Incentive Compatible Referenda and the Valuation of Environmental Goods

Laura O. Taylor

Recent attempts to test the validity of the contingent valuation method have relied on laboratory-type experiments. In these experiments, willingness to pay responses in hypothetical choice experiments are compared with responses from choice experiments requiring actual payments. Often evidence of hypothetical bias is found. Critical for these experimental tests of hypothetical surveys is that the methodology used to elicit willingness to pay from subjects in the real-payment experiment be demand revealing. If it is not, then differences in responses to hypothetical and real valuation questions could be due to *free-riding* in the real-payment survey and not due to *hypothetical bias* in the hypothetical survey. This paper reports on experiments that implement a theoretically incentive-compatible revelation mechanism (a closed referendum) to elicit responses to valuation questions in both hypothetical and real experiments. As in earlier studies, evidence of an upward hypothetical bias is found.

Testing of the validity of contingent valuation (CV) surveys using methods similar to those employed in experimental economics has grown considerably in recent years. These tests employ economic decision experiments where subjects respond to both hypothetical and real valuation questions.[1] Results from such experiments often indicate significant differences between responses to the hypothetical and real valuation questions.[2] Such differences are commonly interpreted as reflecting "hypothetical bias." Critical for these experimental tests of hypothetical surveys is that the methodology used to elicit willingness to pay (WTP) from subjects in the real-payment survey be demand revealing. If a demand revealing mechanism is used, then tests of an identical but hypothetical survey may be conducted by comparing results from the two experimental treatments. However, if a demand revealing mechanism is not used, then the real-payment scenario is not the correct benchmark with which responses to hypothetical surveys should be compared.[3] This point is especially important for public goods valuation. If the real-payment experiment is not demand revealing, then differences in responses to hypothetical and real valuation questions could be due to *free-riding* in the real-payment scenario and not due to *hypothetical bias* in the hypothetical survey.

One particular experimental design, introduced by Cummings et al. (1997; hereafter referred to as CEHM) and employed by Cummings and Taylor (forthcoming; hereafter referred to as CT), Cummings and Taylor (1998), and Bjornstad, Cummings, and Osborne (1997), uses a referendum payment mechanism to test the validity of hypothetical surveys. All of these studies conduct referenda where subjects vote on a proposition that would require all subjects in the experiment to pay money to a public good if the referendum passes (i.e., if more than 50% vote *yes*). They conduct two treatments of this referendum, one in which payments are hypothetical (referred to as the "hypothetical referendum") and one in which actual cash payments are required by the respondents if the referendum passes (referred to as the "real referendum"). The predominant result from these studies is that a significantly larger proportion of respondents votes *yes* in the hypothetical referenda than in the real referenda.[4] This finding is important because the referendum format, under certain conditions, is theoretically an incentive compatible method for valuing public goods.

Evidence of hypothetical bias in an incentive compatible elicitation method such as the referendum is a significant finding for the conduct of CV surveys. However, it must be the case that the real

The author is assistant professor, Department of Economics, and senior associate, Policy Research Center; School of Policy Studies, Georgia State University, Atlanta. Thanks are due to Ronald G. Cummings, Noah Langdale Chair of Environmental Policy, Georgia State University, for financial support of these experiments. Thanks are also due the participants at the 1998 annual meeting of the Northeastern Agricultural and Resource Economics Association for helpful comments and to Kelly Brown and Carol Robinson for helping to conduct the experiments.

referendum is indeed incentive compatible. Implementation of a perfectly incentive compatible elicitation method is often difficult, and the referendum design used by the authors cited above may not have had all the properties required for a theoretically incentive compatible referendum. If this is the case, it would be impossible to determine from their experiments alone whether their results are due to the presence of hypothetical bias in the hypothetical referenda or to the presence of free-riding in the real referenda.[5] This paper reports on experiments that incorporate the conditions for theoretical incentive compatibility in referenda that are missing from previous validation tests using this elicitation method. Results from these experiments suggest that hypothetical bias, and not free-riding behavior, is the likely reason there are differences in voting behavior in the real and hypothetical referenda.

The next section of this paper discusses the properties that a referendum needs for it to be an incentive compatible valuation method for public goods and relates these properties to previous studies that conducted referendum experiments. The following sections describe the experimental design used in CEHM and CT and altered here to encompass all properties for incentive compatibility, report results from the experiments, and offer concluding comments.

