Journal of Experimental Psychology: Applied
1996, Vol. 2, No. 2, 107–125

Copyright 1996 by the American Psychological Association, Inc.
1076-898X/96/$3.00

Determinants of Insensitivity to Quantity in Valuation of Public Goods: Contribution, Warm Glow, Budget Constraints, Availability, and Prominence

Jonathan Baron and Joshua Greene
University of Pennsylvania

Insensitivity to quantity in valuation appears in 3 ways: embedding (when willingness to pay for a good is smaller if assessed after a superordinate good), insensitivity to numerical quantity, and adding up (when willingness to pay for 2 goods is less than inferred from willingness to pay for each good alone). Results of 11 experiments on these effects are generally inconsistent with 3 accounts: People think of the task as a contribution, they get a warm glow from participation, and they have budget constraints on expenditures for certain goods. The results support an explanation in terms of prominence of the type of good as opposed to its quantity. In addition, 1 study supports availability of only the good evaluated rather than other goods of the same type. These findings support the critics of contingent valuation, but they suggest that some of the methods of decision analysis can improve the elicitation of economic values.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Researchers often measure the economic value of public goods by asking people how much money is equivalent to a good. In the contingent-valuation (CV) method, subjects indicate their willingness to pay (WTP) for the good in question. CV is used in several countries for assessing environmental damage from oil and chemical spills and for evaluating environmental and safety effects of government policies. CV is a matter of contentious debate and litigation (NOAA, 1993; Portney, 1994). A large part of the problem is that CV judgments are sometimes remarkably insensitive to the quantity or scope of the good provided. This insensitivity provides ammunition to critics who say that CV does not measure economic value (e.g., Diamond & Hausman, 1994). Defenders of CV have argued that insensitivity does not necessarily imply faulty measurement (e.g., Hanemann, 1994) or that CV can be made more sensitive if properly conducted (e.g., Schuman, 1995). At issue is the explanation of the insensitivity effects themselves. The source of the effects has implications for whether the problem of insensitivity is serious and, if it is, how to make people more sensitive.

First, we summarize the evidence for insensitivity and then discuss some possible explanations and their implications. Insensitivity is a pervasive phenomenon, not dependent on the details of method. Accordingly, although we used a variety of methods to study these effects in experiments we report in this article, we found no reason why our conclusions would not apply just as well to similar methods that we did not use.

Types of Quantity Insensitivity

Several types of effects have been identified.

Jonathan Baron and Joshua Greene, Department of Psychology, University of Pennsylvania. Joshua Greene is now enrolled as an undergraduate at Harvard College.

This research was supported by National Science Foundation Grant SBR92-23015. We thank Jane Beattie and Nicholas Maxwell for comments on an earlier draft and Lisa Chen for assistance in the first study.

Correspondence concerning this article should be addressed to Jonathan Baron, Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, Pennsylvania 19104-6196. Electronic mail may be sent to baron@psych.upenn.edu.

Embedding

Kahneman and Knetsch (1992) asked some subjects their WTP for improved disaster preparedness and other subjects their WTP for improved rescue equipment and personnel. The improved equipment and personnel were thus "embedded in" the improved disaster preparedness, so the preparedness included the equipment and personnel, and other things too. However, WTP was about the same for the larger good and for the smaller good included in it. Kahneman and Knetsch called this the *perfect embedding effect,* presumably because a demonstration of it requires perfect equality of WTP of the two goods. When subjects were asked their WTP for the smaller good after they had just been asked about the larger one, they gave much smaller values for the smaller good than for the larger one, and much smaller values than those given by subjects who were asked just about the smaller good. This order effect is called the *regular embedding effect.* It demonstrates that a good seen as embedded in a larger good has reduced value. Kemp and Maxwell (1993) replicated this regular embedding effect, starting with a broad spectrum of public goods, and narrowing the good down in several steps, obtaining WTPs for an embedded good that were 1/300 of WTP for the same good in isolation.

Adding Up

In a related demonstration, Diamond et al. (1993) asked subjects their WTP values for preventing timber harvesting in federally protected wilderness areas. WTP for the prohibition in three areas was not much (if any) higher than WTP for prohibition in one of the areas alone. This result cannot be explained by assuming that subjects thought that protection of one area was sufficient: When they were asked their WTP to protect one area assuming that another was already protected, or a third area assuming that two were protected, their WTP values were just as high as those for protecting the first area. More generally, in this kind of *adding-up effect,* respondents are asked their WTP for Good A (e.g., a single wilderness area), for Good B assuming that Good A has been provided already, and for Goods A and B together. WTP for A and B together is much lower than the sum of the WTP for A and for B (with A provided).

Quantity Effect

Jones-Lee, Loomes, and Philips (1995) asked subjects to evaluate hypothetical automobile safety devices that would reduce the risk of road injuries. WTP judgments were on the average only 20% higher for a risk reduction of 12 in 100,000 than for a reduction of 4 in 100,000. In general, the rate of substitution between money and the good, the dollars per unit, depends strongly on the amount of the good. This makes it difficult to generalize results to different amounts, a kind of generalization that is almost always required (Baron, 1995). Many other results of the same sort have been reported by investigators who were content on showing that WTP differs, if only by a small proportion, for a difference in goods of several orders of magnitude, resulting in extreme differences in rate of substitution (e.g., Boyle, Desvousges, Johnson, Dunford, & Hudson, 1994; Carson & Mitchell, 1993).

Explanations of Quantity Insensitivity

The present article concerns the explanation of these effects. Several explanations are possible. The first two, contribution and warm glow, derive from Kahneman and his colleagues, who suggested that WTP judgments are expressions of attitudes about the moral satisfaction to be obtained from contributing (Kahneman & Knetsch, 1992; Kahneman, Ritov, Jacowitz, & Grant, 1993). WTP responses are highly correlated with judgments of satisfaction and of importance and with the judged importance of the issues examined and the respondent's feeling of responsibility for them (Guagnano, Dietz, & Stern, 1994).

Contribution

By this account, respondents think of the task as a charitable contribution. The benefit from a contribution depends mainly on its size, not on the size of the total project. For example, there are two ways of contributing to an international charity: In one, your contribution is divided up among many villages around the world, and in the other, it goes to a single village, with other people's contributions going to other villages. Because the same charity is involved, it is reasonable to think that the total good resulting from your contribution is the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

same in both cases, so that size of the total project (one village vs. all villages) does not affect contributers' WTP.

Crucial to the contribution account is the subjects' assumption that each contribution increases the total amount of the good. This account predicts that subjects will be more sensitive to quantity if they understand that the size of their contribution does not affect the size of the good provided. To convey this to subjects, we tell them that the size of the good is fixed and that their WTP will be compared with their fair share of the cost of the good. If more than half of the respondents are willing to pay at least their fair share, then the good will be provided; otherwise it will not.

Warm Glow

This account holds that WTP depends on personal agency or participation rather than on the expected consequences of a contribution. This is, we believe, equivalent to the proposals of Andreoni (1990) and Margolis (1982). Baron and Spranca (1995) discussed other "agent relative" principles that affect WTP. This account predicts that sensitivity to quantity increases if subjects simply provide ratings of the importance of various public goods, because no participation is involved. Kahneman and Ritov (1994) found that ratings of importance and support for interventions were insensitive to quantity, and we have replicated this result with ratings of satisfaction and benefit (e.g., Baron, Chen, & Greene, 1993). Jones-Lee et al. (1995) suggested that their demonstration of the quantity effect for auto safety is difficult to explain in terms of moral satisfaction (presumably the warm-glow version) because the good in question is private.

Budget Constraints

People sometimes say that they will "contribute as much as they can afford." Taken literally, this is implausible, but people may view certain expenditures as setting precedents for similar expenditures, or they may think of their spending in terms of separate, limited accounts (Thaler, 1993), perhaps including accounts for public goods. People may answer with their limit when they are asked their WTP for one unit of a good. The results for ratings (just described) cast doubt on this account. Also, by this account, insensitivity would be found only in WTP judgments, not in judgments of willingness to accept (WTA) money to give up the good in question. We tested this prediction. Dubourg, Jones-Lee, and Loomes (1994) found insensitivity in WTA judgments, although they did not point out the relevance of this finding to the budget hypothesis. We report below three replications of this important result.

