JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT 28, 155–173 (1995)

# Sequencing and Nesting in Contingent Valuation Surveys[1]

RICHARD T. CARSON

*Department of Economics, University of California, San Diego, La Jolla, California 92093*

AND

ROBERT CAMERON MITCHELL

*Graduate School of Geography, Clark University, Worcester, Massachusetts 01610*

Received October 6, 1993; revised December 30, 1994

The term "embedding" is ill-defined and has been applied to distinct phenomena, some predicted by economic theory and others not. This paper lays out a theoretical framework for looking at these phenomena and provides a set of well-defined terms. Included is a discussion of survey design problems which may induce spurious evidence in favor of the hypothesis that respondents are insensitive to the scope of the good being valued. An empirical example of the component sensitivity is provided. This test rejects the hypothesis that respondents are insensitive to the scope of the good being valued. 

## 1. INTRODUCTION

The issue of what effect, if any, a good's placement in a valuation sequence should have on the good's value has recently received a great deal of attention from those interested in environmental benefit and damage assessment. Much of this attention is due to the claim made by the authors of several recent papers on "embedding" experiments that the contingent valuation (CV) method consistently produces results which violate fundamental economic theory.[2] They interpret their experimental findings as showing that respondents are willing to pay the same amount for goods that differ in quantity and different amounts for identical goods valued under a variety of conditions which they argue should not influence the willingness-to-pay amounts.

On the basis of their comparisons, these researchers conclude that these effects are an inherent feature of the CV methodology, which renders it too unreliable for use in a benefit or damage assessment. However, close examination of these experiments raises questions about the nature of these theoretical violations and about the logical consistency of the concept "embedding effects," the term used to describe the range of phenomena under discussion. In addition, there is a plausible alternative explanation for some of the findings of these experiments which raises

[1]The authors appreciate helpful comments on earlier versions of this paper from Nicholas Flores, Michael Hanemann, John Hoehn, Raymond Kopp, V. Kerry Smith and two anonymous referees. Any remaining errors are our own.

[2]See for instance Kahneman and Knetsch [21], Desvousges *et al.* [11], and Diamond *et al.* [12].



0095-0696/95 $6.00

the question of whether the findings might be the result of poor survey design rather than evidence of an inherent flaw in the contingent valuation method.

Like many arguments in economics, the central feature of the embedding debate concerns a common sense empirical relationship economists expect to find in choice behavior. In particular, if an individual is confronted with two quantities of the same good, say $A$ and $B$ where $A > B$, and the good has strictly positive marginal utility, then the individual should be willing to pay more for $A$ than for $B$. Further, if the quantity of $A$ exceeds that of $B$ by a substantial amount, economist's common sense would lead them to feel uncomfortable if the willingness to pay for $A$ exceeded the willingness to pay for $B$ by only a trivial amount.

We wish to accomplish four things in this paper. The first is to argue for limiting the use of the "embedding effect" to a narrow, well-defined phenomenon and to propose a new nomenclature to help clarify the terminological confusion caused by the use of "embedding" to describe a wide variety of nesting and sequencing issues. The second is to clarify which aspect of economic theory is tested in the various "embedding" comparisons which have appeared in the literature. The third is to describe how amenity misspecification biases may produce the "embedding effects" in inadequately designed surveys. The fourth is to present the findings of an experiment conducted as part of a full-scale contingent valuation study which demonstrates that these effects are not inevitable, as some of the method's critics claim on the basis of some of the experiments described above. We conclude that the issues raised by the "embedding" critics need to be taken seriously in the design of valid CV surveys, but that the critics' dismissal of the method on the grounds that it assesses the value of public goods in a way that is "demonstrably arbitrary" [21, p. 58] is unwarranted.

## 2. DEFINING COMPARISONS

Consider two goods, $a$ and $b$. If neither $a$ nor $b$ is a proper subset of the other, we will define the set $\{a, b\}$ to be a set of *conventional* goods. If either $a$ or $b$ is a proper subset of the other we will define $\{a, b\}$ to be a set of *nested* goods.[3] Henceforth, we will designate the elements of a set of conventional goods using lowercase letters $\{a, b\}$ and the elements of a set of nested goods using capital letters $\{A, B\}$. The number of elements in a set of conventional goods, *e.g.*, $\{a, b, c, d, e\}$, can, in principle, be increased to any desired number. It is usually also possible to increase the elements of a set of nested goods $\{A, B, C\}$ by inclusion of a good which encompasses $A$ or divides $C$ in some fashion. While mixed sets, *e.g.* $\{A, B, g, k\}$, are obviously possible, we will not consider the economic relationships between the elements of such a set because they follow directly from those for a set of conventional or nested goods.

There are two types of nesting with respect to the scope of the goods.[4] The first way is for $A$ and $B$ to both be measured along some common scale so that $A$ represents a larger value on that scale than does $B$. We will call this type

[3]An example of a nested environmental good would be to let $A$ represent air quality improvements throughout California and let $B$ represent air quality improvements in Los Angeles.

[4]To keep the notation simple, we will assume without loss of generality that $B$ is a proper subset of $A$.

*quantitative nesting*. [5] The second way is for $A$ to be composed of two objects, possibly sets, $B$ and its complement $B^*$, where neither $B$ nor $B^*$ is an empty set and their intersection is empty. We will call this *categorical nesting*.[6]

The distinction between numerical and categorical nesting is a relative one. At one level of detail, everything becomes categorical. Indeed, quantitative nesting can always be treated as categorical nesting by treating increments of the good of interest as separate goods. At another level of detail, everything might be treated as numerical. The problems with many environmental commodities is that the types of distinctions that respondents perceive as meaningfully different are not known without doing substantial exploratory work.[7]

A valuation sequence consists of a respondent valuing a set of goods in a particular order. We will denote the first good valued in a sequence with the superscript, 1, the next with the superscript 2, and so on, so that a valuation sequence might look like $V(c^1, a^2, b^3)$.[8] For simplicity, we will consider only willingness-to-pay (WTP) sequences which maintain the initial level of utility: in a valuation sequence we will talk about a respondent's willingness to pay for $a^2$ conditional on having already paid for the previous element $c^1$, in the sequence.[9] Particularly with nested goods, it will be useful to define *top-down* valuation sequences, *e.g.*, $V(A^1, B^2, C^3)$, and *bottom-up* valuation sequences, e.g., $V(C^1, B^2, A^3)$.

The elements of a set can be valued by a single respondent or by a subsample of respondents.[10] We will denote the first subsample being asked willingness-to-pay questions with the Roman numerical subscript I, the second with the subscript II, and so on, so that a valuation sequence for subsample II might look like $V(c^1_{II}, a^2_{II}, b^3_{II})$.The summary statistic (e.g., median, mean) for a subsample's willingness to pay for an arbitrary element in a valuation sequence will be noted WTP $(x^i_j)$ if it is a conventional good or WTP$(X^i_j)$ if it is a nested good.

