1-1-2009

Bridging the Gap between the Field and the Lab: Environmental Goods, Policy Maker Input, and Consequentiality

Christian A. Vossler
University of Tennessee

Mary F. Evans
Claremont McKenna College

Bridging the Gap between the Field and the Lab: Environmental Goods, Policy Maker Input, and Consequentiality

Christian A. Vossler[a] and Mary F. Evans[b]

[a] *Department of Economics and Institute for a Secure and Sustainable Environment, University of Tennessee, Knoxville, TN 37996*
[b] *Department of Economics, University of Tennessee, Knoxville, TN 37996 and The Robert Day School of Economics and Finance, Claremont McKenna College, Claremont, CA 91711*

*Correspondence should be directed to Christian Vossler, Department of Economics, 527C Stokely Management Center, The University of Tennessee, Knoxville, TN 37996-0550. E-mail: cvossler@utk.edu. Telephone: 865-974-1699. Fax: 865-974-4601. We thank Bill Neilson, John Lynham, John List, Scott Gilpatric and participants at the CU Environmental and Resource Economics Workshop (2007) and the Triangle Resource and Environmental Economics Workshop (2007) for their helpful comments and suggestions. Steve Cotten provided excellent research assistance.

Bridging the Gap between the Field and the Lab: Environmental Goods, Policy Maker Input, and Consequentiality

Abstract:

This paper explores the criterion validity of stated preference methods through experimental referenda that capture key characteristics of a stated preference survey for a proposed environmental program. In particular, we investigate whether advisory referenda, where participant votes have either known or unknown weight in the policy decision, can elicit values comparable to that of a standard, incentive compatible referendum. When participants regard their votes as consequential, our results suggest there is no elicitation bias with advisory referenda. For advisory referenda where participants view their votes as inconsequential, and for purely hypothetical referenda, we observe elicitation bias.

JEL classification: Q51; C91; D72

Keywords: experimental referenda; consequentiality; stated preferences; willingness to pay; policy maker input

1. Introduction

Stated preference (SP) methods are increasingly used for valuing environmental and other public programs, yet continue to be met with skepticism in many academic circles and policy arenas. There is a wealth of accumulated evidence related to the validity of SP methods. One interpretation of the literature is that the evidence of validity is moderately positive, with the important exception of findings from studies that endeavor to test criterion validity. In particular, myriad field survey studies as well as laboratory and field experiments find evidence that contingent values or choices overstate actual willingness to pay (WTP) or willingness to accept [14, 15, 16]. Carson and Groves [3] argue that theory can explain the purported bias found in many of these studies. For instance, many criterion validity studies involve private goods, base criterion values on incentive incompatible elicitation mechanisms, or use divergent mechanisms to elicit contingent and actual values. Further, Carson and Groves propose conditions under which "advisory referenda" are incentive compatible. This paper helps inform the debate over the validity of SP methods with carefully designed experimental referenda that capture key characteristics of a field survey setting. In particular, we examine whether advisory referenda are demand revealing.

The majority of criterion validity studies involve laboratory or field experiments, in large part due to the difficulty of having an appropriate criterion measure with which to compare field surveys, and the paucity of appropriate naturally-occurring data [11, 21]. Recent papers identify a common characteristic among laboratory and field experiments: they use a purely hypothetical, inconsequential choice setting as the experimental analog to a SP survey [3, 4, 8, 9, 12]. The papers assert, and we agree, that the incentives differ between purely hypothetical choice settings and actual field valuation surveys, and it is thus unclear whether we can use these studies to draw

conclusions about the incentive properties of field surveys. In a field survey respondents may believe, in fact are often led to believe, that their responses have potential influence on agency action, i.e. they are "consequential". Cover letters for SP studies often state that the survey results will be shared with state or local officials (see for example Harrison's [9, pp. 141-142] discussion of [6, 18]). Survey instruments generally provide additional signals that respondents should take seriously the valuation exercise through, for example, reminders of substitute goods and budget constraints. Evidence from a recent SP survey suggests that respondents do believe responses to be consequential, and further that elicited preferences depend on (perceived) consequentiality.[1]

A binding public goods referendum has the theoretical advantage, under certain conditions, of incentive compatibility; further, it reflects the widely-used framing of survey value elicitation questions. Within the context of laboratory and field experiments, there are two main approaches for assessing criterion validity based on experimental referenda. One approach uses a binding referendum on an actual good as a benchmark from which to compare treatments intended to capture the SP survey setting. In this situation, vote proportions or estimated welfare measures from the binding referendum are taken to be an appropriate criterion. A second approach is to use induced values for a virtual good, which allows tests of demand revelation at the individual-level through comparisons between observed votes and theoretical predictions.