A related explanation holds that insensitivity results from substitution effects (Carson & Mitchell, 1995; Haneman, 1994). For example, subjects may think of wilderness areas as substitutable and that preserving one of them is less valuable when another is already preserved. Then they would pay little more for two than for one. Note that this is not the result of Diamond et al. (1993) because subjects were asked how much they would pay for the second assuming they had the first. More generally, the marginal utility of goods of a certain sort may be declining. This explanation, too, predicts that insensitivity would not be found in WTA judgments. Indeed, it predicts oversensitivity. For example, subjects would require a very large amount of money to give up the second of two wilderness areas, although they might require only a little money to give up one of them. Because the substitution account makes the same predictions for the present studies as the budget-constraint account, we lump them together below.

If these accounts explain the results, then insensitivity need not be a problem for certain purposes. People are giving their true WTP for what they are asked about.

Availability

This explanation holds that subjects fail to think of other goods of the same type as the good evaluated, unless these other goods are explicitly presented. It is related to the finding that subjects assign too little probability to "all other problems" when they are given a tree of possible faults in a system (Fischhoff, Slovic, & Lichtenstein, 1978). This account is also similar to the "part–whole" bias described by Mitchell and Carson (1989). This account implies that insensitivity is a real problem, but it might be cured by reminding people about other goods.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Prominence

This explanation of insensitivity is based on a loose analogy with several results from the literature on decision making and decision analysis (Fischer, 1995; Tversky, Sattath, & Slovic, 1988; Zakay, 1990). The essence of the idea is that people assign "importance" to attributes, dimensions, or types of goods, even if they do not know the magnitude or range of variation at issue. For example, when most people are asked whether life or money is more important, they confidently answer that life is. If they are then asked whether they would spend the annual gross national product of the United States in order to reduce someone's chance of death by .0001 this year, they feel tricked. In decision analysis, naive subjects often try to assign numerical weights to dimensions without knowing the range of each dimension. In the case of CV, when people are asked to say something about the importance of cleaning up oil spills, they respond to the type of good rather than the amount of it, even though they are asked to respond with an amount of money. When they are asked about two goods, they respond to their average importance rather than their total. The prominence explanation also implies that insensitivity is a real problem, but one that might be cured by using the methods of decision analysis, where it is apparently not a problem.

Although the prominence hypothesis may seem similar to the warm-glow hypothesis, the latter concerns the actual contribution made. However, perceived prominence implies that WTP may be affected by the way of asking the question. For example, a subject who used WTP to indicate prominence might give the same response when asked for a WTP for 10 units of a good or when asked the WTP per unit for 10 units. However, the warm-glow hypothesis concerns the actual total payment made.

The present studies test these hypotheses as necessary or sufficient accounts of the various forms of insensitivity to quantity. The studies do not test every explanation of every effect, and materials are varied from study to study, sacrificing systematicity for variety. The General Discussion and Table 4 present an overview of the hypotheses and results.

In general, we speak of both sensitivity and insensitivity. *Complete insensitivity* means that subjects ignore quantity; proportional sensitivity means that the subject's price (e.g., WTP) is proportional to the magnitude of the good. We think of a continuum between these two poles. *Sensitivity* is the distance from complete insensitivity, and *insensitivity* is the distance from proportional sensitivity. A little insensitivity need not imply any problem of measurement. However, this departure measure is useful for comparing experimental conditions.

Tests of Contribution

Study 1: Referendum and Public Safety

In this study, we tested the contribution account by comparing a standard condition, in which subjects are free to think of their WTP as a contribution, to a condition in which they are explicitly told that the good is fixed and that their WTP will be compared with their fair share of the good. The good will be provided only if at least half of the subjects are willing to pay their share. We call this a *referendum* condition because it is equivalent to majority rule: If your WTP is more than your share, then you would presumably vote for providing the good. Of course, in a true referendum, a price may be given, and the response is yes or no. McFadden (1994) found that such dichotomous choice did not increase sensitivity to quantity, compared with a free response condition of the sort we used throughout this study (contradicting NOAA, 1993). Moreover, our purpose was not to simulate a real referendum but rather to make clear that the size of the benefit, if provided, is independent of anyone's WTP.)

We tested this hypothesis in the context of tuition increases for increased public safety at the University of Pennsylvania, an urban university where safety is a salient issue for students. Only students took part in the study. Within each group—referendum vs. no-referendum—about half of the subjects were asked about the small good only, and half were asked about the large good and then about the small good. To encourage the contribution interpretation in the no-referendum condition, the amount of increase in the good was not specified, so that subjects were free to think that the increase would depend on their WTP. In the referendum condition, in order to discourage the contribution interpretation, we specified the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

increase as 10%, and subjects were told whether the good would be provided.

For about half of the subjects in each of these subgroups, the good would be provided only after the students graduated, so that the payment could not be seen as a private purchase of additional safety. Although the safety programs were truly public goods, many students in pilot studies tended to think of the payments as providing direct benefit to the individual who paid—although this benefit is a tiny fraction of the overall benefit. By saying that the program would begin only after the student left, we made it entirely a matter of altruism. In summary, we used a 2 × 2 × 2 between-subjects design: Referendum × Order (large-small vs. small only) × Altruism.

Method

Subjects were approached on campus either in residences or on walkways and were offered 50 cents to complete a brief questionnaire. They were told that it was a psychology study, although the situation was so realistic that many had to be convinced that it was not being conducted for the university. There were 204 subjects in all, with at least 25 in each of the eight groups.

In the no-referendum, altruism, large–small condition, the questionnaire read:

> Consider a proposal to increase the security on Penn's campus and the surrounding area. The program would be funded by a student tuition increase. You would pay the tuition increase, but the new program would not begin until after you graduate. For those students with financial aid, the increase will be pro-rated. Answer these questions on the assumption that you pay full tuition, even if you pay less. The exact tuition increase would depend on your response, as well as other Penn students' responses.
>
> 1. The extra amount of tuition would directly increase all safety programs equally—police patrols, escort services, emergency phones, etc. What is the greatest amount in dollars per year that you would be willing to pay? $____
>
> 2. The proposal includes an increase in the amount of money spent on police in patrol cars. If the extra money were spent only on police in patrol cars (with all other spending unchanged), what is the largest amount in dollars per year that you would be willing to pay? (You would pay the tuition increase, but the new program would not begin until after you graduate.) $____

In the small conditions, subjects were asked only about the patrol cars. In the no-altruism conditions, the statements about when the new program would begin were omitted. In the referendum conditions, all increases were specified as 10%, and the following sentence was added immediately after the question about WTP but before the answer blank:

> If the median (middle) response to this question is greater than each student's part of the actual cost of the program, then the program will go into effect, and everyone will pay his or her part of the actual cost. Otherwise, the program will not go into effect.

Subjects were also asked to identify their level in college and their sex.

Results

The embedding effect was found overall, and its magnitude was insensitive to referendum versus no referendum.

Over all conditions, the mean response was $148 ($SD$ = $240) for the large good and $162 ($SD$ = $441) for the small good. However, as is usually found, the response distribution was highly skewed: Many responses were zero (29% for the large good and 26% for the small good over all conditions, including one response of $0.0001). Some responses were unrealistically high (12 responses of $1,000 or more). Therefore, we used a method suggested by McFadden and Leonard (1993; also used by Green et al., 1994), in which the WTPs were log transformed and the numbers of $0 responses were analyzed separately. Table 1 shows the geometric mean responses by condition for nonzero WTPs.

