

How Precise Are Monetary Representations of Environmental Improvements?

Robin Gregory, Sarah Lichtenstein, Thomas C. Brown, George L. Peterson, and Paul Slovic

ABSTRACT. *Subjects valued environmental goods using a response mode expressing willingness to pay as a multiplier or divider of a purported "budgetary unit." Hypothetical willingness to pay was found to be highly dependent on the size of the budgetary unit. Rejecting two other interpretations, we believe the results suggest that people's contingent values are only vaguely represented in monetary terms; any dollar amount within a broad range is considered an acceptable expression of value. Although the procedures differ from those of a typical contingent valuation survey, the results demonstrate some of the difficulties in discovering precise monetary values for environmental resources. (JEL Q21)*

I. INTRODUCTION

Economic theory defines the worth of an item as the maximum amount that an individual will pay for it or the minimum payment acceptable to relinquish it. When markets for goods exist they adequately establish economic worth for marginal changes in an item, depending on the strength of errors introduced by market imperfections and the like. When actual markets for goods do not exist, hypothetical markets can be created; individuals then can be asked the amount of money they would be willing to spend for an item if a market were created.

This is the basis of the contingent valuation (CV) approach, which has been used by economists to value environmental goods for about 30 years. The history of the method has been documented in several recent reviews, including those by Cummings, Brookshire, and Schulze (1986) and by Mitchell and Carson (1989). Environmental amenities such as recreational fishing, hiking, and habitat preservation have received particular attention because they are highly valued by society but are not usually sold through markets (Bromley 1986). CV techniques now are widely used to value a diverse group of

environmental goods whose quantity or quality might be affected by the decisions of a public agency or private developer.

The validity of CV results, however, is a subject of much controversy. Early discussions of response validity and accuracy focused on the importance of judgmental biases (e.g., Schulze, d'Arge, and Brookshire 1981). One prominent source of bias investigated by CV researchers is the so-called "anchoring and adjustment" bias: Individuals in contingent markets who start off with a low initial reference point or bid provide lower final value estimates than do participants who start off with a high initial reference amount (Boyle, Bishop, and Welsh 1985). Other reported sources of bias stem from strategic behavior on the part of respondents or from measurement and sampling errors that may affect the validity of CV results.

In their influential 1986 review of CV methods, Cummings, Brookshire, and Schulze developed the concept of a *reference level of accuracy* equal to ± 50 percent; that is, they concluded that a carefully conducted CV study should in most cases yield

Gregory, Lichtenstein, and Slovic are with Decision Research, Eugene, OR; Brown and Peterson are with U.S.D.A. Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO. The authors thank the U.S.D.A. Forest Service for its funding support, under Cooperative Agreement No. 28-C1-554, the National Science Foundation for its funding support under Grant No. SES-90-22952, and the U.S. Environmental Protection Agency's Office of Research and Development for its funding support under grant No. R81-9832-010. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the U.S.D.A. Forest Service, the National Science Foundation, or the U.S. Environmental Protection Agency.

results that are within 50 percent of the "true" value of the resource. More recent discussions of response validity include both generally optimistic results (e.g., Kealy, Montgomery, and Dovidio 1990; Loomis 1990) and disturbing findings concerning the economic basis for interpreting CV results for public goods (Kahneman and Knetsch 1992), the ability of willingness-to-pay approaches to reflect multidimensional attributes (Gregory, MacGregor, and Lichtenstein 1992), and the importance of the judgment context (Brown and Slovic 1988).

In this article we report an experiment that provides a further test of the contextual sensitivity of dollar value estimates. It also provides information about a more profound question: Are values for nonmarket environmental goods readily expressed in dollar terms, or are they often strongly held but not readily monetized?

Although the experiment is motivated by our concerns about the context sensitivity of CV responses, the procedures used in the experiment were necessarily different from those of a typical CV. For example, we asked participants to value several environmental goods, one after the other, rather than only one good, and questionnaires were self-administered and handed back to the experimenters rather than returned through the mail or completed as part of face-to-face interviews. Although our procedures were not identical to those currently recommended for CV, we believe that our results hold important implications for the interpretation of contingent valuation responses and suggest compelling hypotheses for further testing.

The basis for this study is our belief that although people have strong feelings and strongly held values for many environmental amenities that are not sold through markets (Brown 1984), these feelings and values typically are not quantified by the respondents in specific monetary terms. CV methods have the potential to help people express their strongly held values in monetary terms. However, because people are not accustomed to thinking about many environmental amenities in terms of monetary value, the dollar expressions of value may be quite

sensitive to the specific procedures used as part of the CV elicitation.