Evidence from the three relevant recent studies that use induced-value referenda suggests a distinction between induced value and so-called "homegrown" value experiments [1, 19, 22]. In particular, experiments involving actual goods and homegrown values systematically find significant differences (in terms of aggregate decisions or elicited values) between inconsequential and real payment settings whereas in only one case – one of the three

experiments described in [1] – do induced-value studies find such a difference.[2] People may form different valuations for an actual good in hypothetical versus binding decision settings [19], whereas the value formation task is largely absent in an induced-value experiment regardless of whether the referendum is binding or purely hypothetical. Our study endeavors to capture important features of the survey setting in our experimental design, including the value formation process. This favors the use of homegrown value experiments.

The main contribution of this study is the exploration of consequential decision settings where the referendum outcome is determined both by participant decisions and agency (i.e. experiment moderator) preferences. This captures the situation where SP survey responses are, or are perceived as, an input into a policy decision. In our two Explicit Advisory referenda, a known number of (undisclosed) agency votes are combined with participant votes to determine the outcome. This allows us to investigate whether participant choices are invariant to the relative weight placed on their votes. A third treatment, the Implicit Advisory Referendum, comes closest to representing a valuation survey setting where participants vote in an advisory referendum. The provision rule is undisclosed as well as the process the agency uses to determine an outcome. Participants are told only that the probability the good will be provided is increasing in the number of "yes" votes. Our benchmark treatment is a majority vote referendum with the outcome determined solely by participant votes. A purely hypothetical referendum completes the design.

Other features of this study are intended to capture the field survey environment in which we are interested while avoiding common design issues that may have plagued previous criterion validity studies. First, instead of using a private good, or a publicly provided private good as in some previous experimental referenda studies [4], our experiments ask individuals to vote in a

referendum about an environmental good – in particular the provision and maintenance of a classroom recycling container – with potential passive-use value. Thus, the nature of the good is consistent with most SP studies.

Second, our experimental referenda are "one-shot" in the sense that there are no obvious opportunities for participants in consequential treatments, or others, to purchase the good outside the lab. Cherry et al. [5] provide empirical evidence that in the presence of an outside option auction participants under-reveal demand. Carson and Groves [3] suggest that past laboratory experimental referenda do not have this characteristic, and provide theoretical arguments that this leads to underestimates of elicited values in the actual payment setting due to the opportunity to free-ride off the contributions of others.

2. Theoretical framework

The incentive compatibility of a binding dichotomous choice referendum with a plurality vote implementation rule is well-known [7]. To be clear, we refer to an incentive compatible elicitation mechanism as one where the agent has an incentive to truthfully reveal his preferences. In a binding referendum setting, the basic logic is that for some combinations of votes cast by others, a single voter is pivotal (i.e. the outcome pivots on his vote) in which case voting for his preferred alternative increases utility, whereas for other combinations of votes by others a single voter is not pivotal and utility is thus the same regardless of his vote. Provided the former scenario occurs with some positive probability, the voter has a weakly dominant strategy to vote for his preferred alternative.

Carson and Groves [3] confirm the incentive compatibility of a similarly structured advisory referendum. In general, incentive compatibility of the advisory referendum follows

from the respondent's *perceived* influence on the outcome. Specifically, incentive compatibility of the advisory referendum remains when the following conditions are satisfied:

(I) the referendum consists of a "yes" or "no" vote on a single issue;

(II) participants care about the referendum outcome;

(III) the decision-maker can compel payment from participants; and

(IV) implementation of the proposed outcome is weakly monotonically increasing with the fraction voting "yes".

Carson et al. [4] label condition (I) the take-it-or-leave-it condition since it requires that passage of the referendum serves as the only means by which the proposed outcome can be obtained. When (IV) holds in an advisory referendum, the agent's response *potentially* influences whether the proposed outcome is implemented. Carson and Groves classify settings in which conditions (II) and (IV) are satisfied as "consequential." Note that condition (II) is not an assumption particular to advisory referenda, as a binding referendum is not incentive compatible for voters who are indifferent between voting options.

As described in the next section, we consider mechanisms in which the result of a referendum vote is an input into the decision-making process on whether a proposal is implemented. In particular, we allow other factors, such as agency preferences, to influence the decision in addition to the participant votes. Provided that conditions (I) through (IV) are satisfied, incentive compatibility holds in this setting; incentive compatibility does not require that the decision to implement a proposal be determined exclusively by the referendum results.

