Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method

By RONALD G. CUMMINGS AND LAURA O. TAYLOR*

Several recently published studies concerned with the validation of willingness-to-pay estimates derived with the contingent valuation (CV) method have utilized methods similar to experimental economics' methodologies where experiments are conducted with groups of subjects and responses to both hypothetical and real valuation questions are obtained.[1] It is often—but not always—the case that results from such experiments demonstrate significant differences between responses to the real and hypothetical valuation questions.[2] Such differences are commonly interpreted as reflecting "hypothetical bias." Concern with hypothetical bias has motivated a growing number of researchers to explore techniques that might eliminate such bias, thereby providing methods from which unbiased estimates for willingness to pay might be obtained with the CV method.

One method used in efforts to derive unbiased value estimates relies on calibration techniques. For example, McKinley Blackburn et al. (1994) and Fox et al. (1999) use within-sample techniques where subjects respond to hypothetical and then real valuation questions. A calibration function is then estimated which relates differences in responses obtained in the two treatments to subject characteristics. William Schulze et al. (1999) explore similar techniques where hypothetical willingness to pay is "calibrated" by subjects' self-reports of embedding.[3] While calibration may prove to be a useful tool for *ex post* adjustments of stated preference values, its practicality appears to be limited by the extent to which a calibration function derived for one good can be used to calibrate hypothetical values derived for a different good. Indeed, Fox et al. (1999 p. 2) note that: "As our results show, calibration appears to be commodity-specific, and as such, [the calibration approach] must proceed on a case-by-case basis until enough evidence exists to reveal any systematic bias that can eventually lead to a general calibration function."[4]

A second and very different approach to dealing with the hypothetical bias problem focuses on designs for the CV questionnaire that might *directly* induce subjects to provide responses to hypothetical valuation questions that correspond with responses observed when actual cash payments are involved. To date, efforts to implement this approach have relied on adding to the CV script brief requests that subjects consider budgetary substitutes in responding to the valuation question. Problems encountered with this particular approach are illustrated by Loomis et al. (1994)

* Department of Economics, School of Policy Studies, Georgia State University, Atlanta, GA 30303. The authors acknowledge financial support for this work from the U.S. Environmental Protection Agency Contract No. R824710-01-0. Thanks are due to Nicholas Flores, Bill Schulze, and two anonymous reviewers for helpful comments. The authors also gratefully acknowledge the participation of Glenn Harrison in early stages of this research, and the assistance of Harold Ball, Kelly Brown, Kate Gardner, James Murphy, Tom Stevens, and Melonie Williams in conducting experiments.

[1] See, as examples, John Loomis et al. (1994), Helen R. Neill et al. (1994), Cummings et al. (1995a), Thomas C. Brown et al. (1996), and John Fox et al. (1999).

[2] Two studies report estimates of willingness to pay where hypothetical bias is not identified (David Brookshire and Don Coursey, 1987; Mark Dickie et al., 1987). V. Kerry Smith and Carol Mansfield (1997) fail to find hypothetical bias in an experiment based on willingness to accept. As is discussed later, we also fail to find evidence of hypothetical bias in one particular set of experiments.

[3] The methodology used in Schulze et al. (1999) may be described as "adjustment" rather than "calibration" since "true" values (those observed when actual payments are required) which form the basis for calibration are not obtained in their experiments.

[4] Similar results are commonly reported in the marketing literature. For example, Jordon Louviere (1996 p. 170) observes that: "I think it very unlikely that one simple and generalizable approach to 'adjusting' (stated preference) numbers will be found, despite clear evidence supporting strong monotonic links between stated preferences and actual choices in real markets . . . differences in product awareness, learning, etc. . . ., necessarily imply that no one magic constant can exist across all product categories."

and Neill (1995). In both studies, subjects are reminded of budgetary substitutes by explicit discussions of alternative goods and their costs prior to a valuation question for a specific good. In both cases, these particular "reminding" processes were ineffective in removing hypothetical bias. In addition, Loomis et al. (1994) also directly asked subjects to consider their budget constraint during a valuation experiment and again results suggest that this verbiage was ineffective in removing hypothetical bias.

The challenge remains of finding a hypothetical valuation design that *demonstrably* provides unbiased value estimates for public goods. This paper reports results from the authors' efforts to respond to this challenge. An alternative design for hypothetical CV questionnaires is introduced which takes a somewhat unique approach to the process of inducing subjects to provide unbiased responses to hypothetical valuation questions. Rather than attempting to remove hypothetical bias through references to budget constraints and budgetary substitutes, the design developed here includes an explicit discussion of the hypothetical bias problem—what hypothetical bias is and why it might occur. This design is referred to as the "cheap talk" design following the use of this term in the information, bargaining, and game-theory literature,[5] and it makes the issue of hypothetical bias an integral part of the CV questionnaire.

Results from 16 different experiments are presented which are developed to test the efficacy of the cheap talk design in eliminating hypothetical bias. We begin with a set of eight baseline experiments involving four different public goods. Two sets of referenda are conducted with each good: a real referendum where subjects vote on a proposition that requires the payment of money if the referendum passes; and a hypothetical referendum where different subjects vote on the same proposition but payment for and provision of the good is hypothetical. Using data obtained from these referenda, we test the hypothesis that there is no significant difference between YES responses to the proposition presented in the real and hypothetical referenda. In instances where this hypothesis is rejected—results consistent with hypothetical bias—we conduct hypothetical referenda that use the cheap talk design. The efficacy of the cheap talk design in eliminating any hypothetical bias observed in the baseline experiments is assessed by testing the hypothesis that there is no significant difference between YES responses to the proposition presented in the real and hypothetical-with-cheap talk referenda. In the case of every good for which hypothetical bias is identified in the baseline experiments, we find that the cheap talk design is effective in eliminating the bias. The robustness of this result is also tested by changes in the cheap talk script and by changes in the experimental design.

The paper is organized in the following manner. Section I introduces basic design features of our experiments: the questionnaires, the experimental procedures, and subject recruitment. Results from baseline real and hypothetical experiments are presented in Section II. Results from experiments using the cheap talk design are presented in Section III. Section IV introduces and presents results from experiments that test the robustness of the cheap talk design to changes in the cheap talk script and changes in experimental design. Closing comments are provided in Section V.

I. Experimental Design

Basic design elements for our experiments consist of the questionnaire(s), subject recruitment, and the procedures used in conducting the experiments. Each of these elements are discussed in turn.

A. *Questionnaires*

Questionnaires used in all of our experiments have a structure that is common in CV studies: a "good" is described, the payment mechanism and provision rule is described, and then the willingness-to-pay question is posed. Since we use a referendum format, the willingness-to-pay

[5] Cheap talk refers to the costless of transmission of signals and information (i.e., cheap talk does not *directly* affect the payoffs of players in a game). Many game theorists typically use the term "cheap talk" in referring to nonbinding communication of actions by two or more players in a game prior to their actual binding commitment (Joseph Farrell and Robert Gibbons, 1989; Steven Matthews et al., 1991; Farrell and Matthew Rabin, 1996). As an anonymous reviewer has pointed out, we use the term "cheap talk" in a parallel way, referring to nonbinding communication of actions by two or more players in an experiment prior to their *hypothetical* commitment.

question takes the form of a proposition on which subjects vote. The questionnaire format is the same across all four of the public goods used in our experiments; they differ only in the description of the public good. The goods differ in terms of what is being delivered, where it is being delivered, and the specificity of the connection between how much is donated to the good and "how much" of the good is provided.