In a multiple regression in which the log-transformed WTP for the small good was predicted from altruism versus no altruism, referendum versus no referendum, embedding (large good first) versus no embedding, and the interaction of Referendum × Embedding, only the effect of embedding was significant (p = .009).[1] The

[1] An earlier report of this study (Baron et al., 1993) found a significant interaction between referendum and embedding by using a power transform instead of a log

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 1
Geometric Mean Responses and Zero Percentages by Condition

Condition and order	Geometric mean: Large good	Geometric mean: Small good	Zero: Large good	Zero: Small good
Referendum				
Small only		$87.01		21.6
Large–small	$84.10	$64.46	15.7	25.5
No-referendum				
Small only		$130.71		16.0
Large–small	$96.35	$50.65	42.3	38.5

embedding effect was significantly greater in women than in men (p = .027 for the Sex × Embedding interaction). The triple interaction of Sex × Embedding × Referendum was not significant, nor was the triple interaction of Altruism × Embedding × Referendum. A parallel analysis of the first response made—the large good in the embedded condition, the small good otherwise—found no effect of good size (embedding condition), altruism (whether the subject benefited), referendum, or any interaction. In summary, the WTP measure, excluding subjects with zero WTP, showed both the regular embedding effect and the perfect embedding effect and no effect of altruism.

Log-linear analyses for the zero responses to the small good showed effects of embedding and altruism (simple chi-square tests yielding $\chi^2(1, N = 204)$ of 5.8 and 4.7, p = .030 and .016, respectively), but no effect of referendum and no interactions. In summary, altruism affects the percentage of zero responses (33% in the altruism condition vs. 18% in the no-altruism condition), although it does not affect WTP once the decision is made to pay something. Most importantly, an embedding effect was found here too, and, again, the effect does not differ between referendum and no-referendum conditions.

transform, thus keeping the zero responses. This transform is inappropriate when so many responses are zero: The assumption of normally distributed error is seriously violated.

Study 2: Referendum and Medical Insurance for Mental Health

The second study tested the effect of a referendum format in the context of government medical insurance.

Method

Subjects were 87 people solicited in Philadelphia's main railroad station (36 women and 51 men with a mean age of 37). The questionnaire asked subjects to imagine that in a few years everyone in the United States would have at least a standard medical insurance package costing $1,000 per year on the average, paid through payroll deductions and taxes, with payment adjusted for ability to pay. The mental health coverage in the standard package would be limited to "up to 5 hours of psychotherapy per year and for up to 10 days in a mental hospital." At issue was WTP in order to extend the standard package so that it covered 25 hours of psychotherapy, 30 days in a mental hospital, or both. The questions appeared in three different orders: both, therapy, and hospital; therapy, hospital, and both; and hospital, therapy, and both.

In the referendum condition, the WTP question was followed by:

> If the typical answer to this question is high enough to cover the cost per person of the broader coverage, then the coverage will be increased and everyone will pay their share. Otherwise, the coverage will not be increased. The "typical answer" is the middle: half of the answers are higher, and half are lower.

Results

There was at least one $0 response from 49% of the subjects. Analysis of the remaining subjects yielded no embedding effects, nor were there effects of one versus two coverages. Accordingly, the analysis focused on the number of zero responses. The embedding effect was found in both conditions. That is, a zero response was more likely when a good was embedded than when it was presented first. The main result was that the referendum format did not reduce the effect. This and Study 1 are the first reported demonstrations of an embedding effect for number of zero WTPs.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Specifically, there were more zero responses for hospitalization when it followed "both" (60% in the referendum condition and 50% in the control condition) than when it came first in the questionnaire (15% referendum and 24% control). Likewise, there were more zeros for psychotherapy when it followed "both" (60% referendum and 50% control) than when it came first (31% referendum and 14% control). Log-linear analyses using order (first vs. second), referendum, and zero to classify subjects rejected a model in which zero responses were independent of order ($p = .001$ for hospitalization; $p = .012$ for psychotherapy). A model assuming that zero responses depended on order only was not rejected for either treatment. Hence, the effect of order seemed to be the same for both referendum and control conditions.

Study 3: Referendum and Medical Insurance for Cancer and Transplants

The third study used treatments that subjects were more likely to consider worth the expense of insuring.

Method

Subjects were 210 people solicited in Philadelphia's main railroad station (90 women and 120 men with a mean age of 43) and 106 students (52 women and 54 men with a mean age of 20), solicited by advertisements and paid $6 per hour for filling out questionnaires in a quiet room.

The questionnaire was like that of Study 2, except that the standard package did not cover organ transplants and "certain expensive cancer treatments." Only the rich would be able to afford extra insurance for these. We asked subjects how much extra they would be willing to pay (over $1,000) for extra coverage within the standard package for transplants, cancer treatments, and for both. At issue was whether their WTP for both was as high as the sum of their WTP for each of the two goods.

The questions appeared in three different orders: both, transplant, and cancer; transplant, cancer, and both; cancer, transplant, and both. We folded the questionnaire so that subjects answered the first question without seeing the last two to facilitate between-subjects analysis of the first question.

In the referendum condition, the WTP question was followed by:

> If most people (more than 50%) are willing to pay their share of the cost of the broader coverage, then it will be provided and everyone will pay their share. Otherwise, the broader coverage will not be provided.

Results

Both conditions showed an embedding effect. The referendum format did not reduce it, either in a between-subjects analysis of the first response or a within-subjects analysis of differences in WTP for the three questions. In the between-subjects analysis, log WTP for the first item did not depend significantly on whether it was single or double, in either the referendum or no-referendum condition, and condition did not interact with order. In the no-referendum condition, the geometric mean WTPs were $425 for the single good and $364 for the double good. In the referendum condition, the respective means were $467 and $364. The log of twice the WTP for the single good is significantly greater than the log WTP for the double good; $t(285) = 3.9$, $p < .001$; hence, the procedure is capable of detecting a difference. The "perfect" embedding effect was thus found for the first item presented.

Of 29 responses to the first item, 9.5% were zero; 28 of these responses were from nonstudents ($p < .001$ for the association). Otherwise, students and nonstudents did not differ in any way, nor were any effects of sex or age significant. The proportion of zero responses did not depend on referendum condition, on whether the double item was presented first, or on the interaction of these variables in a log-linear analysis.

In a within-subjects analysis, the log of the sum of WTPs for the two goods presented individually was higher than the log of the WTP for both goods together, significantly for both conditions (geometric means of $727 vs. $513 for the referendum condition and $633 vs. $491 for the no-referendum condition; $p < .001$ for both by a Wilcoxon test). This difference was significantly higher for the referendum condition ($p = .02$, U test)—a result opposite to the hypothesis—but it was unaffected by other variables.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Tests of Budget Constraints and Contribution and Warm Glow

Study 4: Accepting Tax Cuts vs. Paying Tax Increases

The results for ratings cast doubt on the budget-constraint hypothesis. Another test of this hypothesis (and related hypotheses that are based on substitution effects) is to compare WTP and willingness to accept (WTA). All of these hypotheses predict substantial sensitivity in WTA. In Study 4, we examined the quantity effect and the adding-up effect within subjects in both WTP and WTA for government programs. Insensitivity in WTA is also difficult to reconcile with a reverse form of the contribution account, in which subjects believes that the harm done is proportional to their WTA.

Method

Thirty-two students successfully completed a questionnaire; 17 did WTP judgments first, and 15 did WTA judgments first. The questionnaire began:

> This questionnaire concerns the value you place on various programs paid for by the U.S. government. We ask how much you are willing to pay for increases in these programs, or how much tax reduction you are willing to accept in return for cuts.

Subjects were asked to answer in mills, with one mill being 0.1% of current taxes. A table was provided with taxes of $1,000, $5,000, $10,000, and $50,000 as examples. For example, one mill was $10 additional tax for someone paying $10,000.

