Further tests of entreaties to avoid hypothetical bias in referendum contingent valuation

Thomas C. Brown,[a,*] Icek Ajzen,[b] and Daniel Hrubes[c]

[a] Rocky Mountain Research Station, U.S. Forest Service, Fort Collins, CO 80526, USA
[b] Department of Psychology, University of Massachusetts, Amherst, MA 01003, USA
[c] Department of Psychology, College of Mount Saint Vincent, Riverdale, NY 10471, USA

Abstract

Over-estimation of willingness to pay in contingent markets has been attributed largely to hypothetical bias. One promising approach for avoiding hypothetical bias is to tell respondents enough about such bias that they self-correct for it. A script designed for this purpose by Cummings and Taylor was used in hypothetical referenda that differed in payment amount. In comparisons with behavior observed in otherwise identical real payment referenda, the script worked remarkably well at higher payment levels (dropping the likelihood of a yes vote to the level obtained in separate sample real referenda), but less well at a lower payment level. The repeated success of the script at higher payment amounts is a clear sign that additional research is warranted—research that will bring us a more complete understanding of how the entreaty works. This research may need to explore its workings at the individual respondent level.
© 2003 Elsevier Science (USA). All rights reserved.

1. Introduction

Given the goal of efficient management of environmental resources but the difficulty of estimating some benefits of environmental programs using revealed preference methods, hopes have been high that properly designed contingent markets would emulate actual markets. However, evidence has mounted that willingness to pay measured in contingent markets

over-estimates payment observed in otherwise identical real cash markets. Though not a uniform result (e.g., [7,8]), differences between real cash (hereafter just real) and hypothetical willingness to pay, which we will call hypothetical bias, have typically been found both for private [2,6,11–13,15] and public [3,5,7] goods. These studies used a mixture of bidding mechanisms, suggesting that the over-estimation cannot be ignored.

One approach to avoiding hypothetical bias in contingent valuation is to ask respondents before they give their hypothetical bid to carefully consider their answer. An early version of this was to remind respondents of their budget constraint, but this approach proved to be largely ineffective [14,15]. This paper tests an approach similar to the budgetary reminder, in that it attempts to get respondents to avoid over-estimations, but it employs a much more direct attack on hypothetical bias. It instructs people, when answering a contingent willingness to pay question, to respond as if they were *really* spending their money. At least three studies, described next, have used this approach. These studies compared hypothetical and real payment treatments with a hypothetical treatment employing an entreaty urging respondents to answer as if they were really spending their money.

Loomis et al. [13] obtained sealed bids for an art print. In the two-paragraph entreaty used with the hypothetical treatment, respondents were reminded of their budget constraint but also were encouraged to think about what they would "honestly" be prepared to pay if the auction were for real and were told, among other things, "Although the question is hypothetical, we want you to answer as if it were real—as if you were participating in a real sealed-bid auction and would really have to pay your dollar amount if you were the highest bidder" (p. 453). Mean bids from separate samples were $42 in the hypothetical treatment, $26 in the hypothetical-with-entreaty treatment, and $14 in the actual payment treatment. The entreaty helped reduce hypothetical bias, but was insufficient to remove it.[1]

Cummings and Taylor [7] used a much more thorough script, which they called "cheap talk," in their effort to remove bias from a hypothetical referendum on contributing $10 to fund a public good. In the eight-paragraph script, the authors not only encouraged participants to act as if the referendum were for real, but also explicitly addressed the problem of hypothetical bias (of how people in hypothetical bidding situations tend to over-estimate their willingness to pay), presented data demonstrating hypothetical bias, and discussed possible reasons for the phenomenon. Their approach met with notable success. For two of their three goods—the donation to the Nature Conservancy for Costa Rican rain forest protection and a separate Nature Conservancy donation to protect land in the state of Georgia—the percentages of participants voting to donate $10 were nearly identical in the real payment and hypothetical-with-cheap-talk treatments, and these percentages were significantly below those of the standard hypothetical treatment. And for the other good—the Albuquerque citizens' guide—the percent voting yes when the cheap talk script was used was within 5 percentage points of its real referendum counterpart. This study is important not only for its demonstration of a promising approach for adjusting for hypothetical bias, but also because it used a referendum format, which is arguably the most appropriate contingent valuation format for valuing public goods [10].