Incentive Compatible Referenda

Gibbard (1973) and Satterthwaite (1975) independently developed the conditions under which decision rules that map individual preferences over alternatives into a single collective preference ordering for the entire group are strategy-proof, or incentive compatible. A mechanism for eliciting individual preferences is strategy-proof when the individual's optimal choice is to reveal his/her true preference ordering. In other words, the elicitation mechanism is incentive compatible if there are no incentives to misrepresent true preferences. A binary choice referendum (e.g., one may vote for candidate A or B only, or one may vote *yes* or *no* on a project) is strategy-proof when the voting rule is a simple majority rule (i.e., the referendum passes if more than 50% of respondents vote *yes*).[6] In this case, revealing anything other than one's true preferences in an anonymous vote cannot enhance one's chances of receiving one's preferred outcome (see Moulin 1988 for a formal proof).

In the literature using experiments to validate CV surveys, the methods used to elicit preferences are often *not* incentive compatible. For example, a commonly used elicitation method is the voluntary contributions mechanism—a mechanism in which it could be optimal for respondents to reveal less than their true willingness to pay for public goods, as it may be in their best interests to free-ride.[7] In these types of studies, it is impossible to identify whether differences in hypothetical and real WTP for a public good are due to hypothetical bias introducing an *upward* bias of revealed WTP in the hypothetical survey or whether such differences are due to free-riding introducing a *downward* bias of revealed WTP in the real survey. CEHM proposed one experimental design intended to avoid this "observational equivalence" in the data. CEHM conducted binary choice, majority rule referenda in which groups of respondents voted on a proposition that requires *everyone* in the group to donate $10 to a particular organization that provides a quasi-public good (the organization and the good are described in some detail below). The issue that arises here is that the referenda are not "closed." The Gibbard-Satterthwaite theorem, as applied to the referenda conducted by CEHM, relies on the maintained hypothesis that there is no possibility for provision of the good outside the voting group. If this is not the case, then the referendum is no longer closed, and voters who support the project (and would vote *yes* in a closed referendum) may misrepresent their preferences and vote *no* if they feel that the good will be provided by others outside the voting group. Essentially, the voters in the group may choose to free-ride off potential providers of the public good outside the group. Thus, if one finds a higher percentage of voters voting *yes* in the hypothetical referenda as compared with identical real referenda, it could be due to free-riding effects in the real referenda and not hypothetical bias in the hypothetical referenda.[8] The referenda conducted by CEHM, CT, Cummings and Taylor (1998), and Bjornstad et al. (1997), which allow for provision of the good outside the immediate voting group, are not closed referenda. Therefore, their real referenda—the supposed benchmark with which the hypothetical referenda are compared—are not necessarily incentive compatible mechanisms for eliciting WTP for a public good. Thus, their findings that higher percentages of respondents vote *yes* in the hypothetical referenda as compared with the real referenda cannot be unequivocally attributed to the presence of hypothetical bias in their experiments.

The experiments reported here explore the potential effects of closed versus nonclosed referenda on voting behavior. Theoretically, conducting a

closed referendum would be a simple alteration of the experiments conducted by CEHM and CT. However, it is difficult to implement because their experiments rely on actual nonprofit organizations to which respondents can donate money if the referendum passes. It is only in this way that actual "real world" public goods could be used by CEHM to evaluate the incentive compatibility of referenda. To implement a *closed* referendum, it was necessary to find an organization that was willing to provide a public good only in the amount that is funded by the group participating in the referendum experiment. Fortunately, as is described in the next section, such an opportunity—one that was also comparable to the referendum used in CEHM and CT—arose.

Experimental Design

The basic experimental design elements of the closed referendum experiments consist of the referendum description, subject recruitment, and procedures used in conducting the experiments. Each of these elements is discussed below.