One example of the sensitivity of monetary representations of values comes from a study by Slovic, Lichtenstein, and Fischhoff (1979) that asked participants to assign a monetary loss to the seriousness of a death from a specified cause such as smoking, cancer, or alcoholism. Respondents were asked to compare a death from each cause to a standard unit of loss to society. In one condition this standard loss was \$10,000; in a second condition, with a different respondent group, the standard loss was \$1,000,000. Participants were asked to provide a multiplying or dividing factor to indicate the magnitude of the monetary value of the specific death by comparing it to the standard.

The two groups closely agreed on the relative seriousness of the causes of death. Across all 34 items tested, the correlation between the mean responses in the two conditions was 0.94. However, the magnitude of the multiplying or dividing factors for any one cause of death was nearly identical in the two groups, despite the 100-fold difference in the standard of comparison. One interpretation of these results is that the seriousness of deaths from specified causes was represented ordinally in respondents' minds (as shown by the high correlation) but was not represented cardinally (as shown by the identical factors).

This viewpoint emphasizes the constructive nature of preferences (Tversky, Sattath, and Slovic 1988) and the context dependency of CV results (Brown and Slovic 1988). It suggests that studies of reference accuracy (e.g., ± 50 percent deviation from a "true" value) and bias (i.e., systematic distortions from a "true" value) obscure a more fundamental problem: For many types of environmental resources, particularly those that may be unfamiliar or have not previously been the subject of trades, stable economic values may not exist. Instead, they are constructed at the time of the elicitation (Gregory, Lichtenstein, and Slovic 1993). If so, even minor changes in the elicitation process could produce large changes in expressed value (Payne, Bettman, and Johnson 1992). In the absence of a strong nor-

mative basis, there may be scant grounds for selecting a preferred contextual setting for any given resource evaluation situation. Thus, although dollar values for environmental resources may show a consistent ordinal ranking, the significance of specific contextual changes suggests that people may be able to translate their values into money only roughly.

II. METHOD

Subjects

The subjects were 425 people who responded to advertisements, placed in the University of Oregon student newspaper, seeking participants for several different paper-and-pencil judgment tasks. They were paid \$8.00 for the 90-minute session. Because these subjects were primarily students, their average willingness-to-pay values for each resource-improvement project are unlikely to be representative of the general population. However, we are not attempting to generalize these dollar results; instead, we want to test for differences in expressed values across the five experimental conditions (described below). As a result, the critical elements are participants' understanding of the environmental stimuli and the comparability of the five samples.

Stimuli

All participants in the experiment were asked about the dollar amount they would be willing to spend, each year, for 12 "proposed environmental projects." These 12 items are shown in Table 1. The items were selected to be familiar ones, in that each project describes both a good (e.g., roadside picnic areas, stream habitat) and a context (e.g., the state of Oregon, salmon and steelhead populations) that are widely known and common throughout the study area. As a result, extensive descriptions of the goods were not included. Some of the projects were described in very specific terms (e.g., a

TABLE 1
PROPOSED ENVIRONMENTAL PROJECTS

1. Planting daffodils in the median strip along I-5 between Salem and Eugene.
2. Protecting a 300-year-old douglas fir tree in downtown Salem.
3. Preserving the habitat of an endangered wild flower.
4. Winter feeding plan to help hungry elk in the Three Sisters Wilderness area.
5. Funding to maintain roadside picnic areas in Oregon.
6. Funding an emergency rescue team to save beached whales along the Oregon coast.
7. Manual, rather than chemical, clearing of weeds alongside rural roads in Lane County.
8. Payments to help farmers decrease smoke emissions from field burning by 75 percent.
9. 50 percent reduction in the amount of chemical pesticides used on crops in Oregon.
10. Establishing protected zones around bald eagle nesting areas.
11. Enhancing streamside habitat for native fish like salmon and steelhead.
12. Planting 100,000 trees per year to help offset CO₂ emissions.

Note: The 12 items were not presented to subjects in the order shown in Table 1. Instead, the order for all subjects was 4, 1, 9, 2, 3, 12, 5, 10, 8, 7, 11, 6.

300-year-old Douglas fir tree in downtown Salem) and others were described more generally (e.g., bald eagle nesting areas) to reflect the range of descriptions typical of CV studies. Focus group discussions were held to ensure that the final item descriptions were easily understandable.

Design

Each participant completed a six-page questionnaire consisting of instructions (shown in the Appendix), the 12 different proposed environmental projects, and a five-item computational test. The questionnaire design was based on the results of two pretests and several focus groups. The first pretest, involving 32 participants, was conducted in May 1991 to determine if the proposed environmental projects were considered to be familiar and if the proposed range of budgetary units was appropriate

for valuing this set of goods. Based on these results, changes were made in the list of environmental projects to ensure that participants would understand what they were being asked to value. This concern was the subject of in-depth discussions in two focus groups and a second pretest, involving 200 participants, that was conducted in November 1991. In addition to stimulus and payment considerations, this second pretest was designed to determine the effectiveness of a test of participants' computational abilities (described below).