To illustrate the questionnaire format, consider one of the four public goods used in our experiments—the "Nature Conservancy (NC) good." The NC good involves contributions to the Georgia Chapter of the Nature Conservancy, a nonprofit organization that works to protect natural habitats in the state of Georgia. The Chapter has a land acquisition program which directly purchases lands in Georgia that are identified as containing unique and/or endangered flora and fauna to preserve them in their natural state. The NC good is then a contribution to the Georgia Chapter of the Nature Conservancy. With N subjects voting in a referendum requiring all participants to pay \$10 if the referendum passes, $(N \times 10)$ dollars could be contributed to the Conservancy. Respondents are told these monies will be used solely for these land acquisitions. In all referenda using the NC good, subjects vote on the following proposition.

THE NC PROPOSITION: *Everyone here in the room will contribute \$10.00 to the Georgia Chapter of the Nature Conservancy. The contribution is to be used for the purpose of purchasing additional lands in the state of Georgia to be protected and held in stewardship by the Nature Conservancy.*

After presenting the proposition, the payment mechanism and provision rule are explained. A standard majority voting rule is used: if more than 50 percent of the subjects vote YES, then the referendum passes, everyone pays \$10.00, and a $N \times \$10.00$ contribution is immediately mailed to the Nature Conservancy. If 50 percent or less of the subjects vote YES, no one pays \$10.00 and the money is not sent to the Conservancy.

We use three basic questionnaires: one for the real, the hypothetical, and the hypothetical-with-cheap talk referenda. The only difference between questionnaires used for a given good in real and hypothetical referenda is that in hypothetical referenda subjunctive language is used in describing the referendum. In the real referendum active language is used, such as "you *will* vote on the following proposition" and "*all* of you *will* pay \$10.00" (italics added). In the hypothetical referendum the active language of the real referendum is replaced with "I want you to *suppose we were to have* a secret vote," "supposing that we *were* to have such a referendum," and "*all* of you *would* pay \$10.00" (italics added). Notwithstanding subjects' knowledge that the referendum is hypothetical in the sense that even if the provision rule is satisfied no one actually pays \$10.00, they are also asked to vote as they *would vote* if payment were actually required.

In the hypothetical-with-cheap talk referenda (hereafter just "cheap talk referenda") the script is exactly the same as in the hypothetical referenda with one exception: additional verbiage (the cheap talk script) is introduced just prior to the vote. The cheap talk script makes three general points: it describes the hypothetical bias phenomena; it discusses possible explanations for this phenomena—why subjects *might* vote differently in real and hypothetical referenda; and it requests that subjects vote in the upcoming hypothetical referendum as if it were a real referendum. The following excerpts from the cheap talk script illustrate its substance:

> . . . in a recent study, several different groups of people voted on a referendum just like the one you are about to vote on. Payment was hypothetical for these groups, as it will be for you. No one had to pay money if the referendum passed. The results of these studies were that on average, across the groups, 38 percent of them voted "yes." With another set of groups with similar people voting on the *same* referendum as you will vote on here, *but* where payment was real and people *really* did have to pay money if the referendum passed, the results on average across the groups were that 25 percent voted yes. That's quite a difference, isn't it?
>
> We call this a "hypothetical bias." Hypothetical bias is the difference that we continually see in the way people respond to hypothetical referenda as compared to real referenda . . . .

> How can we get people to think about their vote in a hypothetical referendum like they think in a real referendum, where if enough people vote "yes," they'll really have to pay money? How do we get them to think about what it means to really dig into their pocket and pay money, if in fact they really aren't going to have to do it?
>
> Let me tell you why I think that we continually see this hypothetical bias, why people behave differently in a hypothetical referendum than they do when the referendum is real. I think that when we hear about a referendum that involves doing something that is basically good—helping people in need, improving environmental quality, or anything else—our basic reaction in a hypothetical situation is to think: *sure,* I would do this. I *really would* vote "yes" to spend the money . . . .
>
> But when the referendum is real, and we would actually have to spend our money if it passes, we think a different way. We basically still would like to see good things happen, but when we are faced with the possibility of having to spend money, *we think about our options:* if I spend money on this, that's money I don't have to spend on other things . . . we vote in a way that takes into account the limited amount of money we have . . . . This is just my opinion, of course, but it's what I think may be going on in hypothetical referenda.
>
> So if I were in your shoes . . . I would ask myself: if this were a real referendum, and I had to pay $10.00 if the referendum passed: do I *really want* to spend my money this way? If I really did, I would vote yes; if I didn't, I would vote no . . . .
>
> In any case, I ask you to vote just exactly as you would vote if you were really going to face the consequences of your vote: which is to pay money if the proposition passes. Please keep this in mind in our referendum.

As the excerpts indicate, our alternative design for the hypothetical referenda involves the use of cheap talk to make the substance of hypothetical bias an integral and straightforward consideration in the valuation process of subjects.[6] Complete scripts for all experiments are available upon request.

B. *The Goods*

Experiments were conducted using three public goods in addition to the Georgia Nature Conservancy good described above. These goods are described in turn below.

The "Albuquerque (ABQ) good" involves contributions to the Southwest Research and Information Center, a nonprofit organization located in Albuquerque, New Mexico. Contributions are used for the purpose of funding the publication and distribution of a bilingual (English and Spanish) "Citizens Guide" that is to be distributed to low-income, Hispanic households in Albuquerque, New Mexico. The Guide informs households in an area overlying a potentially contaminated aquifer as to whether or not their drinking water wells are likely to be contaminated by toxic substances, how they can have their water tested at no cost to them, and the alternative actions that are available to them if they find that their well is indeed contaminated. The Center's cost for publishing and distributing this Guide is $5.00 per Guide. Thus, with N subjects in a group, each contributing $10.00, $N \times 2$ Guides can be funded by the group. In all referenda using the ABQ good, subjects vote on the following proposition.

THE ABQ PROPOSITION: *Everyone here in the room will contribute $10.00 to the Southwest Research and Information Center. The contribution is to be used for the purpose of preparing and distributing the Citizen's Guide to households in the area affected by groundwater contamination.*

The experimenter reads the proposition aloud at the same time it is shown on an overhead and inserts the number of households that would receive a booklet if the referendum passes (equal to $N \times 2$ households) where the overhead reads "household in the area."

The Albuquerque good was used by Cummings

[6] In an attempt to control for the sincerity and/or intensity with which the cheap talk script was read to subjects, one experimenter was chosen to conduct all cheap talk experiments and all but one of the robustness-test experiments for cheap talk (which are described later). This experimenter also conducted many of the other experiments for each good and each experimental treatment.

et al. (1997) in their study of the incentive compatibility of referendum formats used in CV studies. Cummings et al. (1997) conducted real and hypothetical referenda using this good.[7] Our experimental procedures follow that of Cummings et al. (1997) closely and, importantly, the scripts used to describe the referendum rules in the real and hypothetical referenda were identical. Any differences in experimental procedures between Cummings et al. (1997) and our experiments are incorporated into the analysis conducted in Sections II, III, and IV.

A third good we use in our experiments involves contributions to the Nature Conservancy's *Adopt an Acre* program.[8] This program purchases rain forest acreage that most need protection through partnerships with local agencies in Costa Rica. After describing rain forests to subjects and some of their benefits, the *Adopt an Acre* program is presented. This program allows private parties to "adopt an acre" of the rain forest in Costa Rica at a cost $35.00 an acre. These funds are used to actually purchase the acre of forest and to cover local on-the-ground costs of protection. Subjects are told that 100 percent of the funds will be used solely for the purchase and protection of rain forest acres—none will be used to cover administrative costs of the project. They are also told that they will not actually own the land—they will only be "honorary owners" of the acres whose purchase they fund. With N people participating in a referendum, $(10 \times N)/35$ acres can be adopted by the group. All subjects voting on the "Rain Forest (RF) good" vote on the following proposition.