Referendum Design

For comparability of the closed referendum to those of CEHM and CT, the description of the "good" and the referendum rules follow theirs as closely as possible. The good to which subjects could vote to contribute money in the CEHM and CT experiments involved contributions to the Southwest Research and Information Center (SRIC), a nonprofit environmental organization located in Albuquerque, New Mexico. Contributions were to be used for the purpose of funding the publication and distribution of a bilingual (English and Spanish) *Citizen's Guide*. This guide was to be distributed to low-income Hispanic households in Albuquerque. The guide would inform households in an area overlying a potentially contaminated aquifer as to whether or not their drinking water wells were likely to be contaminated by toxic substances, how they could have their water tested at no cost to them, and the alternative actions that were available to them if they found that their wells were indeed contaminated. The SRIC's cost for publishing and distribution was $5.00 per guide. Thus, with N subjects in a group each contributing $10.00, $N \times 2$ guides could have been funded by the group. The proposition on which all subjects voted in the CEHM and CT experiments was then:

> 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.[9]

Note that at the time the experiments were conducted by CEHM and CT, the *Citizen's Guide* was under development. Thus, at that time, it would have been reasonable to assume that others (outside the experimental group) could also fund the development and distribution of the *Citizen's Guide*. However, since some time had passed since the initial experiments were conducted, it was possible to take advantage of the fact that the SRIC had completed this program (i.e., had already published and distributed the guide) and was no longer distributing this publication to households in the area, even though several hundred households had not received a guide. The SRIC agreed that for the purposes of this study, it would distribute additional guides to households if funds were collected to do so through these experiments, and in the exact amount funded by the experiment ($5 to provide one booklet to one household). Thus, the subjects could be offered a unique, one-time opportunity to fund the distribution of the *Citizen's Guide* to low-income households in Albuquerque.[10] The following excerpts illustrate the language used to describe this opportunity in the real referendum:

> The Southwest Research and Information Center . . . has developed, published and distributed a *bilingual* (English and Spanish) "Citizen's Guide" . . . [which] identifies the areas that have contaminated groundwater and the sources of pollution in the community. While a few hundred families have yet to receive a Citizen's Guide, the Southwest Research and Information Center has terminated this program. However, for this one time only it has agreed to accept contributions from this group here today, and provide Guides to the number of families allowed by the amount of your contribution. They would require ($N \times \$10$) to distribute the guide to ($N \times 2$) households. If everyone in this room were to contribute $10.00, these moneys would be sufficient to cover the Center's cost to prepare and distribute the Citizen's Guide to ($N \times 2$) households in this area.
>
> We are then going to have a secret vote to decide whether or not we will do this: all of us pay $10.00 for this purpose.
>
> If more than 50% of you vote "yes" on this proposition, *all* of you will pay $10.00—I will collect $10.00 from *each* of you—and we will send this money to the Southwest Research and Information Center with instructions that the money is to be used to prepare and distribute the Citizen's Guide to ($N \times 2$) households.

> If 50% or fewer of you vote "yes" on this proposition, *no one* will pay $10.00, we will not send a check to the Center and the Citizen's Guide will not be distributed to these households.

As the survey administrator reads the above aloud, she inserts the value for ($N \times \$10$) and ($N \times 2$), where N is the number of participants in the referendum. The hypothetical referendum is conducted in the exact same manner, except the language used is subjunctive. For example, the script reads: "I want you to suppose that for this one time only [the Southwest Research and Information Center] would accept contributions from this group here today" and "Supposing we were to have such a referendum, we would vote on the following proposition. . . ." In addition, when describing the rules of the referendum, the language is "all of you *would* pay $10.00" and "we *would* send this money . . ." (emphasis added). Lastly, before the vote is taken in the hypothetical referendum, subjects are told, "Remember, that even though payment of money in this referendum is *hypothetical*, we ask that you respond to questions as though they involved real cash payments."

Recruitment and Experimental Procedures

Subjects were recruited from students enrolled in undergraduate courses at Georgia State University during the 1997 academic quarters. All students signed consent forms acknowledging that their participation in the experiment was voluntary and that they agreed to abide by the rules of a referendum. They also acknowledged that they had received a $10.00 participation fee. Subjects then participated in a series of oral double auctions that required approximately fifty to sixty minutes to complete.[11] The rationale for including the market-based experiment prior to describing and conducting the referendum relates to the payment of the participation fee. For control across valuation experiments, each group is paid the same participation fee, regardless of whether the subjects are to participate in a real or a hypothetical referendum (the same procedure is followed in CEHM and CT). In experiments where the subjects' decision-making behavior could be affected by their receipt of a participation fee, it is common to involve the subjects in activities for which they "earn" their income.[12] The use of the oral double auction is an effort to deal with this problem. Results from pretests of this experimental process suggested the absence of "found money" effects. Subjects reported they felt that they had earned the participation fee after the hour to hour-and-a-half required for the experiment.[13]

