studies with familiar public and private goods (local parks, school facilities, housing for the homeless, food distributions) would be enlightening. Real referenda always allow the option of not voting, in a natural way. CV studies have to achieve the same result more deliberately, so there is a need to know if the precise formulation matters very much to the result.

VI. CONCLUSIONS AND RECOMMENDATIONS

The Panel starts from the premise that passive-use loss --

interim or permanent -- is a meaningful component of the total damage resulting from environmental accidents. A problem arises because passive-use losses have few or no overt behavioral consequences. The faintness of the behavioral trail means that a well-designed and adequately sensitive measuring instrument is needed to substitute for conventional observations of behavior. In particular, can the CV method provide a sufficiently reliable estimate of total loss -- including passive-use loss -- to play a useful role in damage assessment?

It has been argued in the literature and in comments addressed to the Panel that the results of CV studies are variable, sensitive to details of the survey instrument used, and vulnerable to upward bias. These arguments are plausible. However, some antagonists of the CV approach go so far as to suggest that there can be no useful information content to CV results. The Panel is unpersuaded by these extreme arguments.

In Section IV above, we identify a number of stringent guidelines for the conduct of CV studies. These require that respondents be carefully informed about the particular environmental damage to be valued, and about the full extent of substitutes and undamaged alternatives available. In willingness to pay scenarios, the payment vehicle must be presented fully and clearly, with the relevant budget constraint emphasized. The payment scenario should be convincingly described, preferably in a referendum context, because most respondents will have had experience with referendum ballots with less-than-perfect

background information. Where choices in formulating the CV instrument can be made, we urge they lean in the conservative direction, as a partial or total offset to the likely tendency to exaggerate willingness to pay.

The Panel concludes that under those conditions (and others specified above), CV studies convey useful information. We think it is fair to describe such information as reliable by the standards that seem to be implicit in similar contexts, like market analysis for new and innovative products and the assessment of other damages normally allowed in court proceedings. As in all such cases, the more closely the guidelines are followed, the more reliable the result will be. It is not necessary, however, that every single injunction be completely obeyed; inferences accepted in other contexts are not perfect either.

Thus, the Panel concludes that CV studies can produce estimates reliable enough to be the starting point of a judicial process of damage assessment, including lost passive-use values.

To be acceptable for this purpose, such studies should follow the guidelines described in Section IV above. The phrase "be the starting point" is meant to emphasize that the Panel does not suggest that CV estimates can be taken as automatically defining the range of compensable damages within narrow limits. Rather, we have in mind the following considerations.

The Panel is persuaded that hypothetical markets tend to overstate willingness to pay for private as well as public goods.

The same bias must be expected to occur in CV studies. To the extent that the design of CV instruments makes conservative choices when alternatives are available, as urged in Section IV, this intrinsic bias may be offset or even over-corrected. All surveys of attitudes or intentions are bound to exhibit sensitivity of response to the framing of questions and the order in which they are asked. No automatic or mechanical calibration of responses seems to be possible.

The judicial process must in each case come to a conclusion about the degree to which respondents have been induced to consider alternative uses of funds and take the proposed payment vehicle seriously. Defendants will argue that closer attention to substitute commodities would have yielded lower valuations. Trustees will argue that they have already leaned over backwards to ensure conservative responses. Judges and juries must decide as they do in other damage cases. The Panel's conclusion is that a well-conducted CV study provides an adequately reliable benchmark to begin such arguments. It contains information that judges and juries will wish to use, in combination with other evidence, including the testimony of expert witnesses.

The Panel's second conclusion is that the appropriate federal agencies should begin to accumulate standard damage assessments for a range of oil spills, as described in Section V.

That process should further improve the reliability of CV studies in damage assessment. It should thus contribute to increasing the accuracy and reducing the cost of subsequent

damage assessment cases. In that sense, it can be regarded as an investment.

The proposals for further research outlined in Section V are an integral part of our recommendations. The Panel believes that the suggestions put forward there could lead to more reliable and less controversial damage assessment at reduced cost. It is not to be expected that controversy will disappear, however. There will always be controversy where intangible losses have to be evaluated in monetary terms.

APPENDIX

GENERAL GUIDELINES

☐ Sample Type and Size: Probability sampling is essential for a survey used for damage assessment.² The choice of sample specific design and size is a difficult, technical question that requires the guidance of a professional sampling statistician.