Consider three different valuation sequences $V(a^1_I, b^2_I, c^3_I)$, $V(b^1_{II}, C^2_{II})$, and $V(c^1_{III})$ and possible comparisons between willingness to pay for different elements of those sequences. The graphical representation in Table 1 (for conventional

[5]For example, $A$ could be 20 days of improved visibility and $B$ improved visibility on 6 of the 20 days represented by $A$.

[6]For example, $A$ could be 20 days of improved visibility and raising the level of water quality in a river basin from fishable to swimmable quality while $B$ 20 days of improved visibility.

[7]To assume, for example, that respondents in a city should be willing to pay for air pollution control to achieve a change from an average of 12 bad quality days a year to an average of 10 days a year assumes that the respondents perceive this as improvement they care about; this may or may not be the case. Contingent valuation surveys rely on the respondents *perception* of the scenario, and the researcher has the responsibility to clearly communicate the nature of the good to the respondent, including its numerical and categorical properties.

[8]$\mathbf{V}(\cdot)$ will be used to simply denote the order in which a respondent is asked questions about the valuation of particular goods. It will not be used to represent the values for the goods themselves. In the example $V(c^1, a^2, b^3)$, the respondent was first asked about the value of good $c$, then about the value of good $a$, and then about the value of good $b$.

[9]The concepts and notations set out here can also be used for constant utility willingness to accept compensation sequences. The direction of income and substitution effects in such a sequence will tend to be the opposite effects of those in a willingness to pay sequence [7].

[10]For simplicity we will assume random assignment of respondents to subsamples. Obviously, more complex statistical designs are possible. We will refer to subsamples rather than individual respondents unless there is a need to make a distinction.

TABLE I
Conventional Goods

| Subsample I | Subsample II | Subsample III |
|---|---|---|
| $a_I^1$ | | |
| $b_I^2$ | $b_{II}^1$ | |
| $c_I^3$ | $c_{II}^2$ | $c_{III}^1$ |

TABLE II
Nested Goods

| Subsample I | Subsample II | Subsample III |
|---|---|---|
| $A_I^1$ | | |
| $B_I^2$ | $B_{II}^1$ | |
| $C_I^3$ | $C_{II}^2$ | $C_{III}^1$ |

goods) and Table II (for nested goods) may be useful.[11] Comparisons between the willingness to pay for different elements by the same respondent or subsample will be termed *within subject tests*. In Tables I and II, these *internal* tests are those performed between the elements of a particular column. Comparisons between elements valued by different respondents or subsamples will be termed *between subject tests*. These *external* tests compare elements across columns, including comparisons between elements along the diagonal.[12]

### *2.1 Predictions from Economic Theory: Conventional Goods*

Economic theory says little about the relationship between the elements of the three valuation sequences of conventional goods depicted in Table 1 without specifying more about the nature of those goods. It is reasonable to assume that most environmental goods are normal goods. It may also be reasonable to assume that the goods in the set of interest are all individual or, in groups, Hicksian substitutes for each other.[13] Under these conditions, it can be shown that two external tests of economic theory are possible:

$$b_I^2 \leq b_{II}^1 \tag{1a}$$

[11] It is possible to contract either of these tables by taking away a column and/or row or to expand them by adding columns and rows. Quantitative nesting, in principle, allows for the possibility of a continuum of rows and columns. Most of the tests put forward in literature can be easily represented as testing the differences between elements in one of these two tables.

[12] For nonnested goods, there is an additional type of test based on the work of Hoehn and Randall [18]. The total value for a package of goods (but not the value of the individual goods) should be invariant to the order in which the individual goods are valued, *e.g.*, $V(a^1, b^2, c^3)$ versus $V(b^1, a^2, c^3)$.

[13] These two conditions guarantee that sequencing effects run in the same direction. There are of course situations of interest where they may not hold. For instance, some forms of recreation appear to be inferior goods and some combinations of environmental amenities may be compliments. The direction of the sequencing effect in such instances will depend on the absolute magnitudes of the individual effects.

and

$$c_{\mathrm{I}}^{3} \leq c_{\mathrm{II}}^{2} \leq c_{\mathrm{III}}^{1}. \tag{1b}$$

Tests of the hypotheses represented by Eq. (1a) and (1b) will be termed *embedding effect tests*.[14] If we make the further assumption that the respondent would always get positive utility from possessing any of these goods regardless of whether the respondent already possesses any or all of the other goods at issue, the weak inequalities in (1a) and (1b) can be replaced with strong inequalities. In general, economic theory does not say anything *a priori* about comparisons between different goods, *e.g.*, $(b_j^i, c_j^i)$.

*2.2 Predictions from Economic Theory: Nested Goods*

Nested goods are a special case of conventional goods so that (1a) and (1b) hold if we make the same assumptions. To get the strong inequality in the nested goods case, one must assume for (1a) that respondents still get positive utility from $B$ after they possess its complement $B^*$, and for (1b) that respondents still get positive utility for $C$ after they possess $C^*$. Testing a null hypothesis of equality in an embedding effect test is essentially testing for consistency with a flat utility function with respect to further increases in the good *e.g.*, WTP ($B1$) conditional on having $B^*$ is zero. Note although a flat utility function with respect to $B$ or ($C$) conditional possessing $B^*$ ($C^*$) might not be considered well behaved, it is not ruled out by economic theory, assuming local nonsatiation does rule out such behavior.

The nesting property also allows predictions about other possible comparisons to be made. Let us first look at the internal tests. Here, since we are using a top-down valuation sequence, it can be shown that

$$A_{\mathrm{I}}^{1} \geq B_{\mathrm{I}}^{2} \geq C_{\mathrm{I}}^{3}, \tag{2a}$$

and

$$B_{\mathrm{II}}^{1} \geq C_{\mathrm{II}}^{2}. \tag{2b}$$

A variant on (2a) and (2b) using a bottom-up approach can be written as

$$C_{\mathrm{I}}^{1} \leq B_{\mathrm{I}}^{2} \leq A_{\mathrm{I}}^{3}, \tag{2c}$$

and

$$C_{\mathrm{II}}^{1} \leq B_{\mathrm{II}}^{2}. \tag{2d}$$

It is important to note that the willingness-to-pay amounts for the goods in the

[14] Kahneman and Knetsch [22] refer to this type of test as a test of "regular" embedding. They use the term "perfect" embedding if the goods involved are nested commodities.

top-down and bottom-up valuation sequences will, in general, not be the same. This is because of differences that respondents perceive in the available implied choice set (initially large or broad depending on the type of nesting in the top-down sequence, and initially small or narrow in the bottom-up sequence) and because of differences in whether the respondent believes he or she already possesses the complement of the good being valued (the top-down case) or has just purchased it (the bottom-up case).[15] Equations (2a) and (2b) result follow from monotonicity assumptions. Again, the weak inequalities may be replaced by strong inequalities, if it is assumed that respondents get utility from the complement of the element being valued. Tests of the hypotheses represented by these equations will be called *nested sequence tests.*

Another comparison which can be made is along the diagonal of Table II:

$$A_{\mathrm{I}}^{1} \geq B_{\mathrm{II}}^{1} \geq C_{\mathrm{III}}^{1}. \tag{3}$$

We will term this comparison a *test of component sensitivity.* It looks at whether respondents are sensitive to differences in levels of the good (quantitative nesting) or the level of inclusion (categorical nesting) in a way that should change its value. Essentially, it is the external test version of Eq. (2) and has the same theoretical properties.