In the final questionnaire of May 1992 (whose results are reported here), subjects were asked how much they would be willing to spend each year, as part of taxes and other payments made to the state of Oregon, to undertake each of the 12 activities. Further, subjects were told to consider the projects separately and one at a time (i.e., as if only that one project were under consideration) and, because payment was to be made as part of state taxes, to assume (as an approximation) that other people also would pay the same amount. Participants in the pretest reported no comprehension difficulties with either of these instructions.

There were five different versions of the questionnaire. Each used the concept of a *budgetary unit* to introduce a standard of comparison without any hint or implication that the standard related in any way to the value of the project, as the more usual "Would you pay x ?" might. The concept of a budgetary unit was discussed in the focus groups and pretest debriefings to determine its acceptance as a payment basis by respondents, and the final wording used in the questionnaire was revised in light of participants' comments.

The five versions of the questionnaire differed in the magnitude of the budgetary unit (BU) that subjects were told was "used by state legislators to plan out yearly expenditures and to decide among competing projects." The five budgetary units were set equal to \$1, \$2, \$5, \$10, and \$20. The first pretest was conducted to ensure that will-

ingness-to-pay amounts for the selected environmental projects would make sense in the context of this range of budgetary units.¹

In each version, the 12 environmental improvement activities were presented alongside four expenditure options, shown in columns. For each item, subjects were asked to choose whether they were "not willing to pay anything at all" for the activity in question (first column) or whether they were willing to pay less than one budgetary unit (second column), an amount equal to one budgetary unit (third column), or more than one budgetary unit (fourth column). If the first or third column was selected, subjects were asked to make a check mark. If the second or fourth column (less or more than one BU) was selected, subjects were asked to indicate their willingness to pay by writing down a factor by which the budgetary unit should be divided or multiplied, respectively. This format was chosen over several alternatives on the basis of pretest results and subject comments, because it allowed participants to respond quickly but, by presenting individuals with multiple columns for marking their evaluation, it also provided a visual reminder that each project required an individualized willingness-to-pay assessment.

The questionnaire ended with a five-item test that served as a check on respondents' computational abilities (i.e., "What response should you give if your willingness to pay for a proposed project is...?"). There were five forms of the test, which corresponded to the five different budgetary units.

¹ Given the pretest results on the magnitude of dollar values assigned to the 12 environmental improvement projects, as well as the reminders to participants to consider each project separately, an explicit statement of a budget constraint was not included. Further support for this position came from a concurrent CV study of fire protection policies for old-growth forests in Oregon, in which inclusion of an explicit budget constraint was found to have no effect on subjects' stated willingness to pay for a proposed program (see Loomis, Gonzalez-Caban, and Gregory 1994).

Each of the five sets of test questions included one question for which the correct answer was equal to the standard, one question requiring an integer answer less than the standard, two questions requiring an integer answer greater than the standard, and one question requiring a noninteger answer less than the standard. Subjects who failed this computational test were excluded from the analysis.

III. RESULTS

Our criterion for including participants was quite cautious. An individual was included only if at least four of the five test questions were answered correctly. About one-third of those participating in each of the five conditions failed to pass the test.² Final sample sizes for the five conditions (\$1, \$2, \$5, \$10, \$20) were 46, 61, 60, 49, and 61, for a total of 277 subjects (i.e., 65 percent of the 425 original participants).

The results are presented in Table 2, which shows the median willingness to pay, in dollars, by item and by condition. The values in Table 2 have been translated from the raw responses to dollars (e.g., a response of "multiply by 2" has been translated to \$2 for BU = \$1 and to \$40 for BU = \$20). The results, shown in Table 2, therefore permit a direct comparison of the reported willingness to pay for the five different groups. Median values were used for the analysis rather than means, due to the skewed distribution of the results.

Median responses to three of the items (#9, 10, 11) are identical to the assigned budgetary units for all five groups. For three additional items (#5, 8, 12), responses increase monotonically across the five budgetary units. Thus for 6 of the 12 test items, monetary responses increase with the magnitude of the budgetary unit. Items 2 and 3 fail to show substantial variation across payment methods. Median willingness-to-pay levels for items 4, 6, and 7 exhibit a steady increase across budgetary units up to \$10 but then collapse when the budgetary unit is raised to \$20.