If a single dichotomous question of the yes-no type is used to elicit valuation responses, then a total sample size of 1000 respondents will limit sampling error to about 3% plus or minus on a single dichotomous question, assuming simple random sampling. However, this or any other sample size needs to be reconceptualized for three reasons. First, if face-to-face interviewing is used, as we suggest above, clustering and stratification must be taken into account. Second, if dichotomous valuation questions are used (e.g., hypothetical referenda), separate valuation amounts must be asked of random sub-samples and these responses must be unscrambled econometrically to estimate the underlying population mean or median. Third, in order to incorporate experiments on

² This need not preclude use of less adequate samples, including quota or even convenience samples, for preliminary testing of specific experimental variations, so long as order of magnitude differences rather than univariate results are the focus. Even then, obvious sources of bias should be avoided (e.g., college students are probably too different in age and education from the heterogeneous adult population to provide a trustworthy basis for wider generalization).

interviewer and wording effects, additional random sub-sampling is required. For all these reasons, it will be important to consult sampling statisticians in the design of a CV survey intended for legal or policy-making purposes.

☐ Minimize Nonresponses: High nonresponse rates would make the survey results unreliable.

To the extent that a CV study is expected to represent the adult population of the United States or a portion of it, minimizing both sample non-response and item non-response are important. The former is unlikely to be below 20% even in very high quality surveys; the latter has also been large in some CV surveys because of the difficulty of the task respondents are being asked to perform. These sources of potential bias can be partially justified on the grounds that they also occur with official referenda, in both cases with the loss especially of the least educated parts of the population. The further reduction of the final sample by elimination of "protest zeros," "unrealistic high values," and other problematic responses may lead to effective final total response rates so low as to imply that the survey population consists of interested and specially instructed quasi-experts. This consideration reinforces the desirability of combining a reasonable response rate with a high but not forbidding standard of information, as discussed in Section III above.

☐ Personal Interview: The Panel believes it unlikely that reliable estimates of values could be elicited with mail surveys. Face-to-face interviews are usually preferable, although telephone interviews have some advantages in terms of cost and centralized supervision.

Assuming a CV survey is to represent a natural population, such as all adults in the United States, or those in a single urban area or a state, it is desirable that it be carried out using either face-to-face or telephone interviews. Mail surveys typically employ lists that cover too small a part of the population (e.g., samples based on telephone directories omit approximately half the U.S. population because of non-listed numbers, incorrect numbers, and non-phone households), and then miss another quarter or more of the remainder through non-response. In addition, since the content of a mail questionnaire can be reviewed by targeted respondents before deciding to return it, those most interested in a natural resource issue or in one side or the other can make their decision on that basis. It is also impossible using mail surveys to guarantee random selection within households or to confine answering to a single respondent, and it is difficult (though not impossible) to control question-order effects. Thus, mail surveys should be used only if another supplementary method can be employed to cross-validate the results on a random sub-sample of respondents.

The choice between telephone and face-to-face administration is less clear. Face-to-face surveys offer practical advantages in maintaining respondent motivation and allowing use of graphic supplements. Both coverage and response rates are also usually somewhat higher than with telephone surveys. However, telephone surveys can cut interviewing costs by between a third and a half; for CV purposes, it may be a disadvantage that most survey investigators believe telephone interviews need to be kept shorter in length than face-to-face interviews because respondent attention and cooperation are more difficult to maintain. In addition, random-digit-dial telephone surveys approximate simple random sampling. Face-to-face surveys must be based on cluster sampling and, therefore, the results provide less precise estimates than do telephone surveys of the same size.

☐ Pretesting for Interviewer Effects: An important respect in which CV surveys differ from actual referenda is the presence of an interviewer (except in the case of mail surveys). It is possible that interviewers contribute to "social desirability" bias, since preserving the environment is widely viewed as something positive. In order to test this possibility, major CV studies should incorporate experiments that assess interviewer effects.

