JOURNAL OF ENVIRONMENTAL ECONOMICS AND MANAGEMENT 22, 57–70 (1992)

Valuing Public Goods: The Purchase of Moral Satisfaction*

DANIEL KAHNEMAN[†] AND JACK L. KNETSCH[‡]

[†] *Department of Psychology, University of California, Berkeley, Berkeley, California 94720; and*
[‡]*School of Resource and Enviromental Management, Simon Fraser University, Burnaby, British Columbia, Canada V5A 1S6*

Received March 2, 1990; revised December 14, 1990

Contingent valuation surveys in which respondents state their willingness to pay (WTP) for public goods are coming into use in cost–benefit analyses and in litigation over environmental losses. The validity of the method is brought into question by several experimental observations. An embedding effect is demonstrated, in which WTP for a good varies depending on whether it is evaluated on its own or as part of a more inclusive category. The ordering of various public issues by WTP is predicted with significant accuracy by independent ratings of the moral satisfaction associated with contributions to these causes. Contingent valuation responses reflect the willingness to pay for the moral satisfaction of contributing to public goods, not the economic value of these goods. © 1992 Academic Press, Inc.

There is substantial demand for a practical technique for measuring the value of non-market goods. Measures of value are required for cost–benefit assessments of public goods, for the analysis of policies that affect the environment, and for realistic estimates of environmental damages resulting from human action, such as oil spills. In recent years the contingent valuation method (CVM) has gained prominence as the major technique for the assessment of the value of environmental amenities. This paper is concerned with a critique of CVM.

The idea of CVM is quite simple: respondents are asked to indicate their value for a public good, usually by specifying the maximum amount they would be willing to pay to obtain or to retain it. The total value of the good is estimated by multiplying the average willingness to pay (WTP) observed in the sample by the number of households in the relevant population. This value is sometimes divided into *use value* and *non-use value* by comparing the WTP of respondents who expect to enjoy the public good personally (e.g., benefit from improved visibility or from the increased number of fish in a cleaned up stream) to the WTP of respondents who have no such expectations. Specific questions are sometimes added to partition non-use value further into the value of retaining an option for future use, a bequest value, and a pure existence value [15].

The accuracy of the CVM is a matter of substantial practical import, not only in cost–benefit assessments but also in litigation over liability and damages. The validity of the technique is take as a rebuttable presumption in environmental cases brought in the United States under the Comprehensive Environmental Response, Compensation and Liability Act of 1980 (CERCLA). The research on the method has been reviewed in two authoritative volumes, which offer detailed

*This research was supported by Fisheries and Oceans Canada, the Ontario Ministry of the Environment, and the Sloan Foundation. Interviews and preliminary statistical analyses were performed by Campbell–Goodell Consultants, Vancouver, British Columbia. We benefited from conversations with George Akerlof, James Bieke, Brian Binger, Ralph d'Arge, Elizabeth Hoffman, Richard Thaler, and Frances van Loo, from a commentary by Glenn Harrison, and from the statistical expertise of Carol Nickerson.

0095-0696/92 $3.00
Copyright © 1992 by Academic Press, Inc.
All rights of reproduction in any form reserved.

guidelines for its use ([8, 17]; see also [11]). Some assessments of CVM have been very favorable, as illustrated by the claims that "the necessary structure for constructing a hypothetical market for the direct determination of economic values within the Hicksian consumers' surplus framework has been developed" [6, p. 173] and that contingent valuation "is potentially capable of directly measuring a broad range of economic benefits for a wide range of goods, including those not yet supplied, in a manner consistent with economic theory" [17, p. 295]. Acceptance of the technique is not universal, however, and some strong reservations about the adequacy of CVM to support specific compensation claims have recently been expressed [7, 9, 19].

The present article reports an experimental investigation of what is perhaps the most serious shortcoming of CVM: that the assessed value of a public good is demonstrably arbitrary, because willingness to pay for the same good can vary over a wide range depending on whether the good is assessed on its own or embedded as part of a more inclusive package. We next provide evidence for a similar difficulty in the response to payment schedules: WTP estimates can be much larger when the payment is described as a long-term commitment rather than as a one-time outlay. Another study suggests that WTP for public goods is best interpreted as the purchase of moral satisfaction, rather than as a measure of the value associated with a particular public good. Lastly, we examine the categories of expenditures from which contributions to public goods are drawn.

PART I. THE EMBEDDING EFFECT

The standard interpretation of CVM results is that the WTP for a good is a measure of the economic value associated with that good, which is fully comparable to values derived from market exchanges and on the basis of which allocative efficiency judgments can be made. However, two related observations that cast doubt on this interpretation have been discussed in the CVM literature. The first is an order effect in WTP responses when the values of several goods are elicited in succession: the same good elicits a higher WTP if it is first in the list rather than valued after others. For example, Tolley and Randall [22] found that estimates of the value of improved visibility in the Grand Canyon differed by a factor of three depending on whether this item appeared first or third in a survey. Because the order in which goods are mentioned in a survey is purely arbitrary, any effect of this variable raises questions about the validity of responses.