Subjects in the WTP-first condition were asked:

> What is the highest tax increase that you would be willing to have added to everyone's taxes for each of the following increases in government spending? Imagine that everyone in the U.S. was asked this question. If more than half said they were willing to have their taxes raised enough to pay for the increase, then government spending would be increased and taxes would be raised enough to pay for it. Also imagine that each item is the only proposal for increases this year.

Subjects then answered a test question about whether they had to write the most they were willing to pay for each increase, the least they were willing to pay for each increase, the least they were willing to accept, and so forth.

The WTA condition read:

> What is the lowest tax reduction in everyone's taxes at which you would be willing to see each of the following cuts in government spending? Imagine that everyone in the U.S. was asked this question. If more than half said they were willing to accept a tax reduction less than what could be saved by cutting the program, then the program would be cut and everyone's taxes would be reduced by the amount saved. Also imagine that each item is the only proposal for cuts this year.

The test question was the same. In addition to the 32 used, 23 subjects did not answer this question correctly, so their data were not used. Some of these subjects were interviewed, and they seemed to understand the task but not the test question itself. Therefore, this procedure was conservative.

The items were then presented in groups of five, as in the following sample configuration.

> (a) 10% increase in acquisition of land for national parks;
> (b) 20% increase in acquisition of land for national parks;
> (c) 10% increase in aid for family planning programs for the poor in the U.S;
> (d) 20% increase in aid for family planning programs for the poor in the U.S;
> (e) 10% increase in acquisition of land for national parks and in aid for family planning programs for the poor in the U.S.

(The word *cut* replaced *increase* in the WTA condition.) Comparison of 10% and 20% measures the quantity effect (insufficient sensitivity to quantity) and comparison of the two 10% items with the double item measures the adding-up effect.

The remaining pairs of items were based on the following programs in the order given here: aid for international family planning and women's health programs and aid for international vaccination programs against polio and the like; aid for vaccination of low-income children in the United States and aid to police departments for fighting crime; AIDS research and cancer research; cleanup of hazardous chemicals and drug treatment programs for users; enforcement of racial discrimination laws by the U.S. Department of Justice and enforcement of sex-discrimination laws by the U.S. Department of Justice; highway safety and low-income

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

housing; maintenance of national parks and public transportation construction; research on climate change and research on education; research on heart disease and support for United Nations peacekeeping activities.

Results

Sensitivity was no greater in the WTA condition than in the WTP condition. To measure insensitivity, we first averaged all the responses in each category for each subject in each condition (WTP and WTA): small (10%), large (20%), and both small goods together. To assess the quantity effect (a measure of insensitivity), we subtracted the log of the mean of the large goods from the log of twice the mean of the small goods. This measure would be zero if the rate of substitution were constant, that is, if subjects were completely sensitive to quantity and positive if they were insensitive. To assess the adding-up effect, we subtracted the log mean of the double good from the log of twice the mean of the two small goods. This measure would also be zero if the price of the double good were the sum of the prices of its components, and it would be positive if subjects are insensitive.

Table 2 shows the relevant geometric means across subjects. Statistical analysis confirmed that both the quantity effect and the adding-up effect occurred in both conditions (WTA and WTP). That is, the values of the large goods were less than twice the value of the small goods, and the value of both goods was less than twice the value of the small good; $t(31) > 4.5, p < .0005$, for all comparisons. The magnitude of the quantity effect (difference of logs) was exactly equal in the WTA and WTP conditions, and the magnitude of the adding-up effect was slightly, but not significantly larger in the WTA condition. The overall WTA versus WTP difference was not quite significant.

The quantity effect for WTP was correlated with the quantity effect for WTA across subjects ($r = .59, p < .0005$), and likewise for the adding-up effect ($r = .60, p < .0005$). Correlations between the two types of effects were also mostly significant but were not so high (WTP adding-up with WTP quantity, $r = .50, p = .004$; WTA adding-up with WTA quantity, $r = .31, p = .09$; WTP adding-up with WTA quantity, $r = .09$, *ns*; WTA adding-up with WTP quantity, $r = .44, p = .011$). In canonical correlation analysis we found only a single significant canonical correlate when the two adding-up measures (WTP and WTA) were predicted from the two quantity measures, but it found two significant correlates when the two WTP measures were predicted from the two WTA measures. In summary, different factors affected whether subjects show adding-up or quantity effects, but there was no evidence for different factors affecting WTA and WTP.

Table 2
Geometric Mean Responses by Condition

Condition	Small good	Large good	Both small
Willingness to pay	6.46	9.15	10.39
Willingness to accept	8.66	12.27	12.99

Note. Calculated in mills (one tenth of a cent).

The results cast further doubt on the budget-constraint hypothesis, but they are consistent with the prominence hypothesis, which holds that people base their response on the type of good rather than the quantity. One subject gave identical responses for both large and small magnitudes of the same good, although her responses were different for different goods. When asked why she gave the same response for the two different magnitudes, she said that if she had paid more for the larger amount of Good X, then that would have been more than what she was willing to pay for the smaller amount of Good Y and that would have been wrong because she considered Y to be more important that X." Respondents could have been making such comparisons implicitly, even when they were not asked about other goods in the same survey.

Study 5: Quantity Sensitivity in WTP vs. WTA

In Study 5, we compared again quantity effects in WTP and WTA. The good was provision of public safety by the University of Pennsylvania.

Method

Subjects were 44 paid volunteers, including 42 students, 18 from the University of Pennsylvania and 24 at the nearby Philadelphia College of Pharmacy and Science.

A questionnaire began with a summary of crime statistics reported by the university: about 150 per 100,000 students and employees were victims of reported crimes against persons (assault, robbery,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

and rape) each year. The questionnaire also summarized the university's public safety system: campus police, security personnel, emergency telephones and so forth. The items in the questionnaire asked about WTP for tuition increases for increased public safety (prorated for those with financial aid) or WTA for reductions resulting from decreased public safety. The expected effects of the increases or reductions were described in terms of the annual rate of crimes against persons per 100,000. For half of the subjects, the first item asked for WTP for a reduction from 150 to 100, or what we called WTP150–100. The other questions were WTA150–200, WTP200–150, WTP200–100, WTA100–150, and WTA100–200. The order was reversed for the remaining subjects.

Results

Subjects were equally insensitive to quantity for WTP and WTA. Also, WTA was larger than WTP. Geometric means were $1,067 for WTA150–200, $791 for WTA100–150, $1,227 for WTA100–200, $472 for WTP200–150, $404 for WTP150–100, and $729 for WTP200–100. Two subjects also said that they would not accept any amount for an increase of 50, and another subject said he would not accept any amount for an increase of 100. WTP insensitivity was defined as the log of (WTP200–150 + WTP150–100)/WTP200–100, and WTA insensitivity was analogous. This measure would be zero if subjects were perfectly sensitive and positive if they were insensitive. The mean insensitivity measures were .226 and .421 for WTP and WTA, respectively. Both were significantly greater than zero ($p < .0005$, t test), but they were not significantly different. WTA was significantly larger than WTP ($p < .0005$, t test) for each of the three comparisons (e.g., WTP150–100 vs. WTA100–150). The results did not depend on whether subjects were students at the University of Pennsylvania.

Study 6: WTP vs. WTA in a Between-Subjects Design

In Study 6, we compared WTP and WTA for quantity effects in a between-subjects design by including several goods at once.

Method

Each of 43 subjects (students as in the previous studies) saw only one amount of each good, either high or low in quantity. Each of 12 items had a large- and small-quantity version. The items (paraphrased with the smaller version in parentheses were saving 40,000 (4,000) acres [1 acre = 4,046.873 m^2] of Alaskan forest; preventing 10,000 (1,000) Americans from passing a kidney stone; saving 30,000 (3,000) of a rare species of fish; preventing 80,000 (8,000) drunk driving accidents; eliminating pollution from rivers in the eastern United States (the Schuylkill River); allowing 100 (10) dying men to see their only children once more; freeing 1,200 (120) innocent political prisoners; see your favorite football team deservedly win the Superbowl for 3 (1) years; restore sight to 90,000 (9,000) blind Americans; allow 1,100 (110) American athletes to participate in the Olympics; inform 10,000 (1,000) people falsely diagnosed as having a fatal disease that they do not have it; allow 15,000 (1,500) alcoholics to recover. Half of the subjects did the WTP condition first, and half did the WTA condition first. For half of the subjects in each order (WTA first and WTP first), high items were odd and low items even; for the other half, the reverse was true. We tested and paid 43 subjects.

