Toward a better understanding of elicitation effects in stated preference studies

Christian A. Vossler and Ewa Zawojska

June, 2018

WORKING PAPER #2018-01

WORKING PAPER SERIES
DEPARTMENT OF ECONOMICS
HASLAM COLLEGE OF BUSINESS
http://econ.bus.utk.edu

HASLAM
COLLEGE OF BUSINESS
THE UNIVERSITY OF TENNESSEE, KNOXVILLE

Toward a better understanding of elicitation effects in stated preference studies

Christian A. Vossler*

Department of Economics and Howard H. Baker Jr. Center for Public Policy,

University of Tennessee, Knoxville, TN, USA

and

Ewa Zawojska

Faculty of Economic Sciences

University of Warsaw, Warsaw, Poland

*Correspondence should be directed to Christian Vossler, Department of Economics, 523 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 Adrienne Sudbury for her excellent research assistance and Robert Johnston for his helpful comments on an earlier draft. This research was undertaken while Ewa Zawojska was a visiting PhD student at the University of Tennessee, under the generous support of the National Science Centre in Poland (Opus 13, grant no. 2017/25/B/HS4/01076; Etiuda 4, grant no. 2016/20/T/HS4/00013).

Toward a better understanding of elicitation effects in stated preference studies

Abstract: Empirical evidence accumulated over several decades suggests that survey-based welfare measures for public goods can be very sensitive to the format of the value elicitation, e.g., an up-or-down vote or an open-ended willingness-to-pay question. The underlying drivers of these effects remain poorly understood. As myriad formats are employed in practice, this raises concerns for both academics and policymakers. We design and implement a controlled experiment to cleanly test for elicitation effects among a set of four oft-used formats: single binary choice, double-bounded binary choice, payment card, and open-ended. The experiment retains important field context properties (e.g., the funding of a public, environmental good) and varies only the elicitation format, while holding fixed ancillary characteristics of the elicitations (such as framing, decision rule, payment method, and incentive compatibility). We find all formats lead to statistically identical welfare estimates. On one hand, this evidence suggests that variance in design characteristics other than the elicitation format may explain some prior results. On the other, to the extent that characteristics of our elicitations can be mirrored in the field, this offers a pathway for mitigating elicitation effects.

Keywords: contingent valuation; mechanism design; experiment; voting; elicitation effects; convergent validity

JEL classification: Q51; C92; D82; H41

1. Introduction

In the context of using stated preference surveys to estimate monetary values for non-market goods, such as a change in air quality or a land conservation project, researchers are charged with the challenging task of providing valid welfare measures to inform decision-making processes. Although many threats to the validity of stated preference measures exist, as discussed in Johnston et al. (2017), one aspect of survey design that has received much attention is the choice of the value elicitation format, for example, whether to use an up-or-down vote or an open-ended willingness-to-pay (WTP) question. The stylized fact in the literature is that estimated values from field surveys are quite sensitive to the elicitation format (e.g., Bateman, Langford and Rasbash, 2001; Cameron et al., 2002; Champ and Bishop, 2006), which is often interpreted as a failure of convergent validity. Myriad formats continue to be used in practice, and this raises concerns for both academics and policymakers (see, for example, Bishop et al., 2017; Kling, Phaneuf and Zhao, 2012). As different formats often lead to different values, this questions whether some or even any of the formats truthfully reveal demand. Moreover, this introduces a source of procedural variance when comparing stated and revealed preferences. This paper proposes a path forward for acquiring a better understanding of elicitation effects, and presents new evidence from an experiment that connects past induced-value lab experiments with field survey studies.

The literature provides many possible explanations for observed elicitation effects (see Bateman et al., 2001). Here we briefly highlight a subset of them. Responses to single binary choice questions and other posted-price formats might be subject to “anchoring” or “reference point” bias, which can arise, for instance, from a respondent’s unfamiliarity with the elicitation method or with the good being valued. This has been suggested as an explanation for observed differences between single binary choice and open-ended responses (e.g., Green et al., 1998;

O'Conor, Johannesson and Johansson, 1999) and for differences between single and double-bounded binary choice responses (e.g., Scheufele and Bennett, 2013; Whitehead, 2002), among others. Hanemann (1995) argues that bids (prices) could be construed as quality signals. This can lead to differences between closed- and open-ended elicitation mechanisms, as for the latter there are no posted prices. Welsh and Poe (1998) show that elicitation effects may be driven by response (un)certainty, with "yes" responses to a single binary choice question associated with a lower (average) level of certainty, and statements of positive values in payment card and open-ended questions associated with a higher certainty level (on average). Similar results of the role of preference certainty for differences in values across formats have been documented in other work, e.g., Ready, Navrud and Dubourg (2001). Frew, Whynes and Wolstenholme (2003) speculate that observed differences in single binary choice valuations relative to open-ended and payment card formats may be in part attributable to "yea-saying." Carson and Groves (2007) argue that the theoretical response incentive properties differ significantly between single binary choice and other formats, giving rise to elicitation effects.

A somewhat lesser known and smaller set of studies examine elicitation effects in a controlled laboratory setting where values over outcomes are induced rather than homegrown, and elicitation mechanisms are designed to be incentive compatible (Carson, Chilton and Hutchinson, 2009; Collins and Vossler, 2009; Messer et al., 2010; Vossler and McKee, 2006). These studies demonstrate that several mechanisms resembling those used in the field are demand revealing in this controlled setting, and that there are no significant elicitation effects. A natural question arises whether these results could be generalized to field settings, which may seem doubtful in light of prevalent field evidence on elicitation effects. Moreover, this motivates additional research to establish what conditions of the field setting are not captured in past lab investigations. In his seminal paper, Smith (1982, p. 937) states: "If a theory is not falsified in several replications, then one can begin to ask whether the results generalize to different subject pools and to field environments."

The induced-value lab experiments and field studies that explore elicitation effects differ in four important dimensions. One dimension is the use of induced versus homegrown values and, in a similar vein, the use of outcomes defined by money rather than by the provision of actual goods. A second dimension is the subject pool, as the lab experiments rely on college students, while field surveys often employ samples from broader populations.[1] Third, in the field setting, many comparisons do not hold fixed important characteristics of the design (such as framing or the payment vehicle) across the formats investigated; alternatively, field studies may be silent about these design components which gives rise to the possibility that opinions about missing information systematically vary across formats. Last, incentive compatibility in the field setting depends crucially on respondent beliefs (for instance, whether responses are viewed as consequential). These beliefs can conceptually be measured by researchers (e.g., Herriges et al., 2010), but so far, no tool has been developed to measure these beliefs precisely and in an incentive compatible manner.

In this study, we design and implement a controlled experiment that incorporates field-study properties to test for elicitation effects among a range of commonly used formats. In doing so, we begin to bridge the gap between induced-value experiments and field studies. Important for the generalizability of our lab results to the field, the valuation task shares characteristics of many contingent valuation surveys. Specifically, our experiment identifies the effects of having

[1] Studies have been conducted where students in a laboratory or classroom setting have filled out stated preference surveys. These investigations usually demonstrate substantial elicitation effects that parallel findings from surveys conducted in the field (e.g., Balistreri et al., 2001; Welsh and Poe, 1998).

participants value an actual environmental public good. This is a logical next step, as possible behavioral drivers, such as anchoring or heterogeneous interactions between uncertainty and elicitation formats, are unlikely to arise with induced values. Our study involves forestation of agricultural land in a distant location; as a result, the elicitation mainly captures passive-use values. Of course, the prominent conceptual advantage of stated preference surveys over observational (revealed preference) studies is the ability to measure welfare changes associated with passive use. Participants in our study are presumed to be unfamiliar with the good, particularly, with the task of evaluating such a good. Further, mirroring the field, whether individual costs differ across participants is ambiguous to participants.

The four elicitation formats we study include single binary choice, double-bounded binary choice, payment card, and open-ended. All four have been demonstrated to be demand revealing within the context of induced-value experiments.[2] Consistent with the related induced-value experiments, we implement these formats so that truthful preference revelation constitutes the single best response strategy; that is, all formats are incentive compatible. In the case of the double-bounded binary choice, payment card and open-ended formats, this is achieved by introducing uncertainty over the individual cost that must be paid in the event the project is funded. The resulting mechanisms can be viewed as repeated binary choice mechanisms with a random cost selection rule. Azrieli, Chambers and Healy (forthcoming) show that the incentive compatibility of such mechanisms arises under a monotonicity assumption. We hold constant ancillary characteristics such as framing and the payment vehicle to rule out confounds to identifying elicitation effects due to question format.