Another problem for CVM is an effect that we call *embedding*, also variously labeled as a part–whole effect, symbolic effect, or disaggregation effect [8, 17]: the same good is assigned a lower value if WTP for it is inferred from WTP for a more inclusive good rather than if the particular good is evaluated on its own. A finding that we obtained some time ago illustrates the embedding effect: the expressed willingness of Toronto residents to pay increased taxes to prevent the drop in fish populations in all Ontario lakes was only slightly higher than the willingness to pay to preserve the fish stocks in only a small area of the province (reported in [14]). It is quite unlikely that the respondents in Toronto viewed saving fish in the Muskoka area as a fully adequate substitute for saving fish in the whole province. The

similar WTP observed for separate regions and for the whole of Ontario therefore appears anomalous. Further, the result raises a question about the proper assessment of WTP for a particular region: should this be estimated by the WTP assessment for that region in isolation, or by allocating to it a share of the sum offered for the cleanup of lakes in the entire province?

The embedding problem was noted long ago by investigators who were concerned with the appropriateness of aggregating WTP for several commodities obtained from different samples into an estimate of WTP for the package ([13], as summarized in [17, pp. 44–46]). Schulze *et al.* stated that "*no* researcher would be willing to defend the summation of CV values that have been obtained in various studies for *many* types of environmental effects; indeed the summation of average CV values for public goods thus far available in the literature would *exhaust* the budget of the average individual" [21, p. 6; emphases in the original].

The effects of order and embedding observed in assessments of value for public goods are difficult to reconcile with standard value theory. To appreciate why this is so, it is useful to consider the conditions under which assessments of the value of private goods would exhibit these effects. Two generic cases can be identified. The first involves goods which are perfect substitutes for one another and for which satiation is attained by the consumption of one unit. Thus, in the absence of opportunities for storage, resale, or altruistic giving, most adults will have zero WTP for a second large ice cream cone offered immediately after the consumption of the first. This is an order effect—the positive value of consuming an initial ice cream cone could be associated with any of those potentially available, but the value of any cone considered immediately thereafter would be zero. The value of ice cream cones under these circumstances also exhibits an embedding effect: WTP for 100 ice cream cones will not be higher than WTP for 1, much as WTP for improved fishing in all of Ontario was little more than WTP for fishing in a small area of that province. Although the notions of substitution and satiation may apply to some environmental goods, they do not readily extend to existence values for beautiful sites, historical landmarks, or endangered species. If it is found that WTP to save all threatened historical landmarks in a region is not much higher than WTP to save any single landmark, this can hardly be because each individual landmark provides as much utility as the whole set. Indeed, the uniqueness of the valued goods is the essence of existence value, as this notion has been discussed since Krutilla [15].

Effects of order and embedding are also expected in another extreme case: goods for which people are willing to pay a large part of their wealth. For example, the sum that an individual will pay to avoid the loss of both an arm and a leg is likely to be much less than the sum of WTP to save each limb separately, because the amount the person is willing to pay to save one limb is almost certain to be high in relation to available wealth, leaving little to prevent the second loss. In this case, order and embedding effects are produced by limited wealth. However, median WTP in CVM studies commonly falls in the range of $40–100 [9], far too small to be severely restrained by wealth.

The problem for the interpretation of CVM results is the following: if the value of a given landmark is much larger when it is evaluated on its own that when it is evaluated as part of a more inclusive package of public goods, which measure is the correct one? The discussions of the problem in the literature provide no agreed principles that would define the proper level of aggregation for the

evaluation of a specific good. In the absence of such principles, the results of CVM become arbitrary. This criticism could be fatal. No measuring instrument can be taken seriously if its permitted range of applications yields drastically different measures of the same object.

Embedding a Public Good

Our first study was conducted to document the embedding effect in a controlled experimental design, focusing on the valuation of a public good that is of personal relevance to respondents: the increased availability of equipment and trained personnel for rescue operations in disasters. Coincidentally, the study was conducted within weeks of the San Francisco earthquake of 1989, a fact that certainly enhanced the relevance of the topic.

Three samples of adults living in the greater Vancouver region in Canada were interviewed by telephone. Samples were evenly split by gender. All calls were made in evening hours. The interviewers introduced themselves as being from a professional polling firm "conducting interviews on behalf of researchers at Simon Fraser University." All respondents were initially told:

> The federal and provincial governments provide a wide range of public services that include education, health, police protection, roads, and environmental services.