The remaining item, planting daffodils (#1), was deemed worthless, or nearly so, by

TABLE 2
MEDIAN WILLINGNESS TO PAY IN DOLLARS,
ALL ITEMS

Item	Budgetary Units				
	\$1	\$2	\$5	\$10	\$20
1	.10	0.00	.05	0.00	0.00
2	.75	2.00	2.50	1.15	2.00
3	1.00	2.00	5.00	2.00	5.00
4	1.00	2.00	5.00	5.00	4.00
5	1.00	2.00	5.00	5.00	5.00
6	1.00	2.00	2.50	10.00	5.00
7	1.00	2.00	5.00	10.00	4.00
8	1.00	2.00	5.00	10.00	10.00
9	1.00	2.00	5.00	10.00	20.00
10	1.00	2.00	5.00	10.00	20.00
11	1.00	2.00	5.00	10.00	20.00
12	5.00	6.00	15.00	20.00	20.00

most subjects. Across all forms, 57 percent were not willing to pay anything at all, and an additional 15 percent were willing to pay some amount less than \$1.00. The results for this item indicate that subjects discriminated among the items and approached the valuation task seriously. However, because of the large number of zero responses, this item did not provide a clear test of differences between the forms. Further analyses excluded this item.

The complete distributions of subjects' responses are shown in Figure 1. Each histogram gives the willingness-to-pay distributions across all 11 items (excluding item 1) for one of the five budgetary units. The overwhelming popularity of the budgetary unit is apparent. Over all individual responses to the 11 items, about one-third (33.6 percent) were equal to the budgetary unit.

The significance of the differences in median willingness to pay for each item (shown

² This performance level was set on the basis of a careful examination of the November 1991 pretest results. At a practical level, nearly all participants (in both the May 1992 experiment as well as the November pretest) got at least three answers right, but few got all five answers right. At a theoretical level, the 80 percent criterion is not dissimilar from that used in many examinations to distinguish performance competence.

in Table 2) was tested using a median test³ on every pair of medians, for each item separately (excluding item 1). The results are shown at the top of Table 3.

One would expect, when making 110 significance tests, that a few would appear significant by chance alone. Moreover, the data tested here are interdependent. Any one item appears in 10 tests; any particular response appears in 4 tests; any one subject is included in 44 tests. Such interdependence has an unknown effect on the significance tests.

Rather than correcting for the unknown effects of chance and interdependence by choosing a more conservative level of significance, we explored the robustness of our results by performing the same 110 median tests on the untranslated data, that is, the raw responses as given by the subjects. Thus a response of "half a budgetary unit" was taken as ".5" for all groups (rather than translating it to, for example, \$.50 when BU = \$1 and \$2.50 for BU = \$5). The results of these tests are shown in the lower half of Table 3.

Whereas the null hypothesis for the first set of tests is that willingness to pay does not differ with a change in budgetary unit, the null hypothesis for these new tests is that the subjects give the same literal response regardless of the budgetary unit (and thus their willingness to pay changes radically).

Whatever the effects of chance and interdependence on the first set of 110 tests, such effects could be expected to be about the same in this new set of tests. A comparison of the two thus will indicate the robustness of the results. The difference between the two sets of tests is striking. For the translated data, 65 percent of all tests were significant (i.e., $p \leq .05$); 31 percent were highly significant (i.e., $p \leq .001$). For the untranslated data, 21 percent were significant; only one was highly significant.

In 12 of the 110 comparisons, significance was found for both the translated data and the raw data. This does not necessarily mean that one or the other test result was due to chance. It is possible that both sets of data are genuinely different; subjects could give

different responses to two forms but fail to make these differences as large as would be required to equate the translated willingness to pay. The most extreme example of this is item 12 (planting 100,000 trees), where the raw medians were 5 for the BU = \$1 group and 1 for the BU = \$20 group, so that the translated medians were \$5 and \$20, respectively. As seen in Table 3, both differences were highly significant. Essentially, the 5-fold greater raw response for

³ The usual convention in doing a median test with one or more observations at the joint median is to group these observations with observations below the median. This convention leads to absurd results with the present data, because of the overwhelming popularity of the median response. For example, consider item 7, BU = \$10 (mdn = \$10) vs. BU = \$20 (mdn = \$4). The joint median is \$10; the data are as follows:

	BU = \$10	BU = \$20	
Below mdn	22	31	53
At mdn	18	3	21
Above mdn	8	27	35
	48	61	109

By convention, the χ^2 for the median test is calculated on the data regrouped as follows:

	BU = \$10	BU = \$20	
At or below mdn	40	34	74
Above mdn	8	27	35
	48	61	109

The median test for these data is significant ($p < .005$), apparently establishing that a median of \$10 is significantly smaller than a median of \$4.

To avoid such absurdities, we assigned a proportionate number of the median observations to the categories "above" and "below," so that the marginal sums were equal. Thus for the data shown above, we regrouped by moving 1.5 of the 21 median observations (7.1 percent) to the "below" category and the rest to the "above" category. After rounding to the nearest integer:

	BU = \$10	BU = \$20	
At or below mdn	23	31	54
At or above mdn	25	30	55
	48	61	109

The χ^2 test is not significant, as is appropriate for these data.