[2] Taylor et al. (2001), Vossler and McKee (2006), and Collins and Vossler (2009) provide evidence on the single binary choice format. Evidence on the double-bounded format comes from Carson et al. (2009). Evidence on the payment card and open-ended format can be found in Vossler and McKee (2006) and Messer et al. (2010), respectively.

The punchline is that we find no statistical evidence of elicitation effects. Estimates of mean WTP as well as empirical WTP distributions are statistically identical across the four elicitation formats considered. This is strong evidence of convergent validity. One implication of this result is that behavioral factors and biases postulated to give rise to elicitation effects, such as anchoring and complexity, may be of second-order importance. To the extent the random cost selection mechanisms we implemented in the lab can be conceptually paralleled in the field, our results further suggest that doing so is likely to mitigate elicitation effects.

2. Experimental design

2.1 Valuation scenario

The experiment elicits preferences toward funding a tree-planting project. The project involves planting and maintaining 160 trees (about ¼-acre) on agricultural land along the Mississippi River Valley. To achieve this outcome, we collaborated with the organization GreenTrees. The organization currently carries out various tree-planting projects across a long stretch of the Mississippi River Valley. Participants receive information on the broad benefits of tree planting. In addition, based on analyses of other projects undertaken by GreenTrees in the region, we present participants with estimates of increased water storage, avoided nutrient runoff and CO_2 capture that would occur if the specific project considered in our experiment were funded. The experiment instructions, which describe additional details of the project, are included in the Appendix.

The tree-planting project was chosen to enhance the generalizability of the study. As in many stated preference surveys, this project is a non-market, public good. Values for the project should be predominantly tied to passive-use, as the Mississippi River Valley is hundreds of miles away from where study participants currently live. Also typical of valuation scenarios in contingent valuation surveys, it is unlikely that our participants have experience funding tree-planting projects. This characteristic is of particular importance in the context of our research in which, as discussed in the introduction, we hypothesize that possible value cues provided by the elicitation format may influence valuations. Further, although individuals can seek out opportunities to fund tree plantings or other carbon offsets, it is improbable that participants envision another opportunity to fund collectively a project of this scope. On their website, GreenTrees does not provide an explicit way for individuals to support tree plantings, which are largely funded by corporations in exchange for carbon offsets.

2.2 Experimental treatments

In designing this test of elicitation effects, aside from inherent differences in elicitation formats, we aim to hold fixed important characteristics of the elicitations as they may influence valuations. In particular, across all formats: (1) value elicitation questions are framed as referenda and rely on a majority-vote implementation rule; (2) it is ambiguous as to whether the individual cost of funding the project, which is displayed to participants privately, varies across participants (similarly to actual field applications, the individual costs in our experiment do vary); and (3) it is common knowledge that neither the total amount paid to GreenTrees (if funded) nor the size of the project (i.e., the number of trees planted) depends on the amount of money collected from participants. The instructions emphasize that the total cost of the project has been pre-negotiated, and that the project will be subsidized by the experimenters if the sum of individual costs collected upon the referendum passing is less than the total cost of the project.[3] While the exact amount of

[3] In the experiment, the sum of individual costs collected upon the referendum passing is always lower than the pre-negotiated total cost of implementing the project.

the total project cost is not disclosed, participants are explicitly informed about the number of trees planted. The design helps avoid notions of "fair share" pricing, possible collection of excess funds, and speculations that the scope of the project depends on the amount collected.

The experiment includes four split-sample treatments, each involving a separate elicitation mechanism: single binary choice, double-bounded binary choice, payment card, or open-ended. The wording used to explain the elicitation mechanisms is as similar as possible across treatments, only varying as required to convey specifics of the mechanisms. We now describe how each mechanism is operationalized, and provide theoretical justification.

2.2.1 Single binary choice

The single binary choice (SBC) treatment involves a simple up-or-down vote on whether to fund the project at a specific individual cost. The exact wording of the value elicitation question is, "If passage of the referendum cost you $x, are you in favor of funding the tree planting project?" As standard in stated preference surveys employing this format, the individual cost varies across voters to make identification of WTP possible. A binding binary choice referendum with a majority-vote implementation rule is well known to be incentive compatible under weak assumptions (Farquharson, 1969).

2.2.2 Double-bounded binary choice

The double-bounded binary choice (DB) treatment extends the SBC elicitation by presenting participants with two separate referenda that differ only in the stated individual cost. That there are two votes is common knowledge, although the two costs are not known in advance. The second referendum is displayed once all votes in the first referendum are submitted. The wording of each value elicitation question is identical to the SBC treatment. To break possible strategic ties between the two votes, one of the two referenda is randomly selected to be binding after all decisions are made. To understand the incentive properties of this mechanism, we draw from the broader theoretical literature that analyzes incentives in experimental games. In particular, attention has been devoted to games involving a sequence of binary choices, one of which is selected at random to be binding. Cox, Sadiraj and Schmidt (2015) prove that such mechanisms are incentive compatible for all theories that satisfy the reduction of compound lotteries axiom and the independence axiom.[4] Azrieli et al. (forthcoming) instead demonstrate that incentive compatibility arises under a weaker monotonicity assumption, and that the independence axiom is only needed when the reduction axiom is also assumed.[5]

2.2.3 Payment card

In the payment card (PC) treatment, participants are provided with several possible individual costs of funding the project, and the elicitation question is cast as an up-or-down vote to each cost. This reflects some field implementations of the mechanism (for example, see Bateman et al., 2005), although it is somewhat more common to ask respondents to circle, from the provided costs, the highest amount they are willing to pay.[6] Framing the elicitation as a set of (independent) votes to each cost amount enhances transparency of the mechanism, while more closely resembling the other mechanisms we study. Conceptually, this generalizes the DB

[4] As illustrated in their paper, this rules out some theories such as rank dependent utility and cumulative prospect theories.

[5] This theory encompasses not only individual-choice mechanisms but also strategic games where payoffs depend on the joint actions of a group of players. This characterizes voting in our experiment. An important requirement is that each game (vote), if analyzed in isolation, represents an incentive compatible elicitation. Therefore, we need for each vote (cost) that may be selected, that a voter perceives her choice can probabilistically determine the outcome.

[6] Vossler and McKee (2006) test the latter form of a payment card using an induced-value experiment, and conclude that it is not demand revealing.

elicitation to more than two referenda. In implementation, and as relayed to voters, one of the referenda is randomly selected to be binding.[7] As in the DB treatment, the random selection process breaks the strategic link across decisions in the sense that participants cannot influence the cost paid. The wording of the referendum is identical to the prior two treatments. Assumptions for incentive compatibility are identical to those for the DB treatment.

2.2.4 Open-ended

The open-ended (OE) format elicits a point estimate of value. The wording of the referendum is revised to ask: "What is the highest amount that you would pay and still vote in favor of funding the tree planting project?" To parallel the other mechanisms, as well as to provide information to enhance truthful demand revelation, the format is described as a concise way to learn the range of possible individual cost amounts for which a person would vote "yes" or "no." It is further emphasized that, as in the PC case, the binding individual cost is randomly determined after all decisions are submitted. Once the cost is drawn, the stated valuation is compared to the cost. If the valuation is equal to or higher than the cost, this is a "yes" vote in an up-or-down referendum at this cost. Otherwise, this is a "no" vote. Theoretically, our OE mechanism is the Random Price Voting Mechanism (RPVM) developed by Messer et al. (2010).

An important property of the OE mechanism for its incentive compatibility is that neither the range nor probability distribution of possible individual costs is made explicit. As discussed

[7] To parallel the SBC and DB mechanisms, the individual cost of the binding referendum in the PC treatment varies across people, and this is not made explicit in the instructions. In their attempt to implement incentive compatible payment card and open-ended mechanisms in the field, Vossler and Holladay (2016) instead emphasize that the individual cost to all respondents is identical. They include this language to avoid beliefs that respondents might manipulate costs through their valuations. As we can commit to an exogenously determined cost, such a restriction is not needed. Importantly, explicitly stating in the PC and OE treatments that the individual cost is identical for all voting participants may lead to treatment effects relative to the SBC and DB treatments, where the cost naturally varies and this aspect of the elicitation is not disclosed.

by Azrieli et al. (forthcoming), it does not matter theoretically if the experimenter uses a randomization device that is ambiguous to participants. Similar to the Becker-DeGroot-Marschak (1964) mechanism, incentive compatibility of the RPVM does require that a voter perceives there is a positive probability that the realized cost is equal to her valuation.[8]

Messer et al. (2010) demonstrate incentive compatibility of the RPVM under expected utility. However, as the stated OE valuation theoretically maps into a continuum of "yes" or "no" votes to referenda distinguished by the cost, the mechanism can be analyzed in a similar manner to the DB and PC mechanisms; that is, it can be interpreted as a sequence of binary choices, one of which is chosen at random to be binding. Therefore, the RPVM should be incentive compatible without expected utility.