To illustrate this result, consider a setting in which a group of n individuals vote in an advisory referendum on a proposal to provide a public good at a specified cost. Let r_i represent voter i's vote with $r_i = 1$ if i votes "yes" and $r_i = 0$ if i votes "no". The aggregate vote is then

$R_n \equiv \sum_{i=1}^{n} r_i$. Let $q(R_n)$ denote the probability the proposal is implemented given the aggregate vote R_n. While $q(\cdot)$ varies with the specific decision rule employed, condition (IV) requires $q(x) \geq q(y)$ for all *x* and *y* where $x > y$, with $q(x) > q(y)$ for some $x > y$.

The regulator or agency considers how the group votes in addition to her preferences when determining whether or not to implement the proposal. Specifically, we assume she combines her preferences with those of the voters by assigning herself a specified number of votes. Thus, we let *n* include regulator votes (if any).

Consider the decision facing voter *i* in this setting. Let v_i and c_i represent the value and cost of the proposal, respectively, to voter *i*. If the proposal is implemented, then voter *i* gains utility $U(v_i - c_i)$ where *U* is increasing and concave in its argument. If not, then assume voter *i* gains utility $U(0)$. Note that condition (II) requires $U(v_i - c_i) \neq U(0)$. Let $R_{n-1} \equiv \sum_{j \neq i} r_j$ represent the aggregate vote <u>independent</u> of *i*'s vote. From *i*'s perspective R_{n-1} is a discrete random variable defined on the set $\{0,1,..., n-1\}$ with probability mass function f_i.

Voter *i*'s expected utility from voting "yes" and "no", conditional on realization $\hat{R}_{n-1}$, are then as follows:

$$EU_{yes} = \left[q\left(\hat{R}_{n-1}+1\right)\right]U(v_i - c_i) + \left[1 - q\left(\hat{R}_{n-1}+1\right)\right]U(0) \quad (1)$$

$$EU_{no} = \left[q\left(\hat{R}_{n-1}\right)\right]U(v_i - c_i) + \left[1 - q\left(\hat{R}_{n-1}\right)\right]U(0). \quad (2)$$

For realizations $\hat{R}_{n-1}$ such that $q\left(\hat{R}_{n-1}+1\right) > q\left(\hat{R}_{n-1}\right)$, the voter affects the probability of implementation, i.e. the voter is pivotal. In this case $EU_{yes} > EU_{no}$ (or $EU_{yes} < EU_{no}$) and voter *i* is better off voting for his preferred alternative $r_i = 1$ (or $r_i = 0$). Otherwise, for $\hat{R}_{n-1}$ such that

$q\left(\hat{R}_{n-1}+1\right)=q\left(\hat{R}_{n-1}\right)$, the voter has no effect on the outcome (i.e. he is not pivotal) and his expected utility from voting "yes" or "no" are equal. Under condition (IV) voter *i* has a strictly positive pivot probability, denoted p_i.[3] It follows that voter *i* has a weakly dominant strategy to vote his preference.

In the absence of explicit information on the preferences of other voters, satisfaction of condition (IV) seems likely. However, holding the number of participant votes constant, the inclusion of regulator votes does affect the pivot probability. To highlight this effect, assume that in the absence of regulator votes $p_i > 0$ as above. The addition of regulator votes has two potential effects on voting incentives. First, the regulator votes increase the total number of votes cast. All else equal, an increase in the total number of votes cast, *n*, decreases the probability mass at each point as it widens the support of the f_i distribution. Thus, the increase in *n* decreases voter *i*'s pivot probability *ceteris paribus*. Second, the presence of regulator votes may alter the shape of f_i. For instance, if voter *i* assigns a relatively high probability on all regulator votes cast being identical, this will move probability mass towards the tails of the f_i distribution.

From this, we hypothesize that the percent in favor when the outcome is completely determined by the results of a binding referendum is equal to that when the outcome is determined only in part by the results of an advisory referendum provided conditions (I) through (IV) hold. Theory suggests our first testable hypothesis:

H1: In consequential advisory referenda, willingness to pay (WTP) is equivalent to WTP in a binding referendum.

For inconsequential referenda, economic theory does not provide a testable hypothesis. However, empirical evidence from previous studies based on homegrown values suggests that

respondents will on average over-report their value (i.e. vote "yes" when they otherwise would not have in a binding referendum). Based on this empirical evidence, we state our second hypothesis:

> **H2**: In *in*consequential referenda, willingness to pay (WTP) is greater than WTP in a binding referendum.

3. Experimental design

Our experimental design ensures that condition (I) holds for all treatments. Condition (III) holds for all but our Hypothetical Referendum. For these same treatments, condition (II) is also likely to hold and we maintain the assumption that it does. The incentives relevant to condition (IV) vary across treatments, as we describe in what follows.