To test for interviewer effects, two modifications might be made to a standard face-to-face CV survey. In one variant on

current practice, respondents would stop when they come to the valuation question, write their "vote" on a ballot, and fold and deposit it in a sealed box. However, since this practice would not mimic the complete anonymity of the voting booth, for a subsample of respondents a second modification should be made. Respondents would be allowed to mail their "ballots" in unmarked envelopes directly to the survey organization, even though that will preclude any but the simplest analysis of responses. Tests of the effect of both these modifications of current practice will indicate whether they are needed routinely or whether at least some calibration should be introduced to compensate for interviewer effects. (The more modest of these proposed modifications -- a simulated ballot box, or even voting on a portable computer -- has few if any disadvantages and might be made standard if it shows any reliable departure at all from answers given orally to the interviewer.)

☐ Reporting: Every report of a CV study should make clear the definition of the population sampled, the sampling frame used, the sample size, the overall sample non-response rate and its components (e.g., refusals), and item non-response on all important questions. The report should also reproduce the exact wording and sequence of the questionnaire and of other communications to respondents (e.g., advance letters). All data from the study should be archived and made available to interested parties (see

Carson et al. (1992), for an example of good practice in inclusion of questionnaire and related details; as of this date, however, the report has not been available publicly and the data have not been archived for open use by other scholars).

- ☐ Careful Pretesting of a CV Questionnaire: Respondents in a CV survey are ordinarily presented with a good deal of new and often technical information, well beyond what is typical in most surveys. This requires very careful pilot work and pretesting, plus evidence from the final survey that respondents understood and accepted the main description and questioning reasonably well.

Parenthetically, the claim sometimes made by CV proponents that particular methods of piloting, such as focus groups, are essential should be viewed with skepticism, since these claims are unsupported by any systematic evidence. Nor is it clear that what are called "state-of-the-art" CV surveys constitute something entirely new or different from other types of serious survey investigations. Thus, although evidence that questionnaire development has been carried out carefully is certainly important, it cannot be taken as a self-sufficient basis of validity -- the more so because we know that many people will answer survey questions without apparent difficulty, even when they do not understand them well. A way of reducing

pressure to give answers of questionable meaningfulness would be to provide respondents an explicit "no opinion" type of alternative when a key valuation question is posed.

GUIDELINES FOR VALUE ELICITATION SURVEYS

The following guidelines are met by the best CV surveys and need to be present in order to assure reliability and usefulness of the information that is obtained.

- ☐ Conservative Design: Generally, when aspects of the survey design and the analysis of the responses are ambiguous, the option that tends to underestimate willingness to pay is preferred. A conservative design increases the reliability of the estimate by eliminating extreme responses that can enlarge estimated values wildly and implausibly.

- ☐ Elicitation Format: The willingness to pay format should be used instead of compensation required because the former is the conservative choice.

In experimental settings, the gap between stated intentions to support a particular referendum and actual behavior in the voting booth can be very great (see Magleby, 1984). This gap might be treated by "calibration" if there were historical data on the relationship between such intentions and behavior. Unfortunately, we are aware of no data that is close enough to

the CV context that could be used to calibrate CV responses. In the absence of historical data that can be used to calibrate the intentions reported in the CV surveys, the survey instrument has to be designed with extraordinary care so that it can stand on its own.

Referendum Format: The valuation question should be posed as a vote on a referendum.

As is now generally recognized by most CV proponents, asking respondents to give a dollar valuation in response to an open-ended question presents them with an extremely difficult task. At the same time, CV proponents also recognize that presenting respondents a set of dollar amounts from which they are to choose is likely to create anchoring and other forms of bias. Thus, we recommend as the most desirable form of CV elicitation the use of a dichotomous question that asks respondents to vote for or against a particular level of taxation, as occurs with most real referenda. As already noted, such a question form also has advantage in terms of incentive compatibility. (If a double-bounded dichotomous choice or some other question form is used in order to obtain more information per respondent, experiments should be developed to investigate biases that may be introduced.)

Accurate Description of the Program or Policy: Adequate

information must be provided to respondents about the environmental program that is offered. It must be defined in a way that is relevant to damage assessment.