## 3. TESTING HYPOTHESES INVOLVING SEQUENCING AND NESTING

Early contingent valuation studies performed nested sequence tests of the hypotheses represented by Eq. (2) and routinely rejected the proposition that different levels of a good were valued equally.[16] There were two competing, although not necessarily contradictory, explanations for these results. The first was that contingent valuation surveys generally elicit responses which are consistent with well-behaved preferences. The second is that respondents simply strive to be internally consistent with the notion that more of something desirable is worth more.

Kahneman and Knetsch [21] contended that the behavior necessary to show that contingent valuation responses are valid was for equality constraints to hold in Eq. (1). This argument was in keeping with Kahneman and other's long-held notion that for contingent valuation to be useful, contingent valuation to be useful, contingent valuation responses should be largely context free. This notion was strongly rejected by resource economists conducting contingent valuation research as inconsistent with the notion of substitution effects and budget constraints

[15]In the top-down sequence, the respondent is asked an allocation question which, to be well-defined, asks how much less $A$ is worth if it does not include $B$. That is the respondent already presumes that he or she has possession of $A$ by virtue of having said how much he or she would pay for it and then is asked what portion of the purchase price for $A$ is for $B$. In the bottom-up sequence, the respondent would have already purchased $B$ and then is essentially asked how much $B^*$ is now worth. Bottom-up sequences have been frequently used by contingent valuation research to value a series of improvements to an existing environmental good. They are generally better defined than top-down sequences.

[16]See for instance, Brookshire *et al.* [5].

[3, 16–18].[17] These economists consider the size and nature of the choice set to be important factors in determining a respondent's value to a particular good and believe that sequencing effects, in general, should be expected. Indeed, some thought that such effects might be sizeable even in common situations.[18] In general, those doing applied contingent valuation research followed the practice of the vast majority of applied benefit–cost studies by treating the change in the good of interest as the only change to the status quo. If the agency was going to undertake related projects, researchers would sometimes suggest that an examination of the "package" might be warranted.

Kahneman and Knetsch [21, 22], noting this admission, argued that since the choice of a valuation sequence is arbitrary, estimates from a contingent valuation study are arbitrary. This argument was sooner taken up by other CV critics, some of whom took the position that respondents should be allowed to change the level of consumption of all public goods.[19] There are two problems with this argument. The first is a theoretical point not specific to contingent valuation; Carson *et al* [7] and Randall and Hoehn [31] have shown that the value of any particular public good generally has to become progressively smaller as it is placed deeper and deeper in a willingness-to-pay sequence. They also show that the opposite is true in the case of a willingness-to-accept sequence. The second is a practical empirical issue specific to contingent valuation: as one tries to value more and more goods in a single survey instrument, the description of each of the goods necessarily becomes so brief as to convey only a vague notion to respondents of the nature of the good to be provided and the method and manner in which it would be provided, violations of good CV practice.

Even assuming agreement on a small set of goods, testing the hypotheses represented by Eq. (1) poses two substantive problems. The first is a fundamental

[17]While Kahneman and Knetsch [21] make some acknowledgment of the possible role of substitution and income effects, they claim, without presenting supporting evidence, that such effects should be small and ignorable. The Diamond *et al.* [12] paper claims to find no substitution effect in a wilderness example. However, Carson and Flores [6] show that Diamond *et al.* should have reached the opposite conclusion—substitution effects were potentially large and not ignorable. They also show, by simply regressing willingness to pay on the number of acres in the wilderness area the respondent was asked to value, that the Diamond *et al.* claim to have demonstrated component insensitivity in their first hypothesis test is incorrect.

[18]It is fairly straightforward to demonstrate that embedding effects must occur even in demand systems for common marketed commodities [31]. Large multiple-site travel cost models, which are sometimes used to estimate nonmarket use values, typically suggest very large embedding effects. A stylized example based on the work of Ward and Fiore [37] involves filling water reservoirs around a large city in the southwestern United States. The water reservoirs vary in size and in their location. Whichever reservoir is filled first typically produces very large recreation benefits while whichever reservoir is filled last typically produces quite small additional recreational benefits.

[19]For an empirical example, see Kemp and Maxwell [23] who offered respondents to a shopping-mall intercept survey the opportunity to say how much extra they would pay in a federal income tax surcharge to "help improve" a wide range of social problems which included "protecting the environment." Using a top-down disaggregation procedure, respondents were next asked to break down, on a percentage basis, the portion of this amount they would pay for "protecting the environment" among nine categories of environmental protection including "protecting wilderness areas and wildlife." Subsequent disaggregation allocated percentages among four major types of harm, and then among five types of "human caused problems" and finally among U.S. coastline regions. After multiplying the original amount given for the environment by the long string of percentage allocations, Kemp and Maxwell conclude that their respondents are willing to pay almost nothing to prevent an Exxon Valdez type oil spill in Alaska.

disagreement between economists and some cognitive psychologists over whether it is "good" or "bad" for contingent valuation to be able to reject the equality in (1). The second is that testing Eq. (1) generally requires a top-down allocation question to get an estimate of WTP($B_1^2$) and/or WTP($C_{II}^2$). The willingness-to-pay questions for the less inclusive goods using a top-down allocation framework typically present respondents with an ambiguous choice situation because the conditions under which these goods are to be provided are not well specified. As a result, different respondents will make different assumptions about those conditions and it is difficult to say what the responses to such questions represent.[20]

### *3.1. Implementation of Component Sensitivity Tests*

Because of problems involved with the embedding effect test represented by Eq. (1) and with consistent acceptance of the theoretical restrictions represented by Eq. (2), attention has shifted to the component sensitivity test represented by Eq. (3) and its *null hypothesis of component insensitivity*. It is important because respondents to CV surveys should be sensitive to perceptually important differences in scope between two nested goods.[21] It is feasible because the appropriate experimental design to test these diagonal relationships can be accommodated by surveys that value a single good; such surveys avoid the problems described above for multiple goods and make it possible to describe the good and its market in sufficient detail to successfully communicate the level of services that will be provided.[22]

It is one thing to conduct component sensitivity experiments to assess the validity of a particular CV instrument; it is another to use them to assess the validity of the *method*. In the latter case researchers who fail to reject a component insensitivity hypothesis must be able to credibly argue that the result cannot be attributed to respondents actually not caring much about the difference between the goods, to poor instrument design, and administration or to a test that lacks sufficient power to reject plausible divergences from the null.[23] Because results that appear to violate economic theory (and common sense) can easily be induced in component sensitivity tests using inadequately designed CV surveys, at a

[20]A dramatic but simplistic example using private goods demonstrates this concern. Assume that our respondent's car coasts into the only gas station on a long stretch of desert road with a leaking radiator and out of gas. Ask the well-defined question, "How much are you willing to pay right now for fixing the radiator and a tank of gas?" Now ask the allocation question, "How much of that amount was for the tank of gas?" The respondent's answer should depend on whether the gas station has already fixed the radiator and been paid; and, if not, whether the gas station can fix the radiator; and, if so, what the cost of fixing the radiator is going to be.