3.1 Treatments

3.1.1. Baseline referendum

The Baseline is a referendum with a majority vote implementation rule: if more than 50% of the votes are "yes" votes, the proposal passes; otherwise it does not pass.[4] This mechanism is well-known to be incentive compatible.

3.1.2. Implicit advisory referendum

The Implicit Advisory Referendum provides the most direct test of whether an advisory referendum, where participant votes potentially influence a policy outcome, is equivalent to a binding, majority vote referendum. In particular, this treatment comes closest to a survey referendum where there is no direct signal on exactly how responses will be used in the policy decision. In this treatment, participants are given the following information:

> Passage of the referendum will not solely be determined by how you and the other participants vote. In particular, we, the experiment coordinators, will use your votes as

> *advice* on whether or not to pass the referendum. While you will not be told how we came to a decision, know that the likelihood the referendum is passed increases with the number of YES votes cast.

As indicated by the above passage, participants are told that their votes have the potential to influence the outcome (i.e. condition (IV) holds). Unknown to participants, in implementation the decision rule is identical to the baseline, i.e. a majority vote implementation rule with no experiment moderator votes.

3.1.3. Explicit advisory referenda

The next two treatments are, like the Baseline Referendum, based on a majority-vote implementation rule. However, both participants and experiment moderators have an explicit number of votes. The number of moderator votes is common knowledge, as is the fact that these votes are predetermined. The instructions indicate that these moderator votes are potentially any combination of "yes" and "no" votes.[5] This language is intended to convey to participants that even with moderator votes, (in expectation) the probability of implementation increases with the number of participant "yes" votes.

We consider two weighting schemes. In the Explicit-Consequential (Explicit-C) Advisory Referendum, the number of experiment moderator votes is equal to 25% of total votes cast (e.g. with a group of 12 participants there are 4 moderator votes), whereas it was 75% of total votes cast in the Explicit-Uncertain Consequential (Explicit-UC) Advisory Referendum (e.g. 12 participant votes and 36 moderator votes). Given the number of moderator votes, the Explicit-C Advisory Referendum is incentive compatible even if a respondent believes all moderator votes are "yes" ("no") votes. The Explicit-UC Advisory Referendum loses incentive compatibility for voters who perceive a sufficiently high number of "yes" ("no") moderator votes. Over this range of votes the participant's pivot probability is zero as the moderator votes are perceived to solely

determine the outcome. Hypothesis H2 applies for Explicit-UC Advisory Referendum participants who regard their vote as inconsequential.

3.1.4. Hypothetical referendum

The last treatment, the Hypothetical Referendum, is included to establish the existence of hypothetical bias in the absence of financial incentives. Clearly for a purely inconsequential vote, conditions (III) and (IV) are violated, and incentive compatibility does not hold. Therefore, the H2 hypothesis is relevant.

3.2. Proposal and procedures

During the experiment participants are visually isolated. Further, it is common knowledge that each individual is assigned an ID number, but that the researchers are not able to link an ID number with the individual's name. These controls for anonymity are intended to minimize the possible impacts of social networks on the voting decision (see [13]), and to better reflect the isolated field survey setting.

Participants receive a written copy of the instructions, which include the proposal followed by a description of the treatment-specific provision rule.[6] The instructions are read aloud by one of the authors. The wording of the proposal is identical in all but the Hypothetical Referendum, which uses slightly different language to make clear that the vote is inconsequential. Participants are asked to vote in a referendum on whether everyone in the group (session) would fund the provision and administration of one on-campus, classroom recycling container at a particular cost. As most participants, given the scope of the experiments, would not ultimately use the recycling bin if provided, this good potentially has passive-use value for some. Further, as there is no clear venue for which students themselves can purchase classroom recycling containers (and have them maintained by the University), this avoids the incentive

problems associated with having an outside purchase option. Also, the University was not currently considering a program that would provide classroom recycling containers. This information is relayed to avoid the notion that the experiments would be used to advise University policy, which could confound our results [3].

After any questions are addressed regarding the proposal and decision rule, participants are directed to their computer, which displays the referendum. The only information provided on screen that does not appear in the instructions is the cost of the proposal. The cost varies across participants, which is not made common knowledge; in fact, we deliberately exclude any cost amount in the instructions to avoid drawing attention to the fact that costs differ. This mimics the SP survey environment, where such information is absent.[7]

Varying the cost across participants enables identification of median WTP in all treatments. In determining the costs, we took into consideration the cost incurred by us to fund the recycling bin in the event the proposal passes, as well as the credibility of different costs given the good and group size. The first two sessions, which correspond with our Baseline majority vote referendum, were used to determine a reasonable distribution. In particular, the cost amounts used in all but these two sessions are $1, $3, $6, and $8. The first two sessions also included costs of $2, $4, $5 and $9.