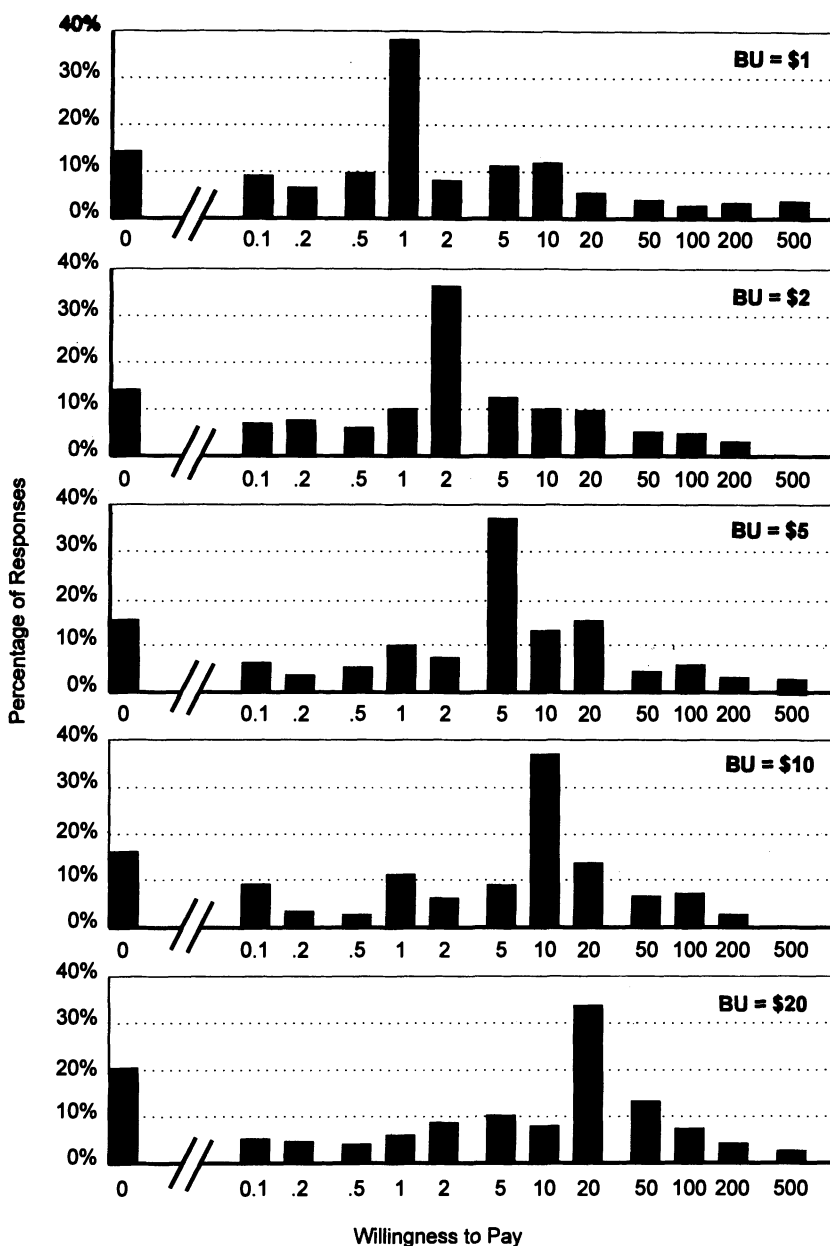


FIGURE 1

Histograms of WTP responses (transformed) for all questions except Question 1. The responses are grouped in groups of equal width on a log scale and each bar is centered on the geometric mean of the responses in the group. The ranges of the responses were 0; .01-.10; .20-.25; .33-.50; 1.00-1.33; 1.67-3.00; 3.33-6.67; 7-12; 15-30; 35-60; 70-125; 175-300; 375-2,500.

the BU = \$1 group did not match the 20-fold difference in the size of the budgetary unit.

There is still no way of knowing how much noise has been introduced into both

sets of tests from chance factors and interdependence among the data. However, the stark difference between the two sets of tests (which are both vulnerable to such noise) indicates that the use of different

TABLE 3
RESULTS OF MEDIAN TESTS

Comparison of Forms with Different Budgetary Units										
Item	1 v 2	2 v 5	5 v 10	10 v 20	1 v 5	2 v 10	5 v 20	1 v 10	2 v 20	1 v 20
Responses Translated to WTP										
2	*				**					*
3	**	***			***			*		***
4	*	***			***	***		**		**
5	*	*			***			**	**	***
6	***				*	*		***		**
7		*			**	*		**		**
8	**	***			***	*		***	**	***
9		**	*	***	*	***	***	***	***	***
10	*	***	*	**	***	***	***	***	***	***
11		***	*		***	**	**	**	***	***
12		*			**	**	**	***	***	***
Raw, Untranslated Responses										
2							*		*	*
3			*				**	*	**	**
4				**			**		*	**
5										*
6		*							*	
7										
8							*			
9										
10								*		*
11										
12					*		**	**	*	***

* = $p \leq .05$; ** = $p \leq .01$; *** = $p \leq .001$.

budgetary units led to real and widespread differences in willingness to pay. The ± 50 percent standard proposed by Cummings, Brookshire, and Schulze (1986) is generally violated by these data.