---

[1] Cummings and Taylor [7] cite two working papers that also unsuccessfully used short scripts to remove hypothetical bias.

Recently, List [12] used the Cummings and Taylor cheap talk script in a second-price auction to purchase a baseball card. Respondents from two groups of card collectors (professional card dealers and nondealers) were interviewed individually at a sports card show and given an opportunity to bid (either hypothetically or for real) on a card with a book value of about $200. For nondealers, the real payment and hypothetical-with-cheap-talk treatments each produced a mean bid of $26, compared with a mean bid of $49 in the hypothetical treatment. However, the professional dealers were largely unaffected by the cheap talk script, bidding a mean of $116 in the hypothetical treatment and $108 in the cheap talk treatment, compared with a mean of $60 in the actual auction. As in the Cummings and Taylor study, the cheap talk script worked remarkably well when bidders were unfamiliar with bidding for the good at issue. The lack of success with professional dealers raises the question: Are there other circumstances in which the script does not correct effectively for hypothetical bias?

We further tested the Cummings and Taylor script. Our bidding format, subject pool, recruitment enticement, and good were similar to theirs, in that students were paid $10 and participated in actual or hypothetical referenda on donating to a worthy cause. But whereas Cummings and Taylor used only one donation amount ($10), we varied the amount across referenda to investigate the effect of payment amount on the success of the cheap talk script. The practical reason for such a test was to explore the script's utility for application to dichotomous choice contingent valuation, for if the script is to work in that context it must work at a range of bid levels. The more fundamental objective was to explore the script's limitations, for by finding situations where it does not work well—if indeed there are other such situations—and comparing those with situations where it does work, we can begin to form hypotheses about how it works.

2. Approach and assumptions

We accept as given that when placed in a real payment situation, each participant forms an estimate of willingness to pay (or casts a vote reflecting that willingness to pay). The question is, in a hypothetical situation, will the cheap talk script lead participants to the response that would be forthcoming in a real payment situation, or will it create a separate response that only equals the real payment response in certain circumstances? We suspected that the latter possibility was more likely, and specifically that the cheap talk script—a very direct and lengthy attack on hypothetical bias—would perform best when hypothetical bias was greatest. Our hunch was that a script that worked well when the bias was substantial would be less effective or over-correct when the bias was smaller. Loss of effectiveness is possible in situations of small bias if, for example, respondents do not think hypothetical bias is plausible. Over-correction is possible if the script alters responses when no alterations are needed.

One may question whether the referenda we and Cummings and Taylor used, which were about providing donations to worthy causes, are incentive compatible. A concern with referenda about pubic goods is that some voters may suspect that other options exist for providing the good, leading them to vote no when, in the absence of such suspicions, they would vote yes.[2]

---

[2] One alternative provision option is a future, more attractive referendum. Voters may suspect that other referenda could follow the current one [9]; indeed, in some states repeated referenda on the same funding issue are common [16].

Table 1
Sample size of 12 referenda

	Payment amount			
	$1	$3	$5	$8
Hypothetical	31	37	33	52
Cheap talk	33	43	31	52
Real	28	37	30	17
Total	92	117	94	121

We attempted to lessen the likelihood of such suspicions by informing participants of the great need for funds for the cause described (scholarships) and that whatever the participants donated would indeed be used, but we can make no claims of having avoided the potential concern. Rather, we take our experimental participants' votes in the real payment referenda on face value, as reflecting real willingness to pay in the context presented. We test whether the cheap talk script avoids hypothetical bias given whatever assumptions participants may have made about other options for providing the good. Because real referenda about public goods that completely avoid suspicions about other possible options for funding the good are probably rare, we suspect that our test results are relevant to many real-life referenda.