2.3 Design parameters

The bid designs (i.e., procedures for selecting individual costs) used in the SBC, DB and PC elicitations are informed by results from a pre-test and a pilot session (n = 46), both involving OE elicitations. These results yield an observed distribution of WTP reasonably approximated by a normal distribution with a mean of \$3 and a standard deviation of 1.75. Although research on bid designs for SBC and DB elicitations suggests that a small number of bids placed, loosely speaking, sufficiently away from the tails of the distribution is most efficient (Alberini, 1995), such designs assume that the underlying WTP is known. As we anticipated treatment effects, we instead utilize a large set of bids that (roughly) span the 20th to 80th percentiles of the OE WTP distribution observed in the pre-test and pilot. Specifically, for the SBC treatment, we draw bids

[8] To illustrate this, suppose a voter's valuation is \$5 and she believes with certainty that the highest possible individual cost is \$3. In this case, the voter is indifferent between stating any amount greater than or equal to \$3. To help mitigate such beliefs, the instructions explicitly state that the range of possible costs vary from "very low to very high amounts."

randomly from the vector {$1, $2, $3, $4, $5, $6} with equal probabilities. This corresponds with the general rule-of-thumb given in Kanninen (1995) that bids should not be outside the 15th and 85th percentiles of the distribution.

As standard in the literature, to enhance the efficiency of the DB design, the individual cost in the second referendum depends on whether a participant votes "yes" or "no" in the first referendum. Mechanically, if a voter responds "no" to the first cost, the second vote involves a cost randomly drawn from a set of lower cost amounts available (e.g., if a voter says "no" to $3 in the first referendum, a cost of $1 or $2 is drawn with equal probability in the second referendum). In the event of a "yes" vote, a cost is randomly selected from a set of higher cost amounts available. For comparability to the SBC treatment, the first cost is a random draw from the vector {$2, $3, $4, $5} with equal probabilities. To accommodate "yes" responses to the highest cost or "no" responses to the lowest cost, the second cost vector is identical to that used for the SBC treatment.

Research on PC bid design (Rowe, Schulze and Breffle, 1996) emphasizes the importance of including a range of bids that sufficiently spans the underlying WTP distribution. The pre-test and the pilot session yield a range of $0 to $10, and we set the lowest and highest values of the PC to match these amounts. To allow for comparisons with the SBC and DB formats, the PC includes all integers between $0 and $10 (inclusive).

The bid distributions, and sample sizes, are informed by power calculations based on Monte Carlo simulations. The simulations suggest that for a wide range of possible treatment effects (i.e., mean differences between treatments), the PC design yields virtually the same power as an OE elicitation. This is not entirely surprising, as the PC elicits a reasonably tight WTP interval. Mirroring the literature, the simulations confirm that smaller, rather than larger, bid designs are more efficient for a SBC elicitation. Nevertheless, these designs lead to considerable power loss when the assumed mean of the distribution is reasonably far away from the true mean. Our bid design is a compromise that, based on the simulations, performs well "on average." The larger bid distribution we employ has a small but acceptable loss in power relative to the efficient design when the true distribution is known, but performs considerably better across a wide range of possible effect sizes. Using bids from the SBC design, the simulations suggest that the power of the DB design is nearly as good as that of the PC and OE elicitations.

To determine sample sizes, we set as an objective the ability to identify an effect size of 70 cents with 80% probability. The simulations assume the data are analyzed by an interval regression model (which we in fact use), and that the null hypothesis of equal means is evaluated with a t-test based on a 5% significance level.[9] This requires sample sizes of roughly 100 for each of the OE, PC and DB treatments, and a sample of 130 for the SBC treatment. We note that these sample sizes are rather large by laboratory experiment standards. Nevertheless, when controlled experiments rely on induced values, this reduces the variation in choices considerably.

2.4 Experimental procedures

A typical session proceeds as follows. Participants are randomly assigned a seat in the lab. The same experiment moderator summarizes lab protocols (e.g., no deception), and reads the experiment instructions aloud while participants follow along on their printed copy. To ensure anonymity, participants are informed that each is randomly assigned an ID number and that decisions are anonymous. Questions are encouraged by the moderator. The experiment consists of two stages: an earnings stage and a voting stage. All decisions are made on the computer. The experiment is programmed and conducted using the software z-Tree (Fischbacher, 2007).

[9] An effect size of roughly 60 cents can be detected with 80% probability based on a 10% significance level.

In the earnings stage, participants earn money by scoring points in two tasks. Both tasks are designed to be "real effort" tasks in the sense that exerting more effort leads to higher earnings, and performance in these tasks is not heavily dependent on a cognitive ability or experience. The earnings tasks provide participants with an amount of money more than sufficient to cover the potential cost to them of funding the tree-planting project. Moreover, earnings tasks help to enhance external validity by avoiding possible "house money" effects. To determine earnings, scores from the tasks are summed, and all participants are rank-ordered according to their scores. Based on her place in the score distribution, a participant earns either $25 (top 20%), $22.50 (top 21-40%), $20 (top 41-60%), $17.50 (top 61-80%), or $15 (bottom 20%). This procedure induces competition among participants and holds fixed the earnings distribution across sessions.

The first task, developed by Abeler et al. (2011), involves counting the number of zeros in tables that contain randomly generated zeros and ones. Participants have three attempts to enter the correct number of zeros for a given table, after which a new table is generated. For each correctly counted table, a participant earns four points. Five minutes are given for this task. In the second task, developed by Erkal, Gangadharan and Nikiforakis (2011), participants are provided with a table that assigns a number value to each letter of the alphabet, and are asked to encode words into numbers. Participants cannot move to a next word until the currently displayed word is encoded correctly. Similar to the first earnings task, participants have five minutes to encode as many words as possible. For each correctly encoded word, a participant earns one point.[10]

In the next stage of the experiment, that is, in the voting stage, participants are provided information on the tree-planting project. Instructions are identical across treatments, including not

[10] The conversion of correctly counted tables and correctly encoded words into points is based on evidence from previous experiments (Abeler et al., 2011; Erkal et al., 2011) how much time was needed to solve each of the tasks. The actual mean number of points scored across the two tasks in our experiment is virtually identical.

only the description of the good, but also the procedure for verifying that money collected will in fact be used to fund the project. The only variation across treatments is the explanation of the elicitation mechanism ("the voting process" in the instructions). After all decisions are collected, DB, PC and OE participants learn their random individual cost draw. In all treatments, the results screen displays a participant's vote in the binding referendum, the percentage of "yes" and "no" votes in the group, and whether the referendum passes. If the referendum passes, the individual cost of funding the project is subtracted from earnings obtained in the first stage of the experiment. A volunteer is asked to place in the mail a sealed envelope containing a check to GreenTrees along with a letter describing the purpose of the check and that students from the University of Tennessee are funding the project. Upon receipt of payment, participants are emailed a letter from GreenTrees to verify the transaction.

The experiment continues with a short questionnaire that probes participants about their voting decision(s) and collects basic socio-demographic information. The experiment concludes by paying participants their cash earnings. This is done privately for each participant.

2.5 Participants

Four hundred and ten students of the University of Tennessee took part in the experiment from December 2017 to February 2018. All sessions took place at the UT Experimental Economics Laboratory. Participants were selected from a large pool of students registered as potential participants in economic experiments. The pool resembles the general population of students of the University with respect to gender, age and academic major. Thirty-eight percent of individuals had previously participated in an (unrelated) economics experiment. Participants were not allowed to attend more than one session of the experiment.

Of the 18 sessions conducted, there are six sessions for the SBC treatment and four sessions for the DB, PC and OE treatments each. The number of participants per session ranged from 16 to 24. A single session lasted about 40 minutes. Average earnings were $19.79, and the referendum passed in seven of the 18 sessions.

3. Data Analysis

Table 1 summarizes, by treatment, data obtained from the post-experiment questionnaire. It also includes pre-vote earnings from the first stage of the experiment. The average participant age is 21 years, 43% of the participants are female, 49% are currently employed, and the average reported GPA is 3.27 on a 4-point scale. As anticipated due to the random treatment assignment, the distributions of these socio-demographic variables are similar and not statistically significantly different across treatments.