The WTP instructions read:

> Imagine that you are an average American, paying $10,000 per year in taxes. How much would you be willing to pay in extra taxes to prevent each of the following events? That is, what is the largest tax increase that would be justified by prevention of each event? Please take every question seriously even if you find it ridiculous. Most people find these questions very difficult. Please take the time to think about them and give the most accurate and consistent answer you can.

The introduction to the WTA condition read:

> Now please indicate how much you would accept in reduced taxes to allow each of these events to happen. That is, how great would the tax reduction have to be in order for you to feel that each additional event is justified by the saving?

Results

Sensitivity to quantity was computed for each subject in each condition (WTA and WTP) as the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

log of the ratio of the geometric mean of the high-quantity items to the geometric mean of the low-quantity items divided by log(10). This is the value that the log ratio would have if subjects were proportionally sensitive. This measure would be zero if the subject were completely insensitive to quantity and one if subjects were proportionally sensitive. Subjects with more than 2 out of 8 zero responses or "no amount is enough" responses in any of the four conditions were omitted from analyses including that condition. Subjects were insensitive in the WTP condition (M = 0.04, not significantly greater than zero) but were somewhat sensitive in the WTA condition; $M = 0.10$, $t(29) = 2.81$, $p = .005$. However, both conditions were undersensitive, that is, less than 1.0 ($p < .0005$, t test). Sensitivity in the WTA condition was greater than in the WTP condition, $t(25) = 1.77$, $p = .045$, one-tailed.

No-amount-enough responses were confined to the WTA condition and did not depend on the size of the good (M = 0.77 out of eight possible large-good responses and 0.88 for small goods). Zero responses were more frequent in the WTP condition and did not depend on size (large WTP = 1.14; small WTP = 1.32; large WTA = 0.93 and small WTA = 0.79). Unlike Study 5, this study showed a large and highly significant ($p < .0005$, t test) difference between WTA and WTP values: geometric mean of $66 for WTP and $303 for WTA.

The evidence is consistent with the possibility that some of the subjects thought in terms of declining marginal utility of goods or budget constraints (increasing marginal disutility) for expenditures. Yet even the WTA conditions show very little sensitivity to quantity (.10), and these hypotheses would both predict considerable sensitivity, specifically, more than 1, assuming some insensitivity in the WTP condition.

Study 7: Matching Money to Risk vs. Risk to Money

Another test of the budget constraint hypothesis is to ask subjects to do the task in reverse, indicating the quantity of the good for which they would be willing to pay some given amount of money, rather than the amount they would pay for a given quantity of the good. If insensitivity to quantity is the result of budget constraints (real or perceived) or of declining marginal utility for the good, then subjects should be over sensitive in the new task. For example, they should want more than 10 times as much of the good for $50 than for $5. The same prediction seems to follow from the contribution and warm-glow hypotheses, both of which hold that judgments are accurate representations of the rate of substitution between the good and the subject's money within the given task.

Method

Each of 53 student subjects answered four pairs of questions after being told to assume that they were "a U.S. citizen living alone, making $40,000 per year." The introductions to the respective pairs were as follows:

> 1. Suppose that you are a typical driver. Statistics about death rates apply to you. Imagine a device that reduces death from a certain kind of auto accident. (Rates of injury are not affected.) Installation of the device is free, but each device requires a yearly maintenance fee for testing and adjustment.
> 2. Suppose that a flu epidemic is approaching. The flu is mild, but it can be fatal to some people who get it. A vaccine is available that prevents the flu.
> 3. Suppose there are small amounts of a cancer-causing chemical in your drinking water. The type of cancer has a 50% cure rate. The chemical can be removed by a filter, which must be replaced once a year.
> 4. Suppose you are in a foreign country and your work requires you to fly from one place to another several times. There are two airlines, A and B. Airline A has a poor safety record. Airline B has a perfect record but costs more. Your employer will pay only the cost of airline A, so you would have to pay the extra cost of B yourself. You must take all your trips on the same airline.

Half of the subjects were asked for their WTP for two levels of risk reduction after the first and third item, half after the second and fourth. An example follows:

> What is the largest amount that you would pay per year for the device if it prevented a type of accident that caused 1 (or 10) death per year for every 1,000,000 drivers?

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

The high and low risks differed by a factor of 10, and the order of the two items in each pair was counterbalanced. For the other two items, subjects got risk questions such as the following:

> How high would the chance of death from this kind of accident have to be in order for you to be willing to pay $500 ($50) per year for the device? Answer in number of deaths per year for every million drivers.

Again, amounts of money differed by a factor of 10, and order was counterbalanced. Note that for each subject, each scenario was given only with one type of question, WTP or risk.

Results

To measure insensitivity to quantity, we computed the log ratio of 10 times the small-amount answer to the large-amount answer for each item. The mean of these measures would be zero if the rate of substitution were constant, and they would be one if subjects ignored quantity. For the money questions, the means of these measures across subjects were positive and significant ($p < .001$, t test) for all four items (1.32 overall—on the basis of the means of the two items for each subject). This is just a replication of the quantity effect. For the risk questions, the measures were not significantly different from zero for any items or overall.

If subjects were basing their responses on a consistent marginal rate of substitution between risk and money, then the sum of the measures for the risk items and the money items would be zero. In fact, it was greater than zero across subjects; $M = 0.50$, $t(50) = 3{:}44$, $p = .001$. Insensitivity was uncorrelated with (geometric mean) WTP (dollars per unit of risk), so this result cannot be explained in terms of any dependence of rate of substitution on WTP. Thus, it appears that the insensitivity to quantity does not result from the slope of the indifference function for money and risk, as implied by the idea of budget constraints or declining marginal utility.

The fact that the log ratios for risk questions did not differ from zero was the result of a few subjects who were highly oversensitive in the direction predicted by budget constraints and many subjects who were oversensitive: 18% were oversensitive and 45% were undersensitive. The difference was significant by a sign test ($p = .026$, one-tailed). Thus, a plurality of subjects was undersensitive for risk, just as almost all subjects (88%) were undersensitive for WTP (none oversensitive). The fact that some subjects were oversensitive for risk suggests that people without special training can understand the concept of marginal rates of substitution, even though most people do not spontaneously apply this concept to judgment tasks.

The results of this study are analogous to those of Delquié (1993). Although he explains his results in terms of compatibility effects, the prominence explanation can also explain them. Such an explanation is consistent with the large number of subjects who simply gave the same response to both members of the pair, ignoring quantity totally (26% for WTP and 6% for risk responses did this at least once).

Study 8: Unit-Price WTP

In Study 8, we examined the quantity effect in two different methods of eliciting WTP. In the whole-good method, which is the usual one, the subject gives a WTP for 1 unit of a good and for 10 units. Sensitivity to quantity is the log ratio of the two prices. In the unit-price method, the subject gives a WTP for 1 unit and then a WTP per unit for 10 units. The subject then calculates WTP for the 10 units by multiplying the unit price × 10. The budget-constraint hypothesis implies that these two methods should give the same rate of substitution: to determine their unit price, subjects would first determine their whole-good price for 10 units and then divide by 10. Subjects with an implicit budget constraint would pay less per unit when the number of units in the purchase was larger. Likewise, contribution and warm-glow accounts make the same prediction.