THE RF PROPOSITION: *Everyone here in the room will contribute $10.00 to the Nature Conservancy. This contribution will be used to adopt acres of the rain forest in Costa Rica.*[9]

The final good we use in our experiments is the "Path Foundation (PF) good," which involves contributions to the Path Foundation—a nonprofit organization located in Atlanta, Georgia. The Path Foundation is building and maintaining a system of pedestrian and bicycling trails throughout the city of Atlanta and its surrounding areas. Subjects are shown maps which indicate where trails have been completed and where they were still under construction at the time of the experiment. The Path Foundation good is then contributions to the Path Foundation to help finish a specific three-mile segment of a greenway. With N subjects voting in the referenda, $(N \times 10)$ dollars could be donated to the Path Foundation. All subjects in these referenda vote on the following proposition.

THE PF PROPOSITION: *Everyone here in the room will contribute $10.00 to the Path Foundation. The contribution is to be used specifically for construction costs associated with completing the greenway extending from Freedom Park to Ponce de Leon Avenue.*

C. *Recruitment and Procedures*

Subjects for our experiments were recruited from students enrolled in undergraduate courses at Georgia State University during the 1995 and 1996 academic quarters. Subjects sign a consent form in which they acknowledge their voluntary participation in the experiment, agree to abide by the rules of a referendum, and acknowledge their receipt of a $10.00 participation fee. Subjects then participate in a market-based experiment—a series of oral double auctions[10]—that require approximately 50 to 60 minutes to complete. The rationale for including the oral double auctions as a part of our experimental procedures relates to our payment of a participation fee. For control across valuation experiments, we pay every group the same participation fee, regardless of whether they are to participate in a real or hypothetical referendum. We must acknowledge that subjects' decision-making behavior can be affected by their receipt of a participation fee. Such effects are described as "endowment"

[7] We gratefully acknowledge the willingness of the authors to allow us to include their data in this study. Two-hundred and eighty-eight observations from hypothetical and real referenda are pooled from the Cummings et al. (1997) study.

[8] Experiments using the Rain Forest good were conducted as a part of a companion project examining cultural differences in valuation. A full description of this project is contained in Cummings et al. (1996).

[9] The experimenter reads the value for $(10 \times N)/35$ acres where the proposition reads ". . . to adopt acres . . . ."

[10] See Douglas D. Davis and Charles A. Holt (1993 Appendix A.1) for an outline of how a oral double auction is conducted.

and "found money" effects (see E. Elisabet Rutström, 1998) and may arise in instances where subjects view the participation fee as money that must be spent in the experiment or, as relevant to our experiments, subjects do not accept the fee as becoming a part of their disposal income. Our use of the oral double auction is an effort to deal with this problem. Results from debriefing sessions following pretests of the experimental process used here suggested the absence of found money effects. Subjects reported they felt that they had earned the participation fee after the hour to an hour and a half required for the experiment. Of course, we cannot be positive that this design succeeds in eliminating behavior consistent with found money effects in all cases. However, we are encouraged by results from robustness tests (described below) which address this experimental design issue.[11]

Following the oral double auctions, subjects complete a brief questionnaire that requests sociodemographic information and then the referendum is introduced and conducted.

II. Results from the Baseline Experiments

We refer to "baseline experiments" as the initial set of experiments that conduct real and hypothetical referenda on each public good (see Table 1 for a summary of all experiments). The referenda were conducted with group sizes of no greater than 30 students and no subject participated in more than one referendum.[12] Voting outcomes and selected data summaries from the baseline experiments are presented in Table 2. Note the variation in the percentage of YES votes across goods. The percent of YES votes was highest for the Rain Forest good and lowest for the Path Foundation across each referenda design. The differences in voting outcomes across goods may be viewed as suggesting that our respondents distinguished between goods as opposed to responding to a generic "good cause" proposition (see Smith, 1996; Smith and Mansfield, 1997).

To identify whether or not voting behavior across goods is consistent with the presence of hypothetical bias, we begin the analysis with simple contingency tables and measures of association. As indicated in Table 2, a greater percentage of respondents voted YES in the hypothetical referenda as compared to the real referenda for three of the four goods. As compared to the real referenda, YES responses in hypothetical referenda were 17.3, 16.7, and 19.4 percentage points higher for the NC, ABQ, and RF goods, respectively. Pearson χ^2 and Fisher's exact tests are used to test the null hypothesis that the YES responses in these real and hypothetical referenda (for each good) are independent of payment condition. The results of these tests are presented in Table 3. For the NC, ABQ, and RF goods, we *reject* the hypothesis that there is no significant difference between voting behavior in the real and hypothetical referenda at no less than the 98-percent level of confidence (see Table 3). Thus for these three goods, we find evidence consistent with hypothetical bias.[13]

However, there was little difference between the percentage of YES votes in the real and hypothetical referenda for the Path Foundation good (only a 2.9-percentage-point difference). We *fail to reject* the hypothesis that there is no significant difference between YES responses in the hypothetical and real referenda for this good (the Fisher's 2-sided exact p-value is 0.682). Because we failed to find evidence of hypothetical bias in the referenda for the Path Foundation good, we did not conduct cheap talk referenda using this good as part of our initial research design. However, a related robustness test is discussed in Section V.

[11] See David Bjornstad et al. (1997) for a more detailed discussion of the rationale for using the oral double auction in this type of valuation experiment.

[12] The maximum group size was 40 students for the rain forest experiments and 35 in the robustness experiments which involved no oral double auction (these experiments are described in Section V).

[13] Tests presented in Section III using probability models also support this conclusion. We note that the evidence presented here, leading us to conclude there is hypothetical bias present, could also be consistent with free-riding behavior since the referenda we conduct are not a strictly closed referendum (see Alan Randall, 1996). To address this issue, Taylor (1998) conducts *closed* referenda on the ABQ good. Her results from 153 observations indicate that YES responses in the closed-hypothetical referenda were 18 percentage points higher than in the closed-real referenda—a difference commensurate with that in the experiments reported here.

TABLE 1—SUMMARY OF EXPERIMENTS

Abbreviation	Description
	Baseline experiments:
NC-R, ABQ-R, RF-R, PF-R	Real referenda using the Nature Conservancy good, the Albuquerque good, the Rain Forest good, and the Path Foundation good, respectively
NC-H, ABQ-H, RF-H, PF-H	Hypothetical referenda using the Nature Conservancy good, the Albuquerque good, the Rain Forest good, and the Path Foundation good, respectively
	Cheap talk experiments:
NC-CT, ABQ-CT, RF-CT	Hypothetical referenda with the cheap talk script using the Nature Conservancy good, the Albuquerque good, and the Rain Forest good, respectively
	Robustness experiments:
ABQ-MCT	Hypothetical referenda with the *modified* cheap talk script using the Albuquerque good
NC-CT$, ABQ-CT$	Hypothetical referenda with the cheap talk script where the referenda are *not* preceded by the oral double auction and subjects are *not* paid a participation fee, using the Georgia Nature Conservancy and Albuquerque goods, respectively
NC-MCT$, ABQ-MCT$	Hypothetical referenda with the *modified* cheap talk script where the referenda are *not* preceded by the oral double auction and subjects are *not* paid a participation fee, using the Georgia Nature Conservancy and Albuquerque goods, respectively

III. Results from the Cheap Talk Experiments

Results from our cheap talk experiments are given in Table 2. Testing the effectiveness of the cheap talk design in eliminating hypothetical bias from the hypothetical valuation experiments is accomplished through a series of nonparametric and parametric tests. First, using measures of association, we test the hypothesis that YES responses in the two types of hypothetical referenda—those with and without cheap talk—are independent of treatment. For each of the goods used in cheap talk experiments we *reject* the null hypothesis that voting responses are the same in the hypothetical referenda and the hypothetical referenda with cheap talk (see Table 3). These two hypothetical survey instruments, which vary only by the insertion of the cheap talk script, yield responses that are statistically different: the responses to the cheap talk script are significantly *lower* than those in the hypothetical survey without the cheap talk script.