[21]The Blue Ribbon Panel convened by NOAA identified embedding [insensitivity to scope] in this sense as a potentially "perhaps the most important internal argument against the reliability of the CV approach" [1, p. 4607].

[22]This test, with random assignment, also has the advantage of ensuring that the initial level of utility (*i.e.*, income, perceived prices of private commodities, and perceived portfolio of public goods) in the different treatments are at least asymptotically equivalent, thus controlling one of the major factors which might be changing in an unobserved manner in the embedding effect test of Eq. (1).

[23]This statement should not be taken as saying anything about the general validity of contingent valuation as a valuation technique. Rather we are saying that before claiming that there is a problem with a measurement tool one should determine how that tool was used.

minimum, component sensitivity tests of the CV *method* should do the following:

1. Use goods $A$, $B$ and $C$, the differences between which are meaningful to respondents.[24]

2. Conduct in-depth exploratory research to identify the relevant design issues, especially the types of amenity misspecification which may bias the findings.[25]

3. Use the findings of this preliminary research to develop a description of each of the goods that respondents understand and a method of provision they find plausible.

4. Use a mode of survey administration that promotes respondent cooperation and attention during the interview process.

5. Use a sample of sufficiently large size that ensures that the hypothesis test has the power to reject reasonable differences of the magnitude.

## 4. SURVEY DESIGN ISSUES IN THE EMBEDDING DEBATE

From a survey perspective, the measurement issues raised by the "embedding" controversy involve a well recognized and potentially serious class of biases which we have elsewhere called amenity misspecifications [28]. Although these biases pose a methodological challenge to CV researchers and require careful attention in the design phase of a study, they are often avoidable if the scenario is plausible and the good is carefully described.

Among the amenity misspecification errors which are particularly relevant when testing hypotheses represented by Eqs. (1), (2), and (3) are symbolic, part–whole, metric, and probability of provision biases.[26] The presence of the first three biases tends to produce the appearance that people are insensitive to the characteristics of the good that they are asked to pay for. The last one can produce dramatic violations of economic theory by conveying the impression that people are willing to pay more for less. In what follows, we briefly describe the conditions which promote each type of bias and the implications for the "embedding" experiments conducted by the CV critics.

*Symbolic bias* occurs if respondents react to an amenity's general symbolic meaning instead of to specific levels of provision. It is promoted by scenarios in which respondents value relatively small changes, scenarios that use brief descriptions of the good and the market, and studies that employ impersonal survey methods [28, p. 250]. Under these conditions respondents are likely to assume that

[24]Meaningful differences between goods is of course a relative concept. The distinction is necessary because there is nothing in economic theory that specifies which differences in the categorical or quantitative features of goods should result in consumers' giving different willingness-to-pay amounts for them. The reason that nested goods are so suitable for this type of experiment is that the directionality of the effect can be predicted for these goods even though the point where one good nested in the other should begin to command a different value cannot be predicted with certainty.

[25]Tolley *et al.* [35] provide an example of this type of exploratory research. Upon finding that their respondents were willing to pay almost as much for a distributional shift that improved visibility on 10 days as they were for visibility improvements for an entire season, they revised the scenario to present a much more thorough presentation of the good. After doing this and administering the survey to a new group of respondents they note: "the strikingly smaller bids for the modest distributional changes, in contrast to the large bids obtained in the original survey, suggests that the refined questionnaire imparted a more careful understanding of the programs being considered."

[26]For a more detailed discussion of these biases see Mitchell and Carson [28].

the small damages described to them are symbolic for a good of greater magnitude; otherwise, why would anyone be taking the trouble to ask them to place a value on it.

*Part–whole biases* are the result of the willingness of respondents to answer questions even though the good itself may be vaguely described. In such instances, respondents fill in the details and make assumptions about the good that they think the interviewer has in mind. Vague descriptions are those which merely identify a good without specifying its important qualities and how it would be provided. Two examples from the study by Kahneman and Knetsch [21] are "provide protection for endangered birds" and "restore museums in the rural areas of British Columbia." Because these goods are vaguely defined and their provision implausible,[27] respondents are unlikely to make the effort necessary to give a considered value for their provision which takes into account all of the services they claim to provide (*e.g.*, all endangered birds, all rural museums of all types). Put another way, ask a vague question and you are likely to get a vague, uninterpretable answer.

*Metric bias* occurs when the respondent values an amenity according to a metric different from the one intended by the researcher [28, p. 252]. Often this involves treating the numbers as ordinal instead of interval data. Another important metric consideration involves the baseline: do the 2000 dead birds belong to an endangered species or to one that has more than half a million members living in the general area where the deaths occurred? Preventing metric bias requires understanding the relevant default assumptions a naive respondent is likely to hold about the relevant numbers and carefully framing the numerical information in the scenario to properly communicate its properties.

*Probability of provision* bias occurs when the perceived probability that the good will be provided differs from the researcher's intended probability. Most CV studies want the respondent to assume that the good will be provided with certainty. To the extent that respondents discount this, they will undervalue the good. The types of goods that are most vulnerable to this effect are those that involve large changes (*e.g.*, 50% reduction of air pollution nationwide), or are vaguely defined, or are valued in a scenario which does not contain a plausible method of providing such a large improvement.[28]

Under these conditions, probability of provision bias has the potential to cause people to appear to pay more for less, especially in combination with part–whole bias. For example, in one experiment Kahneman and Knetsch [21] found small that split samples were willing to pay $125 and $59, respectively, to "provide protection for the Peregrine Falcon, an endangered bird" and "provide protection for endangered birds." A possible reason that people were willing to pay less for the second, seemingly more inclusive good, is that they discounted the likelihood that protection of all endangered birds will be provided more than they discounted the likelihood of protecting a single named species.

[27] For example, respondents were not provided information about which museums, what kinds of restoration, who would ensure that restoration took place, the time period over which restoration would take place, and the details of the payment scheme for the program.

[28] In a series of experiments which presented subjects with brief CV-like scenarios, Fischhofff *et al.* [13] found that larger goods were viewed as less likely to be provided than smaller goods and that the lower the likelihood of provision, the less people were willing to pay.