Respondents in one sample were then told to focus on environmental services, which were described as including "preserving wilderness areas, protecting wildlife, providing parks, preparing for disasters, controlling air pollution, insuring water quality, and routine treatment and disposal of industrial wastes." They were then asked the following question:

> If you could be sure that extra money collected would lead to significant improvements, what is the most you would be willing to pay each year through higher taxes, prices, or user fees, to go into a special fund to improve environmental services?

The evaluation questions were concluded at this point if the respondent's answer was zero. Other respondents were then asked:

> Keeping in mind the services just mentioned, including those related to providing parks, pollution control, preservation of wilderness and wildlife, and disposal of industrial wastes, I would like to ask you in particular about improved preparedness for disasters. What part of the total amount that you just mentioned for all environmental services do you think should go specifically to improve preparedness for disasters?

Subjects were allowed to answer by stating a dollar amount, a fraction, or a percentage. Where necessary, the interviewer immediately computed the dollar amount of the offered contribution and recorded that value. A third question was asked after some aspects of preparedness for disasters were listed (emergency services in hospitals; maintenance of large stocks of medical supplies, food, fuel, and communication equipment; ensuring the availability of equipment and trained personnel for rescue operations; and preparing for cleanup of oil, toxic chemicals, or radioactive materials):

> Keeping in mind all aspects of preparedness for disasters, what part of the total amount you allocated to improving preparedness do you think should go specifically to improve the availability of equipment and trained personnel for rescue operations?

TABLE I

Willingness to Pay for Selected Classes of Goods and Allocations of Totals to Less Inclusive Groups

Public good		Sub-sample Group 1 (N = 66) ($)	Group 2 (N = 78) ($)	Group 3 (N = 74) ($)
Environmental services	Mean	135.91		
	Median	50.00		
Improve disaster preparedness	Mean	29.06	151.60	
	Median	10.00	50.00	
Improve rescue equipment, personnel	Mean	14.12*	74.65**	122.64
	Median	1.00	16.00	25.00

*Two respondents did not answer this question, reducing N to 64.
**Four respondents did not answer this question, reducing N to 74.

The same procedure was followed with the second sample, except that the initial question they answered referred to "a special fund to improve preparedness for disasters," with a subsequent allocation to "go specifically to improve the availability of equipment and trained personnel for rescue operations." Respondents in the third sample were told to focus on preparedness for disasters and were asked to state their willingness to pay "into a special fund to improve availability of equipment and trained personnel for rescue operations."

Table I presents the medians and means of the willingness to pay responses for each of the questions in the three surveys. Zero responses are included in the calculations; respondents who stated a zero response to the initial question were assigned zero responses to subsequent allocation questions.[1] As in other applications of CVM, the data included extremely high responses, in some cases up to 25% of reported household incomes, which probably reflect a misunderstanding of instructions. These responses have considerable effect on the means of WTP, but there is no agreed way to draw a line beyond which responses will be rejected. To avoid this problem, our analyses of WTP results are based on medians, using all responses. The qualitative conclusions are unaffected by this choice of statistics.

The WTP for the public good mentioned in the first question posed to respondents was hardly affected by the inclusiveness of this good. The percentages of positive contributions were 61% for improvements in "the availability of equipment and trained personnel for rescue operations," 63% for "preparedness for disasters," and 65% for "all environmental services." The median WTP was $25 at the lowest level of inclusiveness and $50 at the two higher levels, but the difference

[1]Glenn Harrison has suggested that this procedure could bias the results, because of the theoretical possibility that a respondent might be willing to pay for a good but not for a bundle that includes it. Note that this objection can be eliminated by informing respondents in advance that they will have an opportunity to allocate each contribution to an inclusive good among its separate constituents. It seems highly implausible that this minor procedural change would significantly alter results. The reasons for refusals to contribute in CVM surveys are commonly quite general (rejection of responsibility, opposition to extra taxes, etc.) and therefore likely to apply to the constituents as well as to more inclusive goods.

was not significant in a Mann–Whitney test in which the two higher levels were combined. The means of WTP across levels of embedding were also very close, and the differences among them did not approach statistical significance by F test. The pattern is the same as that which we observed in our previous study, in which WTP to preserve fish in all Ontario lakes was only slightly higher than WTP to maintain but a particular few.

The bottom rows of Table I display the effect of position in the embedding structure on stated WTP for a particular good. The median amounts allocated to "equipment and trained personnel" vary from $25 when that good is evaluated on its own to $1.50 when the initial question concerns WTP for "environmental services." The three values shown in the last row of the table differ markedly from each other in the predicted direction. Group 1 differs significantly from Groups 2 and 3, by both parametric and nonparametric tests. The difference between Groups 2 and 3 only approaches significance.[2]

As in other studies, WTP values were small relative to reported incomes. Pooling over the three samples, the median WTP stated in response to the first question was $37.50 for respondents stating a family income under $20,000 (23% of total sample), $50 for income between $20,000 and $40,000 (39% of total), and $100 for families with incomes in excess of $40,000 (38% of total). The corresponding means were $97, $131, and $230. Clearly, these values are not in the range in which the embedding effect could be explained by constraints of wealth or income.