Method

The items were the WTP versions of those used in Study 7. For each good in each condition, subjects provided their WTP for 1 and 10 units. The unit-price questions (with Item 1 as an example) read as follows:

> What is the largest amount that you would pay per year for each death prevented out of 1,000,000 for a device that prevented a type of accident that caused 10 deaths per year for every 1,000,000 drivers? How much is this for 10 deaths prevented out of 1,000,000? (Hint: Multiply your last answer by 10.)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

We used two whole-good conditions. The basic condition was identical to that used in Study 7. The divide-whole-good condition was identical except that immediately after pricing 10 units subjects were asked in one example, "How much is this for each death prevented out of 1,000,000?" The three conditions—unit-price, basic whole-good, and divide whole-good—were between-subjects. Within each condition, half of the subjects were asked about 1 unit, and then 10 for each item, and half were asked about 10 units and then 1. There were 27 subjects who completed the questionnaire. Subjects were asked to indicate any sources of difficulty in answering the questions.

Results

Sensitivity was defined for each item as log[WTP for 10)/(WTP for 1)]/log(10), yielding a value of zero for no sensitivity and one for proportional sensitivity. Sensitivity for each subject was the mean across items. Mean sensitivity was 0.36 for the basic condition, showing the usual insensitivity, 0.54 for the divide-whole-good condition and 1.04 for the unit-price condition. The unit-price condition differed significantly from the basic condition ($p = .001$, t test) and from the divide-whole-good condition ($p = .018$), but the basic condition and divide-whole-good conditions did not differ significantly. This result was identical to that obtained in a follow-up study in which subjects matched risk to money for half of the items. That study is not reported because it simply replicated the findings of Studies 7 and 8. Geometric mean WTPs, 1 and 10 units, respectively, were $24 and $196 for the unit-price condition, $17 and $50 for the divide-whole-good condition, and $33 and $79 for the basic condition. Order had no significant effect.

In the question about difficult aspects of the questionnaire, 2 subjects (out of 8 in the unit-price condition) mentioned the difficulty of answering the unit-price question. However, most complaints were about the lack of any standard for making the judgment.

In summary, the unit-price method induced perfect proportional sensitivity to quantity. The result casts doubt on any hypothesis in which responses are taken to reflect true economic value, such as the budget-constraint hypothesis. It is consistent with the prominence account, in that responses are dependent on the type of good rather than the amount of it. A similar result was reported by Kemp and Willetts (1995): Judgments of the value of government services correlated highly with judgments of value per dollar spent, despite large differences in expenditures.

Test of Availability

Study 9: Effect of Fuller Context

In two unreported studies, we found perfect sensitivity to quantity for medical insurance or risk reduction (Baron & Greene, 1995). Goods were presented in tabular form or on the same page, so that subjects could see all the goods before evaluating any of them. These findings suggested that putting the goods in a larger context can increase sensitivity. This supports the availability explanation of insensitivity. Loomis, Lockwood, and DeLacy (1993) also found what they considered to be only a small embedding effect in a CV study of forest preservation in Australia. The effect may have been small because the full context, all the possible forests, was presented at the beginning of the questionnaire. Even though subjects did not have to price the full context, its presence might have reduced the prices assigned to parts of the whole, compared with cases (not presented) in which the part was presented on its own, without any mention of the full context. In Study 9, we examined sensitivity to quantity with and without prior presentation of the full context.

Method

We tested 142 subjects. There were 51 women and 91 men in the study, 69 of whom were not students. We tested 45 in the laboratory, and 97 responded by electronic mail (e-mail) after volunteering in response to postings on various news groups. Laboratory subjects were paid $6 per hour. For each e-mail subject, $1.50 was put into a kitty, and 1 subject who was selected at random received the kitty either when it reached $99 or when the study was over.

The questionnaire asked about expensive cancer treatments and transplants, as in Study 3. Examples were given of transplants and cancer treatments, but no other details were provided, except that the standard policy cost $1,000 per year. Three orders were used: both, cancer, and trans-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3
Geometric Mean Responses by Condition

Condition	Both treatments	Sum of separate treatments
Single first with no preview	$155	$180
Both first	$182	$198
Single first with preview	$163	$165

plants; transplants, both, and cancer; cancer, transplants, and both. For half of the subjects who got cancer or transplants first, no other treatments were mentioned on the first page (or screen) of the questionnaire. The other half received a preview in which both treatments were described before the subjects were asked to evaluate one of them. Of course, the subjects who got both treatments first necessarily got the preview.

Results

There were 11 subjects, 9 of them from the e-mail nonstudent group, who gave zero answers either to all cases or all but one case. The student–nonstudent difference and the e-mail–lab difference did not quite reach significance by two-tailed Fisher tests. Otherwise students and nonstudent subjects did not differ either in overall WTP or in any effects. The same was true for e-mail and lab subjects.

Preview (presenting the context of both possible treatments) did increase sensitivity, but only in the within-subject measures. Between-subjects comparisons yielded perfectly proportional sensitivity. Table 3 shows the geometric mean WTP responses for the three conditions.

For within-subject analysis, the log of both was less than the log of the sum of the two separate treatments for the no-preview condition, $t(49) = 3.45$, $p = .001$, and for the both-first condition, $t(41) = 2.84$, $p = .007$, but not for the preview condition with one item given first. The single-first preview and no preview conditions differed significantly in the magnitude of this effect, $t(89) = 2.48$, $p = .008$, one-tailed).

Between-subjects analysis showed no difference between both-first and twice the WTP for the first item in any condition or overall, although the both-first WTP was larger than the raw WTP for the first item. In summary, no insensitivity was found between subjects.

The increase of within-subject sensitivity with preview suggests that the embedding effect does not require pricing of the larger good in order for prices assigned to the embedded good to be reduced. Subjects may need only to be reminded of the existence of the larger good as a context, as implied by the availability hypothesis. However, for practical purposes, asking subjects to price the superordinate good (as done by Kemp & Maxwell, 1993) may be more effective than simply reminding subjects of its existence (as done by Loomis et al., 1993).

Tests of Prominence

Study 10: Good–Money Versus Good–Good

The prominence explanation suggests that sensitivity can be increased by inducing subjects to attend more to the amount of the good. One way to do this might be to ask for responses in a more compatible medium (Tversky et al., 1988). For example, people might find it easier to trade off the size of a budget cut in government program A with the size of a cut in program B than to trade off a cut in A with their taxes. Such tradeoffs are often requested in multiattribute decision analysis, which has been suggested as an alternative to CV (Gregory, Lichtenstein, & Slovic, 1993). In Study 10, we compared sensitivity in this kind of good–good comparison with the usual good–money comparison.

Method

Twenty-six students completed a questionnaire, which began as follows:

> This questionnaire concerns the value you place on various programs paid for by the U.S. government. Some of the programs must be cut to reduce the budget deficit. But each of them could be saved from cuts if taxes are raised or if other programs are cut. Our concern is how much you are willing to pay to prevent reductions in each of these programs, or how much you are willing to give up of one program in order to save another.

The WTP condition, which half of the subjects

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

completed first, asked:

> What is the highest tax increase that you would be willing to pay in order to prevent each of the following reductions in government spending? Imagine that everyone in the U.S. was asked this question. If more than half said they were willing to pay enough to prevent the cut, then everyone's taxes would be raised by what was needed and the cut would not be made. Also imagine that each item is the only proposal for tax reductions this year. Feel free to use fractions or decimals in your answer.

Responses were in mills, as in Study 4. The questionnaire then presented the 20 goods used in Study 4, with 10% reductions and 50% reductions alternating. Then it presented the same goods with 50% and 10% switched.

The good–good condition asked subjects

> to evaluate one program in terms of another, in pairs. In each case, please fill in the blank so that the reduction in one program by the amount we give is just as bad as the reduction in the other program by the amount you write into the blank.
>
> Notice that, when you write a very small percent, that means you think that the program is very important, because a small cut in it is just as bad as a larger cut in the other program. You may write zero if you think that it is better to cut the other program than to cut the program with the blank by any amount.
>
> On the other hand, if you write a greater percent in the blank than the one we give you, that means you think that the program is less important. You cannot cut a program by more than 100%. So we will take answers of 100% to mean that you prefer the program with the blank to be cut completely rather than cut the other program by the amount we give.