*4.1. Amenity Misspecification Issues in Desvousges et al.*

The methodological issue raised in the embedding debate is whether CV surveys can be designed so that they avoid or reduce these biases, *i.e.*, whether these effects are inevitable. The CV critics [11, 12, 21] are confident that their experiments show the latter. However, their position assumes that the findings of their experiments can be unequivocally attributed to a flawed methodology rather than their CV instrument and administration, an assumption that seems unwarranted when one looks closely at their studies. To illustrate this point, we discuss briefly some of the weaknesses of the Desvousges *et al.* [11] study which conducted two split-sample experiments using questionnaires that were self-administered by respondents recruited by mall intercepts in Atlanta, Georgia.[29]

The Desvousges *et al.* first set of experiments found that respondents gave statistically similar WTP amounts in treatments which asked split samples to value preventing 2000, 10,000, and 200,000 birds from being killed by waste-oil holding ponds in the U.S. Central Flyway. While the difference in the number of birds killed appears dramatically large, this effect is diminished by other information in the scenario about the size of the total bird population that frames the loss in such a way that respondents could rationally perceive the three injury levels as relatively similar in scope.[30] Further, no information was given about whether any of the species affected by the ponds were endangered, something that many respondents might plausibly have assumed since at least three of the five specifically mentioned in the scenario as affected–white fronted geese, snow geese, and greater sandhill cranes—would not be familiar to many Atlanta residents.

Several aspects of their scenario are potentially implausible to their respondents. These include: (1) the idea that Atlantans would actually have to pay the amounts they said they were willing to pay for a good involving birds and waste-oil holding ponds located more than 1000 miles away from their house and (2) the idea that 250,000 waste-oil holding ponds were killing only 2000 or even 20,000 birds each year out of over 8 million passing over them. That such a large number of oil ponds would kill so few birds might have also encouraged some respondents to assume that the actual numbers stated in the survey were symbolic of more substantial injuries. In addition, it is unfortunate that the study did not ask debriefing questions to allow one to help evaluate whether respondents were sufficiently motivated to pay attention to the details of the scenario.[31]

[29]For a more detailed discussion of serious design flaws in the survey instruments used in another of the major embedding experiments, see the Smith [34] critique of Kahneman and Knetsch [21, 22].

[30]In the text of scenario, the respondents are told that there are 8.5 million migratory waterfowl in the central flyway and that the number of birds whose deaths would be prevented amount to: "much less than 1% (for the 2000-bird scenario), "less than 1%" (20,000) and "about 2 percent" (200,000) of total birds. Desvousges *et al.* find equivalent medians for each of these three treatments ($25), although the study's estimated mean value for the 2 and 1% scenarios were 71 and $59, respectively. Their experiment lacked sufficient power to detect differences of this magnitude.

[31]Schkade and Payne [33] working with Desvousges *et al.* found, using the same shopping mall setting and questionnaire, that respondents had a median willingness to pay for the 200,000-bird treatment that was twice as large as that for the 2000-bird treatment, as well as a substantial difference in the two means. The difference between their work and that of Desvousges *et al.* was that Schkade and Payne asked respondents to slow down and tell them what they were thinking while answering the questionnaire, thereby more closely approximating an in-person interview.

The second set of experiments reported by Desvousges *et al.* asked one split sample to value reducing the environmental effects of oil spills of less than 50,000 gallons and the other to value preventing all spills, including those under and over 50,000 gallons. Respondents expressed similar levels of value for the two levels of prevention which Desvousges *et al.* contend represent large differences that should have elicited different size WTP amounts. The first problem with this survey instrument is that its scenario does not clearly describe the aggregate amount of environmental injury that would be prevented by each program.[32] The second design problem is the complexity of the description of how the program would work, which obscures the difference between the two goods.[33] and uncertainty about whether the good will be provided.[34] These design problems raise questions about the nature of the injury that respondents were asked to value and may well have increased the prospect for part–whole biases.

## 5. A NEW TEST OF THE COMPONENT INSENSITIVITY HYPOTHESIS

The component insensitivity hypothesis has been tested in some fashion in a number of other CV studies conducted for other purposes, including several in which we have been involved. In the first of these, Mitchell and Carson [27] tested whether respondents valued different levels of risk reduction with respect to trihalomethane (THM), a common drinking water contaminant, using random assignment and a split-sample approach. We found significant differences in the three comparisons made. In Mitchell *et al.* [29], using a similar approach, we tested whether respondents valued air quality programs which provided health and visibility benefits significantly more than programs which provided only health benefits and found that they did. In Carson and Mitchell [8], again using a split-sample approach, we tested whether respondents were willing to pay different amounts to prevent residential water shortages in California of different severities and frequencies. We found the amounts for the four programs asked about to be significantly different from each other. We have also carried out a comparison [9] which looked at willingness to pay for improvements in national water quality versus a particular large river system from two independent studies which used very similar survey instruments except for the exact scope of the good. We strongly rejected that the willingness to pay estimates were the same. Carson and Mitchell [10] discuss these tests in more detail.

[32]Two pages before the valuation question, respondents are told that "The damage from spills of less than 50,000 gallons is usually limited to sea birds and shoreline habitats in the immediate vicinity of the spill. Large offshore spills may cover miles of shoreline, contaminate shellfish beds, and kill thousands of seabirds and some marine mammals." Missing from this description are several types of information which respondents would need to determine the value of preventing such a spill to them. These include the types and endangered status of the birds, the number of miles of shoreline, whether the effect on shellfish poses possible harm to humans, and the number of types of marine mammals at risk.

[33]Both subsamples were told that 95% of oil spills are less than 50,000 gallons and that the remainder are the few very large spills that can exceed 1,000,000 gallons. One subsample was asked how much they would pay in higher prices to prevent 90% of the environmental damage from spills of 50,000 gallons or less. The other subsample was asked how much they would pay to prevent 90% of the damage from spills under 50,000 gallons and 75% of the damage from larger spills.

[34]The scenario relies primarily on citing "experts" who "think" that the programs will prevent the specified percentage of the spills.

In what follows we present the findings of a new experiment which offers a clear test of the component insensitivity hypothesis. It comes from a contingent valuation study [19] conducted by the Australian Resource Assessment Commission to measure willingness to pay to prevent possible injuries from a large proposed open pit mine in the Kakadu Conservation Zone, an area surrounded by the Kakadu National Park, one of Australia's two major national parks. Substantial effort was put into identifying and addressing potential sources of confusion and bias prior to fielding the CV instrument. The good that was valued, preservation of some of the amenities provided by the Kakadu National Park in Australia, is likely to provide mostly passive use values for most of the respondents since they live far away from the park but are aware of its reputation as one of Australia's premiere national parks.

### *5.1. Kakadu Conservation Zone Study*

The policy issue under consideration was whether the conservation zone should be added to the Park to protect the area from the harm that might be caused by mining or whether mining should be allowed to proceed. The large resources that are necessary to conduct split-sample experiments of this magnitude were committed in this case because policy makers needed to know the value of different levels of this good. A national sample of Australian households was selected according to a multistage design. Two thousand in-person interviews lasting an average of 30 minutes were conducted by professional interviewers from one of Australia's major survey research firms.