The results of this study demonstrate a large embedding effect. The key finding is that WTP is approximately constant for public goods that differ greatly in inclusiveness. The inevitable consequence of the insensitivity of WTP to inclusiveness is that estimates of WTP for the same particular good differ—by a factor of 16 for medians or 8 for means—depending on the scope of the initial question. An even larger embedding effect could probably be obtained by asking respondents to make explicit allocations to all the sub-categories at each level of embedding: the procedure of the present study, in which respondents make an allocation to a single subordinate good, appears likely to enhance its importance.

The specific good evaluated in the present study is fairly well defined, and the answer is interpretable as a quantity choice: how much extra equipment and personnel would you be willing to pay for? The good also has personal use value for most respondents, because improved availability of equipment and personnel for rescue operations would contribute to their safety and that of their families. The findings therefore extend the evidence for embedding: unlike demonstrations of embedding for existence value, the present results cannot be explained by invoking a concept of symbolic response [14; 17, p. 250].

Temporal Embedding of Payments

The embedding effects discussed so far apply to the specification of the good that is to be acquired. A related effect can arise in the specification of the schedule of payments. The question is whether respondents in a CVM survey are likely to make the appropriate discriminations between a one-time payment and a long-term commitment to a series of payments for a good. The issue is of some importance to the practical implementation of CVM. For example, willingness to pay was assessed in one study by asking people to state the amount that they would be willing to pay annually for 10 years in order to acquire a good, and conventional discount factors were applied to obtain an estimate of the present value of WTP [20].

[2]An earlier draft mistakenly stated that all three groups differed significantly from one another.

The issue of whether the participants in a CVM survey actually perform the discount calculations that are imputed to them has not been systematically examined, to our knowledge. The observations of embedding effects in which respondents did not discriminate between goods that vary in inclusiveness suggested the hypothesis that a similar failure of discrimination could be found between payments that vary in temporal inclusiveness. A small study was carried out to examine this hypothesis.

After completion of the main part of the interview, participants in the three groups of the first study were asked one of two versions of the following question. The version presented to each individual was selected at random:

> Now we would like to ask you how much you would be willing to pay (as a one-time payment/every year for a period of five years) to a fund to be used exclusively for a toxic waste treatment facility that would safely take care of all chemical and other toxic wastes in British Columbia.

Median WTP was $20, both in the group that considered a one-time payment ($N = 106$) and in the group that considered a five-year commitment ($N = 100$). The corresponding means were $141 and $81. The difference between the means was produced almost entirely by a few extremely high responses in the one-year group. There were five responses stating WTP of $1000 or more in that group, averaging $1800. There were only two such responses in the five-year group, averaging $1300. These extreme responses contribute approximately $54 to the difference of $60 between the means of the two groups. The results provide no reliable indication that the respondents discriminated between payment schedules that differed greatly in total present value.

The Significance of Embedding

What can be learned from these demonstrations of embedding? It may be useful to state the obvious qualification, that the present results have not established that insensitivity to inclusiveness is a *universal* characteristic of the valuations elicited in CVM studies. No single study could do so. The conservative conclusion from our findings is that future applications of CVM should incorporate an experimental control: the contingent valuation of any public good should routinely be supported by adequate evidence that the estimate is robust to manipulations of embedding, both in the definition of the good and in the specification of the number of payments. Whether this challenge can be met by appropriate CVM techniques is a question that will likely be the subject of further research.

Another defense of CVM against the embedding problem should be mentioned: the observation that different embeddings lead to different valuations of the same good would not be as troubling if there were a way of selecting one of these valuations as the correct one. As noted earlier, however, we were unable to identify in the existing CVM literature any compelling principles that could guide the choice of the appropriate embedding level for the good to be valued, or of a duration for the schedule of payments. Indeed, it is far from obvious that such principles can be found. Should the value of the damaged Alaska shoreline be assessed by WTP to clean up the damage done to it, by aggregating separate estimates of WTP to clean up parts of it, or by allocating to the cleanup a fraction of total WTP for environmental improvements? In the absence of agreed answers

to such questions, our results suggest that current standards for the use of CVM may allow estimates of the value of a good that differ by more than an order of magnitude, all with an a priori equal claim to validity. As illustrated by the two examples reported here, the designer of a contingent valuation survey may be able to determine the estimated value of any good by the choice of a level of embedding. This potential for manipulation severely undermines the contingent valuation method.