Test questions followed, and 12 subjects (not included in the 26) did not answer these questions or answered them incorrectly.

The first two pairs read as follows:

> A. 10% reduction in acquisition of land for national parks.
> B. __% reduction in aid for family planning programs for the poor in the U.S.
> A. 50% reduction in aid for international family planning and women's health programs.
> B. __% reduction in aid for international vaccination programs against polio, etc.

Then the same 10 pairs were listed, with 50% and 10% switched. Finally, the 20 pairs were listed again with the first member left blank and the second member marked with a percentage (10% on odd items for the first 10 pairs and on even for the second 10).

Results

We defined sensitivity to quantity in the WTP condition as $\log(WTP_{50}/WTP_{10})/\log(5)$, for each pair, where the subscript indicates the size of the cut. Then we found the mean of these values for each subject. This measure is zero when the subject is completely insensitive and one when the subject is proportionally sensitive. We defined sensitivity in the good–good condition analogously by using the equivalent percentage reduction instead of WTP, excluding items with answers of 100%. Subjects might have wanted to indicate a higher ratio, but cuts of greater than 100% were impossible.

The mean sensitivity was 0.266 in the WTP condition and 0.773 in the good–good condition. The difference was significant; $t(25) = 8.12$, $p < .001$. In both conditions, sensitivity was less than one and greater than zero ($p < .001$). In summary, sensitivity to quantity was greater in good–good comparisons than in WTP judgments. In principle, this effect could be explained by budget constraints, but in Study 4 we ruled out this account for the same stimuli. More likely, subjects found good–money tradeoffs difficult, so, as a result, they paid less attention to the quantity of the good by using global importance as a cue for WTP.

Study 11: Producing Quantities and Prices

A second way to force attention to quantity of the good is to ask subjects to produce the quantity information about the good as well as about the money. In Study 11, we asked subjects to provide an amount for the good, their WTP for that amount, a second amount (larger or smaller than the first), and their WTP for the second amount. The study thus compared a one-blank condition, in which the subject writes WTP in one blank, to a two-blank condition, in which the subject writes the size of the cut in one blank and WTP in the other.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Method

We admininstered a questionnaire, as in Study 10, to 30 subjects. Instructions for the one-blank condition were also essentially identical. Instructions for the two-blank condition read:

> In this part, we give you each program twice, with two blanks for you to fill in each time. One blank is the size of the cut, in percent. The other blank is the the highest tax increase that you would be willing to pay in order to prevent it. You are to fill in both blanks. The first case in each pair should have a lower percent or a lower number of mills than the second. Try to imagine very different numbers in the two cases. If this is not clear, please ask.

The rest of the instructions were identical to the one-blank condition.

Half of the subjects did the one-blank condition first, and half did the two-blank condition first. For the former half, the small cut was on odd-numbered items and the large cut on even; for the latter half the reverse was true.

Results

We defined sensitivity to quantity for each pair of items as

$$\log(\mathrm{WTP}_{\mathrm{large}}/\mathrm{WTP}_{\mathrm{small}})/\log(\mathrm{size}_{\mathrm{large}}/\mathrm{size}_{\mathrm{small}}),$$

where size refers to the size of the cut and large and small refer to the larger and smaller cuts in each pair, respectively. Of course, for the one-blank condition, the size ratio was always 50:10. Sensitivity was 0 when responses were the same for both size cuts. Dividing by the log of the size ratio means that sensitivity is 1 when WTP is proportional to quantity. For each subject, we computed the geometric mean sensitivity in each condition. Of course, pairs on which one or both responses were 0 were omitted from this mean. Most commonly, both responses in a pair were zero rather than just one.

Mean sensitivity was 0.570 for the one-blank condition and 0.885 for the two-blank condition. These two means differed from each other; $t(28) = 3.10$, $p = .004$, two-tailed. In addition, the one-blank mean was significantly less than 1; $t(28) = 5.01$, $p < .0005$, but the two-blank mean was not significantly less than 1. The fact that mean sensitivity in the one-blank condition was higher than in Study 10 may be partly ascribed to a (nonsignificant) order effect: Sensitivity was higher when the two-blank condition was first (0.68 vs. 0.45). In the two-blank condition, subjects generally used ratios of cut sizes less than 5 to 1 (geometric $M = 3.34$, significantly <5); although the ratio was negatively correlated ($r = -.29$) with two-blank sensitivity, the correlation was not significant.

In summary, the two-blank condition made WTP almost proportionally sensitive to quantity. Presumably, asking subjects to fill in the size of the cut made them pay more attention to it.

General Discussion

Table 4 summarizes the results and their implications, indicating specifically where an hypothesized mechanism could not explain a result. Note that insensitivity might have more than one cause; therefore, a mechanisms inconsistent with one result might still explain some other result and might still produce insensitivity.

All the results but one are consistent with the prominence hypothesis, and most are explained by this hypothesis only. The one remaining result, the effect of context in Study 9, indicates that availability is operating as well. Although findings in Study 9 supported the availability hypothesis, findings in Studies 4, 10, and 11 indicated insensitivity even with full context (or with all questions in view). These results suggest that availability is not the only explanation. Still, the results of Study 9 support the practice of reminding subjects of the broader context (as recommended by NOAA, 1993).

In Studies 1–3, the use of a referendum format, with the amount of good specified so that subjects did not think it depended on their contribution, did not increase sensitivity. This result contradicts the contribution hypothesis. Moreover, this hypothesis does not mesh with other results. First, as noted earlier, insensitivity has been found for ratings as well as for pricing and for private goods as well as public goods.

Second, insensitivity for WTA as found in Study 4 is also a problem for the contribution hypothesis, according to which the subject would believe that the amount of harm increases with the amount of money accepted. In Study 4, we used a referendum format in both WTP and WTA conditions, specify-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 4
Results and Hypotheses That Failed to Explain Them (Indicated by Letter X)

Result and study	Hypothesized mechanism				
	Contribution	Warm glow	Budget	Availability	Prominence
Studies 1–3					
Sensitivity not increased by referendum or trigger price	X				
Studies 4–6					
Insensitivity in WTA as well as WTP	X		X		
Study 7					
No hypersensitivity with reverse match	X	X	X	X	
Study 8					
Good sensitivity with unit price	X	X	X	X	
Study 9					
Increased sensitivity with full context	X	X	X		X
Study 10					
Increased sensitivity with good-good match	X	X	X	X	
Study 11					
Increased sensitivity with production of good quantity	X	X	X	X	

Note. WTA = willingness to accept; WTP = willingness to pay.

ing the sizes of the cuts to be made (and hence the amount of harm) in the latter. It found little sensitivity in either condition.

An extension of the warm-glow hypothesis might explain insensitivity in WTA. The warm-glow hypothesis for WTP says that people pay for the good feeling of contributing, which depends on the amount of money they pay but not on the amount of the resulting benefits. The analog of this for WTA is that people accept money as compensation for guilt feelings arising from accepting money to allow harm, which depend on the amount of money but not on the amount of harm.

Both moral-satisfaction hypotheses, contribution and warm glow, hold that the overall size of the contribution is crucial. Therefore, neither form is consistent with the increase in sensitivity from unit pricing (see Study 7). Likewise, neither hypothesis can explain that many subjects were insensitive to the amount of money when they matched goods to money in Study 8. These studies thus discriminate the warm-glow hypothesis from the (similar) prominence hypothesis, which can explain them all. The differences is that the warm-glow hypothesis does not predict any effects of form of presenting the question.

If moral satisfaction is a problem, then efforts should be made to make the judgment independent of the subject's action of contributing or participating. This seems to be unnecessary. However, greater effects of participation may occur with different goods than those used here. Baron and Spranca (1995) found that certain goods, such as prevention of the destruction of nature, seem to evoke "protected values" that relate specifically to personal participation.