Our primary concern is with the relationship between voting responses in the real referenda and those observed in the cheap talk referenda. The percentage of YES responses to the cheap talk referenda differ from those obtained in real referenda by only 1.7, 4.8, and 0.4 percentage points for the NC, ABQ, and RF goods, respectively. We *fail to reject* the hypothesis that there is no significant difference between YES responses in the real and cheap talk referenda for all three public goods at any conventional level of significance (see Table 3). In this case, the measures of association suggest that the cheap talk script is effective in providing responses to *hypothetical* referenda that comport with responses observed when *actual cash payments* may be required by the respondent as a result of their voting choice.[14]

[14] Note that the failure to reject a hypothesis does not provide statistical inference, and we cannot *prove* that there are no differences between the responses to the real and cheap talk referenda. However, considering the results indicating that

TABLE 2—REFERENDA RESULTS AND SELECTED DATA SUMMARIES FOR THE BASELINE EXPERIMENTS

Referenda treatment[a]	Number of participants	YES responses (percent of total)	NO responses (percent of total)	Mean age[b,c]	Mean income[b,d]	Percent male[c]	Percent married[c]
			Nature Conservancy good				
NC-H	115	49 (42.6)	66 (57.4)	24.8 (6.2)	35.6 (19.2)	40.7	20.3
NC-R	71	18 (25.3)	53 (74.6)	25.1 (4.6)	30.9 (20.7)	45.6	19.1
NC-CT	74	20 (27.0)	54 (73.0)	25.6 (7.6)	37.1 (19.2)	45.7	18.6
			Albuquerque good				
ABQ-H	211	98 (46.4)	113 (53.6)	36.4 (15.3)	38.8 (18.3)	50.0	37.0
ABQ-R	182	54 (29.7)	128 (70.3)	35.1 (15.9)	37.3 (19.8)	48.3	38.1
ABQ-CT	84	29 (34.5)	55 (65.5)	22.4 (3.5)	34.0 (20.3)	50.6	9.6
			Rain Forest good				
RF-H	63	43 (68.3)	20 (31.7)	22.1 (4.8)	14.9 (17.7)	41.3	9.5
RF-R	90	44 (48.9)	46 (51.1)	26.2 (5.0)	27.5 (18.8)	58.9	32.6
RF-CT	33	16 (48.5)	17 (51.5)	23.2 (5.6)	13.5 (14.1)	45.5	15.1
			Path Foundation good				
PF-H	97	21 (21.6)	76 (78.4)	26.8 (6.8)	28.1 (20.7)	52.3	24.4
PF-R	49	12 (24.5)	37 (75.5)	21.0 (2.4)	40.1 (17.7)	50.0	8.3

[a] See Table 1 for variable definitions.
[b] Standard deviations are in parentheses.
[c] Means or percentages for some groups are based on less than the full sample due to nonresponses.
[d] Income is reported in thousands and is based on the midpoint of an interval response to a question asking the monthly after-tax income of the household. Intervals were 0–300, 301–400, 401–500, 501–600, 601–800, 801–1,000, 1,001–2,000, 2,001–3,000, 3,001–4,000, over 4,000.

To incorporate respondents' socioeconomic characteristics and features of the experimental design into the analysis, we present a set of tests based on probability models. The socioeconomic characteristics considered in the analysis are the respondents' age, marital status, gender, income, and race. Experimental design variables that could vary across respondents are the number of subjects in the experimental session, the earnings of the respondent from the oral double auction,[15] and the referendum treatment (e.g., hypothetical, real, or cheap talk). In addition, because we pool data from Cummings et

responses to the cheap talk referenda *are not* the same as in the baseline *hypothetical* referenda, these "failures to reject" across each set of cheap talk experiments provides reasonable support for the conclusion that the responses to hypothetical referenda with the cheap talk script inserted are consistent with those observed in the baseline *real* referenda.

[15] The mean earning in the oral double auction was $1.39 (standard deviation is $0.49), with a minimum value of $0.00 and a maximum value of $3.20.

TABLE 3—NONPARAMETRIC TEST RESULTS FOR THE BASELINE AND CHEAP TALK EXPERIMENTS

Null hypothesis[a]	Pearson χ^2 (p-value)	Fisher's exact p-value 1-sided	Fisher's exact p-value 2-sided	Conclusion
		Nature Conservancy good		
NC-R = NC-H	5.672 (0.017)	0.012	0.019	Reject null hypothesis
NC-H = NC-CT	4.716 (0.030)	0.021	0.032	Reject null hypothesis
NC-R = NC-CT	0.053 (0.819)	0.484	0.852	Cannot reject null hypothesis
		Albuquerque good		
ABQ-R = ABQ-H	11.593 (0.001)	0.000	0.001	Reject null hypothesis
ABQ-H = ABQ-CT	3.483 (0.062)	0.041	0.069	Reject null hypothesis
ABQ-R = ABQ-CT	0.6307 (0.427)	0.256	0.477	Cannot reject null hypothesis
		Rain Forest good		
RF-R = RF-H	5.666 (0.017)	0.013	0.021	Reject null hypothesis
RF-H = RF-CT	3.5731 (0.059)	0.048	0.078	Reject null hypothesis
RF-R = RF-CT	0.002 (0.968)	0.565	1.000	Cannot reject null hypothesis
		Path Foundation good		
PF-R = PF-H	0.150 (0.698)	0.425	0.682	Cannot reject null hypothesis

[a] See Table 1 for variable definitions.

al. (1997) for the Albuquerque good, in those models we include variables describing whether payments in the oral double auction were real or hypothetical, whether or not the respondent is a student, and the version of the script used to describe the oral double auction script.[16]

Probability models for the baseline experiments are reported in Table 4. In each model, whether or not the respondent voted YES is the dependent variable. Standard errors are computed using the P.J. Huber (1967) formula for robust standard errors, generalized to allow for observations arising from cluster sampling (each experimental group may be considered a cluster). The categorical variable indicating that a respondent participated in a hypothetical referendum is the categorical variable not included in each model. The parameter estimates are reported in columns (1), (2), (3), and (4). The models are evaluated at the mean of the observed data to estimate the marginal effects and these are reported in columns (1a), (2a), (3a), and (4a).

Overall, the results suggest that the payment conditions of the referendum—whether real, hypothetical, or cheap talk—significantly affect

[16] Cummings et al. (1997) conducted some "hypothetical" oral double auctions (respondents were not be paid their earnings from the auction), recruited approximately half of their sample from civic organizations (all subjects in our experiments were recruited by asking for voluntary participation during a class period), and used a different script to describe the rules of the oral double auction to the subjects.