Following the oral double auctions, subjects completed a brief questionnaire that requested demographic information, and then the referendum was introduced and conducted. After the ballots were collected, the vote was tallied. If the referendum passed, and it was an experimental treatment requiring actual cash payments, the money was collected and a check was written and mailed to SRIC by a volunteer subject immediately.[14] Following the vote, one last double auction period was completed, and subjects were then paid the earnings from all periods of the double auctions.[15]

Experimental Results

The closed referendum experiments were conducted with group sizes of no greater than thirty students, and no subjects participated in more than one referendum. Voting outcomes and selected data summaries are presented in table 1. To identify whether or not voting behavior is consistent with the presence of hypothetical bias, the analysis begins with simple contingency tables and measures of association. As indicated in table 1, a greater percentage of respondents voted *yes* in the closed hypothetical referenda as compared with the closed real referenda. Specifically, the *yes* responses in hypothetical referenda were 17.9 percentage points higher than in the real referenda. CEHM and CT find a 18.2 and 16.7 percentage

Table 1. Referenda Results and Selected Data Summaries

	Hypothetical Referenda	Real Referenda
Number of participants	77	79
Yes responses	45	32
(percentage of total)	(58.4)	(40.5)
No responses	32	47
(percentage of total)	(41.6)	(59.5)
Mean age[a,b]	29.1	29.2
	(7.0)	(5.2)
Mean income[a,b,c]	38.6	40.1
	(18.9)	(18.7)
Percentage male[b]	50.6	48.7
Percentage married[b]	33.8	36.8

[a]Standard deviations are in parentheses.
[b]Means or percentages are based on less than the full sample due to nonresponses.
[c]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.

point difference between hypothetical and real referenda votes, respectively.[16] Pearson and Fisher's exact tests are used to test the null hypothesis that the *yes* responses in the real and hypothetical referenda are independent of payment condition. Both tests *reject* the hypothesis that there is no significant difference between voting behavior in the closed real and closed hypothetical referenda at the 96% level of confidence or better.[17] Thus, based on measures of association alone, evidence consistent with hypothetical bias is found in the closed referenda.

To incorporate respondents' socioeconomic characteristics and experimental design features into the analysis, probability models were also estimated. The socioeconomic characteristics of the respondents considered in the analysis are the respondent's 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, and the referendum treatment (e.g., hypothetical or real). Results are presented in table 2. 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

Table 2. Probit Models[a]

	Model 1		Model 2		Model 3	
Variable	Coefficient (z-statistic)	Marginal Effect[b]	Coefficient (z-statistic)	Marginal Effect[b]	Coefficient (z-statistic)	Marginal Effect[b]
Referendum was real = 1	−0.526	−0.197	−0.489	−0.193	−0.554	−0.214
	(2.57)		(6.47)		(7.59)	
Referendum was closed = 1					0.052	0.020
					(0.47)	
Age in years	0.013	0.005	0.014	0.006	0.017	0.007
	(0.91)		(0.92)		(2.65)	
Gender = 1 if male	0.168	0.067			−0.038	−0.015
	(0.50)				(0.29)	
Married = 1 if married	0.021	0.010			0.099	0.039
	(0.13)				(0.94)	
Yearly income in	0.007	0.003	0.007	0.003	0.001	0.002
thousands of dollars	(3.22)		(5.38)		(0.23)	
Earning from the double oral	−0.017	0.001			−0.082	−0.032
auction (in dollars)	(0.07)				(0.78)	
Number of participants	0.006	0.001			−0.003	−0.001
in the experiment	(0.36)				(0.31)	
Double auction instructions					0.354	0.138
were *not* CEHM's = 1[c]					(2.79)	
Subject was a student = 1[d]					0.281	0.108
					(1.18)	
Subject was paid auction					0.082	0.032
earnings = 1[e]					(0.53)	
Caucasian = 1[f]	−0.090	−0.173	−0.024	−0.010	−0.043	−0.017
	(0.30)		(0.10)		(0.31)	
Asian = 1	0.523	−0.123	0.557	0.216	0.349	0.138
	(2.12)		(2.26)		(2.22)	
Hispanic = 1	−0.269	−0.236			0.114	0.045
	(0.27)				(0.30)	
Intercept	−0.531		−0.553		−0.082	
	(0.18)		(1.15)		(0.06)	
	n = 153		n = 153		n = 531	
	ln L = −100.34		ln L = −100.74		ln L = −345.77	
	pseudo R^2 = 0.0538		pseudo R^2 = 0.0501		pseudo R^2 = 0.0453	