3. Methods

In groups, students at the University of Massachusetts participated in a session of three or more stages, for which they were each paid $10 in cash at the outset. The referendum stage of the sessions was preceded by an unrelated task and followed by a related but quite different task, a survey intended to measure tendency to respond in a socially desirable manner.

Each of 424 subjects participated in one of 12 different experimental versions including a secret ballot referendum. The 12 versions were created by crossing three levels of "reality" of payment (hypothetical, hypothetical with cheap talk script, and real payment) with four payment levels ($1, $3, $5, and $8). The hypothetical with cheap talk (hereafter just cheap talk) versions of the referenda differed from the comparable hypothetical versions only in that the former included an extensive entreaty to respond as if the vote were for real. Sample sizes for the 12 versions are listed in Table 1.

The referendum stage of a session began with a description of and justification for the proposition to be voted on. It was described as an effort to make additional scholarship funds available to deserving students wishing to attend the university.[3] The proposition ($8 version) was

---

[3] Complete texts of this and other scripts we used are available at the following website: http://www.fs.fed.us/rm/value/cheaptalkscript.html.

then stated as follows:

> *Everyone* here in this room will contribute $8.00 to the *University of Massachusetts—Amherst Scholarship Fund*. The contribution is to be used to provide scholarship aid to needy and deserving students at the University of Massachusetts—Amherst.

The potential outcomes of the voting procedure were then explained, to the effect that if more than 50% of the participants in the room voted yes, all participants would donate the stated amount, and otherwise no one would pay. The real and hypothetical versions differed only in the tense used to describe the outcomes. For example, the first sentence of the real version read "If more than 50% of you vote YES on this proposition, *all* of you pay $8.00—I will collect $8.00 from *each* of you..."; whereas the first sentence of the hypothetical version read "If more than 50% of you were to vote YES on this proposition, *all* of you would pay $8.00—I would collect $8.00 from *each* of you...". The process of collecting the money and writing and sending the check for the total amount was then explained.

At this point in the cheap talk versions a lengthy discussion of hypothetical bias was presented. The script we used is identical to that used by Cummings and Taylor except for differences related to changes in the good. In this script, participants are told of the problem of hypothetical referenda—that when money payments are at issue people tend to vote differently in a hypothetical referendum than they do in a real referendum. Results of past studies are presented showing a higher percentage voting yes in a hypothetical referendum than in a real referendum. The reasons for this difference are then discussed, including the notion that in a referendum to do something good for others, people tend to focus on doing good and tend to ignore the cost to them when it is hypothetical, but they temper that inclination with considerations of other options for their money when the referendum is for real. The cheap talk script ended with the following request: "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."

The vote was recorded on individual printed ballots that were placed in a ballot box. Before the vote, participants were reminded that their $10 participation fee was their money, which they earned by spending their time helping a research project.

4. Results

Percentages voting yes varied widely among the referenda, from a low of 21 percent to a high of 82 percent (Table 2, Fig. 1). Voting in the four hypothetical referenda was insensitive to payment

Table 2
Percent voting yes in 12 referenda

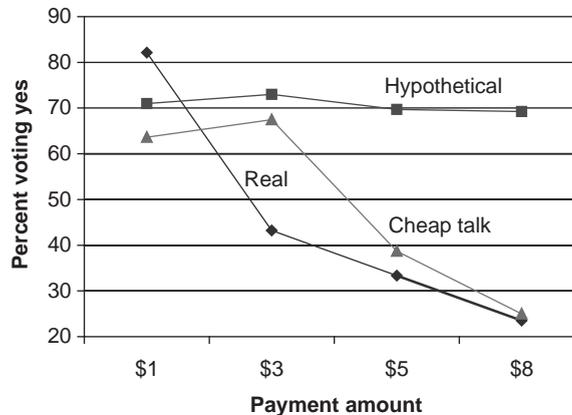|	Payment amount			
	$1	$3	$5	$8
Hypothetical	71	73	70	69
Cheap talk	64	67	39	25
Real	82	43	33	24

Fig. 1. Referenda on contribution to scholarship.