IV. DISCUSSION

We are not recommending a Budgetary Unit technique for eliciting willingness-to-pay values; it is a difficult task for some people, as evidenced by the 35 percent of our subjects who failed the test at the end. Nor do we recommend the use of these values for policy purposes; our subjects' average willingness to pay, for both methodological and design reasons, is unlikely to be representative of the general population.⁴ Instead, we are asserting that if individuals held reasonably precise (within ± 50 percent) monetary values for environmental improvements like these, then those who

were not defeated by our technique would express such values when asked. The resulting medians would not, then, show any systematic effect across the forms. The translated median dollar responses would be the same across the five groups for any one item. That is, if an environmental good has a stable, well-defined monetary value of \$4, then subjects presented with a BU of \$2 would multiply this standard by 2 whereas subjects given a BU of \$10 would divide this value by 2.5.

However, these data (as shown in Table 2) show a systematic effect of the change in budgetary unit. Together with the strong popularity of the budgetary unit as a re-

⁴In addition, resource values that are used for policy purposes should be elicited using procedures, such as debriefing sessions, that permit respondents to more carefully consider the personal and social implications of their answers.

sponse, these results suggest that people have no clear monetary value for these goods. Instead, subjects' responses are sensitive to the arbitrarily assigned change in budgetary units because they lack precise monetary representations of their values. For some items, any amount between \$1 and \$20 is judged to be as good a translation of value into money as any other amount.

We began this experiment with the hypothesis that people do not hold readily accessible monetary values for many environmental goods. Our data have led to a refinement in our thinking: People do have a sense of their monetary value for these items, as shown most clearly by the differences in responses to items 1 and 12 and by the generally consistent ordinal ranking in median willingness to pay across the items. However, people's monetary values are diffuse, distributed across a surprisingly broad range.

The results suggest an unusual interpretation of the term value. We do not observe evidence supporting the concept of a "true" value, characterized by a point estimate, for these environmental goods. Instead, we observe results that support the concept of a broad range, such that any value within this range could be deemed appropriate. The popularity of the arbitrary budgetary unit suggests that many participants will select it as an appropriate value if it is within this broad range; their concept of monetary value is sufficiently imprecise that the budgetary unit cannot be improved on.

Several alternative explanations of these data are possible. One explanation is in terms of the anchoring-and-adjustment heuristic whereby respondents, when asked for a numerical judgment, start with any convenient number (even a notably irrelevant number; see Tversky and Kahneman 1974) and adjust it upwards or downwards. Inadequate adjustment leads such judgments to be unduly influenced by the anchor.

The anchoring-and-adjustment strategy, if used in this experiment, would lead to a systematic dependence of the medians on the budgetary units, yielding a pattern of results somewhat like ours. However, the

details of our data do not support an explanation via anchoring and adjustment. As shown in Figure 1, a response equal to the budgetary unit was common; this could provide evidence of anchoring, but it fails to suggest subsequent adjustment in participants' values. Such anchoring without adjustment is not a strategy one would expect from people who have a clear monetary value for an item.

One further possible explanation of the results is laziness. The experiment was clearly hypothetical; perhaps subjects just didn't bother to think very hard.

The explanation of extreme laziness may be rejected. Subjects obviously spotted the one item most of them found worthless (#1, daffodils) and responded accordingly. At the other extreme, subjects given $BU = \$1$ or $BU = \$2$ took the trouble to indicate that their value for item 12 (100,000 trees) was greater than \$1 or \$2; 69 percent of their responses exceeded the budgetary unit.

The hardest response to make in this task is to report a nonzero value less than the budgetary unit. A respondent had to (a) think of a value, (b) form a fraction of that value to the budgetary unit, (c) convert that fraction to the form $1/x$, and (d) report x in the second of the four columns showing expenditure options. Subjects given $BU = \$20$ did not avoid this complexity. For items 2 through 7 (the items, we assert, for which \$20 was most often outside the acceptable range for the vague translation of value into money), 38 percent of the $BU = \$20$ responses were of this hardest type.

Thus, a critic of our interpretation would have to argue some sort of "semi-laziness." Such a tortuous explanation would have to explain why some groups were more "lazy" than others and why some items elicited more "lazy" responses than others. Our view of the data is more parsimonious: There exists for each subject a range of money within which any particular dollar amount is apparently deemed as correct as any other to express the respondent's value for the good.