[a]In all models the dependent variable is whether the respondent voted *yes* (= 1) or *no* (= 0) in the referendum.
[b]Marginal effects are computed at the mean of the observed data.
[c]Instructions used to describe the double auction to subjects were different in CT and in CEHM. The instructions used by CT were shorter and clearer, so these instructions were also used here in the closed referendum experiments.
[d]CEHM conducted some experiments on groups of adults (nonstudents).
[e]CEHM conducted some experiments where respondents were not paid their double auction earnings (i.e., the auction was hypothetical).
[f]African-American is the category not included in the model.

arising from cluster sampling (each experimental group may be considered a cluster). The variable indicating that a respondent participated in a hypothetical referendum is the categorical variable not included in each model.

Consider first models 1 and 2. Model 1 reports a "full" model containing all socioeconomic characteristics and experimental design features that varied across experiments. Model 2 is more parsimonious: it retains only parameters that are significant at the 75% or better level of confidence. Also reported are the estimated marginal effects, which are the models evaluated at the mean of the observed data. The results suggest that the payment conditions of the closed referendum—whether real or hypothetical—significantly affects voting behavior. As reported in table 2, the parameter estimates for the variable indicating whether or not the referendum was real is negative and highly significant. The probability that a respondent votes *yes* to the proposition is almost 20% lower in a real referendum as compared with the hypothetical referendum (19.7% in model 1 and 19.3% in model 2). Similarly, CEHM and CT find that participating in a real referendum reduces the probability that a respondent votes *yes* to the Albuquerque proposition by 18.9% and 21.8%, respectively, amounts commensurate with the results reported here. Other than the variable indicating the referendum treatment, the income of the respondent and whether or not the respondent is Asian (as compared with African American—the category left out of the model) are the only other significant predictors of voting behavior in models 1 and 2.

For comparison purposes, the referenda results for the Albuquerque booklet-distribution good used in CT are pooled with the data from the closed referenda. The percentage of respondents who voted *yes* in CT was 46.4% in the hypothetical referenda and 29.7% in the real referenda. Note that these percentages are different from what is observed in the closed referenda. However, it is not surprising that the voting responses might be different since the "good" in CT is different from that of the closed referenda. In CT, the proposition was to donate money to help *develop* and distribute the *Citizen's Guide*, while the proposition in the closed referenda was for donations to distribute additional booklets (already developed) to households that had not yet received them. In the closed referenda, the booklets had already been prepared and distributed to some households, and so the perceived connection between donations and the actual provision of booklets may have been different in these referenda as compared with those conducted by CEHM and CT.

Model 3 in table 2 presents the results from pooling the CT data with the closed referenda. In their analysis, CT pool 211 observations from CEHM. Thus, the data here are essentially pooled from both CEHM and CT. Because the experimental methods used in CEHM differed somewhat from those of CT and from those used here, additional variables had to be included in model 3 to capture these differences. Specifically, the experimental design elements that varied were whether or not subjects were paid their earnings from the double auction, whether or not subjects were students, and which of two versions of a script was used to describe the double auction. These variables are described in more detail in the notes to table 2. Interestingly, the results indicate that the categorical variable indicating whether or not the referendum was closed is *not* a significant predictor of voting behavior. However, whether or not the referendum was real has a strong effect on the probability that a respondent would vote *yes*, reducing that likelihood by 21% in the real referendum—an amount consistent with the results from models 1 and 2. The age of the respondent and whether or not the respondent was Asian are also significant predictors of voting behavior. The version of the script that was used to describe the double auction was the only experimental design variable that was a significant predictor of voting behavior in model 3. The script used by CEHM significantly reduced the probability that a respondent would vote *yes* as compared with the script used in CT (and used here in the closed referenda).