Table 3
Tests of equal proportions voting yes by payment amount (Fisher's exact p-value, two-sided test)

Null hypothesis	Hypothetical	Cheap talk	Real
$1 = $3	1.000	0.809	0.002
$1 = $5	1.000	0.079	0.000
$1 = $8	1.000	0.001	0.000
$3 = $5	0.796	0.019	0.458
$3 = $8	0.814	0.000	0.229
$5 = $8	1.000	0.221	0.529

amount, with 69–73% of participants voting yes.[4] At the 95% confidence level we cannot reject the hypothesis of no difference in voting among the four hypothetical referenda (Table 3).[5] However, in the real referenda, percent voting yes decreased monotonically as payment amount increased, from 82% at the $1 level to 24% at the $8 level (percent voting yes is significantly greater at the $1 payment level than at the other three payment levels, Table 3). Clearly, participants were sensitive to cost when spending real money. In the cheap talk referenda, percent voting yes decreased with payment amount across the $3, $5, and $8 amounts (with the $3 percentage yes being significantly greater than the $5 and $8 percentages yes, Table 3), but the $1 percentage yes was slightly below and not significantly different from the $3 percentage yes.

Examining the tendency toward hypothetical bias, percentages voting yes in the hypothetical referenda were significantly greater than in the real referenda at the $3, $5, and $8 levels (Table 4).

---

[4] To test whether a yes vote of about 70% was an invariable result for public goods with our subject population, we also ran a hypothetical referendum with 19 participants on donating $8 to a different scholarship fund, one usable at universities throughout the U.S., and observed a 47% yes vote. This extra referendum suggests sensitivity to the nature of the good in the hypothetical condition. The difference in percentages voting yes between the UMass scholarship fund and this other scholarship fund was marginally significant (the Fisher's exact two-sided test yields a p-value of 0.065).

[5] Two-sided tests of hypotheses are reported herein using Fisher's exact test. One-sided tests, of course, yield lower probability levels, but reach the same conclusions for our data at the 0.05 significance level, except for the comparison of $1 and $5 in the cheap talk treatment, for which the probability level drops from 0.079 with a two-sided test to 0.040 with a one-sided test.

Table 4
Tests of equal proportions voting yes by treatment (Fisher's exact p-value, two-sided test)

Null hypothesis	Payment amount			
	$1	$3	$5	$8
Real = hypothetical	0.370	0.018	0.006	0.001
Real = cheap talk	0.154	0.042	0.791	1.000
Hypothetical = cheap talk	0.601	0.632	0.023	0.000

At the $1 level, the percentage voting yes was actually greater in the real referendum than in the hypothetical referendum, though not significantly so.

The cheap talk script had the desired effect of eliminating hypothetical bias at the higher payment levels, dropping percent voting yes to within one percentage point of what it was in the real referendum at the $8 level, and to within 6 percentage points at the $5 level.[6] At both of these payment amounts, the hypothetical percentages yes are significantly greater than the corresponding cheap talk percentages (Table 4). But at the $3 payment level the cheap talk script had a very modest effect, producing a percentage yes only 6 points below the hypothetical percentage yes, and significantly greater than the real percentage yes (Table 4). Finally, at the $1 level no correction was needed and the correction with the cheap talk script was not statistically significant.

5. Discussion

This study joins a number of prior split-sample studies demonstrating hypothetical bias in elicitations of willingness to pay by comparing responses from hypothetical and real payment conditions. Two of these studies [5,7] used referenda to estimate values of public goods, but employed only a $10 payment amount. We extended their work by testing for bias in several referenda differing in payment amount. Among the four payment amounts ($1, $3, $5 and $8), the percent voting in favor in the hypothetical referenda remained nearly constant, whereas when the referenda were for real the percent in favor steadily dropped as the amount was raised, causing hypothetical bias to increase with payment amount.