The table further reports summary statistics from a set of three questions included in the questionnaire to gauge how well participants understood experimental procedures. The vast majority of participants (87%) stated that the experiment instructions were overall "well understood." Further, 86% and 75% of participants, respectively, indicated disagreement with the statements "I was confused about the procedure used to determine whether the referendum passed" and "I did not have enough information to make a comfortable decision in the referendum." Using Pearson's chi-squared test of equality of means, we find some statistically significant differences in the response distributions of these questions, which point to some differences in the level of understanding of the experimental procedures across the treatments. Specifically, when compared to the SBC treatment, the PC and OE treatments are characterized by statistically lower comprehension of the instructions (at the 5% significance level for each comparison) and by statistically higher confusion about the voting procedure (at the 1% significance level for the comparison of SBC and PC and at the 10% significance level for the comparison of SBC and OE). Observing these differences to be significant aligns with expectations given the higher complexity of the voting procedure in the PC and OE treatments relative to the SBC treatment. Further, relative to DB participants, the PC participants appear to have been in need for more information to make a comfortable decision in the referendum (a difference significant at the 10% level). Although statistical differences arise, the magnitudes of the differences highlighted above are modest.

We also asked about certainty over a participant's decision in the referendum, as common in stated preference surveys. The mean response of 4.01, on a scale from 1 ("very uncertain") to 5 ("absolutely certain"), suggests the average participant was "certain"; the modal response was 5. The average certainty level is lowest in the OE sample at 3.78, and highest in the DB sample at 4.14. Using Pearson's chi-squared test, the OE sample distribution of responses to the certainty question is statistically different at the 5% level when compared with the SBC or DB sample. This result is somewhat expected, as the OE elicitation requires participants to provide a point estimate of WTP rather than a "yes" or "no" response to explicit cost levels. No other statistically significant differences are found for other pairwise comparisons of treatments.

3.1 Nonparametric tests of WTP distributions

Table 2 presents the observed WTP distributions (survival functions) for each treatment. The SBC and PC survival functions are the observed percentages of "yes" votes recorded at each cost amount. The OE survival function is constructed by calculating the percentage of participants indicating a WTP at least as high as a specific cost amount. The survival function for the DB data is computed in a similar fashion.[11] As expected by construction, the survival functions are monotonically decreasing.

The survival functions look reasonably similar based on visual inspection. To nonparametrically test for differences in empirical distributions, we use two-sample Kolmogorov-Smirnov (K-S) tests. The test statistic is the absolute value of the largest difference in the observed probabilities across two distributions. A rejection of the null hypothesis can result from differences in the shapes or locations of the distributions. The largest observed difference in probabilities across all pairwise comparisons is 0.1286, which occurs at $1 when comparing the DB and PC treatments. The K-S statistics for other comparisons are: 0.0815 (SBC vs. DB); 0.0851 (SBC vs. PC); 0.0798 (SBC vs. OE); 0.0811 (DB vs. OE); and 0.0957 (PC vs. OE). We fail to reject equal distributions in any pairwise comparison.[12] In order to estimate mean WTP values, and to further condition them on participants' socio-demographic characteristics and on the money earned from the pre-vote tasks, we now turn to an econometric analysis.

3.2 Econometric Analysis

The data generated from the four types of value elicitation mechanisms gives rise to continuous, left-censored, right-censored or interval-censored signals of participants' WTP. The maximum likelihood estimator we specify, which nests the estimators of Cameron and James (1987) and Cameron and Huppert (1989), accommodates the different types of information obtained across elicitation formats in a unified way, enabling tests of treatment effects that are not

[11] For the DB survival function, when calculating the percentage of "yes" votes at $x, we omit observations for which $x falls within the elicited WTP interval. For example, if a participant votes "yes" to $2 and "no" to $5, this is interpreted as a "yes" vote to $1 and $2, missing values for $3 and $4, and a "no" vote to $5 and $6.

[12] Critical values for the K-S test depend on sample sizes. For these comparisons, the 5% critical values are in the 0.18 to 0.20 range, and 10% critical values span from 0.16 to 0.18.

driven by differences in statistical assumptions. The estimation allows for the possibility that some participants have negative WTP, as suggested by the finding that 17% voted "no" at a cost of \$0 in the PC treatment. Further, 21% of the OE sample indicated \$0 WTP, which may also signal negative WTP (noting that negative valuations were not permissible).

Let WTP_i denote participant i's willingness to pay for the project. WTP_i is not directly observed, except for OE participants, but instead can be treated as a censored dependent variable. For the PC elicitation, we obtain the signal $c_{i,l} \leq WTP_i < c_{i,u}$, where $c_{i,l}$ is the highest cost for which participant i votes "yes" and $c_{i,u}$ is the next higher amount. For the case where the participants votes "no" to the lowest amount, $c_{i,l} = -\infty$ and $c_{i,u}$ is equal to the lowest amount; similarly, $c_{i,l}$ is equal to the highest amount and $c_{i,u} = +\infty$ if she votes "yes" to the highest amount. For the SBC elicitation, we obtain the signal $WTP_i < c_i$ if the participant votes "no" to the stated cost c_i, and the signal $WTP_i \geq c_i$ for a "yes" vote. As such, SBC data represents a special case of PC data, where c_i defines only an upper or lower bound on the WTP interval (e.g., $c_{i,l} = c_i$ and $c_{i,u} = +\infty$ for a "yes" vote). For DB responses, a well-defined interval emerges in cases where a "yes" vote is observed for just one of the two votes. Otherwise, for two "no" votes, the data is left-censored (that is, $c_{i,l} = -\infty$) and the lower amount offered forms the upper bound $c_{i,u}$. For two "yes" votes, the higher of the two amounts offered forms the lower bound $c_{i,l}$ and the data is right-censored (that is, $c_{i,u} = +\infty$). Finally, for OE responses, WTP_i is directly observed, with the exception of zero valuations, which we interpret to be left-censored to allow for possibly negative WTP, which is consistent with our treatment of the other data types. These left-censored observations are accommodated by defining a WTP interval with $c_{i,l} = -\infty$ and $c_{i,u} = 0$.

Assume WTP_i is a linear function of a row vector of covariates, $\mathbf{x}_i$, such that $WTP_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i$, where $\boldsymbol{\beta}$ is a column vector of unknown parameters and ε_i is a normally distributed mean-zero error term with a standard deviation σ_i. With the linear conditional mean function, assuming the error term has a normal distribution is analogous to assuming a normal distribution for WTP_i. Moreover, interpretation of estimated parameters is the same as for a standard linear regression model that treats WTP_i as a directly observed (i.e., uncensored) dependent variable. Let $D_i = 1$ if the response is censored (that is, SBC, DB, PC and zero OE responses), and $D_i = 0$ if the response is uncensored (that is, OE responses larger than zero). Then, the log-likelihood function for the WTP regression is

$$\ln \mathcal{L} = \sum_{i=1}^{N} \left\{ D_i \cdot \ln\left(\Phi\left(\frac{c_{i,u} - \mathbf{x}_i\boldsymbol{\beta}}{\sigma_i}\right) - \Phi\left(\frac{c_{i,l} - \mathbf{x}_i\boldsymbol{\beta}}{\sigma_i}\right) \right) + (1 - D_i) \cdot \ln\left(\frac{1}{\sigma_i} \phi\left(\frac{WTP_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma_i}\right) \right) \right\},$$

where Φ and ϕ denote the CDF and PDF of the standard normal distribution, respectively. The first term corresponds with the log-likelihood for (interval) censored data, whereas the second term corresponds to that of a normal regression model for uncensored data.[13] The OE data are a mix of censored and uncensored data, and applying the above estimator is equivalent to a using Tobit with left-censoring at zero.

Studies such as Haab, Huang and Whitehead (1999) highlight the importance of allowing for different error variances when pooling preference data from different experimental treatments. To allow for possibly different error variances across elicitation formats, we define $\sigma_i = \sigma_0 + \sigma_1 DB_i + \sigma_2 PC_i + \sigma_3 OE$, where OE_i is an indicator that equals 1 for OE observations and DB_i and PC_i are similar indicator variables for DB and PC observations, respectively. In this formulation, σ_0 is the standard deviation of the error for the SBC data.

[13] Given how we code the upper and lower bounds, the contribution to the log-likelihood for a SBC participant is mathematically equivalent to that of a probit model.

Table 3 reports the results of the WTP regressions. Model I allows the mean WTP to vary across treatments, but constrains the error variances to be equal. Model II extends the specification to allow for unequal variances of WTP across treatments. Model III additionally includes variables that controlling for participants' socio-demographic characteristics and the amount of earned income in the pre-vote experiment tasks. These variables may explain some of the variation in WTP across participants, while also adjusting estimates for unintended differences due to sampling.

For Models I and II, there are no statistically significant differences in either the means or standard deviations across treatments, which corroborates the findings from the K-S tests. For Model I, mean WTP estimates range from \$3.69 (OE) to \$3.94 (SBC). For Model II, these estimates vary from \$3.66 (OE) to \$3.84 (SBC and DB). For Model III, the range of mean WTP estimates across treatments remains narrow, ranging from \$3.53 (OE) to \$3.96 (PC), and there are no statistical differences. This is overall very strong evidence of convergent validity. There is a marginal statistical difference in error variances between the SBC and OE treatments (*p*-value = 0.09). However, as this difference does not arise in Model II, this may simply imply that less heterogeneity in elicited OE preferences is explained by the control variables. Model III also shows that older participants and females have a higher WTP for the tree-planting project.