PART II. THE PURCHASE OF MORAL SATISFACTION

The results presented in Part I do not support the interpretation of WTP for a public good as a measure of the economic value of this good. It remains a fact, however, that respondents express a willingness to contribute for the acquisition of many public goods, and there is no reason to doubt their sincerity or seriousness. Indeed, some elegant experiments have confirmed the willingness of people to pay for existence value—subjects actually paid to prevent a plant from being destroyed [5]. What is the good that respondents are willing to pay to acquire in such experiments or in CVM surveys? We offer the general hypothesis that responses to the CVM question express a willingness to acquire a sense of moral satisfaction (also known as a "warm glow of giving"; see [1, 2]) by a voluntary contribution to the provision of a public good. In attaining this satisfaction, the public good is a means to an end—the consumption is the sense of moral satisfaction associated with the contribution. An interesting feature of the warm glow of moral satisfaction is that it increases with the size of the contribution; for this unusual good, the expenditure is an essential aspect of consumption [18]. The interpretation of the responses to the hypothetical questions used in CVM in terms of moral satisfaction is consistent with Andreoni's economic analyses of actual donations to public goods, both in the field [1, 2] and in experimental situations [3], which distinguish the utility derived from increasing the total supply of the good from the utility gained in the act of giving.

Public goods differ in the degree of moral satisfaction that they provide to the individual making a contribution. Saving the panda may well be more satisfying for most people than saving an endangered insect and cancer research may be a better cause than research on gum disease. The quality of causes as sources of moral satisfaction will reflect individual tastes and community values. Our first hypothesis is that differences in WTP for various causes can be predicted from independent assessments of the moral satisfaction associated with these causes.

The results of Part I can be explained by invoking the additional hypothesis that moral satisfaction exhibits an embedding effect: the moral satisfaction associated with contributions to an inclusive cause extends with little loss to any significant subset of that cause. A closely related idea is that people may be willing to "dump their good cause account" on any valued cause [7]. Thus, contributing to the provision of rescue equipment may be as satisfying as contributing to the more inclusive cause of disaster preparedness. Indeed, a narrowly defined cause can be even more satisfying than a cause that includes it: it could be the case, for example, that saving the panda is more appealing than saving endangered species. Different subsets of a cause may vary in their appeal. In general, however, moral satisfaction

could be expected to be about the same for an inclusive cause and for representative subsets of it.

An experiment was conducted to test these hypotheses. For the purpose of the experiment, a set of 14 pairs of public goods was constructed (see Table II). Each pair consisted of two causes, one of which was embedded in the other. The items were chosen to include two types of embedding: geographical embedding (e.g.,

TABLE II

Maximum Willingness to Pay for Various Causes and Ratings of Satisfaction from Making Contribution

Cause	Satisfaction Mean	WTP ($) Median	WTP ($) Mean
Reduce acid rain damage in Muskoka, Ont.	7.18	20	40.91
Reduce acid rain damage in eastern Canada	7.25	50	214.55
Restore rural B.C. museums	4.67	10	32.78
Restore rural Canada museums and heritage buildings	5.79	20	113.47
Improve sport fish stocks in B.C. fresh water	5.25	10	41.89
Improve sport fish stocks in Canada fresh water	6.61	10	147.16
Protection for marmot, a small animal in B.C.	5.48	1.5	33.27
Protection for small animals in B.C.	6.42	10	141.75
Research on dengue fever, a tropical disease	4.57	0	52.42
Research on tropical diseases	4.97	4	17.83
Protection Peregrine falcon, an endangered bird	6.46	25	125.00
Protect endangered birds	6.98	20	59.07
Improve sport facilities in small communities in B.C.	6.22	10	209.75
Improve sport facilities in small communities in Canada	5.42	10	55.96
Rehabilitate recently released young offenders	5.78	50	233.16
Rehabilitate all recently released criminals	4.97	0	25.04
Habitat for muskrats, wild N. American rodent	4.70	0	51.60
Habitat for muskrats, squirrels, and other wild N. American rodents	4.59	4	52.28
Improve literacy of recent adult B.C. immigrants	6.30	10	190.53
Improve literacy of adults in B.C.	7.10	10	56.61
Replant trees in cutover areas in B.C.	7.80	20	151.70
Replant trees in cutover areas in western Canada	7.53	20	54.74
Increase research on toxic waste disposal	7.87	50	234.12
Increase research on environmental protection	7.44	50	98.77
Famine relief in Ethiopia	6.38	20	157.67
Famine relief in Africa	5.57	25	72.68
Research on breast cancer	8.12	50	243.14
Research on all forms of cancer	8.38	50	162.09

famine relief in Ethiopia or in Africa) and categorical embedding (e.g., research on breast cancer or research on all forms of cancer). The 14 issues were arbitrarily divided into two sets, labeled A and B, respectively, including the first 4 and the last 10 issues in Table II.