The sensitivity of the vote to payment amount in the real referenda was expected, but the insensitivity to payment amount in the hypothetical referenda was not, as previous studies of hypothetical responses have found that proportion responding yes drops as bid level rises. Such sensitivity has been found for a private good at bid levels ($1 and $5) similar to those used here (e.g., [1]). And dichotomous choice contingent valuation studies of public goods regularly show a significant negative relation between stated willingness to pay and bid level [for contingent

---

[6] The effect of the script was also tested in three additional $8 referenda, which differed from the standard $8 referenda in that the respondents were informed before the vote that they would be asked after the vote to reveal and explain their votes to others in the room. In these referenda, 63%, 45% and 43% voted yes in the hypothetical, cheap talk, and real referenda, respectively. Thus, anticipation that the votes would be discussed tended to narrow the difference between the hypothetical and real referenda, compared with the secret ballots, but the effect of the script was again to correct for the bias. For more detail, see our discussion paper "Obtaining Unbiased Contingent Values: Further Tests of Entreaties to Avoid Hypothetical Bias," available at: http://www.fs.fed.us/rm/value/docs/dp-01-1.pdf.

referenda examples, see [4,14]. However, unlike our study, most contingent valuation studies use a wide range of payment amounts. Thus, the lack of sensitivity to payment amount that we found cannot be generalized to higher payment amounts.

Why was the hypothetical vote invariant to payment amount, thereby causing the substantial hypothetical bias we observed? Four possibilities come to mind, the first three of which are suggested in the cheap talk script. First, some participants in hypothetical referenda may try to estimate their real willingness to pay but may over-estimate. Second, some may ignore their budget constraint, thereby responding as if they faced none. Third, some may use their vote to indicate whether or not they think the cause is worthy of support. Fourth, some may use their vote to indicate whether or not they think such a referendum should be held for real. In essence, participants acting according to the latter three of these possibilities are answering different questions from the one they were asked in the hypothetical referendum.

The Cummings and Taylor cheap talk script worked remarkably well at the two larger payment amounts, lowering the percent voting in favor to within 1 ($8 payment) and 6 ($5 payment) percentage points of what it was in the real referenda. This result is similar to those obtained by Cummings and Taylor in their public good referenda, which used a $10 payment with a similar subject pool. However, the script significantly under-corrected at the $3 payment level, lowering percentage voting in favor only 6 percentage points below that of the hypothetical condition. What could explain these conflicting results?

The cheap talk script tries to bring to the participant's attention factors that are presumably considered in real referenda but that may not readily come to mind in hypothetical referenda. In discussing these factors the script may help participants to focus on the question asked (rather than on ancillary questions that they might prefer to answer), and then helps them to think realistically. There is no evidence that the script did not achieve this at the $8 and $5 payment levels. One explanation for the script's limited effect at the $3 payment level is that some participants may have thought it implausible that they would vote against the referendum at such a low payment. That is, they may not have envisioned significant bias at that level, leaving them less susceptible to the script.

The Cummings and Taylor script is long, and would thus be difficult to use on the phone or in the mail. A much shorter version of the entreaty used by Loomis et al. [13], which did not specifically educate respondents about hypothetical bias, helped reduce hypothetical bias but under-corrected on average. Based simply on the differences between the Loomis et al. and the Cummings and Taylor scripts, it appears that the entreaty has a greater impact if it not only instructs participants to act as if the response were for real, but also reports that responses to a hypothetical payment question are often very different from those to a real payment question and then explains why such a difference might occur.

To summarize, ours and earlier tests of the entreaty method suggest that: (1) entreaties have an effect, although the effect may be modest depending on its wording and on the details of the valuation question; (2) just urging respondents to act as if they were really spending their money helps, but is insufficient; (3) to be effective the script must make hypothetical bias salient to the respondent; (4) the cheap talk script appears to increase in effectiveness as payment level, and correspondingly hypothetical bias, increase—though the evidence of this is limited to the relatively modest payment levels used here and by Cummings and Taylor; and (5) under-correction can happen at low payment amounts.