The experiment is followed by a two-page questionnaire, which probes respondents about their vote as well as elicits basic demographic information. Participants are told the outcome of the referendum, and if it passes a volunteer is asked to place in the mail an envelope containing the appropriate payment.

3.3. Participants

Two hundred and fifty-six undergraduate students at the University of Tennessee participated in the experiment during a two-week period of September, 2007. These individuals were drawn from a large group who were registered to be potential participants in economics experiments. The participant pool was similar to the general undergraduate population in terms of age, gender, and academic major. Experiments were conducted in a designated experimental laboratory. There are four groups of size twelve for each of the five experiment treatments. For our baseline treatment there was an additional session with sixteen participants.

Prior to the voting experiment, everyone participated in an unrelated first-price auction (induced-value) experiment. Experiment parameters were chosen to have pre-vote earnings in the $15 to $20 range.[8] Earnings were (discreetly) supplemented for seven participants to make sure that participants had at least $10 prior to voting in the referendum. Sessions lasted about an hour.

4. Results

The results of the experimental referenda are summarized in Tables 1 and 2. Table 1 presents the raw vote percentages by treatment and cost amount. These results suggest that the probability of voting "yes" decreases monotonically with cost for the Baseline, Implicit Advisory, and Hypothetical referenda. For the Explicit-C Advisory and Explicit-UC Advisory referenda, the percentage of "yes" votes is slightly higher at $8 than at $6. However, the sample sizes for the $8 cost are small, and certainly the raw voting results do not account for preference heterogeneity. All treatments enjoy at least 50% approval at the lowest cost amount, $1, and in only the Explicit-UC Advisory Referendum (at $8) and Hypothetical Referendum (at $3) do we see such high approval at another cost. The previous observations taken together suggest that our bid design adequately identifies median WTP.

We begin our analysis with simple nonparametric tests of equivalence between the Baseline proportion of "yes" votes and the vote proportions from other treatments. Using a Fisher exact (two-sided) test, we fail to reject the null hypothesis for the Explicit-C (p = 0.842) and Implicit Advisory referenda (p = 0.439). The Explicit-UC (p = 0.053) and Hypothetical (p = 0.082) referenda are statistically different from the Baseline.[9] These results coincide with casual observations based on Table 1, as the proportion of "yes" votes for the Explicit-UC (54%) and Hypothetical (52%) are noticeably higher than for the Baseline (34%). These statistical results are only suggestive, however, as they do not control for proposal cost or fundamental differences in preferences across participants and treatment groups. To include such controls as well as to facilitate estimation of median WTP, we turn to a formal parametric analysis.

Following the maximum likelihood approach of Cameron and James [2], we treat willingness to pay as a censored dependent variable for which we obtain the signal $WTP_i \geq c_i$ if participant *i* votes "yes" to cost c_i and the signal $WTP_i < c_i$ if participant *i* votes "no" to cost c_i. In particular, let WTP_i be a linear function of a column vector of covariates, x_i, such that $WTP_i = x_i' \beta + \varepsilon_i$, where β is a column vector of unknown parameters and ε_i is a normally distributed mean-zero error term with standard deviation σ. Assuming the error term has a normal distribution here is analogous to assuming a normal distribution for WTP_i. With our functional form and error distribution assumptions, interpretation of estimated parameters is analogous to that of a standard linear regression model that treats *WTP* as a directly observed (i.e. uncensored) dependent variable.

WTP is assumed to be a function of treatment-specific indicators, as well as control variables hypothesized to be correlated with participant preferences for the recycling program. These include (pre-referendum) experimental auction earnings, an indicator variable that equals

unity if the participant has class in the building designated as the location for the proposed recycling container, an indicator for environmental group membership, age, and gender.[10] Bootstrap estimates for standard errors are used, with assumed clustering at the group level.[11]