The decision to use a split-sample design suitable to test the sensitivity hypothesis was made for policy purposes rather than for the sake of conducting a methodological experiment. The commission knew from submissions by interested parties and a series of public hearings held around the country that the likely impact the mine would have on and off the site was a matter of considerable scientific and political debate and that the use of a single scenario would be of limited use if the impacts it presented were later shown substantially different from the commission's final assessment of the actual impacts. For this reason, two scenarios were developed—a minor impact scenario and a major impact scenario.[35] The basic facts underlying the scenarios were drawn in large part from the environmental impact statement prepared on behalf of the mining companies involved. The major interest groups on both sides of the issue were consulted on the wording of the scenarios: one of the scenarios tilted toward the mining industry's view of the risk and the other toward the environmental groups' view of the risk.[36] With the exception of the variation in the degree and location of impact and risk of occurrence, the survey instruments were identical.

The questionnaire began by asking the respondents their views about several types of nonenvironmental problems followed by several environmental attitudinal measures (including a question that posed a trade-off between protecting the Australian environment or developing its economy). Then respondents were asked

[35]The scenarios were not described as "major" or "minor" to the respondents. These terms are used here as a convenient label for the two scenarios.

[36]See Imber *et al.* [19] and Australian Resource Assessment Commission [2] for details. Neither interest group was completely happy with the characterization made of their position.

a series of questions about their awareness of and their experience with the Kakadu National Park. Maps, photographs, and artist renderings were used to convey basic background information such as the Parks and conservation zone's size, location, and characteristics. The respondents were also told about previous mining and other resource use activities in the area.

The next portion of the interview described the impact scenario and showed respondents an artist's rendering of what the proposed mine would look like. Four types of environmental effects were highlighted: (1) mine-related traffic, (2) chemicals used to extract minerals, (3) mine process water, waste rock material, and (4) impact on wildlife. The difference between the major and minor impact scenarios was the degree of risk of off-site environmental damage. In the minor impact scenario respondents were informed:

> While there would be some toxic chemicals used in the mining process, these chemicals would pose little risk to the environment. This is because environmental controls based on best practicable technology would be used. A range of stringent measures and safefguards will be taken to ensure that storage and on-site disposssal of toxic substances is strictly controlled.

The parallel information in the major impact scenario states:

> The use of toxic chemicals, mainly cyanide, hydrochloric acid, and lime, in the mining process poses a small risk to the environment. Although environmental controls based on best practicable technology and a range of stringent measures and safeguards would be used, there is always a small chance of an accident. Any such accident could harm plants and animals that live close to or well away from the mine.

The next paragraph in both scenarios began with a sentence elaborating on the safeguards: "The water storage arrangements for the proposed mine are designed to allow natural decontamination and prevent release of water and chemicals used in the mining process." The possibility of a toxic leak centered on the tailings pond and differed between the two scenarios.

The major impact scenario presented a photograph depicting one of the more important ecological areas that might be at risk from the mining activities. The major impact scenario also told respondents that "It's possible that the mine could upset the natural balance of Kakadu National Park. The mere existence of the mine could spoil the natural value of Kakadu National Park."

The Kakadu survey used a double-bounded discrete choice elicitation framework in the context of a referendum.[37] Respondents were randomly assigned to different monetary amounts as well as to the major or minor impact scenario. The monetary amounts used were chosen to a large degree to facilitate comparison with the value of mining outputs but were adjusted upward somewhat on the basis of pretest results. The payment vehicle was a reduction in take-home pay.

A test of the component insensitivity hypothesis is whether respondents who were asked about the major impact scenario give higher WTP amounts to prevent the mine than those who were presented with the minor impact scenario. The component insensitivity hypothesis can be tested nonparametrically by fitting the Peto–Turnbull version [36] of the Kaplan–Meier estimator to the combined major and minor impact scenarios data set and to the data from each of the two scenarios

[37]See Nelson [30] or Hanemann *et al.* [14] for a detailed discussion of this estimator.

TABLE III
Weibull Equation for Willingness to Pay for Kakadu Conservation Zone

| | Parameter | Standard error | *T*-statistic |
|---|---|---|---|
| Location ($\alpha$) | 5.4391 | 0.1404 | 38.75 |
| Shape ($\beta$) | 2.9323 | 0.1048 | 27.98 |
| MAJOR | 0.6731 | 0.1941 | 3.47 |

*Note.* Log likelihood $-2236.80$; $N = 2034$.
Median major impact scenario A$154; 95% confidence interval [A$116–A$204].
Median minor impact scenario A$79; 95% confidence interval [A$60–A$102].

separately.[38] For the double-bounded data, this likelihood ratio test statistic is 29.046, which has a $\chi^2$ distribution with six degrees of freedom under the null, indicating rejection of the hypothesis that responses are insensitive to scope of the good being valued at the 0.01 level. The same test statistic based on only the response to the first willingness to pay question is 26.296 which has a $\chi^2$ distribution with four degrees of freedom under the null, again indicating rejection of the null at the 0.01 level.

Fitting a parametric distribution allows one to obtain a smooth curve rather than simply a step-function as in the non-parametric approach. Parametric tests based upon Weibull, log-normal, log-logistic, and normal likelihood functions all also reject the null hypothesis at the 0.01 level, using both willingness to pay responses and only the first response. Fitting a Weibull distribution to the data (Table III), as an illustration, yields results in median WTP estimate for the major impact subsample of A$154 with a 95% confidence interval of A$116–A$204. This compares with a median value for the minor impact subsample of A$79 with a 95% confidence interval of A$60–A$102. The component insensitivity hypothesis can be tested by including a dummy variable for being assigned to the major impact version (MAJOR) in the Weibull likelihood equation. The *t*-statistic for the MAJOR coefficient of 3.47 indicates rejection of the component insensitivity hypothesis at the 0.01 level. The Weibull distribution fits the observed data well except for the extreme tails. Using it, Fig. 1 displays the estimated percentage willing to pay as a function of the cost in Australian dollars.

In assessing the quality of any study it is useful to consider whether the willingness-to-pay amounts can be predicted by using the covariates available in

[38]The data for a separate and smaller sample from the Northern Territory (NT) where the Kakadu National Park is located suggests no difference in willingness to pay between the major and minor risk scenario. This finding is not surprising given the extensive news coverage and political debate over the mine proposal which took place in the Northern Territories. The nature of the two scenarios was such that neither was likely to directly contradict prior information held by NT respondents and neither contained information which directly contradicted the other. Consistent with large differences between the two samples in the distribution of most of the key covariate, the WTP estimates for the NT sample are on average smaller. The NT sample, however, also has a clear bimodal WTP distribution. This is also consistent with the distribution of covariates values in the NT sample and the perception that some NT residents would directly gain financially from the mine while some would lose.