TABLE 4—PROBIT MODELS FOR THE BASELINE AND CHEAP TALK EXPERIMENTS[a]

	Columns (1)–(4) are the model coefficients (z-statistics are in parentheses) Columns (1a)–(4a) are the marginal effects (models are evaluated at the mean of the observed data)							
	Nature Conservancy good		Albuquerque good		Rain Forest good		Path Foundation good	
Variable	(1)	(1a)	(2)	(2a)	(3)	(3a)	(4)	(4a)
Referendum was real = 1	−0.521	−0.174	−0.594	−0.218	−0.662	−0.256	0.053	0.015
	(4.84)		(5.31)		(2.83)		(0.21)	
Referendum was hypothetical with cheap talk = 1	−0.526	−0.175	−0.444	−0.160	−0.536	−0.211		
	(5.20)		(4.84)		(1.25)			
Age in years	0.008	0.003	0.018	0.007	0.006	0.002	0.070	0.019
	(0.64)		(2.37)		(0.24)		(5.59)	
Gender = 1 if male	0.061	0.022	−0.150	−0.057	−0.245	−0.096	0.073	0.020
	(0.46)		(1.22)		(1.75)		(0.36)	
Married = 1 if married	−0.158	−0.055	0.142	0.054	0.207	0.080	−0.795	−0.173
	(1.22)		(1.19)		(0.60)		(2.49)	
Yearly income in thousands of dollars	0.004	0.001	−0.001	−0.0005	−0.003	−0.001	0.023	0.006
	(1.04)		(0.42)		(0.43)		(3.33)	
Earnings from the oral double auction (in dollars)	−0.170	−0.061	−0.054	−0.021	0.595	0.234	−0.079	−0.022
	(0.87)		(0.48)		(3.08)		(0.18)	
Auction = 1 if actually paid their earnings			0.164	0.062				
			(0.96)					
Number of participants in experiment	0.023	0.008	0.003	0.001	0.005	0.002	0.001	0.0004
	(2.71)		(0.38)		(0.57)		(0.12)	
Student = 1 if a student			0.302	0.112				
			(1.34)					
Script = 1 if our oral double auction instructions are used			0.266	0.102				
			(2.71)					
Psych = 1 if Psychology majors were recruited for experiment					0.390	0.151		
					(1.20)			
Caucasian = 1	0.403	0.143	−0.062	−0.023	0.625	0.244	0.299	0.083
	(2.35)		(−0.44)		(1.95)		(0.68)	
Asian = 1	0.515	0.195	−0.020	−0.007	0.103	0.040	1.104	0.365
	(1.95)		(−0.09)		(0.23)		(1.59)	
Hispanic = 1	−0.790	−0.223	−0.042	−0.016	1.482	0.406		
	(1.85)		(−0.13)		(6.34)			
Intercept	0.596		−0.362		−6.903		−2.780	
	(0.27)		(−0.26)		(2.95)		(0.58)	
χ^2 statistic[b] (p-value)	0.01 (0.9171)		1.67 (0.1959)		0.09 (0.7642)		0.04[c] (0.8354)	
Likelihood ratio test (p-value)	0.00 (0.9856)		0.51 (0.4736)		0.10 (0.7535)		0.03[c] (0.8667)	
	n = 251		n = 461		n = 170		n = 134	
	ln L = −149.48		ln L = −296.19		ln L = −105.66		ln L = −60.95	
	pseudo R^2 = 0.0660		pseudo R^2 = 0.0367		pseudo R^2 = 0.0960		pseudo R^2 = 0.1724	

[a] In all models the dependent variable is whether the respondent voted yes (=1) or no (=0) in the referendum.

[b] Each statistic tests the hypothesis that the effect of participating in a real referendum is identical to that of participating in a hypothetical referendum with cheap talk. The first chi-squared statistic is computed using the inverse matrix of second partials as the estimate of the covariance matrix, the second chi-squared statistic results from a likelihood ratio test.

[c] These statistics test the null hypothesis that the effect of participating in a real referendum is equal to zero.

voting behavior. As reported in Table 4, the parameter estimates for the variables indicating whether or not the referendum was real or hypothetical with cheap talk are negative and highly significant for the NC and ABQ goods. The probability that a respondent votes YES to the NC proposition is 17.4 percent lower in a real referenda and 17.5 percent lower in a cheap talk referenda as compared to a hypothetical referenda. Chi-squared and likelihood ratio tests indicate that there is *not* a significant difference between these two parameter estimates (see Table 4). For the Albuquerque good, participating in a real referenda reduces the probability that a respondent votes YES by 21.8 percent, and participating in the cheap talk referenda

reduces it by 16.0 percent. Again, we *cannot* reject the null hypothesis that these parameter estimates are equal (Table 4). For these two goods, the models suggest that participating in a hypothetical referenda with the cheap talk script inserted affects voting behavior in the same manner as participating in a referendum that involves actual cash payments.[17]

Results for the rain forest proposition indicate that participating in a real referendum reduces the probability that a respondent will vote YES by 25.6 percent as compared to participating in a hypothetical referendum. Participating in a cheap talk referendum reduces the probability that a respondent will vote YES by 21.1 percent, however this parameter estimate is not precisely estimated. The imprecision in the estimate of this parameter is most likely due to the small percentage of the sample that participated in the cheap talk referenda (only 18 percent of the sample participated in a cheap talk referendum). Finally, the results from the probability model for the Path Foundation good indicate that participating in a real referendum did not affect voting behavior significantly as compared to participating in a hypothetical referendum, which is in accord with the tests based on measures of association alone.

Other than the variables indicating the referendum in which respondents participate, there are no variables that are consistently significant predictors of voting behavior across goods. For the NC good, the experiment's group size and the race of the respondent are the only other significant predictors of voting behavior (at the 90-percent level of confidence or better). In this case, respondents who characterize themselves as Caucasian and Asian are more likely to vote YES for the NC proposition and Hispanic respondents are less likely to vote YES (as compared to African-American respondents, the category left out of the model). The age of the respondent and the version of the script used in the oral double auction are the only two significant predictors of voting behavior for the ABQ proposition (other than the referenda treatments). For the Rain Forest good, Caucasian and Hispanic respondents were more likely to vote YES as compared to African-American respondents, males were less likely to vote YES, and respondents with higher earnings in the oral double auction were more likely to vote YES. Finally, the age, marital status, and income of the respondent are significant predictors of voting behavior for the PF good (although the referendum treatment was not).

In summary, tests based on both contingency tables and probability models support our conclusion: for goods in which respondent voting behavior is found to be consistent with hypothetical bias, the cheap talk script appears effective in eliminating that bias. We find that responses to *hypothetical* referenda that include the cheap talk script are statistically indistinguishable from responses to referenda that would involve *actual cash payments* by the respondents if the referendum were to pass.

IV. Robustness Tests and Results

Given the robustness of the cheap talk design across goods offered to respondents, we now test the efficacy of our cheap talk to two major design changes. The first change involves the cheap talk script itself. Recall that part of the cheap talk script describes actual numerical results from real and hypothetical referenda "... just like the one you are about to vote on." To explore any potential effects of this information on respondent behavior, the cheap talk script is modified by eliminating the discussion of the numerical results. The relevant portion of the modified script reads as follows.[18]

> ... in a recent study, several different groups of people voted on a referendum just like the one you are about to vote on. Payment was hypothetical for these groups, as it will be for you. No one had to pay money if the referendum passed. Another set of groups with similar people were also used in this study. These people voted on the *same* referendum as you will vote on here, *but* payment was real and these people *really* did have to pay money if the referendum passed. On average, more people voted "yes" when the referendum was *hypothetical* than when it was *real.*

[17] Sensitivity tests indicate that these results are not sensitive to model specification or assumptions regarding the model error term.

[18] There are of course any number of other changes in the cheap talk script that might warrant examination. See Cummings et al. (1995b) for a discussion of some of our earlier pretests with shorter versions of the cheap talk script.