The estimated WTP regression (Model I), as well as descriptions of included variables, are presented in Table 2. An indicator variable corresponding with the Baseline Referendum is omitted, such that treatment effects are measured relative to the Baseline. Insignificance of parameters on the *Explicit-C* and *Implicit Advisory* indicators thus suggests statistical equivalence of median WTP between each of these treatments and the Baseline Referendum. This is consistent with theoretical predictions, as all three referenda are incentive compatible under reasonable assumptions; we cannot reject H1 for the Explicit-C and Implicit Advisory referenda. Further, the coefficients on *Explicit-UC* and *Hypothetical* suggest demand is over-revealed for these treatments, and statistically different than Baseline demand at the 5% significance level. In fact, estimates suggest that, on average, elicited WTP is roughly $2.50 higher in these treatments. To put this in perspective, the estimated model suggests that median WTP for the Baseline, Explicit-C Advisory and Implicit Advisory referenda (evaluated at the mean of the data) is about $2. Thus, elicited Explicit-UC Advisory and Hypothetical referenda WTP suggest a bias of over 100%. According to participant control variables, current use of the classroom building where the proposed recycling bin would be located, membership in an environmental organization, and age have positive and significant effects on WTP.

Certainly the finding of hypothetical bias for our Hypothetical Referendum is of no real surprise and is consistent with past experiments and our hypothesis H2. Yet, the result is still important as the absence of hypothetical bias in this study would cast doubt on the general relevance of other treatment effects. What is somewhat unexpected is the apparent elicitation

bias in the Explicit-UC Advisory Referendum. However, as suggested previously, some participants in this treatment may regard their responses as inconsequential. We explore this possibility below.

Given the high number of experiment moderator votes in the Explicit-UC Advisory Referendum, under a range of beliefs about the proportion of "yes" moderator votes a participant may perceive his vote to be inconsequential (i.e. his pivot probability equals zero). In particular, the participant would perceive his vote to be inconsequential if he believes that: (1) at least 25 of 36 or 69% of moderator votes are "yes"; or (2) less than 11 of 36 or 31% are "yes" votes. Some insight on whether this is true can be gleaned from the experiment questionnaire. In particular, respondents were asked what percentage of moderator votes they perceived to be "yes" votes. Eleven out of forty-eight indicated that they believed that 70% or more moderator votes were "yes", while seven stated that 30% or less were "yes" votes. Of these eighteen participants, two-thirds voted "yes" in the referendum, consistent with the observed direction of elicitation bias.

To more formally assess whether the observed elicitation bias pivots upon the participants' beliefs about moderator votes, we add to the previous WTP regression model a dummy variable (labeled *Inconsequential*) that equals unity for the 18 respondents who may have viewed their vote as inconsequential. This model is presented as Model II in Table 2. The inclusion of *Inconsequential* renders the *Explicit-UC* coefficient statistically insignificant, which suggests there is no elicitation bias for those in the Explicit-UC Advisory Referendum who did perceive their vote to be consequential (i.e., we cannot reject H1). Further, the *Inconsequential* coefficient is significant at the 5% level and suggests that elicited WTP is $3.70 higher for those in the Explicit-UC Advisory Referendum who perceive their vote to be inconsequential relative to those that do not. In sum, evidence from the experiment questionnaire suggests that bias in

elicited WTP based on the Explicit-UC Advisory Referendum is driven by a violation of the consequentiality assumption (i.e. condition (IV)) for 38% of participants.

5. Conclusion

The results of our experiment, which was designed to capture key characteristics of a stated preference survey for a proposed environmental program, provide support for the theoretical predictions regarding voter behavior in advisory referenda. The main contribution of this study is an exploration of situations where respondent decisions are merely an input into the policy process. Our focus on the relative weight of participant votes in the policy outcome differs from that of other carefully designed experimental referenda studies that have examined the effect on voting behavior of the probability that the referendum results are binding [4, 12]. Although we argue that our design comes closer to replicating the decision faced by respondents in field surveys, our take-home message is the same: if SP survey respondents engaged in an advisory referendum view their responses as consequential, then SP surveys have criterion validity. This message is further corroborated by the survey work of Herriges et al. [10]. Additional evidence that respondents view SP surveys as consequential would serve to strengthen the findings from these experimental studies while casting doubt on the myriad studies that assume an inconsequential decision setting is the analog to a field SP survey.

Future experimental referenda studies that relate to ours might benefit from more carefully designed questions or elicitation procedures for determining whether participants perceive their decisions to be consequential. In particular, the questionnaire results used to characterize respondents in our Explicit-UC treatment may be best interpreted as suggestive as there was no financial incentive for truth-telling. One possible approach, suggested by the voting

experiment of Tyran [20], is to have respondents predict the percentage of “yes” votes and pay them based on the accuracy of the prediction.