Ideally a CV survey would elicit attitudes toward three alternative (future) recovery scenarios: (A) "immediate" restoration, (b) accelerated restoration, and (c) natural restoration. Damages would be the difference between (a) and (b) on the assumption that accelerated restoration is provided by the responsible party. Unfortunately, respondents may not find "immediate" restoration very plausible and they may resist the notion that they should be expected to contribute to accelerated restoration when it is an oil company that is at fault. If respondents are unable or unwilling to deal hypothetically with the most relevant "clean-up" scenarios, alternative "prevention" scenarios will have to be used in the survey instrument. For example, respondents may be asked to vote for a referendum that offers reduced risk of another spill for a specified period of time.³ The weaker is the linkage between the "prevention" scenarios and the "clean-up" scenarios, the more unreliable are the survey results. Rhetorically: Is a decade of prevention equal in value to the difference in value between accelerated and immediate clean-up?

³ As in the survey actually performed by the State of Alaska after the *Valdez* spill (See Carson et al. (1992)).

- ☐ Pretesting of Photographs: The effects of photographs on subjects must be carefully explored.

One effective means for conveying information and holding interest in a CV interview has been the use of large and impressive photographs. However, this technique is a two-edged sword because the dramatic nature of a photograph may have much more emotional impact than the rest of the questionnaire. Thus it is important that photographs be subjected to even more careful assessment than verbal material if the goal is to avoid bias in presentation.⁴

- ☐ Reminder of Undamaged Substitute Commodities: Respondents must be reminded of substitute commodities, such as other comparable natural resources or the future state of the same natural resource. This reminder should be introduced forcefully and directly prior to the main valuation question to assure that respondents have the alternatives clearly in mind.

- ☐ Adequate Time Lapse from the Accident: The survey must be conducted at a time sufficiently distant from the date of the environmental insult that respondents regard the scenario of complete restoration as plausible. Questions

⁴ Failure to test the effects of photographs on responses is one shortcoming of Carson et al. (1992).

should be included to determine the state of subjects' beliefs regarding restoration probabilities.

Survey respondents who would not suffer interim passive-use loss may not regard full restoration as very plausible; therefore, they may report substantial passive-use loss even if told that full restoration in some reasonable amount if time is certain. Misunderstanding of the restoration probability is most acute when the accident has recently occurred and before any substantial restoration takes place. It would be ideal to assess steady state passive-use loss after natural and human restoration is complete or nearly so, since then presumably respondents would believe in the restoration. If that is not a possibility, surveys might be conducted over time until the reported willingness to pay settles down (assuming that it does), as the respondents come to believe more and more in the probable success of the restoration effort. Alternatively, respondents might be asked to value a menu of alternative possible scenarios, without being told explicitly which is applicable for the environmental insult under study. The menu should be designed to force them to consider the difference between interim and steady-state passive-use value.

☐ Temporal Averaging: Time dependent measurement noise should be reduced by averaging across independently drawn samples taken at different points in time. A clear and substantial

time trend in the responses would cast doubt on the "reliability" of the finding.

- ☐ "No-answer" Option: A "no-answer" option should be explicitly allowed in addition to the "yes" and "no" vote options on the main valuation (referendum) question. Respondents who choose the "no-answer" option should be asked nondirectively to explain their choice. Answers should be carefully coded to show the types of responses, for example: (i) rough indifference between a yes and a no vote; (ii) inability to make a decision without more time or more information; (iii) preference for some other mechanism for making this decision; and (iv) bored by this survey and anxious to end it as quickly as possible.

- ☐ Yes/no Follow-ups: Yes and no responses should be followed up by the open-ended question: "Why did you vote yes/no?" Answers should be carefully coded to show the types of responses, for example: (i) It is (or isn't) worth it; (ii) Don't know; or (iii) The oil companies should pay.

- ☐ Cross-tabulations: The survey should include a variety of other questions that help to interpret the responses to the primary valuation question. The final report should include summaries of willingness to pay broken down by these categories. Among the items that would be helpful in

interpreting the responses are:

Income

Prior Knowledge of the Site

Prior Interest in the Site (Visitation Rates)

Attitudes Toward the Environment

Attitudes Toward Big Business

Distance to the Site

Understanding of the Task

Belief in the Scenarios

Ability/Willingness to Perform the Task

We believe that these cross tabulations will prove useful in interpreting and lending credibility to the responses and possibly also in forming adjustments that can enhance reliability.

☐ Checks on Understanding and Acceptance: The above guidelines must be satisfied without making the instrument so complex that it poses tasks that are beyond the ability or interest level of many participants.