The suggestion that people's maximum willingness to pay for some goods is a range of acceptable values has particular implica-

tions for the dichotomous choice form of CV. In this method, currently preferred by many practitioners and by the recent panel that reviewed CV for NOAA (Arrow et al. 1993), the respondent says "yes" or "no" to a posited amount. Our results suggest that any posited amount within the range will receive a "yes." Thus the dichotomous choice format will yield an upper bound on maximum willingness to pay rather than a midpoint or other measure of central tendency of the value range. This supposition is supported by several recent studies (Desvousges et al. 1992; Johnson et al. 1990; Kealy and Turner 1993; Loomis, Lockwood, and DeLacy 1993) that found dichotomous-choice estimates of willingness to pay exceeded open-ended estimates. In the light of the NOAA panel's suggestion that contingent valuation estimates are more likely to be accepted if conservative methods were used, further research is needed on whether the dichotomous choice format really produces conservative estimates.

V. CONCLUSION

This article follows from a long-standing interest in the environmental evaluation community regarding the accuracy of participants' responses to survey questions. The issue is important because approaches such as CV increasingly are being used as the basis for determining damage payments and other resource policies.

We have good reasons to be modest about the generalizability of the study results presented in this paper, which derive from a student sample and reflect responses to 12 specific environmental items presented under hypothetical payment conditions. We also recognize that our budgetary-unit approach precluded following conventional CV procedures. Nevertheless, we believe that our results add to the evidence that people have only vague notions of the monetary value of many nonmarket environmental items. Further, the results suggest that any value within a rather broad range is acceptable to respondents. Thus, when people are asked how much they would pay for a good,

contextual effects may easily move respondents to one place or another within the acceptable range. Specific tests of this value-range hypothesis are needed to understand the contribution of different factors to the width of the value range.

The more fundamental question raised by our results is whether holistic monetary estimates, as required by CV elicitation, provide appropriate measures of value for many nonmarket environmental goods. Our conclusion, based on the results of this study and supported by earlier work (e.g., Brown and Slovic 1988; Gregory, Lichtenstein, and Slovic 1993; Irwin et al. 1993) is that often they do not. The imprecision in monetary estimates reported here arises because statements of monetary value for these goods are constructed by participants in the course of the elicitation process. As a result, we recommend that valuation strategies used as part of resource policy decisions recognize the need for value-structuring techniques to assist people, first to think about the multiple dimensions of value and then to integrate these values in a theoretically defensible and personally meaningful way.

APPENDIX

INSTRUCTIONS FOR BU = \$1

Decision makers often have a difficult time determining the value of natural resources. Values need to be determined when resources have been damaged due to a natural event (e.g., a severe winter storm) or a technological accident (e.g., an oil spill). Values also need to be determined when environmental improvements are considered, such as programs to improve air quality or to plant trees and improve habitat for animal species.

Listed on the following pages are brief descriptions of several types of possible activities that could be undertaken to protect or to enhance the environment of Oregon.

We'd like to know *how much more you would be willing to spend each year*, as part of the taxes and other payments you make to the State of Oregon, to undertake each of these projects. *Consider the projects separately*, as if only that one project were under consideration. When you are thinking about the additional payment you would

make assume that other people would pay the same amount as part of their state payments.

To make the payment as realistic as possible, we want you to think about it in terms of a BUDGETARY UNIT. BUDGETARY UNITS are used by state legislators to plan out yearly expenditures and to decide among competing projects. *Each BUDGETARY UNIT represents \$1.*

Please tell us, in terms of BUDGETARY UNITS, how much more money in taxes and other payments you would be willing to spend on each project. You can express your value for the project in one of four ways, as shown by the four columns A–D:

1. If you believe that you do not want to pay anything at all, put an "X" in the *left-hand* column, column A.
2. If your willingness to pay is equal to one budgetary unit, put an "X" in the *third* column, column C.
3. If your willingness to pay is more than one budgetary unit, write a number greater than 1 in the *right-hand* column, column D. This number is the number by which the budgetary unit must be *multiplied* to express your willingness to pay. For example, if your willingness to pay for a particular project is \$100, the budgetary unit would have to be multiplied by 100 to get \$100 (because \$1 multiplied by 100 equals \$100). So for this example you would write "100" in column D.
4. If your willingness to pay is less than one budgetary unit, write a number greater than 1 in the *second* column, column B. This number is the number by which the budgetary unit must be *divided* to express your willingness to pay. For example, if your willingness to pay for a particular project is 10¢, the budgetary unit would have to be divided by 10 (because \$1 divided by 10 equals 10¢). So for this example you would write "10" (not "1/10") in column B.

Remember to think about *your willingness to pay, on an annual basis*, to undertake each of the environmental projects described on the next pages. Remember also to *consider each project separately*, as if only that one were being considered.