Finally, the experimental design may be modified to invoke closer correspondence with the field survey setting. For instance, participants could come to the lab one-by-one to vote, without explicit knowledge of the number of participants or exactly when the outcome would be determined. Certainly in the survey setting these two factors are unknown. Further, given that many researchers, usually out of concern for statistical efficiency or scenario plausibility, continue to use alternative elicitation formats such as the payment card or dichotomous choice with follow-up certainty question, sometimes with a voluntary contributions payment vehicle, exploration of these formats in controlled but consequential decision settings is warranted.

References

1. A.C. Burton, K.S. Carson, S.M. Chilton, W.G. Hutchinson, Resolving questions about bias in real and hypothetical referenda, Environ. Resource Econ. 38 (2007) 513-525.
2. T.A. Cameron, M. D. James, Efficient estimation methods for use with "closed-ended" contingent valuation survey data, Rev. of Econ. Statist. 69 (1987) 269-276.
3. R. Carson, T. Groves, Incentive and information properties of preference questions, Environ. Resource Econ. 37 (2007) 181-210.
4. R. Carson, T. Groves, J. List, M. Machina, Probabilistic influence and supplemental benefits: a field test of the two key assumptions underlying stated preferences, Working paper, Department of Economics, University of California, San Diego, 2004.
5. T.L. Cherry, P. Frykblom, J.F. Shogren, J.A. List, M.B. Sullivan, Laboratory testbeds and non-market valuation: the case of bidding behavior in a second-price auction with an outside option, Environ. Resource Econ. 29 (2004) 285-294.
6. D.A. Dillman, Mail and telephone surveys: the total design method, Wiley-Interscience, New York, 1978.
7. R. Farquharson, Theory of voting, Yale University Press, New Haven, CT, 1969.
8. G.W. Harrison, J.A. List, Field experiments, J. Econ. Lit. 42 (2004) 1009-1055.
9. G.W. Harrison, Experimental evidence on alternative environmental valuation methods, Environ. Resource Econ. 34 (2006) 125-162.
10. J. Herriges, C. Kling, C.-C. Liu, J. Tobias, What are the consequences of consequentiality?, Working paper, Department of Economics, Iowa State University, 2007.
11. R.J. Johnston, Is hypothetical bias universal? Validating contingent valuation responses using a binding public referendum, J. Environ. Econ. Manage. 52 (2006) 469-481.

12. C.E. Landry, J.A. List, Using ex ante approaches to obtain credible signals for value in contingent markets: evidence from the field, Amer. J. Agr. Econ. 89 (2007) 420-432.
13. J.A. List, R. Berrens, A. Bohara, J. Kerkvliet, Examining the role of social isolation in stated preferences, Amer. Econ. Rev. 94 (2004) 741-752.
14. J.A. List, C.A. Gallet, What experimental protocol influence disparities between actual and hypothetical stated values? Evidence from a meta-analysis, Environ. Resource Econ. 20 (2001) 241-254.
15. J. Little, R. Berrens, explaining disparities between actual and hypothetical stated values: further investigation using meta-analysis, Econ. Bull. 3 (2004) 1-13.
16. J.J. Murphy, T.H. Stevens, D. Weatherhead, A meta-analysis of hypothetical bias in contingent valuation, Environ. Resource Econ. 30 (2005) 313-325.
17. N.A. Powe, I.J. Bateman, Investigating insensitivity to scope: a spilt-sample test of perceived scheme realism, Land Econ. 80 (2004) 258-271.
18. R.D. Rowe, W.D. Schulze, W.D. Shaw, D. Schenk, L.G. Chestnut, Contingent valuation of natural resource damage due to the Nestucca Oil Spill. Final Report RCG/Hagler, Bailly, Inc., Boulder, 1991.
19. L.O. Taylor, M. McKee, S.K. Laury, R.G. Cummings, Induced-value tests of the referendum voting mechanism, Econ. Letters 71 (2001) 61-65.
20. J-T. Tyran, Voting when money and morals conflict: an experimental test of expressive voting, J. Public Econ. 88 (2004) 1645-1664.
21. C.A. Vossler, J. Kerkvliet, A criterion validity test of the contingent valuation method: comparing hypothetical and actual voting behavior for a public referendum, J. Environ. Econ. Manage. 45 (2003) 631-649.

22. C.A. Vossler, M. McKee, Induced-value tests of contingent valuation elicitation mechanisms, Environ. Resource Econ. 35 (2006) 137-168.