Since CV interviews often present information that is new to respondents, the questionnaire should attempt at the end to determine the degree to which respondents accept as true the descriptions given and assertions made prior to the valuation question. Such an inquiry should be carried out in detail but

non-directively, so that respondents feel free to reject any part of the information they were given at earlier points.

GOALS FOR VALUE ELICITATION SURVEYS

The following items are not adequately addressed by even the best CV surveys. In the opinion of the Panel, these issues will need to be convincingly dealt with in order to assure the reliability of the estimates.

☐ Alternative Expenditure Possibilities: Respondents must be reminded that their willingness to pay for the environmental program in question would reduce their expenditures for private goods or other public goods. This reminder should be more than perfunctory, but less than overwhelming. The goal is to induce respondents to keep in mind other likely expenditures, including those on other environmental goods, when evaluating the main scenario.

Consumers can be expected to make expenditure decisions that are adequately sensitive to other expenditure possibilities with which they are familiar. But environmental referenda of the type presented in CV surveys are unfamiliar and respondents may not be aware of the large set of other expenditure possibilities that might be offered in future CV surveys or future referenda. Unless informed otherwise, respondents may suppose that there is

only one environmental scenario that will ever be offered and they may overspend on it.

It is not at all clear how exhaustive should be the list of alternative public goods that are explicitly presented. If the list is too brief, overspending can be expected. If the list is too long, respondents will be encouraged to spread expenditures to public goods for which there is not adequate total demand and which therefore cannot really be offered to them. Also, if the list gets large enough to encompass a significant fraction of income, the gap between willingness to pay and willingness to accept may widen.

It is also not clear what form the reminder should take. It does not seem enough merely to list other environmental goods since respondents would then have to guess the level of expenditure that would be necessary to pay for the alternatives.

The survey should probably include some statement about the price of the alternatives, for example, the per capita expenditure that would be required to provide the items.

☐ Deflection of Transaction Value: The survey should be designed to deflect the general "warm-glow" of giving or the dislike of "big business" away from the specific environmental program that is being evaluated. It is possible that the referendum format limits the "warm glow" effect, but until this is clear the survey design should explicitly address this problem.

Economic models of consumer behavior generally are based on the assumption that value derives from the goods and services that are consumed, not from the process by which these goods are allocated. But happiness that derives from charitable giving may come mostly from the act of giving rather from the material changes that follow from the gift. To give another example, consumers may get pleasure from the act of shopping as well as from ownership of the goods they purchase. Words that might be useful to distinguish between these utility-producing events are "consumption value" and "transaction value," the latter referring to the process or transaction that establishes ownership.

We do not question the validity of "transaction value" or differentiate it from "consumption value" as far as damage assessment is concerned. But for both forms of value, respondents need to be thinking clearly about the substitutes, since the closer are the substitutes the less the damage that is done. In the case of "transaction value," there are many close substitutes to cleaning up oil spills since there are many other charitable activities that can generate the same "warm glow" and there are many other ways to express hostility toward big business and modern technology.

☐ Steady State or Interim Losses: It should be made apparent that respondents can distinguish interim from steady-state

losses.

The quality of any natural resource varies daily and seasonally around some "equilibrium" or "steady state" level. Active-use value of a resource depends on its actual state at the time of use (and at other times), not on its equilibrium. But passive-use value of a natural resource may derive only or mostly from its steady state and not from its day-to-day state. If so, full restoration at some future date eliminates or greatly reduces passive-use loss. Surveys accordingly need to be carefully designed to allow respondents to differentiate interim from steady state passive-use loss.

☐ Present Value Calculations of Interim Losses: It should be demonstrated that, in revealing values, respondents are adequately sensitive to the timing of the restoration process.

As discussed in Section III above, the time profile of restoration following an accident potentially is an important determinant of active-use loss and interim passive-use loss, but respondents may have little ability to distinguish between and to evaluate different profiles.

☐ Advance Approval: Since the design of the CV survey can have a substantial effect on the responses, it is desirable

that -- if possible -- critical features be preapproved by both sides in a legal action, with arbitration and/or experiments used when disagreements cannot be resolved by the parties themselves.