A special telephone survey of adult residents of the Vancouver region was conducted, with respondents randomly assigned into four groups. Respondents in groups 1 ($N = 60$) and 2 ($N = 61$) judged the moral satisfaction associated with the various causes (group 1 judged the inclusive items of set A and the embedded items of set B; group 2 judged the remaining items). After an introduction similar to that used in the surveys described in Part I and an indication that the questions were about "various causes to which people might be willing to make voluntary contributions," the instructions given to these two groups were:

> Please consider each of the causes separately and independently; that is, assume you are only being asked about the one cause. Indicate the degree of satisfaction you would receive from contributing to each cause on a scale from 0 to 10, with 0 indicating no satisfaction at all and 10 indicating a great deal of personal satisfaction.

Groups 3 ($N = 61$) and 4 ($N = 60$) were matched respectively to groups 1 and 2, but they provided measures of WTP for the same sets of causes. After the same introduction and indication of what the questions were about, these respondents received the following instructions:

> Please consider each of the causes separately and independently; that is, assume you are only being asked about the one cause. Indicate the most that you would be willing to pay for each.

The order in which the causes were presented was randomly determined separately for each respondent.

Table II presents the mean ratings of moral satisfaction and the medians and means of WTP for the 28 public goods included in the study, arranged in pairs. In each case, the embedded good is the first member of the pair. There is as usual a large discrepancy between mean and median WTP, due to extremely large WTP reported by a few individuals in each group. As before, we chose to focus on medians, without discarding any responses. The moral satisfaction ratings, which were made on a bounded scale, are not susceptible to large effects of a few aberrant responses, and the means of the satisfaction ratings were accordingly used in the analysis.

The hypothesis that WTP is predictable from assessments of moral satisfaction was tested by ranking the 14 issues evaluated by each group and by comparing the ranking of these issues by WTP and by moral satisfaction. The rank correlations between the means of satisfaction ratings made by group 1 and the median WTP of group 3 was 0.78 ($p < 0.01$). The corresponding correlation between the responses of groups 2 and 4 was 0.62 ($p < 0.02$). The general hypothesis that WTP can be predicted by ratings of moral satisfaction is strongly supported. As may be seen in Table II, there was only one striking discrepancy in the rankings of issues by WTP and by moral satisfaction: the rehabilitation of young offenders was one of the four causes eliciting the largest monetary contributions, but it ranked very low as a source of moral satisfaction. The discrepancy was not predicted, and any account of it must be speculative. One hypothesis is that the illegitimate context in which the need for public contributions arises makes it difficult to describe these contributions as yielding any kind of satisfaction—including moral satisfaction.

The data of Table II also allow a test of the effects of inclusiveness (embedding) on both moral satisfaction and WTP. The more inclusive causes have a very slight advantage on both dimensions overall, but the effect is weak and inconsistent: the more inclusive cause is associated with a higher rating of moral satisfaction for 8 of the 14 pairs of causes and with a lower rating for the other 6. Median WTP is higher for the inclusive than for the embedded cause in 6 pairs, identical in 6 others, and inferior in the remaining 2 pairs. On the other hand, mean WTP is higher for the embedded cause in 9 of the 14 pairs.

The results of this experiment support the proposed interpretation of willingness to pay for public goods as an expression of willingness to pay to acquire moral satisfaction. With only one salient exception, causes that were judged to provide little moral satisfaction also elicited relatively low WTP. Overall, there was a close correspondence between the rankings of issues by the two measures. Furthermore, the interpretation of WTP as an index of moral satisfaction helps explain the embedding effect: if the inclusiveness of the cause does little to enhance moral satisfaction, increasing inclusiveness should have little effect on WTP, as was indeed observed both in this study and in Part I.

The Sources of Contributions to Public Goods

The question posed to respondents in a CVM survey is an unusual one, which has some features of a market survey, an opinion poll, and an appeal on behalf of a new charity. Respondents who follow instructions consider the possibility of a significant financial commitment to the provision of a public good. If they are serious about it, such a commitment to a new expenditure entails a corresponding reduction in other categories of spending. To understand the decisions of respondents in CVM surveys it is useful to identify the categories of spending from which they would expect to draw their contributions.

In the budget of most households there already exists a category of spending that is dedicated to obtaining moral satisfaction—voluntary contributions to charity. Spending on charity is far from negligible. For example, in fiscal 1988 donations by individuals in the United States totaled $86 billion, approximately $350 per capita. The respondent in a CVM study is likely to consider a new contribution in the context of the existing pattern of voluntary donations by the household. It is of interest to find out whether respondents view the proposed payment as a substitute to current charitable giving or as an addition to the moral satisfaction budget, requiring a reduction in other categories of spending. To test these possibilities, participants in the survey of Part I who had stated a positive WTP for the cause presented to them were asked a series of questions in the following format:

> Suppose you were actually called on to make the contribution to environmental services you indicated earlier. Which expenditure categories do you think this money would mainly come from? Would you spend less on _____?