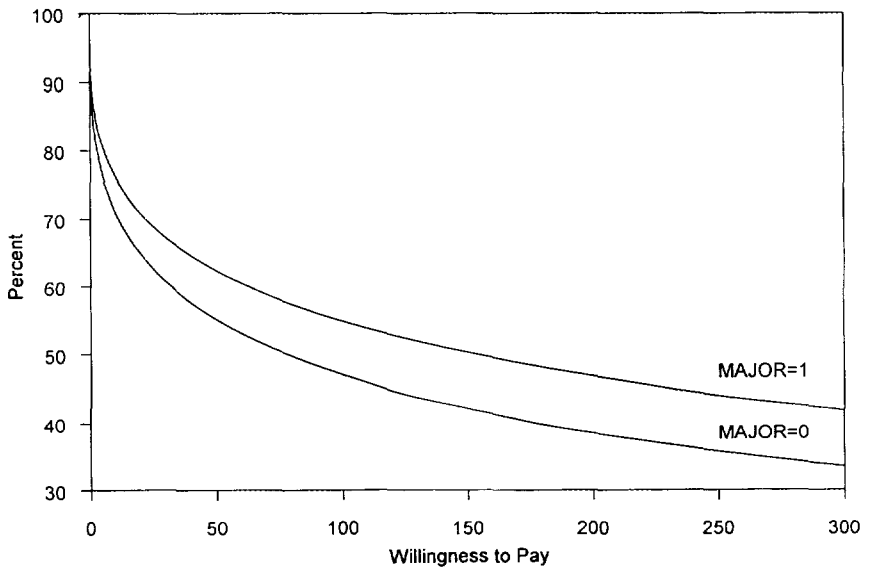


FIG. 1. Willingness to pay for Kakadu Conservation Zone.

the survey. Here, parameterizing the location coefficient in the Weibull survival model, it can be shown that income, concern over aboriginal issues related to the park, feeling that mining within a national park reduces it value, desiring the government to provide more parks, visiting parks, and engaging in conservation behavior are all positively (and significantly) related to willingness to pay. Concern over jobs from mining, belief that risk from mining activities in general was low, belief that consideration of financial benefits to the country is important in making natural resource decisions, and age are all negatively (and significantly) related to willingness to pay. An indicator variable for the scenario the respondent received remains a significant predictor of willingness to pay. In contrast, studies showing a lack of scope effects (*e.g.*, Desvousges *et al.* [11]) typically show little ability to predict respondent willingness to pay.

The Australian Kakadu data clearly reject the component insensitivity hypothesis, and therefore, challenge the strong assertions Kahneman and Knetsch [21] and others [15] make about the fundamental inability of respondents in a contingent valuation survey to take into account the characteristics of the good they are asked to value. We should note though that greater resources were available to design and conduct this study than is the case for many CV studies. This made it possible, for example, to use in-person interviews and extensive visual aids. It is possible that CV studies conducted under tighter budgets would find it more difficult to demonstrate component sensitivity.

A search of recent CV literature finds a number of other studies which conduct split-sample comparisons that can be shown to be component sensitivity tests. These studies vary in the techniques used and valued a wide variety of goods including some whose total value is likely to be derived mostly from direct use and others whose total value is likely to be attributable to mostly passive use considerations. Among those that clearly demonstrate a significant scope effect in one or

more comparisons are the Boyle, *et al.* [4] white-water rafting and Colorado River flows, Jakus' [20] gypsy moth programs, Krieger's [24] sport fishing toxics information programs, the Loomis, *et al.* [25] forest area protection programs, Magnussen's [26] Norwegian pollution programs, Römer and Pommerehne's [32] hazardous waste risks, Whitehead's [38] sea turtle protection programs, Whitehead and Blomquist's [39] wetlands programs, and Wu's [40] Big Darby Creek improvement programs.

## 6. CONCLUSION

A major source of confusion in the embedding debate has been the indiscriminate use of the term embedding to describe issues or findings that relate to completely different types of relationships. To clarify the linkage between embedding and the economic theory governing consumer behavior, we have proposed an analytical framework based on the concepts of nesting and sequencing. By classifying the relationships that multiple goods can have, the framework should help facilitate systematic testing of hypotheses that follow from the interrelationships between public goods. Of course the relevance of this framework is not limited to CV analyses. In principal these tests of sequencing and nesting properties can also be applied to revealed preference studies of multiple public goods.

A number of the tests described, including the component sensitivity test, may be useful in assessing the validity of a *particular CV* study. These tests can serve as useful diagnostic tools in the design phase of a contingent valuation study and may help researchers determine whether respondents are focusing on a CV scenario's key features. In order to enhance testing efficiency, we devote considerable attention to the development of criteria for a well-designed CV component sensitivity test. The criteria we recommend include: (1) careful attention to potential amenity misspecifications, (2) the use of a plausible market scenario, (3) a mode of survey administration that promotes respondent cooperation and attention, (4) the choice of nested goods whose differences are meaningful to respondents, and (5) a sample large enough to reject reasonable differences. These criteria are consistent with the methodological guidelines the NOAA Panel [1] proposed to address the "embedding problem" and other possible weaknesses of the CV method.

Conducting a single-component sensitivity test cannot, however, test the validity of the method. This follows from the realization that failure to reject the hypothesis of component insensitivity may simply be the failure of respondents to interpret the particular scenario as researchers intended. To evaluate the credibility of the claims of CV critics [11, 12, 21], it is desirable to consider evidence from as wide a variety of contingent valuation studies as possible. If lack of sensitivity to scope is an unavoidable problem with the method *per se*, there should be wide-spread failure to reject the component insensitivity hypothesis. Evidence from the Kakadu study presented here, as well that from a number of other recent studies, suggests otherwise. Such a judgement, should it prove to be robust, should not be taken as license to ignore the ever present threat of amenity misspecification biases. Research efforts can profitably be directed toward helping to understand the conditions under which these biases are likely or not likely to occur.