TABLE 5—REFERENDA RESULTS AND SELECTED DATA SUMMARIES FOR ROBUSTNESS EXPERIMENTS

Referenda treatment[a]	Number of participants	YES responses (percent of total)	NO responses (percent of total)	Mean age[b,c]	Mean income[b,d]	Percent male[c]	Percent married[c]
ABQ-MCT	56	19 (33.9)	37 (66.1)	23.5 (6.4)	34.8 (18.4)	52.7	10.9
NC-CT$	58	18 (31.0)	40 (69.0)	23.1 (8.4)	35.8 (19.6)	43.1	10.3
NC-MCT$	59	16 (27.1)	43 (72.9)	27.4 (7.1)	33.7 (14.5)	50.0	25.9
ABQ-CT$	53	17 (32.1)	36 (67.9)	23.0 (5.2)	34.9 (20.1)	51.0	7.8
ABQ-MCT$	45	15 (33.3)	30 (66.7)	23.7 (6.0)	34.2 (18.2)	47.7	11.4

[a] See Table 1 for variable definitions.
[b] Standard deviations are in parentheses.
[c] Means or percentages for some groups are based on less than the full sample because of nonresponses by some participants.
[d] See Table 2 for more on the definition of the income variable.

> We call this a "hypothetical bias." "Hypothetical bias" is the difference that we continually see in the way people respond to hypothetical referenda as compared to real referenda—people seem to respond differently when they really don't have to pay money as a result of their vote.
>
> In the real referendum, where people knew they would have to pay money if the referendum passed, fewer voted "yes" than when payment was hypothetical and people knew they would not pay anything if the referendum passed.

Following these statements, the original cheap talk script continues beginning with "How can we get people . . . ." Referenda using this modified cheap talk design are conducted using the Albuquerque good. Results are reported in Table 5.

Results from experiments using the modified cheap talk script (ABQ-MCT in Table 5) were that 33.9 percent of respondents voted YES, which compares favorably with the 34.5 percent obtained with the original cheap talk script (ABQ-CT in Table 2). As Table 6 reports, this YES response rate *is* significantly different than the response rate to the hypothetical referenda and *is not* significantly different from the response rate to the real referendum for the Albuquerque good. Likewise, the model reported in columns (1) and (1a) of Table 7 indicates that participating in a referendum with this modified cheap talk script reduces the probability that a respondent will vote YES by 19.5 percent, while participating in a real referendum reduces the probability that a respondent will vote YES by 22.9 percent (as compared to participating in a hypothetical referendum). The null hypothesis that these two parameter estimates are equal *cannot* be rejected at any conventional level of significance (see Table 7).

The second experimental design change involves the oral double auction that was included in the experiments in an effort to eliminate "found money" effects. This process is tedious and time-consuming, and it may prove to be unwieldy or impossible for many field applications, e.g., telephone surveys. To examine the effects of these experimental design elements, "surveys" are conducted which *do not* include a oral double auction and the subjects *do not* receive a participation fee. Recruitment methods, descriptions of the goods and the referenda rules, and the cheap talk scripts remain unchanged. Surveys using the original and the modified cheap talk script are conducted using the Nature Conservancy and Albuquerque goods. Table 1 summarizes each of these surveys and Table 5 presents their results and a few selected data summaries.

Again, results indicate that our cheap talk is robust to changes in the experimental design. For both the NC and ABQ goods, there *is not* a significant difference between the voting responses in the *surveys* using the cheap talk script and the *experiments* using the cheap talk script. This is true regardless of whether or not the original or modified cheap talk

TABLE 6—NONPARAMETRIC TEST RESULTS FOR ROBUSTNESS EXPERIMENTS

Null hypothesis[a]	Pearson χ^2 (p-value)	Fisher's exact p-value 1-sided	Fisher's exact p-value 2-sided	Conclusion
		Modified cheap talk experiment		
ABQ-MCT = ABQ-CT	0.005 (0.942)	0.545	1.000	Cannot reject null hypothesis
ABQ-MCT = ABQ-R	0.365 (0.546)	0.327	0.619	Cannot reject null hypothesis
ABQ-MCT = ABQ-H	2.816 (0.093)	0.063	0.098	Reject null hypothesis
		Surveys with Nature Conservancy good		
NC-CT$ = NC-CT	0.2547 (0.614)	0.377	0.699	Cannot reject null hypothesis
NC-CT$ = NC-R	0.512 (0.474)	0.301	0.555	Cannot reject null hypothesis
NC-CT$ = NC-H	2.177 (0.140)	0.094	0.186	Inconclusive
NC-MCT$ = NC-CT	0.000 (0.991)	0.572	1.000	Cannot reject null hypothesis
NC-MCT$ = NC-R	0.052 (0.820)	0.488	0.843	Cannot reject null hypothesis
NC-MCT$ = NC-H	3.998 (0.046)	0.032	0.049	Reject null hypothesis
		Surveys with Albuquerque good		
ABQ-CT$ = ABQ-CT	0.087 (0.768)	0.458	0.853	Cannot reject null hypothesis
ABQ-CT$ = ABQ-R	0.113 (0.737)	0.430	0.737	Cannot reject null hypothesis
ABQ-CT$ = ABQ-H	3.558 (0.059)	0.041	0.064	Reject null hypothesis
ABQ-MCT$ = ABQ-CT	0.018 (0.892)	0.526	1.000	Cannot reject null hypothesis
ABQ-MCT$ = ABQ-R	0.229 (0.632)	0.378	0.718	Cannot reject null hypothesis
ABQ-MCT$ = ABQ-H	2.586 (0.108)	0.074	0.137	Inconclusive

[a] See Table 1 for variable definitions.

script is used in the survey (see Table 6). In addition, there *is not* a significant difference in voting behavior in the surveys (conducted with either cheap talk script) and the baseline real referenda for both goods. In general, we may conclude that there *is* a significant difference between the voting behavior in the cheap talk surveys and the baseline hypothetical referenda; the strength of this conclusion varies across goods, however (see Table 6).[19]

The probability models presented in Table 7 indicate that participating in a *survey* for the

[19] Note that the null hypotheses NC-CT$ = NC-H and ABQ-MCT$ = ABQ-H are rejected at the 90-percent level of confidence only with Fisher's exact 1-sided tests.

TABLE 7—PROBIT MODELS FOR THE ROBUSTNESS EXPERIMENTS[a]