Future research must determine (1) which parts of the Cummings and Taylor script are essential and which are not; (2) what the effect of the script is at higher referendum payment levels; and (3) whether a different script would be more effective at very low payment levels. Of course, development of cheap talk scripts would be aided by a theory of hypothetical bias, explaining why it occurs. Development of such a theory may require moving from studies of aggregate behavior to research at the individual level. Work at the individual level would determine whether hypothetical bias occurs in all respondents, or only in some. It would then determine whether entreaties affect only respondents subject to bias (causing accurate correction for each such respondent), or all respondents (causing a set of under-corrections balanced by a set of over-corrections).

References

[1] K. Blumenschein, M. Johannesson, G.C. Blomquist, B. Liljas, R.M. O'Connor, Experimental results on expressed certainty and hypothetical bias in contingent valuation, Southern Econom. J. 65 (1998) 169–177.

[2] R.R. Boyce, T.C. Brown, G.D. McClelland, G.L. Peterson, W.D. Schulze, Experimental evidence of existence value in payment and compensation contexts, in: K.J. Boyle, T. Heekin (Eds.), Western Regional Research Project W-133: Benefits and Costs in Natural Resources Planning, University of Maine, Orono, ME, 1989.

[3] T.C. Brown, P.A. Champ, R.C. Bishop, D.W. McCollum, Which response format reveals the truth about donations to a public good?, Land Econom. 72 (1996) 152–166.

[4] R.T. Carson, W.M. Hanemann, R.J. Kopp, J.A. Krosnick, R.C. Mitchell, S. Presser, P.A. Ruud, V.K. Smith, M. Conaway, K. Martin, Referendum design and contingent valuation: the NOAA panel's no-vote recommendation, Rev. Econom. Statist. 80 (1998) 484–487.

[5] R.G. Cummings, S. Elliott, G.W. Harrison, J. Murphy, Are hypothetical referenda incentive compatible?, J. Polit. Econom. 105 (1997) 609–621.

[6] R.G. Cummings, G.W. Harrison, E.E. Rutstrom, Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive-compatible?, Amer. Econom. Rev. 85 (1995) 260–266.

[7] R.G. Cummings, L.O. Taylor, Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method, Amer. Econom. Rev. 89 (1999) 649–665.

[8] M. Dickie, A. Fisher, S. Gerking, Market transactions and hypothetical demand: a comparative study, J. Amer. Statist. Assoc. 82 (1987) 69–75.

[9] C.L. Eavey, G.J. Miller, Bureaucratic agenda control: imposition or bargaining, Amer. Polit. Sci. Rev. 78 (1984) 719–733.

[10] J.P. Hoehn, A. Randall, A satisfactory benefit cost indicator from contingent valuation, J. Environ. Econom. Management 14 (1987) 226–247.

[11] M.J. Kealy, J.F. Dovidio, M.L. Rockel, Accuracy in valuation is a matter of degree, Land Econom. 64 (1988) 158–171.

[12] J.A. List, Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards, Amer. Econom. Rev. 91 (2001) 1498–1507.

[13] J. Loomis, T. Brown, B. Lucero, G. Peterson, Improving validity experiments of contingent valuation methods: results of efforts to reduce the disparity of hypothetical and actual WTP, Land Econom. 72 (1996) 450–461.

[14] J. Loomis, A. Gonzalez-Caban, R. Gregory, Do reminders of substitutes and budget constraints influence contingent valuation estimates?, Land Econom. 70 (1994) 499–506.

[15] H. Neill, R. Cummings, P. Ganderton, G. Harrison, T. McGuckin, Hypothetical surveys and real economic commitments, Land Econom. 70 (1994) 145–154.

[16] T. Romer, H. Rosenthal, Bureaucrats versus voters: on the political economy of resource allocation by direct democracy, Quart. J. Econom. 93 (1979) 563–587.