Evidence from field surveys suggests that the DB format may invoke behavioral responses or updating that leads to differences in the mean or variance of WTP across the value elicitation questions. To explore this issue, following Cameron and Quiggin (1994), we estimate a bivariate normal model for interval-censored data, allowing for differing means and variances across the two DB valuation questions. This results in a mean WTP of \$4.03 (std. err. = 0.32) for the first question and \$3.53 (0.75) for the second. The estimated standard deviations are 2.20 (0.60) and

5.88 (2.47), respectively, and the correlation coefficient is 0.90 (0.10). Neither the means (*p*-value = 0.40), nor the standard deviations (*p*-value = 0.18) are statistically different.

4. Discussion

This research is motivated by contradictory evidence: while myriad field studies document meaningful differences in welfare estimates based on the value elicitation format used, no such effects are observed for a subset of formats investigated in controlled, induced-value experiments. With this puzzle in mind, we take the next step by bringing field context to the experimental laboratory. In particular, we elicit homegrown values for an environmental public good. Reflecting common field applications of contingent valuation surveys, the good is most likely strongly tied to passive-use values and unfamiliar to participants. We ask participants whether they want to fund the good using earned, rather than endowed, money. Our results corroborate those from induced-value tests, as we find that the four elicitation formats examined lead to statistically identical willingness-to-pay (WTP) distributions.

The literature argues that, in a setting where people are presented with an unfamiliar task of valuing non-market goods, valuations are likely to be influenced by various behavioral factors related to the stated cost(s) (e.g., "anchoring" on the bid(s) in a close-ended elicitation) and decision complexity, among others. The elicitation formats we study vary considerably, making it possible for behavioral factors to play a role in the value elicitation process, and trigger elicitation effects. In fact, this was what we expected to find. While our results do not definitively rule out the importance of behavioral factors, our findings imply that behavioral factors may be of second-order importance. Differences in incentive properties across formats, as implemented in field settings, may instead be the primary driver of elicitation effects.

There are several possible next steps in the line of research we have undertaken, which involve identifying differences between our lab experiment and the field survey setting and systematically modifying either the design of the experiment or the information provided in field surveys. The most important disconnect, of course, is that we use incentive-compatible elicitation mechanisms with direct financial consequences. Prior studies provide ways in which our experimental design can be modified to reflect the field. For instance, Carson, Groves and List (2014) vary the probability that a vote is binding. Vossler et al. (2012) frame the elicitation as a vote but keep the decision rule undisclosed. In a future test of the double-bounded format, for example, one possibility is to have participants vote "yes" or "no" to two possible costs, but provide no information on how this information will map into a binding outcome.

Accumulated evidence suggests that contingent valuation surveys match well the outcomes of binding, public referenda in settings where single binary choice elicitations are believed to adhere to incentive compatibility assumptions (e.g., Johnston, 2006). It follows logically that if researchers uncover design modifications that mitigate elicitation effects, this in turn should enhance the criterion validity of other formats. This is to say that results from the lab can be used to inform field studies. Aspects of our elicitation mechanisms, such as framing all formats as referendum votes, and using cost uncertainty to break strategic links, can potentially translate to field environments. Vossler and Holladay (2016) implement a field survey that incorporates these features, and demonstrate convergent validity between single binary choice and open-ended formats. Nevertheless, they continue to find differences between payment card and single binary choice formats. Aside, of course, from the stated preference aspect of the field elicitation, our experiment differs from their implementation in two respects. First, we do not explicitly mention (which Vossler and Holladay, 2016, do) that the proposed cost will be the same everyone for the payment card treatment upon the project realization. Second, in our experiment only those costs presented on the payment card are possible to be binding, whereas their field survey remains silent about possible cost levels to be actually introduced.

As standard in laboratory investigations, we utilize college students as participants. The accumulated evidence from broader experimental economics research suggests that important differences between students and more targeted subject pools are the exception rather than the rule (Fréchette, 2016). Specific to the study of elicitation effects, prior studies, such as Welsh and Poe (1998), effectively administers a stated preference survey in the experimental laboratory (or classroom) with college students. As these studies provide similar results to stated preference studies conducted in the field, we suspect that the subject pool effects might not be of first-order importance. Even so, if the complexity of the decision environment is in fact an important driver, broadening the participant pool could lead to insight. College students are presumably more acclimated to deal with new and complex decision environments.

On a final note, our examination is limited to four elicitation formats, whereas many others are used in practice (see, for example, Carson and Louviere, 2011). The variation in the formats we investigate coincides well with key characteristics of extant formats: the precision at which values are elicited and the number of value questions used. Our research involves both open-ended and close-ended formats, considers both single and repeated choice valuation formats, and includes repeated formats that span sequential (i.e., double-bounded) and simultaneous (i.e., payment card) decision settings. Nevertheless, there is merit in using controlled experiments with field context to study other important formats, especially those that are proven to reveal demand in an induced-value setting.

References

Abeler, Johannes, Armin Falk, Lorenz Goette and David Human. 2011. Reference points and effort provision. *American Economic Review* 101(2), 470-492.

Alberini, Anna. 1995. Testing willingness-to-pay models of discrete choice contingent valuation survey data. *Land Economics* 71(1), 83-95.

Azrieli, Yaron, Christopher P. Chambers and Paul J. Healy. Forthcoming. Incentives in experiments: A theoretical analysis. *Journal of Political Economy*.

Balistreri, Edward, Georg McClelland, Gregory L. Poe and William D. Schulze. 2001. Can hypothetical questions reveal true values? A laboratory comparison of dichotomous choice and open-ended contingent values with auction values. *Environmental and Resource Economics* 18(3), 275-292.

Bateman, Ian J., Philip Cooper, Stavros Georgiou, Ståle Navrud, Gregory L. Poe, Richard C. Ready, Pere Reira, Mandy Ryan and Christian A. Vossler. 2005. Economic valuation of policies for reducing acidity in remote mountain lakes. *Aquatic Sciences* 67(3), 274-291.

Bateman, Ian J., Ian H. Langford and Jon Rasbash. 2001. Willingness-to-pay question format effects in contingent valuation studies. In Ian J. Bateman and Kenneth G. Willis (eds.), *Valuing environmental preferences: Theory and practice of the contingent valuation method in the US, EU, and developing countries*. Oxford: Oxford University Press.

Becker, Gordon M., Morris H. DeGroot and Jacob Marschak. 1964. Measuring utility by a single-response sequential method. *Behavioral Science* 9(3), 226-232.

Bishop, Richard C., Kevin J. Boyle, Richard T. Carson, David Chapman, W. Michael Hanemann, Barbara Kanninen, Raymond J. Kopp, Jon Krosnick, John List, Norman Meade, Robert Paterson, Stanley Presser, V. Kerry Smith, Roger Tourangeau, Michael Welsh, Jeffrey M. Wooldridge, Matthew De Bell, Colleen Donovan, Matthew Konopka and Nora Scherer. 2017. Contingent valuation: Flawed logic? – Response. *Science* 357(6349), 363-364.

Cameron, Trudy A. and Daniel D. Huppert. 1989. OLS versus ML estimation of non-market resource values with payment card interval data. *Journal of Environmental Economics and Management* 17(3), 230-246.

Cameron, Trudy A. and Michael D. James. 1987. Efficient estimation methods for use with "closed-ended" contingent valuation survey data. *The Review of Economics and Statistics* 69(2), 269-276.

Cameron, Trudy. A., Gregory L. Poe, Robert G. Ethier and William D. Schulze. 2002. Alternative non-market value-elicitation methods: Are the underlying preferences the same? *Journal of Environmental Economics and Management* 44(3), 391-425.

Cameron, Trudy A. and John Quiggin. 1994. Estimation using contingent valuation data from a "dichotomous choice with follow-up" questionnaire. *Journal of Environmental Economics and Management* 27(3), 218-234.

Carson, Katherine Silz, Susan M. Chilton and W. George Hutchinson. 2009. Necessary conditions for demand revelation in double referenda. *Journal of Environmental Economics and Management* 57(2), 219-255.

Carson, Richard T. and Theodore Groves. 2007. Incentive and informational properties of preference questions. *Environmental and Resource Economics* 37(1), 181-210.

Carson, Richard T., Theodore Groves and John A. List. 2014. Consequentiality: A theoretical and experimental exploration of a single binary choice. *Journal of the Association of Environmental and Resource Economists* 1(1-2), 171-207.