	Columns (1)–(5) are the model coefficients (z-statistics in parentheses) Columns (1a)–(5a) are the marginal effects (models are evaluated at the mean of the observed data)									
	ALB-MCT		NC-CT$		NC-MCT$		ALB-CT$		ALB-MCT$	
Independent variable	(1)	(1a)	(2)	(2a)	(3)	(3a)	(4)	(4a)	(5)	(5a)
Referendum was real = 1	−0.621	−0.229	−0.507	−0.173	−0.532	−0.179	−0.425	−0.159	−0.419	−0.158
	(4.91)		(−4.81)		(−4.98)		(−3.79)		(−3.64)	
Referendum was hypothetical with a type of cheap talk = 1	−0.561	−0.195	−0.346	−0.120	−0.583	−0.192	−0.365	−0.132	−0.265	−0.098
	(3.95)		(−1.68)		(−6.52)		(−3.50)		(−1.34)	
Age in years	0.019	0.007	0.027	0.010	0.032	0.012	0.016	0.006	0.014	0.005
	(2.30)		(2.41)		(2.27)		(2.08)		(1.96)	
Gender = 1 if male	−0.226	−0.086	0.037	0.013	−0.170	−0.061	−0.104	−0.040	−0.027	−0.011
	(1.79)		(0.25)		(−0.96)		(−0.85)		(−0.19)	
Married = 1 if married	0.237	0.091	−0.092	−0.033	−0.227	−0.079	0.077	0.029	0.016	0.006
	(1.86)		(−0.82)		(−2.53)		(0.62)		(0.12)	
Yearly income in thousands of dollars	−0.0002	−0.0001	0.008	0.003	0.005	0.002	−0.002	−0.001	−0.002	−0.001
	(0.07)		(2.85)		(1.22)		(−0.76)		(−0.83)	
Earnings from the oral double auction (in dollars)	−0.097	−0.037								
	(0.82)									
Auction = 1 if actually paid their earnings	0.155	0.058								
	(0.95)									
Number of participants in experiment	−0.005	−0.002	0.011	0.004	0.023	0.008	0.008	0.003	−0.007	−0.003
	(0.41)		(0.92)		(1.73)		(0.78)		(−0.61)	
Student = 1 if a student	0.402	0.148					0.300	0.112	0.224	0.084
	(1.61)						(1.24)		(−0.92)	
Script = 1 if our oral double auction instructions are used	0.359	0.138								
	(2.51)									
Caucasian = 1	−0.071	−0.027	0.505	0.181	0.508	0.180	−0.151	−0.058	−0.088	−0.034
	(0.47)		(3.26)		(3.19)		(−0.89)		(−0.64)	
Asian = 1	0.209	0.081	0.433	0.164	0.389	0.146	0.004	0.001	0.243	0.095
	(1.26)		(1.91)		(1.67)		(0.02)		(1.28)	
Hispanic = 1	0.137	0.053	−0.622	−0.191	−0.606	−0.184	0.326	0.128	0.185	0.072
	(0.36)		(−1.36)		(−1.37)		(0.86)		(0.58)	
Intercept	0.133		−1.741		−1.905		−0.834		−0.747	
	(0.09)		(−6.34)		(−5.03)		(−1.47)		(−1.37)	
χ^2 statistic[b] (p-value)	0.27 (0.6021)		1.11 (0.2926)		1.10 (0.2933)		0.45 (0.5036)		0.66 (0.4183)	
Likelihood ratio test (p-value)	0.07 (0.7918)		0.36 (0.5494)		0.04 (0.8513)		0.07 (0.7871)		0.47 (0.4937)	
	n = 433		n = 239		n = 239		n = 429		n = 422	
	ln L = −276.02		ln L = −142.25		ln L = −141.71		ln L = −276.35		ln L = −273.95	
	pseudo R^2 = 0.04		pseudo R^2 = 0.08		pseudo R^2 = 0.07		pseudo R^2 = 0.03		pseudo R^2 = 0.03	

[a] In all models, the dependent variable is whether the respondent voted yes in the referendum (=1) or no (=0).
[b] Statistics test the hypothesis that the coefficients for the referenda treatments present in the model are equal.

NC good reduces the probability (relative to the baseline hypothetical referendum) that a respondent would vote YES by 12 and 19 percent, depending on where the survey uses the original (NC-CT$) or the modified (NC-MCT$) cheap talk script, respectively [see columns (2a) and (3a)]. Neither of these parameter estimates are significantly different than the parameter estimate for the effect of participating in a real referendum. The probability models for the ABQ good indicate that participating in a survey reduces the probability that a respondent would vote YES by 13 and 10 percent, again depending on where the survey uses the original (ABQ-CT$) or the modified (ABQ-MCT$) cheap talk script, although the parameter estimate for ABQ-MCT$ is not precisely estimated. Again, neither of the parameter estimates for ABQ-CT$ or ABQ-MCT$ are significantly different than the parameter estimates for the variable indicating the referendum was real.[20]

[20] Models pooling all data for the NC good or all data for the ABQ good were also estimated. In these models, joint

V. Conclusions

In summary, the goods used in our experiments involve contributions to public goods that, by design, vary in terms of: what is being delivered; where it is being delivered; and the specificity of the connection between how much is donated to the good and "how much" of the good is provided. For the three of our four goods where evidence consistent with hypothetical bias was identified (the NC, ABQ, and RF goods), the cheap talk design was successful in eliciting responses to hypothetical valuation questions that were indistinguishable from responses to valuation questions involving actual payments. This finding was robust across changes in the cheap talk script and changes in the experimental design that may be of importance for field applications of the method. We have interpreted these results as suggesting that the cheap talk design can be effective in eliminating hypothetical bias.

A possible explanation for the success of our cheap talk script in these regards is suggested by social psychologists' research related to correction processes in social judgement (e.g., Leonard L. Martin et al., 1990; Duane Wegener and Richard Petty, 1995). When applied to the CV method, this research points to the different relationships between the context of the referendum and the subject's valuation (judgement) of the public good ("target") in real and hypothetical referenda as explanations for what we refer to as hypothetical bias: subjects' judgements are "primed" differently in the two types of referenda (see Wegener and Petty, 1995 pp. 36–37). The cheap talk script makes subjects *aware* of the potential influence of the context of a hypothetical referendum on their valuation of a good. In such cases, social psychologists find that subjects may ". . . effortfully subtract or partial out reactions toward the context from reactions toward the target" (Wegener and Petty, 1995 p. 37). In other words, subjects may "effortfully" attempt to correct for the hypothetical nature of the referendum. Of course, such "correction" is the intended purpose of our cheap talk design.

An important question arises: Are corrections induced by the cheap talk script appropriate corrections? We consider two issues related to this question and the hypotheses implied by them. First, social psychologists' research suggest that when advised of sources of bias, a subject's correction in response to this information may take the form of an *over*correction. We have effectively tested this hypothesis for the three goods for which evidence of hypothetical bias was found. As discussed above (see Table 3), we find no significant difference between behavior in the cheap talk referenda and the real referenda for any of our experiments. Hence we reject the hypothesis that the cheap talk script induced overcorrections, i.e., that a lower percentage of respondents voted YES in the cheap talk referenda as compared to the real referenda. Secondly, as hypothesized by an anonymous reviewer, it could be the case that the cheap talk script simply introduces a bias of its own which is independent of hypothetical bias and which just happens to offset the hypothetical bias found in this study's particular experiments. We can test this hypothesis in the following way. The reader will recall that for the PF good, we found no significant difference between the YES responses obtained with the real (24.5 percent) and hypothetical (21.6 percent) referenda. If it is that case that the cheap talk script introduces corrections that are *not* systematically related to hypothetical bias, then applying the cheap talk script to the PF good should result in YES responses that are significantly less than those observed in the hypothetical and real referenda for this good. Using procedures and scripts as described above,[21] we then conduct a final set of referenda for the PF good using the cheap talk

tests on the coefficient estimates indicating the referenda treatments (for instance, NC-R, NC-CT, NC-CT$, NC-MCT$) indicated no significant differences between the coefficient estimates for the real referenda, the cheap talk referenda, and any of the cheap talk surveys.

[21] Given our earlier-reported finding of no significant difference between responses to the CT and MCT$ scripts, the MCT$ script was used in these robustness tests with the PF good. A few modifications of the script were required. Since hypothetical bias was *not* found with the PF good, but was found for our other three goods, it was necessary to change the cheap talk script wording: ". . . just like the referendum you are about to vote on . . ." to ". . . similar to the referendum you are about to vote on." In addition, our earlier experiments were conducted prior to the Olympic games held in Atlanta, and so it was necessary to delete references to completion dates for some segments of the trail and the "upcoming Olympics" when describing the PF good.

script to test this hypothesis. With 70 subjects participating in the experiments, 17 (24.3 percent) voted YES on the PF proposition as compared to 24.5 and 21.6 percent in the PF-R and PF-H referenda, respectively. Using measures of association, we find no significant difference between YES responses in the real, hypothetical and cheap talk referenda for this good.[22] These results, together with our earlier findings, indicate that the cheap talk script does not lead to overcorrections for any of our four goods—regardless of whether or not there was evidence of hypothetical bias in the baseline experiments for the good. Furthermore, these results may be interpreted as suggesting that corrections induced by the cheap talk script are indeed systematically related to the degree of hypothetical bias observed.