Conclusions

To date, the experimental evidence is far from definitive in showing that hypothetical surveys eliciting WTP for public goods generate different results from those involving actual cash payments because of a "pure" hypothetical bias effect. Indeed, a perfectly demand-revealing mechanism for eliciting WTP in actual payment scenarios for "real-world" public goods has yet to be implemented. This is because mechanisms for eliciting WTP for public goods that are theoretically demand revealing (such as the Groves-Ledyard 1977 mechanism) tend to be very unwieldy or impossible to implement with actual public goods.[18] The practical issues of implementing experiments with field public goods typically results in the violation of some of the maintained hypotheses required for the mechanism to be incentive compatible. This is not unique to the use of experiments to validate hypothetical surveys. In practically all inquiries

into individual choices, the ideal conditions upon which theoretical predictions of individual choices rely are violated. In the referendum experiments discussed in this paper, as in most demand studies, it cannot be unequivocally stated whether the testable hypothesis has been shown to be false, or whether one of the maintained hypotheses underlying the *test* of the hypothesis has failed.

This paper highlights the "observational equivalence" in the results from past tests for hypothetical bias that relied on binary choice, majority rule referendum experiments. Because the *ideal* conditions under which the referendum is a demand revealing mechanism were violated in these past studies, it is not possible to determine whether the differences in responses to hypothetical and real referenda were due to free-riding or hypothetical bias. The experiments presented here attempt to incorporate the theoretically desirable feature for a referendum that was not included in earlier experiments (that it be a closed referendum). Results indicate that the percentage of *yes* responses to the hypothetical referenda were greater than those of the real referenda—evidence consistent with the earlier studies and consistent with hypothetical bias. However, as noted earlier, even the experimental design used here is not ideal. Indeed, the experiments presented here highlight the difficulty of implementing seemingly simple mechanisms that, under ideal conditions, are demand revealing mechanisms for valuing public goods. Nonetheless, it is hoped that the evidence presented, in conjunction with the evidence from earlier referenda experiments, are compelling enough to suggest that *hypothetical* binary choice, majority rule referenda are *not* fully incentive compatible mechanisms for eliciting WTP for public goods.

References

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

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

Brown, Thomas, Patricia Champ, Richard Bishop, and Daniel McCollum. 1996. "Response Formats and Public Good Donations." *Land Economics* 72(2):152–66.

Carson, Katherine. 1997. "Operationalizing Incentive Compatible Mechanisms: A Pilot Experiment." Working paper, United States Air Force Academy.

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

Cummings, Ronald G., Glenn Harrison, and Elisabet Rutström. 1995. "Homegrown Values and Hypothetical Surveys: Do Dichotomous Choice Questions Elicit Real Economic Commitments?" *American Economic Review* 85:260–66.

Cummings, Ronald, and Laura Taylor. Forthcoming. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review.*

_____. 1998. "Does Realism Matter in Contingent Valuation Surveys?" *Land Economics* 74(2):203–15.

Davis, Douglas D., and Charles Holt. 1993. *Experimental Economics.* Princeton: Princeton University Press.

Dickie, Mark, Anne Fisher, and Shelby Gerking. 1987. "Market Transactions and Hypothetical Demand Data: A Comparative Study." *Journal of the American Statistical Association* 82:69–75.

Fox, John, Jason Shogren, Dermot Hayes, and James Kleibenstein. Forthcoming. "CVM-X: Calibrating Contingent Values with Experimental Auction Markets." *American Journal of Agricultural Economics.*

Gibbard, Allan. 1973. "Manipulation of Voting Schemes: A General Result." *Econometrica* 41:587–601.

Groves, Theodore, and John Ledyard. 1977. "Optimal Allocation of Public Goods: A Solution to the 'Free Rider' Problem." *Econometrica* 45(4):783–809.

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

Johannesson, Magnus, Bengt Liljas, and Richard O'Connor. 1997. "Hypothetical Versus Real Willingness to Pay: Some Experimental Results." *Applied Economic Letters* 4:149–51.

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

Moulin, Hervé. 1988. *Axioms of Cooperative Decision Making.* Cambridge: Cambridge University Press.

Neill, Helen, and Laura Taylor. 1998. "Substitution Effects and Hypothetical Bias in Contingent Valuation Surveys: Evidence from the Laboratory." Working paper, Georgia State University. August.