The expenditure categories mentioned in the questions included food, charities, holidays (vacations), entertainment, savings, and "other things." After answering this series of questions, respondents who had listed more than one category were asked: "From which category do you think most of the money would come?" and the relevant list of categories was repeated to them. Table III presents the results.

TABLE III

Percentage of Respondents Indicating Reduced Spending in Various Expenditure Categories (N = 137)

Category	"Reduction"	"Greatest reduction"
Food	19.1	2.3
Charities	34.6	9.3
Holidays	64.2	15.5
Entertainment	76.1	41.1
Savings	46.2	22.5
Other things	65.9	9.3

The results indicate that added spending on environmental and disaster services would be drawn from discretionary spending, and especially from entertainment. Respondents would not expect to alter their eating habits. More important, they would not withdraw the contribution from current charitable giving. Most respondents apparently viewed the contribution as an addition to the "good cause" budget, not as a substitute for existing items in that spending category. The observed pattern of responses is the same as would be expected in answers to the question: "If you made an *extra* contribution to charity, where would it come from?" In the terms of the present analysis, the respondents appear to have considered the contingent valuation question as an opportunity to acquire additional moral satisfaction. Note that if households contribute only to causes that yield high moral satisfaction, the only way to increase the consumption of this good is by increasing contributions—to the currently favored causes or to equally satisfying causes.

CONCLUDING REMARKS

The research reported here had two objectives: to examine the proposition that CVM results are susceptible to an embedding effect that could render them largely arbitrary and consequently useless for practical purposes and to advance the interpretation of what people do in answering CVM questions.

The central result of the first study was that willingness to pay was almost the same for a narrowly defined good (rescue equipment and personnel) and for vastly more inclusive categories (all disaster preparedness, or even all environmental services). Correspondingly, the value assigned to the more specific good varied by an order of magnitude depending on the depth of its embedding in the category for which WTP was initially assessed. This result appears to invalidate a basic assumption of CVM: that standard value theory applies to the measures obtained by this method. As the choice of embedding structure is arbitrary, the estimates of value obtained from CVM surveys will be correspondingly arbitrary.

Our assessment of the validity of the CVM is in marked contrast to that reached by Mitchell and Carson [17] in their comprehensive review of the literature on this method. Mitchell and Carson recognized the potential severity of the embedding effect, but sounded a hopeful note in their discussion of it (p. 250), arguing that such an effect is not inevitable. The evidence cited for this conclusion was the

observation that WTP to improve water quality nationwide, assessed in a national survey [16], was approximately twice as high as an estimate of WTP to raise the quality of water in the Monongahela River system in Pennsylvania, assessed in a local survey conducted by other investigators using a similar instrument [10]. Given the uncertainties of comparisons across studies and sampling areas, this evidence against embedding is not persuasive.

There was some prior reason to hope that the embedding effect might be restricted to non-use values, but the present results show that estimates of use value are not generally immune to embedding effects: an essentially complete embedding effect was obtained for disaster preparedness, a public good for which respondents have use value. Our tentative conclusion is that the factor that controls the magnitude of the embedding effect is not the distinction between public goods that have use value and those that only have non-use value. A more important distinction could be between public goods for which private purchase is conceivable and other goods for which it is not. Access to clean air and the right to fish in a stream could be privately purchased in a market, and sometimes are. The respondents in contingent valuation surveys have some experience in the purchase of such goods and could rely on this experience to determine their willingness to pay [6]. On the other hand, few respondents have experience in individual purchases of improvements in disaster preparedness, air traffic control, maintenance of species, or expansion of parks. The only way to procure such goods is by concerted public action, and the decision to make a voluntary contribution to such action has more in common with charity than with the purchase of consumption goods. Note that we do not assert that the CVM is necessarily valid for public goods that could be purchased by individuals. Our point is that the purchase of moral satisfaction is especially plausible as an interpretation of WTP for goods that could not be so purchased, even when these goods have use value.[3]

Students of CVM have long known that the respondents' answers to questions about their willingness to accept compensation for the loss of public goods (WTA) are strongly affected by moral considerations. Participants are prone to respond with indignation to questions about the compensation they would require to accept pollution of the Grand Canyon National Park, or of an unspoiled beach in a remote region. The indignation is expressed by the rejection of the offered transaction as illegitimate, or by absurdly high bids. The practitioners of contingent valuation hoped to avoid the difficulties of assessing WTA by substituting WTP even where WTA is the theoretically appropriate [8, 12]. The present results suggest that the adoption of the WTP measure does not really avoid moral concerns because the voluntary contribution to the provision of such goods can be morally satisfying. A treatment that interprets contributions to public goods as equivalent to purchases of consumption goods is inadequate when moral satisfaction is an important part of the welfare gain from the contribution [2]. The amount that individuals are willing to pay to acquire moral satisfaction should not be mistaken for a measure of the economic value of public goods.