(Note: For the other forms, the multiplication example was always willingness to pay \$100, with the multiplier changed as appropriate; the division example was always willingness to pay 10¢, with the divisor changed as appropriate.)

References

- Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, and Howard Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." *Federal Register* 58 (10):4602–14.
- Boyle, Kevin, Richard Bishop, and Robert Welsh. 1985. "Starting Point Bias in Contingent Valuation Bidding Games." *Land Economics* 61 (May):188–94.
- Bromley, Daniel W., ed. 1986. *Natural Resource Economics: Policy Problems and Contemporary Analysis*. Boston: Kluwer.
- Brown, Thomas C. 1984. "The Concept of Value in Resource Allocation." *Land Economics* 60 (Aug.):231–46.
- Brown, Thomas C., and Paul Slovic. 1988. "Effects of Context on Economic Measures of Value." In *Integrating Economic and Psychological Knowledge in Valuations of Public Amenity Resources*, eds. G. L. Peterson, B. L. Driver, and R. Gregory, 23–30. State College, PA: Venture.
- Cummings, Ronald G., David S. Brookshire, and William D. Schulze. 1986. *Valuing Environmental Goods: Assessment of the Contingent Valuation Method*. Totowa, NJ: Rowman and Allanheld.
- Desvousges, W. H., F. R. Johnson, R. W. Dunford, K. J. Boyle, S. P. Hudson, and K. N. Wilson. 1992. "Measuring Natural Resource Damages With Contingent Valuation: Tests of Validity and Reliability." In *Contingent Valuation: A Critical Assessment*. Cambridge, MA: Cambridge Economics.
- Gregory, Robin, Sarah Lichtenstein, and Paul Slovic. 1993. "Valuing Environmental Resources: A Constructive Approach." *Journal of Risk and Uncertainty* 7:177–97.
- Gregory, Robin, Donald MacGregor, and Sarah Lichtenstein. 1992. "Assessing the Quality of Expressed Preference Measures of Value." *Journal of Economic Behavior and Organization* 17:277–92.
- Irwin, J., Paul Slovic, Sarah Lichtenstein, and G. McClelland. 1993. "Preference Reversals and the Measurement of Environmental Values." *Journal of Risk and Uncertainty* 6:5–18.
- Johnson, Rebecca L., N. Stewart Bergenzer, and Bo Shelby. 1990. "Contingent Valuation Question Formats: Dichotomous Choice Versus Open-Ended Responses." *Economic Valuation of Natural Resources*, eds. R. L. Johnson and G. V. Johnson, 193–204. Boulder, CO: Westview.

- Kahneman, Daniel, and Jack Knetsch. 1992. "Valuing Public Goods: The Purchase of Moral Satisfaction." *Journal of Environmental Economics and Management* 22 (1):57-70.
- Kealy, Mary Jo, Mark Montgomery, and John Dovidio. 1990. "Reliability and Predictive Validity of Contingent Values: Does the Nature of the Good Matter?" *Journal of Environmental Economics and Management* 19 (3):244-63.
- Kealy, Mary Jo, and Robert W. Turner. 1993. "A Test of the Equality of Closed-Ended and Open-Ended Contingent Valuations." *American Journal of Agricultural Economics* 75 (May):321-31.
- Loomis, John. 1990. "Comparative Reliability of the Dichotomous Choice and Open-Ended Contingent Valuation Techniques." *Journal of Environmental Economics and Management* 18 (1):78-85.
- Loomis, John, Armando Gonzalez-Caban, and Robin Gregory. 1994. "Do Reminders of Substitutes and Budget Constraints Influence Contingent Valuation Estimates?" *Land Economics* 70 (Nov.):499-506.
- Loomis, John, M. Lockwood, and T. DeLacy. 1993. "Some Empirical Evidence on Embedding Effects in Contingent Valuation for Forest Protection." *Journal of Environmental Economics and Management* 24 (1):45-55.
- Mitchell, Robert C., and Richard T. Carson. 1989. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington, DC: Resources for the Future.
- Payne, John W., James R. Bettman, and Eric J. Johnson. 1992. "Behavioral Decision Research: A Constructive Processing Perspective." *Annual Review of Psychology* 43:87-132.
- Schulze, William, Ralph d'Arge, and David Brookshire. 1981. "Valuing Environmental Commodities: Some Recent Experiments." *Land Economics* 57 (May):151-72.
- Slovic, Paul, Sarah Lichtenstein, and Baruch Fischhoff. 1979. "Images of Disaster: Perception and Acceptance of Risks From Nuclear Power." *Energy Risk Management*, eds. G. Goodman and W. Rowe, 223-45. London: Academic.
- Tversky, Amos, and Daniel Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185:1124-31.
- Tversky, Amos, Shmuel Sattath, and Paul Slovic. 1988. "Contingent Weighting in Judgment and Choice." *Psychological Review* 95:371-84.