Carson, Richard T. and Jordan J. Louviere. 2011. A common nomenclature for stated preference elicitation approaches. *Environmental and Resource Economics* 49(4), 539-559.

Champ, Patricia A. and Richard C. Bishop. 2006. Is willingness to pay for a public good sensitive to the elicitation format? *Land Economics* 82(2), 162-173.

Collins, Jill P. and Christian A. Vossler. 2009. Incentive compatibility tests of choice experiment value elicitation questions. *Journal of Environmental Economics and Management* 58(2), 226-235.

Cox, James C., Vjollca Sadiraj and Ulrich Schmidt. 2015. Paradoxes and mechanisms for choice under risk. *Experimental Economics* 18(2), 215-250.

Erkal, Nisvan, Lata Gangadharan and Nikos Nikiforakis. 2011. Relative earnings and giving in a real-effort experiment. *American Economic Review* 101(7), 3330-3348.

Farquharson, Robin. 1969. *Theory of Voting*. New Haven, CT: Yale University Press.

Fischbacher, Urs. 2007. Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2), 171-178.

Fréchette, Guillaume R. 2016. Experimental economics across subject populations. In John H. Kagel and Alvin E. Roth (eds.), *The Handbook of Experimental Economics, vol. 2*. Princeton, NJ: Princeton University Press.

Frew, Emma J., David K. Whynes and Jane L. Wolstenholme. 2003. Eliciting willingness to pay: Comparing closed-ended with open-ended and payment scale formats. *Medical Decision Making* 23, 150-159.

Green, Donald, Karen E. Jacowitz, Daniel Kahneman and Daniel McFadden. 1998. Referendum contingent valuation, anchoring, and willingness to pay for public goods. *Resource and Energy Economics* 20(2), 85-116.

Haab, Timothy C., Ju-Chin Huang and John C. Whitehead. 1999. Are hypothetical referenda incentive compatible? A comment. *Journal of Political Economy* 107(1), 186-196.

Hanemann, W. Michael. 1995. Contingent Valuation and Economics. In K.G. Willis and J.T. Corkindale (eds.), *Environmental Valuation New Perspectives*. Oxon: CAB International.

Herriges, Joseph, Catherine Kling, Chih-Chen Liu and Justin Tobias. 2010. What are the consequences of consequentiality? *Journal of Environmental Economics and Management* 59(1), 67-81.

Johnston, Robert J. 2006. Is hypothetical bias universal? Validating contingent valuation responses using a binding public referendum. *Journal of Environmental Economics and Management* 52(1), 469-481.

Johnston, Robert J., Kevin J. Boyle, Wiktor L. Adamowicz, Jeff Bennett, Roy Brouwer, Trudy A. Cameron, W. Michael Hanemann, Nick Hanley, Mandy Ryan, Riccardo Scarpa, Roger Tourangeau and Christian A. Vossler. 2017. Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists* 4(2), 319-405.

Kanninen, Barbara J. 1995. Bias in discrete response contingent valuation. *Journal of Environmental Economics and Management* 28(1), 114-125

Kling, Catherine L., Daniel J. Phaneuf and Jinhua Zhao. 2012. From Exxon to BP: Has some number become better than no number? *Journal of Economic Perspectives* 26(4), 3-26.

Messer, Kent D., Gregory L. Poe, Daniel Rondeau, William D. Schulze and Christian A. Vossler. 2010. Social preferences and voting: An exploration using a novel preference revealing mechanism. *Journal of Public Economics* 94(3-4), 308-317.

O'Conor, Richard M., Magnus Johannesson and Per-Olov Johansson. 1999. Stated preferences, real behaviour and anchoring: Some empirical evidence. *Environmental and Resource Economics* 13(2), 235-248.

Ready, Richard C., Ståle Navrud and W. Richard Dubourg. 2001. How do respondents with uncertain willingness to pay answer contingent valuation questions? Land Economics 77(3), 315-326.

Rowe, Robert D., William D. Schulze and William S. Breffle. 1996. A test for payment card biases. *Journal of Environmental Economics and Management* 31(2), 178-185.

Scheufele, Gabriela and Jeff Bennett. 2013. Effects of alternative elicitation formats in discrete choice experiments. *Australian Journal of Agricultural and Resource Economics* 57(2), 214-233.

Smith, Vernon L. 1982. Microeconomic systems as an experimental science. *American Economic Review* 72(5), 923-955.

Taylor, Laura O., Michael McKee, Susan K. Laury and Ronald G. Cummings. 2001. Induced-value tests of the referendum voting mechanism. *Economics Letters* 71(1), 61-65.

Vossler, Christian A., Maurice Doyon and Daniel Rondeau. 2012. Truth in consequences: Theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics* 4(4), 145-171.

Vossler, Christian A. and J. Scott Holladay. 2016. Alternative value elicitation formats in contingent valuation: A new hope. Working Paper #2016-02, Department of Economics, University of Tennessee.

Vossler, Christian A. and Michael McKee. 2006. Induced-value tests of contingent valuation elicitation mechanisms. *Environmental and Resource Economics* 35(2), 137-168.

Welsh, Michael P. and Gregory L. Poe. 1998. Elicitation effects in contingent valuation: Comparisons to a multiple bounded discrete choice approach. *Journal of Environmental Economics and Management* 36(2), 170-185.

Whitehead, John C. 2002. Incentive incompatibility and starting-point bias in iterative valuation questions. *Land Economics* 78(2), 285-297.

Table 1. Summary Statistics by Treatment

	Single Binary Choice (SBC)	Double-Bounded Binary Choice (DB)	Payment Card (PC)	Open-Ended (OE)
Age	20.65	20.80	20.53	20.79
	(3.31)	(2.79)	(2.29)	(1.51)
Female	0.45	0.41	0.37	0.48
	(0.50)	(0.50)	(0.49)	(0.50)
Earned Income	19.77	19.84	19.79	19.79
	(3.54)	(3.49)	(3.49)	(3.49)
Employed	0.46	0.58	0.47	0.48
	(0.50)	(0.50)	(0.50)	(0.50)
GPA	3.19	3.34	3.22	3.36
	(0.57)	(0.50)	(0.50)	(0.43)
Comprehension	4.90	4.88	4.79	4.79
	(0.30)	(0.36)	(0.55)	(0.48)
Confusion	1.35	1.51	1.70	1.54
	(0.73)	(0.88)	(1.13)	(0.99)
Need Information	1.89	1.77	2.10	1.96
	(1.12)	(0.89)	(1.31)	(1.15)
Certainty	4.08	4.14	4.02	3.78
	(0.99)	(0.91)	(1.14)	(1.13)
N	130	92	94	94

Notes: Standard deviations are reported in parentheses. *Earned Income* corresponds with experiment earnings obtained prior to the voting task. GPA was assessed on a scale from 1 to 4. *Comprehension* is a 1 (“poorly understood”) to 5 (“well understood”) indication of instruction clarity. *Confusion* is a 1 (“completely disagree”) to 5 (“completely agree”) indication of whether participant was confused about the voting process. *Need Information* is a 1 (“completely disagree”) to 5 (“completely agree”) indication of whether participant had enough information to make an comfortable decision in the referendum. *Certainty* is a 1 (“very uncertain”) to 5 (“absolutely certain”) indication of a participant’s certainty about her voting decision(s).

Table 2. Empirical Survival Function – Percentage of "Yes" Votes by Cost

Cost	Single Binary Choice (SBC)	Double-Bounded Binary Choice (DB)	Payment Card (PC)	Open-Ended (OE)
$0			82.98	
$1	79.17	87.32	74.47	84.04
$2	72.73	75.00	67.02	71.28
$3	61.90	56.58	56.38	59.58
$4	50.00	50.67	41.49	42.55
$5	33.33	31.94	36.17	35.11
$6	25.00	20.55	20.21	17.02
$7			17.02	13.83
$8			12.77	9.58
$9			12.77	8.51
$10			12.77	8.51

Table 3. Willingness-to-Pay Regressions

	I	II	III
Double-Bounded Binary Choice (DB)	-0.10	0.00	-0.09
	(0.68)	(0.62)	(0.62)
Payment Card (PC)	-0.13	-0.03	0.07
	(0.65)	(0.56)	(0.55)
Open-Ended (OE)	-0.25	-0.18	-0.36
	(0.65)	(0.61)	(0.62)
Age			0.25^{***}
			(0.09)
Female			1.05^{**}
			(0.44)
Earned Income			-0.07
			(0.06)
Employed			0.15
			(0.44)
GPA			0.50
			(0.42)
Constant	3.94^{***}	3.84^{***}	3.89^{***}
	(0.48)	(0.38)	(0.39)
Standard deviation function (σ)			
Double-Bounded Binary Choice (DB)		0.89	0.81
		(0.99)	(0.96)
Payment Card (PC)		0.65	0.47
		(0.81)	(0.78)
Open-Ended (OE)		1.24	1.36^{*}
		(0.81)	(0.79)
Constant	4.15^{***}	3.23^{***}	3.19^{***}
	(0.23)	(0.73)	(0.71)
Log-*L*	-669.13	-667.92	-659.55
N	410	410	410

Notes: All socio-demographic variables are demeaned so that the Constant can be interpreted as the estimated mean WTP for the SBC treatment in all models. *** implies significance at the 1% level, ** significance at the 5% level and * significance at the 10% level.