The budget-constraint hypothesis and related hypotheses have been put forward to argue that we need not worry about quantity insensitivity when it occurs in CV. The present results give almost no support to this hypothesis. Insensitivity is found for WTA, and it disappears when subjects are asked for unit prices, which suggests that the whole phenomenon depends on how the questions are asked, not on the true value of the goods in question. Subjects are not committed to the tradeoffs they express through standard WTP questions. However, this conclusion need not generalize to other methods of value elicitation, such as those used in decision analysis.

The prominence hypothesis is supported by the effects of manipulations designed to influence prominence: good–good matching and the two-blank condition, both of these drawn from standard practice in decision analysis. This hypothesis about the nature of insensitivity is consistent with all other results except the effect of providing full context. In all other studies, answers are governed

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

largely by perceived importance of issues, which is independent of quantity. In the adding-up effect, some averaging of importance may occur. Further research should address the source of importance judgments. In the case of public goods, subjects may answer according to whether they think more or less of the good should be provided than is currently being provided, taking into account the cost of provision as they perceive it. Of course, this is not the judgment that is needed for policy decisions, because people are supposed to judge the benefits of proposals alone, not their benefit/cost ratio.

The prominence hypothesis has two practical implications. First, subjects must be induced to attend to quantity of the good. Perhaps the simplest way to do this is to extend the method of Study 11 by asking subjects for an entire function relating WTP (or some other attribute) to the quantity of the good. This is standard practice in decision analysis. Second, it is necessary to check to see that subjects have taken quantity into account adequately. One way to do this is to reverse the judgment task, as in Study 7. Another is to ask for tradeoff judgments between two goods and between each good and money, then checking to see that the three functions are consistent. If the checks fail, then the subjects must be asked to repeat the judgments. These methods are also commonly used in decision analysis.

It may be argued that these methods increase the time required for CV analysis per subject. However, current CV practice nevertheless involves spending substantial amounts of time with each subject (NOAA, 1993). Moreover, the errors that result from insensitivity to quantity are large, sometimes several orders of magnitude. When such large errors can result from the procedure itself, the advantage of large samples of subjects is reduced. Much greater accuracy may be achievable by use of more extensive consistency checking of the judgments of fewer subjects.

References

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *Economic Journal, 100,* 464–477.

Baron, J. (1995). Rationality and invariance: Response to Schuman. In D. J. Bjornstad & J. Kahn (Eds.), *The contingent valuation of environmental resources: Methodological issues and research needs* (pp. 145–163). London: Edward Elgar.

Baron, J., Chen, L., & Greene, J. (1993, November). *Determinants of insensitivity to quantity in valuation of public goods.* Poster presented at the Annual Meeting of the Judgment and Decision Making Society, Washington, DC.

Baron, J., & Greene, J. (1995). [Perfect sensitivity to quantity]. Unpublished raw data.

Baron, J. & Spranca, M. (1995). *Protected values.* Manuscript in preparation. University of Pennsylvania.

Boyle, K. J., Desvousges, W. H., Johnson, F. R., Dunford, R. W., & Hudson, S. P. (1994). An investigation of part–whole biases in contingent valuation studies. *Journal of Environmental Economics and Management, 27,* 64–83.

Carson, R. T., & Mitchell, R. C. (1993). The issue of scope in contingent valuation. *American Journal of Agricultural Economics, 75,* 1263–1267.

Carson, R. T., & Mitchell, R. C. (1995). Sequencing and nesting in contingent valuation surveys. *Journal of Environmental Economics and Management, 28,* 155–173.

Delquié, P. (1993). Inconsistent trade-offs between attributes: New evidence in preference assessment bias. *Management Science, 39,* 1382–1395.

Diamond, P. A., & Hausman, J. A. (1994). Contingent valuation: Is some number better than no number? *Journal of Economic Perspectives, 8,* 45–64.

Diamond, P. A., Hausman, J. A., Leonard, G. K., & Denning, M. A. (1993). Does contingent valuation measure preferences? Some experimental evidence. In J. A. Hausman (Ed.), *Contingent valuation: A critical assessment* (pp. 3–38). Amsterdam: North Holland Press.

Dubourg, W. R., Jones-Lee, M. W., & Loomes, G. (1994). Imprecise preferences and the WTP–WTA disparity. *Journal of Risk and Uncertainty, 9,* 115–133.

Fischer, G. W. (1995). Range sensitivity of attribute weights in multiattribute value models. *Organizational Behavior and Human Decision Processes, 62,* 252–266.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 330–334.

Green, D. P., Kahneman, D., & Kunreuther, H. (1994). How the scope and method of public funding affects willingness to pay for public goods. *Public Opinion Quarterly, 58,* 49–67.

Gregory, R., Lichtenstein, S., & Slovic, P. (1993). Valu-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

ing environmental resources: A constructive approach. *Journal of Risk and Uncertainty, 7,* 177–197.

Guagnano, G. A., Dietz, T., & Stern, P. C. (1994). Willingness to pay for public goods: A test of the contribution model. *Psychological Science, 5,* 411–415.

Hanemann, W. M. (1994). Valuing the environment through contingent valuation. *Journal of Economic Perspectives, 8,* 19–43.

Jones-Lee, M. W., Loomes, G., & Philips, P. R. (1993). Valuing the prevention of non-fatal road injuries: Contingent valuation vs. standard gambles. *Oxford Economic Papers, 47,* 676.

Kahneman, D., & Knetsch, J. L. (1992). Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management, 22,* 57–70.

Kahneman, D., & Ritov, I. (1994). Determinants of stated willingness to pay for public goods: A study of the headline method. *Journal of Risk and Uncertainty, 9,* 5–38.

Kahneman, D., Ritov, I., Jacowitz, K. E., & Grant, P. (1993). Stated willingness to pay for public goods: A psychological perspective. *Psychological Science, 4,* 310–315.

Kemp, M. A., & Maxwell, C. (1993). Exploring a budget context for contingent valuation estimates. In J. A. Hausman (Ed.), *Contingent valuation: A critical assessment.* Amsterdam: North Holland Press.

Kemp, S., & Willetts, K. (1995). Rating the value of government-funded services: Comparison of methods. *Journal of Economic Psychology, 16,* 1–21.

Loomis, J., Lockwood, M., & DeLacy, T. (1993). Some empirical evidence on embedding effects in contingent valuation of forest protection. *Journal of Environmental Economics and Management, 24,* 45–55.

Margolis, H. (1982). *Selfishness, altruism, and rationality: A theory of social choice.* New York: Cambridge University Press.

McFadden, D. (1994). Contingent valuation and social choice. *American Journal of Agricultural Economics, 76,* 689–708.

McFadden, D., & Leonard, G. (1993). Issues in the contingent valuation of environmental goods: Methodologies for data collection and analysis. In J. A. Hausman (Ed.), *Contingent valuation: A critical assessment* (pp. 165–208). Amsterdam: North Holland Press.

Mitchell, R. C., & Carson, R. T. (1989). *Using surveys to value public goods: The contingent valuation method.* Washington, DC: Resources for the Future.

National Oceanic and Atmospheric Administration. (1993). Report of the NOAA panel on contingent valuation. *Federal Register, 58*(10), 4602–4614.

Portney, P. R. (1994). The contingent valuation debate: Why economists should care. *Journal of Economic Perspectives, 8,* 3–17.

Schuman, H. (1995). The sensitivity of CV outcomes to CV survey methods. In D. J. Bjornstad & J. Kahn (Eds.), *The contingent valuation of environmental resources: Methodological issues and research needs* (pp. 75–96). London: Edward Elgar.

Thaler, R. H. (1993, August). Paper presented at SPUDM 14, Aix-en-Provence, France.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95,* 371–384.

Zakay, D. (1990). The role of personal tendencies in the selection of decision-making strategies. *Psychological Record, 40,* 207–213.

Received February 6, 1995
Revision received October 18, 1995
Accepted October 26, 1995 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.