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

Rose, Steven, Jeremy Clark, Gregory Poe, Daniel Rondeau, and William Schulze. 1997. "The Private Provision of Public Goods: Tests of a Provision Point Mechanism for Funding Green Power Programs." Working paper, Cornell University.

Rutström, E. Elisabet. Forthcoming. "Home-Grown Values and Incentive Compatible Auction Design." *International Journal of Game Theory.*

Satterthwaite, Mark. 1975. "Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions." *Journal of Economic Theory* 10:187–217.

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

Notes

1. *Real valuation questions* are those where actual payments are required on the part of the respondent if the provision rule is met.
2. See, as examples, Brown et al. (1996), Cummings et al. (1995), Fox et al. (forthcoming), Loomis et al. (1994), and Neill et al. (1994). Two studies where hypothetical bias is not identified in willingness to pay estimates are Brookshire and Coursey (1987) and Dickie, Fisher, and Shelby (1987). Smith and Mansfield (1997) fail to find hypothetical bias in an experiment based on willingness-to-accept questions.
3. For instance, with validity tests using private goods, the real payment surveys may suffer from field-censoring wherein subjects reveal only a WTP that is their perception of the market price for the commodity plus some premium for the reduction in transactions costs resulting from having the opportunity to purchase the product in the experiment. See Johannesonn, Liljas, and O'Connor (1997) as an example where this may be the case.
4. CT fail to find evidence of hypothetical bias in one particular set of referenda involving funding the completion of a local recreational project.
5. Thanks are due to Richard Carson for bringing this issue to our attention. This issue also applies to the referenda conducted by Bjornstad, Cummings, and Osborne (1997) and Cummings and Taylor (1998) since they use the same referendum design as CEHM.
6. More correctly, if voters have strict preference orderings (i.e., they cannot be indifferent to the two choices) and the choice is binary, then a voting rule that is monotonic is strategy-proof. Monotonicity implies that a new supporter can do no harm—which is the case in a binary choice, majority rule referendum.
7. See Brown et al. (1996) as an example. Neill and Taylor (1998) conduct experiments using a voluntary contribution method and correct for free-riding effects in their analysis.
8. This conclusion requires the assumption that there is no incentive (or less of an incentive) to free-ride in the hypothetical referenda—an assumption that is an open empirical question.
9. 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."
10. Note that it is necessary to run multiple experiments to achieve the necessary sample size for analysis. While subjects in any one experiment are not told that there will be other similar experiments (and the experiments are not advertised, as is sometimes done for recruiting purposes), it is possible that subjects could think that other experiments were going to be conducted and that those groups might choose to fund the *Citizen's Guide*. Unfortunately, we have no way to determine to what degree this was the case.
11. See Davis and Holt (1993, appendix A.1) for an outline of how an oral double auction is conducted.
12. Such effects are described as "endowment" and "found money" effects (see Rutström, forthcoming) and may arise in instances where subjects view the participation fee as money that must be spent in the experiment. In these experiments, it could be the case that subjects do not treat the fee as part of their disposal income.
13. Of course, there is no way to be positive that this design succeeds in eliminating behavior consistent with found money effects in all cases. See Bjornstad, Cummings, and Osborne (1997) for a more detailed discussion of the rationale for using the oral double auction in this type of valuation experiment.
14. Because the experiments were conducted on the campus at Georgia State University, there was a mailbox nearby for each experiment.
15. The mean earning in the oral double auction was \$1.34 (standard deviation is \$0.53), with a minimum value of \$0.00 and a maximum value of \$2.45.
16. CT, who extended the CEHM data by 105 observations, found that 46.4% of respondents voted *yes* in the hypothetical referenda (211 observations) and 29.7% voted *yes* in the real referenda (182 observations)—a difference of 16.7 percentage points.
17. The Pearson χ^2 statistic is 5.0181 (p-value = 0.025), and the Fisher's exact p-value is 0.037 for the two-sided test, and 0.019 for the one-sided test.
18. See Carson (1997) for an example of the Groves-Ledyard mechanism in a laboratory environment. Rose et al. (1997) test a provision point mechanism with a field public good as a potentially demand revealing mechanism that is easily implemented. However, the mechanism they employ has been tested only in the laboratory under conditions of actual payments and not hypothetical payments.