Appendix

Experiment Instructions

You are about to participate in an experiment in economic decision making. Please follow the instructions carefully. At any time, please feel free to raise your hand if you have a question. At the end of today's session, you will be paid your earnings privately and in cash.

You have been randomly assigned an ID number for this experiment. You will never be asked to reveal your identity to anyone. Your name will never be associated with any of your decisions. In order to keep your decisions private, please do not reveal your choices or otherwise communicate with any other participant. Importantly, please refrain from verbally reacting to events that occur during the experiment.

Today's session consists of three parts: Experiment 1, Experiment 2 and a short questionnaire. In Experiment 1, you will have the opportunity to earn money. In Experiment 2, you will be asked whether you are willing to use some of your earnings from Experiment 1 to support an actual tree planting project.

Instructions for Experiment 1

In this experiment, you earn points based on your performance in two tasks. The tasks are:

(1) Counting zeros; and
(2) Encoding words.

You have a limited time to work on each task. A timer will be shown in the upper right-hand corner of your computer screen.

After the two tasks are completed, the computer will rank-order all players in the room according to their total number of points scored in both tasks. When doing so, the computer will randomly break any ties. Your cash earnings for Experiment 1 will depend on your rank as follows:

If your rank is among the …	**You earn …**
top 20% of all players	$25
top 21% to 40% of all players	$22.50
top 41% to 60% of all players	$20
top 61% to 80% of all players	$17.50
bottom 20% of all players	$15

Counting zeros

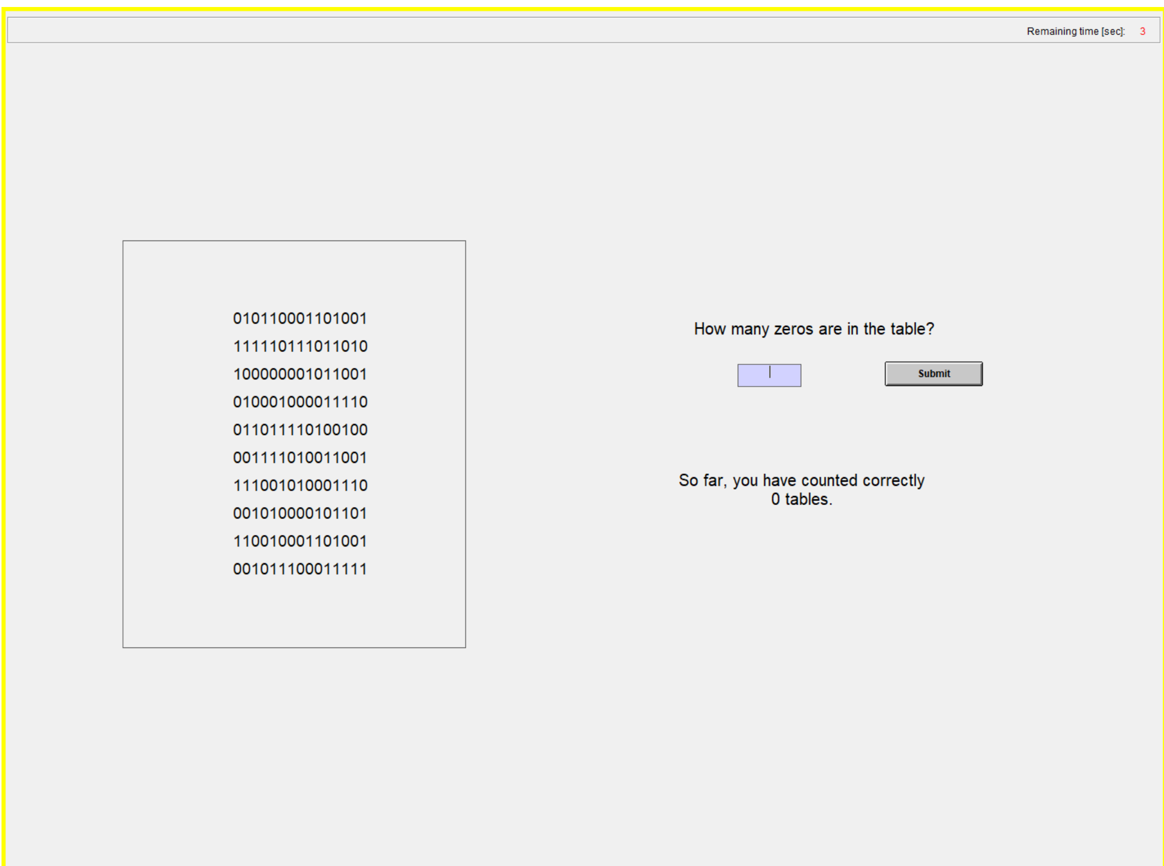
In this task, you are shown tables containing zeros (0s) and ones (1s), as illustrated below. Your job is to count correctly the number of zeros in the table. In the example below, there are 75 zeros.

You enter the number of zeros into the box on the right-hand side of the screen and click the "Submit" button. If your entry is correct, a new table will appear. If your entry is incorrect, you have two additional tries to enter the correct number into the box. After three incorrect attempts, a new table will appear. When a new table appears, your prior entry may remain in the answer box. If this occurs, please erase the number, as it is unlikely to be correct for the new table.

Your objective is to count correctly the number of zeros in as many tables as possible within the allotted time. For <u>each</u> table you count correctly, you earn <u>4 points</u>.

You will have 5 minutes in total. There will <u>not</u> be any practice before we begin – when we proceed, you will see a screen similar to the one below. Once the task begins, the timer will immediately start counting down.

Are there any questions?



Encoding words

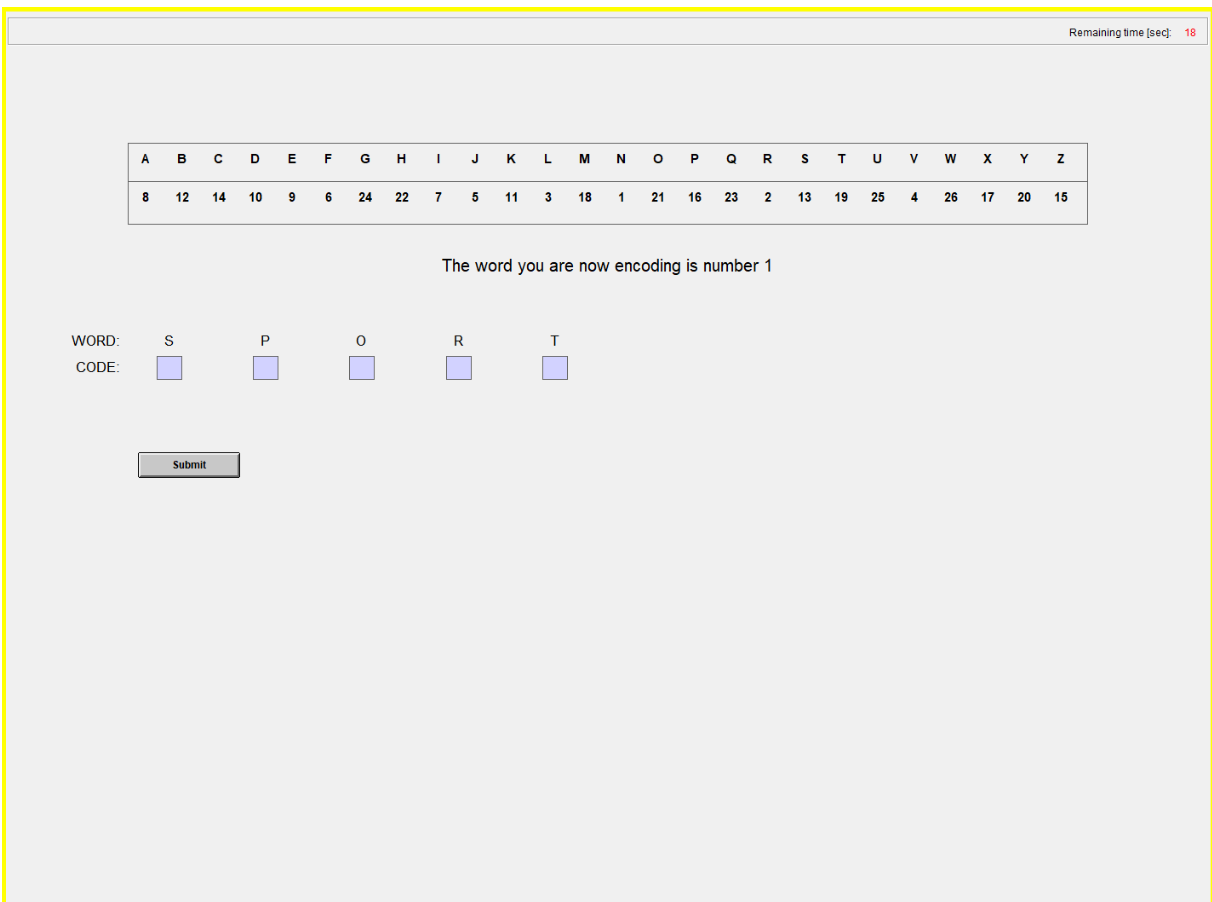
In this task, you are shown a table that assigns a number value to each letter of the alphabet, as illustrated below. Your job is to encode the word below the table by assigning to each letter the number value provided in the table. In the example below, the word is S-P-O-R-T. From the table, the correct numbers for this word are 13-16-21-2-19.