Table 1. Voting results ("yes" votes / total votes)

Cost	Baseline[#]	Implicit Advisory	Explicit-C Advisory	Explicit-UC Advisory	Hypothetical
$1	7/12 (58%)	11/16 (69%)	8/16 (50%)	13/16 (81%)	11/16 (69%)
$3	5/13 (38%)	5/13 (38%)	5/13 (38%)	6/13 (46%)	8/13 (62%)
$6	3/13 (23%)	4/13 (31%)	3/13 (23%)	4/13 (31%)	5/13 (38%)
$8	0/6 (0%)	0/6 (0%)	2/6 (33%)	3/6 (50%)	1/6 (17%)
Overall	**15/44 (34%)**	**20/48 (42%)**	**18/48 (38%)**	**26/48 (54%)**	**25/48 (52%)**

Note: The "Overall" measure for the Baseline Referendum is not entirely comparable with the other treatments given the fewer responses at $1. Based on simple extrapolation, it is expected that 9.33/16 would vote "yes" in the Baseline at $1, which yields a comparable overall expected vote percentage of 36% (17.33/48).

[#] Additional results for this treatment: 3 of 8 "yes" votes at cost of $2; 2/4 at $4; 0/4 at $5; 0/4 at $9.

Table 2. Willingness to pay regressions

Variable / Parameter	Variable Description	Sample Mean	Model I	Model II
Implicit	=1 if Implicit Adivsory Referendum; =0 otherwise	0.188 [0.391]	0.71 (1.68)	0.77 (1.69)
Explict-C	=1 if Explicit-C Advisory Referendum; =0 otherwise	0.188 [0.391]	0.11 (2.79)	0.11 (2.85)
Explicit-UC	=1 if Explicit-UC Advisory Referendum; =0 otherwise	0.188 [0.391]	2.64** (1.15)	1.20 (1.49)
Hypothetical	=1 if Hypothetical Referendum; =0 otherwise	0.188 [0.391]	2.47** (1.10)	2.46** (1.14)
Earnings	Experiment earnings prior to vote ($)	19.768 [7.672]	0.09 (0.07)	0.10 (0.08)
Use	=1 if uses building designated for recycling bin; =0 otherwise	0.746 [0.436]	2.21* (1.16)	2.24* (1.17)
EnvOrg	=1 if member of an environmental organization; =0 otherwise	0.293 [0.456]	2.05* (1.11)	1.80 (1.13)
Age	Age (years)	20.859 [2.656]	0.48** (0.22)	0.51** (0.24)
Male	=1 if male; =0 otherwise	0.516 [0.501]	-1.52 (1.03)	-1.69 (1.08)
Inconsequential	= 1 if perceived vote to be inconsequential; = 0 otherwise	0.070 [0.256]		3.70** (1.88)
Constant	Model Intercept		-11.34** (5.61)	-12.19** (6.17)
σ	Standard deviation of WTP		4.85** (1.11)	4.78** (1.10)
Log-Likelihood			-144.96	-143.33
Pseudo-R^2			0.1725	0.1818
N			256	256

Note: standard deviations in brackets; bootstrap standard errors are in parentheses.
*, and ** denote parameter is statistically different from zero at the 10% and 5% significance levels, respectively.

NOTES

[1] Herriges et al. [10]. These findings are consistent with earlier evidence on the importance of perceived scheme realism in explaining observed insensitivity to scope among SP respondents [17].

[2] In the particular experiment described in [1], this statistical difference occurs both in terms of aggregate vote distributions as well as in the pattern of individual-level deviations from theoretical predictions. The authors provide evidence that the result is driven by differences in other-regarding preferences.

[3] Given the actual votes cast by others, voter *i* is either pivotal or not. However, the pivot probability here refers to voter *i*'s perception that his vote will be pivotal.

[4] The assumed tiebreaking rule gives preference to the status quo outcome and in that respect parallels a supermajority vote rule. Several states require a supermajority vote for ballot issues; many states require a supermajority rule of the legislature to amend state constitutions, to increase taxes, and to pass general appropriations bills.

[5] In all cases, the experiment moderator votes were split down the middle (i.e. 50% "yes"), such that in actuality all outcomes were determined by participant votes. This information was never divulged to participants.

[6] The experiment instructions and questionnaire are contained in an Appendix that is available at JEEM's on-line archive of supplementary material, which can be accessed at http://www.aere.org/journals/.

[7] In a few instances a voter asked about the cost to others and we replied that "costs may or may not differ across participants."

[8] Average earnings were $19.78 prior to the vote.

[9] The p-value corresponding with one-sided test is 0.029 and 0.046 for the Explicit-UC and Hypothetical referenda, respectively.

[10] We initially ran regressions that included other covariates constructed from the experiment questionnaire (e.g. student status, major, #economics courses, etc.). These additional parameters are jointly insignificant and excluded from the final model.

[11] We also estimated cluster-robust standard errors, which lead to the same statistical conclusions.