☐ Burden of Proof: Until such time as there is a set of reliable reference surveys, the burden of proof of reliability must rest on the survey designers. They must show through pretesting or other experiments that their survey does not suffer from the problems that these guidelines are intended to avoid. Specifically, if a CV survey suffered from any of the following maladies, we would judge its findings "unreliable":

- A high nonresponse rate to the entire survey instrument or to the valuation question.
- Inadequate responsiveness to the scope of the environmental insult.
- Lack of understanding of the task by the respondents.
- Lack of belief in the full restoration scenario.
- "Yes" or "no" votes on the hypothetical referendum that are not followed up or explained by making reference to

the cost and/or the value of the program.

- ☐ Reliable Reference Surveys: In order to alleviate this heavy burden of proof, we strongly urge the government to undertake the task of creating a set of reliable reference surveys that can be used to interpret the guidelines and also to calibrate surveys that do not fully meet the conditions.

Table of References

Andreoni, James; "Giving With Impure Altruism: Applications to Charity and Ricardian Equivalence;" *Journal of Political Economy* 97 (1989); pp. 1447-1458.

Bishop, Richard C., and Thomas A. Heberlien; "Measuring Values of Extra-Market Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics* 61 (1979); 926-930.

Carson, Richard T, Robert Cameron Mitchell, W. Michael Hanemann, Raymond J. Kopp, Stanley Presser, and Paul A. Ruud; "A Contingent Valuation Study of Lost Passive Use Values Resulting from the Exxon Valdez Oil Spill;" A Report for the Attorney General of the State of Alaska; November 19, 1992.

Cummings, Ronald G., and Glenn W. Harrison; "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive Compatible;" Department of Economics, University of New Mexico, submitted to Office of General Counsel, National Oceanic and Atmospheric Administration, 1992, 18 pp.

Desvousges, William H., F. Reed Johnson, Richard W. Dunford, Kevin J. Boyle, Sara P. Hudson, and K. Nicole Wilson; "Measuring Natural Resource Damages with Contingent Valuation: Tests of Validity and Reliability;" Paper presented at the Cambridge Economics, Inc., Symposium, Contingent Valuation: A Critical Assessment; Washington, D.C., April 1992.

Diamond, P.A., and J.A. Hausman; "On Contingent Valuation Measurement of Nonuse Values;" Paper presented at the Cambridge Economics, Inc. Symposium, Contingent Valuation: A Critical Assessment; Washington, D.C., April 1992.

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

Dickie, Mark, Ann Fisher, and Shelby Gerking; Market Transactions and Hypothetical Demand Data: A Comparative Study; *Journal of American Statistical Association*, Vol. 82, March 1987, pp. 69-75.

Duffield, John W., and David A. Patterson; "Field Testing Existence Values: An Instream Flow Trust Fund for Montana Rivers;" Paper presented at the annual meeting of the American Economic Association; New Orleans, January 1991.

Kahneman, Daniel; "Comments" in Valuing Environmental Goods, edited by Ronald G. Cummings, David S. Brookshire and William D. Schulze; Totowa, N.J.: Rowman and Allanheld, 1986.

Kahneman, Daniel, and Jack Knetch; "Valuing Public Goods: The Purchase of Moral Satisfaction;" 22 JEEM 57-70; 1992.

Kemp, M.A., and C. Maxwell; "Exploring a Budget Context for Contingent Valuation Estimates;" Paper presented at the Cambridge Economics, Inc. Symposium, Contingent Valuation: A Critical Assessment; Washington, D.C., April 1992.

Magleby, David B.; Direct Legislation, Johns Hopkins Press, 1984.

Mitchell, Robert Cameron, and Richard T. Carson; *Using Surveys to Value Public Goods: The Contingent Valuation Method*; Resources for the Future; Washington, DC, 1989; 499 pp.

Rowe, Robert D., W. Douglass Shaw, William Schulze; "Nestucca Oil Spill;" Chapter 20 in *Natural Resource Damages: Law and Economics* (eds. Kevin M. Ward and John W. Duffield); New York: John Wiley & Sons; 1992.

Schuman, Howard, and Stanley Presser; *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*; New York: Academic Press, 1981.

Seip, Kalle, and Jon Strand; *Willingness to Pay for Environmental Goods in Norway: A Contingent Valuation Study with Real Payment*; Paper prepared for the SAF Center for Applied Research, Department of Economics, University of Oslo, 26 pp.

State of Ohio v. Department of the Interior, 880 F.2d 432 (D.C. Cir. 1989)