## REFERENCES


1. K. Arrow, R. Solow, P. R. Portney, E. E. Leamer, R. Radner, and H. Schuman, Report of the NOAA Panel on Contingent Valuation, *Fed. Regist.* **58**, 4601–4614 (1993).
2. Australian Resource Assessment Commission, "Commentaries on the Resource Assessment Commission's Contingent Valuation Survey of the Kakadu Conservation Zone Report," Canberra (1991).
3. R. C. Bishop, and M. P. Welsh, Existence values in benefit–cost analysis and damage assessment, *Land Econom.* **68**, 405–417.
4. K. Boyle, M. P. Welsh, and R. C. Bishop, The role of question order and respondent experience in contingent valuation studies, *J. Environ. Econom. Management* **25**, S80–S99 (1993).
5. D. S. Brookshire, B. C. Ives, and W. D. Schulze, The valuation of aesthetic preferences, *J. Environ. Econom. Management*, **3**, 325–346 (1976).
6. R. T. Carson, and N. E. Flores, A closer look at "Does contingent valuation measure preferences?: experimental evidence." How evident is the evidence? Unpublished manuscript, Department of Economics, University of California, San Diego (1993).
7. R. T. Carson, N. E. Flores, and W. M. Hanemann, "On the Creation and Destruction of Public Goods: The Matter of Sequencing," Paper presented at the European Association of Environmental and Resource Economists Meeting, Cracow, Poland (1992).
8. R. T. Carson and R. C. Mitchell, "The Value of Diamonds and Water," Paper presented at the European Association of Environmental and Resource Economists Meeting, Stockholm, (1991).
9. R. T. Carson, and R. C. Mitchell, The value of clean water: The public's willingness to pay for boatable, fishable, and swimmmable quality water, *Water Res.* **29**, 2445–2454 (1993).
10. R. T. Carson and R. C. Mitchell, The issue of scope in contingent valuation studies, *Amer. J. Agric. Econom.* in press.
11. W. H. Desvousges, F. R. Johnson, R. W. Dunford, K. J. Boyle, S. P. Hudson, and K. N. Wilson, Measuring natural resource damages with contingent valuation: Tests of validity and reliability, *in* "Contingent Valuation: A Critical Assessment" (J. A. Hausman, Ed.), North-Holland, Amsterdam (1993).
12. P. A. Diamond, J. A Hausman, G. K. Leonard, and M. A. Denning, Does contingent valuation measure preferences?: experimental evidence, *in* "Contingent Valuation: A Critical Assessment" (J. A. Hausman, Ed.), North-Holland, Amsterdam (1993).
13. B. Fischhoff, M. J. Quadrel, M. Kamlet, G. Loewenstein, R. Dawes, P. Fishbeck, S. Klepper, J. Leland, and P. Stroh, Embedding Effects: Stimulus representation and response mode, *J. Risk Uncertainty* **6**, 211–234 (1993).
14. W. M. Hanemann, J. B. Loomis, and B. Kanninen, Statistical efficiency of double-bounded dichotomous choice contingent valuation, *Amer. J. Agric. Econom.* **73**, 1255–1263 (1991).
15. J. A. Hausman, (Ed.), "Contingent Valuation: A Critical Assessment," North-Holland, Amsterdam (1993).
16. J. P. Hoehn, "Valuing the multidimensional impacts of environmental policy: Theory and methods, *Amer. J. Agr. Econom.* **73**, 289–299 (1991).
17. J. P. Hoehn, and J. B. Loomis, Substitution effects in the valuation of multiple environment programs, *J. Environ. Econom. Management*. **25**, 56–75 (1993).
18. J. P. Hoehn, and A. Randall, Too many proposals pass the benefit cost test, *Amer. Econom. Rev.* **79**, 544–551 (1989).
19. D. Imber, G. Stevenson, and L. Wilks, "A Contingent Valuation Survey of the Kakadu Conservation Zone," Australian Resource Assessment Commission Research Paper No. 3, Canberra (1991).
20. P. Jakus, "Valuing the Private and Public Dimensions of a Mixed Good: An Application to Pest Control," Ph.D. dissertation, North Carolina State University (1992).
21. D. Kahneman, and J. Knetsch, Valuing public goods: the purchase of moral satisfaction, *J. Environ. Econom. Management*. **22**, 55–70 (1992).
22. D. Kahneman, and J. Knetsch, Reply: Contingent valuation and the value of public goods, *J. Environ. Econom. Management* **22**, 90–94 (1992)
23. M. A. Kemp, and C. Maxwell, Exploring a budget context for contingent valuation estimates, *in* "Contingent Valuation: A Critical Assessment" (J. A. Hausman, Ed.), North-Holland, Amsterdam (1993).

24. D. J. Krieger, "Economic Value of Environmental Risk Information: Theory and Application to Michigan Sport Fisher," Ph.D. Dissertation, Michigan State University (1994).
25. J. M. Lockwood Loomis, and T. DeLacy, Some empirical evidence on embedding effects in contingent valuation of forest protection, *J. Environ. Econom. Management*. **25**, 45–55 (1993).
26. K. Magnussen, Valuation of reduced water pollution using the contingent valuation method: testing for amenity misspecification, *in* "Pricing the European Environment" (S. Navrud, Ed.), Oxford Univ. Press, New York (1992).
27. R. C. Mitchell, and R. T.Carson, "Valuing Drinking Water Risk Reductions using the Contingent Valuation Method: A Methodological Study of Risks from THM and Giardia," Report to the U.S. Environmental Protection Agency, Washington, D. C. (1986).
28. R. C. Mitchell, and R. T. Carson, "Using Surveys to Value Public Goods: The Contingent Valuation Method," Resources for the Future, Washington, DC (1989).
29. R. C. Mitchell, R. T. Carson, and P. A. Ruud, "Cincinnati Visibility Study: Pilot Study Findings," Report to the Electric Power Research Institute, Palo Alto, CA (1989).
30. W. Nelson, "Applied Life Analysis," Wiley, New York (1982).
31. A. Randall, and J. P. Hoehn, "Embedding Effects in Contingent Valuation: Implications for Natural Resource Damage Assessment," Staff paper 92-14, Department of Agricultural Economics, Michigan State University (1992).
32. A. U. Römer, and W. W. Pommerehne, Valuing reductions of public risks: the case of hazardous waste, *in* "Models and Measurement of Welfare and Inequality" (P. Eichhorn, Ed.), Springer, Heidelberg, Germany, in press.
33. D. A. Schkade, and J. W. Payne, Where do the numbers come from?: How people respond to contingent valuation questions, *in* "Contingent Valuation: A Critical Assessment" (J. A. Hausman, Ed.), North-Holland, Amsterdam (1993).
34. V. K. Smith, Comment: Arbitrary values, good causes, and premature verdicts, *J. Environ. Econom. Management*. **22**, 71–89 (1992).
.35. G. Tolley, M. Brien and R. Fabian, Refinement of visibility value function for seasonality and distribution of days, *in* "The Economic Value of Visibility" (G. Tolley and R. Fabian, Eds.), Blackstone, Mount Pleasant, MI (1988).
36. B. W. Turnbull, The empirical distribution function with arbitrarily grouped, censored, and truncated data, *J. Roy. Statist. Soc. Ser. B* **38**, 290–295 (1976).
37. F. A. Ward and J. Fiore, "A Model of the Welfare Effects of Changes in Quality: New Mexico Fishing," Paper presented at the W-133 Conference, Monterey, CA (1991).
38. J. C. Whitehead, Total economic values for coastal and marine wildlife: Specification, validity and valuation issues, *Mar. Resour. Econom.* **8**, 119–132 (1993).
39. J. C. Whitehead, and G. C. Blomquiust, Measuring contingent values for wetlands: Effects of information about related environmental goods, *Water Resour. Res.* **27**, 2523–2531 (1991).
40. P. I. Wu, "Benefit evaluation of Multidimensional Environmental Policy: An Application to Ohio's Big Darby Creek," Ph.D. dissertation, Ohio State University (1991).