[3]The application of CVM to goods such as hunting licenses in limited supply [4] is perhaps best viewed as a special case of market research, because these goods are in all essential respects conventional private goods.

REFERENCES

1. J. Andreoni, Giving with impure altruism: Applications to charity and ricardian equivalence, *J. Polit. Econom.* **97**, 1447–1458 (1989).
2. J. Andreoni, Impure altruism and donations to public goods: A theory of warm-glow giving, *Econom. J.*, 100 (June 1990).
3. J. Andreoni, "An experimental test of the public goods crowding-out hypothesis," Working paper, University of Wisconsin (1990).
4. R. C. Bishop and T. A. Heberlein, Measuring values of extramarket goods: Are indirect measures biased? *Amer. J. Agr. Econom.* **61**, 926–930 (1979).
5. R. R. Boyce, T. C. Brown, G. D. McClelland, G. L. Peterson, and W. D. Schulze, "An Experimental Examination of Intrinsic Environmental Values," Working Paper, University of Colorado (1990).
6. D. S. Brookshire, M. A. Thayer, W. D. Schulze, and R. C. d'Arge, Valuing public goods: A comparison of survey and hedonic approaches, *Amer. Econom. Rev.* **72**, 165–176 (1982).
7. R. G. Cummings, Letter to Office of Environmental Project Review, Department of the Interior, November 10, 1989.
8. R. G. Cummings, D. S. Brookshire, and W. D. Schulze (Eds.), "Valuing Environmental Goods: An Assessment of the Contingent Valuation Method," Rowman and Allanheld, Totawa, NJ (1986).
9. R. C. d'Arge, A practical guide to economic valuation of the environment, *in* "Thirty-fourth Annual Rocky Mountain Mineral Law Institute Proceedings," Matthew Bendier and Co., New York (1989).
10. W. H. Desvousges, V. K. Smith, and M. P. McGivney, "A Comparison of Alternative Approaches for Estimating Recreation and Related Benefits of Water Quality Improvements," Report to the U.S. Environmental Protection Agency, Washington, DC (1983).
11. B. Fischhoff and L. Furby, Measuring values: A conceptual framework for interpreting transactions with special reference to contingent valuation of visibility, *J. Risk Uncertainty* **1**, 147–184 (1988).
12. A. M. Freeman, "The Benefits of Environmental Improvement: Theory and Practice," Johns Hopkins Press, Baltimore (1979).
13. J. P. Hoehn and A. Randall, "Aggregation and Disaggregation of Program Benefits in a Complex Policy Environment: A Theoretical Framework and Critique of Estimation Methods," Paper presented at the American Agricultural Economics Association Meetings, Logan, UT (1982).
14. D. Kahneman, Comments on the contingent valuation method, *in* "Valuing Environmental Goods" (R. G. Cummings, D. S. Brookshire, and W. D. Schulze, Eds.), Rowman and Allanheld, Totawa, NJ (1986).
15. J. V. Krutilla, Conservation reconsidered, *Amer. Econom. Rev.* **57**, 787–796 (1967).
16. R. C. Mitchell and R. T. Carson, "A Contingent Valuation Estimate of National Freshwater Benefits," Report to the U.S. Environmental Protection Agency, Washington, DC (1984).
17. R. C. Mitchell and R. T. Carson, "Using Surveys to Value Public Goods: The Contingent Valuation Method," Resources For the Future, Washington, DC (1989).
18. H. Margolis, "Selfishness, Altruism and Rationality," Cambridge Univ. Press, New York (1982).
19. C. V. Phillips and R. J. Zeckhauser, Contingent valuation of damage to natural resources: How accurate? How appropriate? *Toxics Law Reporter*, 520–529 (1989).
20. R. D. Rowe, W. D. Schulze, and D. Hurd, "A Survey of Colorado Residents' Attitudes about Cleaning Up Hazardous Waste-Site Problems in Colorado," Report for the Colorado Attorney General's Office, Denver (1986).
21. W. D. Schulze, R. G. Cummings, and D. S. Brookshire, "Methods Development in Measuring Benefits of Environmental Improvements," Vol. II, Report to the U.S. Environmental Protection Agency, Washington, DC (1983).
22. George S. Tolley and A. Randall, "Establishing and Valuing the Effects of Improved Visibility in the Eastern United States," Report to the U.S. Environmental Protection Agency, Washington DC (1983).