While we find the results of our robustness tests on the cheap talk design for hypothetical valuation questions encouraging, we are of course aware of the considerable challenge that remains for future research if this design, or modifications of this design, are to be validated in a more monolithic manner. As just one example of an important issue that remains as an open empirical question is the length of the cheap talk script required to achieve desired corrections on the part of subjects. The importance of this issue derives from two related observations. First, the length of the script used in our experiments may be infeasible for telephone applications of the CV method. Second, while admittedly limited in nature, two independent efforts to obtain unbiased valuation responses with a much shorter version of the cheap talk script were unsuccessful (see Cummings et al., 1995b; Gregory L. Poe et al., 1997). A particularly important avenue for future research might then be determining the minimum amount of discussion in the cheap talk script (the minimum amount of "stimulus" *a la* Wegener and Petty, 1995) that is required to induce the kinds of corrective behavior desired for applications of the CV method. Overall, our hopes are that we have provided enough evidence here to warrant such further research.

[22] The Pearson chi-square statistic testing the null hypothesis that voting outcome is independent of referenda treatment (PF-CT or PF-R) is 0.2230, which is not significant at any conventional level of confidence. Results are similar for the test PF-CT = PF-H.

REFERENCES

Bjornstad, David; Cummings, Ronald and Osborne, Laura. "A Learning Design for Reducing Hypothetical Bias in the Contingent Valuation Method." *Environmental and Resource Economics*, October 1997, *10*(3), pp. 207–21.

Blackburn, McKinley; Harrison, Glenn W. and Rutström, E. Elisabet. "Statistical Bias Functions and Informative Hypothetical Surveys." *American Journal of Agricultural Economics*, December 1994, *76*(5), pp. 1084–88.

Brookshire, David and Coursey, Don. "Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures." *American Economic Review*, September 1987, *77*(4), pp. 554–66.

Brown, Thomas C.; Champ, Patricia; Bishop, Richard and McCollum, Daniel. "Which Response Format Reveals the Truth about Donations to a Public Good?" *Land Economics*, May 1996, *72*(2), pp. 152–66.

Cummings, Ronald G.; Elliott, Steven; Harrison, Glenn W. and Murphy, James. "Are Hypothetical Referenda Incentive Compatible?" *Journal of Political Economy*, June 1997, *105*(3), pp. 609–21.

Cummings, Ronald G.; Harrison, Glenn W. and Rutström, E. Elisabet. "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible?" *American Economic Review*, March 1995a, *85*(1), pp. 260–66.

Cummings, Ronald G.; Harrison, Glenn W. and Taylor, Laura O. "Can the Bias of Contingent Valuation Surveys Be Reduced? Evidence from the Laboratory." Working paper, Georgia State University, 1995b.

Cummings, Ronald G.; Williams, Melonie and Bjornstad, David. "The Value of the Global Environment: A Cross-Cultural Analysis Employing the Contingent Valuation Methodology in Support of Integrated Policy Analysis." Working paper, Policy Research Center, Georgia State University, 1996.

Davis, Douglas D. and Holt, Charles A. *Experimental economics*. Princeton, NJ: Princeton University Press, 1993.

Dickie, Mark; Fisher, Anne and Gerking, Shelby. "Market Transactions and Hypothetical Demand Data: A Comparative Study." *Journal*

of the American Statistical Association, March 1987, *82*(397), pp. 69–75.

Farrell, Joseph and Gibbons, Robert. "Cheap Talk Can Matter in Bargaining." *Journal of Economic Theory*, June 1989, *48*(1), pp. 221–37.

Farrell, Joseph and Rabin, Matthew. "Cheap Talk." *Journal of Economic Perspectives*, Summer 1996, *10*(3), pp. 103–18.

Fox, John; Shogren, Jason; Hayes, Dermot and Kleibenstein, James. "CVM-X: Calibrating Contingent Values with Experimental Auction Markets." *American Journal of Agricultural Economics*, August 1999, *80*(3), pp. 455–65.

Huber, P.J. "The Behavior of Maximum Likelihood Estimates Under Non-standard Conditions." *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967, (1), pp. 221–23.

Loomis, John; Gonzalez-Caban, Armondo and Gregory, Robin. "Substitutes and Budget Constraints in Contingent Valuation." *Land Economics*, November 1994, *70*(4), pp. 499–506.

Louviere, Jordan. "Relating Stated Preference Measures and Models to Choices in Real Markets: Calibration of CV Responses," in David J. Bjornstad and James R. Kahn, eds., *The contingent valuation of environmental resources: Methodological issues and research needs*. Cheltenham, U.K.: Elgar, 1996, pp. 167–88.

Martin, Leonard L.; Seta, John J. and Crelia, Rick A. "Assimilation and Contrast as a Function of People's Willingness and Ability to Expend Effort in Forming an Impression." *Journal of Personality and Social Psychology*, July 1990, *59*, pp. 27–37.

Matthews, Steven; Okuno-Fujiwara, Masahiro and Postlewaite, Andrew. "Refining Cheap-Talk Equilibria." *Journal of Economic Theory*, December 1991, *55*(2), pp. 247–73.

Neill, Helen R. "The Context for Substitutes in CVM Studies: Some Empirical Observations." *Journal of Environmental Economics and Management*, November 1995, *29*(3), pp. 393–97.

Neill, Helen R.; Cummings, Ronald G.; Ganderton, Philip; Harrison, Glenn W. and McGuckin, Thomas. "Hypothetical Surveys and Real Economic Commitments." *Land Economics*, May 1994, *70*(2), pp. 145–54.

Poe, Gregory L.; Clark, Jeremy and Schulze, William. "Can Hypothetical Questions Predict Actual Participation in Public Programs?: A Field Validity Test Using a Provision Point Mechanism." Working paper, Cornell University, 1997.

Randall, Alan. "Calibration of CV Responses: Discussion," in David J. Bjornstad and James R. Kahn, eds., *The contingent valuation of environmental resources: Methodological issues and research needs*. Cheltenham, U.K.: Elgar, 1996, pp. 198–207.

Rutström, E. Elisabet. "Homegrown Values and Incentive Compatible Auction Design." *International Journal of Game Theory*, October 1998, *27*, pp. 427–41.

Schulze, William; McClelland, Gary; Lazo, Jeffrey and Rowe, Robert. "Embedding and Calibration in Measuring Non-use Values." *Resources and Energy*, 1999 (forthcoming).

Smith, V. Kerry. "Can Contingent Valuation Distinguish Economic Values for Different Public Goods?" *Land Economics*, May 1996, *72*(2), pp. 139–51.

Smith, V. Kerry and Mansfield, Carol. "Buying Time: Real and Hypothetical Offers." Working paper, Duke University, December 1997.

Taylor, Laura O. "Incentive Compatible Referenda and the Valuation of Environmental Goods." *Agricultural and Resource Economics Review*, October 1998, *27*(2), pp. 132–39.

Wegener, Duane and Petty, Richard. "Flexible Correction Processes in Social Judgment: The Role of Naive Theories in Corrections for Perceived Bias." *Journal of Personality and Social Psychology*, January 1995, *68*(1), pp. 36–51.