You enter the number associated with each letter in the box below the letter. To type a number into a box, you first move your mouse pointer to the box, and then use a left mouse click. You can also use the tab key to move to different boxes, although know that this will move the cursor from left to right. After encoding a word, click the "Submit" button. If your entry is correct, a new word for encoding will appear. You cannot move to a new word until your entry is correct.

Your objective is to encode as many words as possible within the allotted time. For <u>each</u> word you encode correctly, you earn <u>1 point</u>.

You will have 5 minutes in total. There will <u>not</u> be any practice before we begin – when we proceed, you will see a screen similar to the one below. Once the task begins, the timer will immediately start counting down.

Are there any questions?



Instructions for Experiment 2

In this experiment, you will be asked to vote in a referendum on whether all participants in the room will fund an actual tree planting project. If the referendum passes, some of your earnings from Experiment 1 will be used to fund the project.

The project involves planting and maintaining 160 trees along the Mississippi River Valley, which includes the state of Tennessee. To carry out the tree planting, we have partnered with the organization GreenTrees. GreenTrees contracts with farmers to create forests that landowners permanently maintain.

The Mississippi River is a critical body of water in North America for commerce, climate and energy. It is the largest river in the United States and the third largest in the world. The Mississippi River Valley is a vital habitat for migratory birds and numerous plant and animal species. 40% of North America's waterfowl and 60% of all bird species migrate along the Mississippi River, although their population has declined from habitat loss.

There are many benefits of reforestation, including:

- **Improved water quality**. 180 million Americans depend on forest watersheds for their drinking water. The natural water filtration trees provide can lower costs associated with drinking water treatment.
- **Improved air quality**. Trees remove pollution from the atmosphere, improving air quality.
- **Riparian buffers**. Trees help improve water quality in streams, rivers and lakes, and protects these waterways from the impact of adjacent land uses.
- **Flood control**. Forests reduce floods, therefore minimizing sediment, nitrates and phosphorus runoff into critical waterways.
- **Soil Stabilization**. Trees reduce the effects of erosion caused by water and wind.
- **Wildlife habitat**. Large populations of wildlife rely on forests for food, shelter and water.

Based on related projects carried out by GreenTrees in the Mississippi River Valley, the estimated impacts from the proposed project of planting 160 trees, over a 15 year period, are as follows:

- 4,275 gallons of water stored.
- 62 pounds of nitrogen and phosphorous out of runoff and groundwater before it reaches the Mississippi River.
- 30 metric tons of carbon dioxide (CO_2) captured.

Payment procedures

If the referendum passes, we will subtract a specified amount from your prior earnings in today's session and set this cash aside. We will take this cash and write a check to GreenTrees to pay for the project. The check will be mailed directly to GreenTrees. Along with the check, we will include a letter that briefly describes the project and recognizes that students from the University of Tennessee are funding the project. We have appended a copy of this letter to the instructions. If the referendum passes, we will ask a volunteer to place the stamped envelope containing the check and letter in the mail immediately after this session. As soon as we receive receipt of payment from GreenTrees, we will email it to all participants in this session.

If the referendum does not pass, no money will be subtracted from your earnings. No check will be sent to GreenTrees and the tree planting project will not be funded.

We have already negotiated a price for the project with GreenTrees. This amount may be higher than the money that would be collected from you and the other participants should the referendum pass. If the referendum passes, money from a research grant will be used to pay the difference.

The voting process

<Binary choice treatment>
You will now vote on whether to fund the tree planting project. In the referendum, we will ask you to place a YES or NO vote on whether you are in favor of funding the project if it cost you a specific amount.

The referendum passes if a majority, more than half of the votes, are YES votes. Otherwise, the referendum does not pass. If the referendum passes, each participant will pay the specified cost and the tree planting project will actually be funded. If the referendum does not pass, no money will be collected and the tree planting project will not be funded.

Before we proceed to the referendum, where you will be presented with the specific cost to you for funding the project, are there any questions?

<Double-bounded binary choice treatment>
You will now vote on whether to fund the tree planting project. In this context, a referendum is usually framed as a YES or NO vote on whether you would fund the project if it cost you a specific amount. In this experiment, however, the per person cost is uncertain. For this reason, we will ask you to vote YES or NO separately for two possible cost amounts. You will be first presented with one possible cost amount and asked to vote YES or NO. Then, you will be presented with another possible cost amount and again asked to vote YES or NO.

To determine the cost to you, the computer has been programmed to select randomly one of the two stated cost amounts. You will not know the selected cost prior to entering your decisions.

Your YES or NO vote for the randomly selected cost will be used to determine whether the referendum passes.

The referendum passes if a majority, more than half of the votes, are YES votes. Otherwise, the referendum does not pass. If the referendum passes, each participant will pay the randomly selected cost and the tree planting project will actually be funded. If the referendum does not pass, no money will be collected and the tree planting project will not be funded.

Before we proceed to the referendum, where you will be presented with two possible cost amounts to consider, are there any questions?

<Open-ended treatment>
You will now vote on whether to fund the tree planting project. In this context, a referendum is usually framed as a YES or NO vote on whether you would fund the project if it cost you a specific amount. In this experiment, however, the cost to each person is uncertain. Rather than ask you to vote YES or NO separately for each possible cost amount, we will ask: "What is the highest amount that you would pay and still vote in favor of funding the tree planting project?" By knowing the highest amount that you would pay, we will know all the possible costs for which you would vote NO (i.e. any cost that is higher than the amount you entered) and all the possible costs for which you would vote YES (i.e. any cost that is equal to or lower than the amount you entered).

To determine the cost to you, the computer has been programmed to select randomly from a wide range of possible cost amounts. These range from very low to very high amounts. You will not know the selected cost prior to entering your decision.

After all participants have made a decision, the amount you entered will be compared to the randomly selected cost and converted to either a YES or a NO vote. If you entered an amount that is greater than or equal to the cost, this will become a YES vote. If the amount you entered is less than the cost, this will become a NO vote.

The referendum passes if a majority, more than half of the votes, are YES votes. Otherwise, the referendum does not pass. If the referendum passes, each participant will pay the randomly selected cost and the tree planting project will actually be funded. If the referendum does not pass, no money will be collected and the tree planting project will not be funded.

Before we proceed to the referendum, are there any questions?

<Payment card treatment>
You will now vote on whether to fund the tree planting project. In this context, a referendum is usually framed as a YES or NO vote on whether you would fund the project if it cost you a specific amount. In this experiment, however, the cost to each person is uncertain. For this reason, we will ask you to vote YES or NO separately for several possible cost amounts.

To determine the cost to you, the computer has been programmed to select randomly one of the stated cost amounts. You will not know the selected cost prior to entering your decisions.

Your YES or NO vote for the randomly selected cost will be used to determine whether the referendum passes.

The referendum passes if a majority, more than half of the votes, are YES votes. Otherwise, the referendum does not pass. If the referendum passes, each participant will pay the randomly selected cost and the tree planting project will actually be funded. If the referendum does not pass, no money will be collected and the tree planting project will not be funded.

Before we proceed to the referendum, where you will be presented with several possible cost amounts to consider, are there any questions?

HASLAM COLLEGE OF BUSINESS
THE UNIVERSITY OF TENNESSEE, KNOXVILLE

Christian A. Vossler
Department of Economics
523 Stokely Management Center
Knoxville, TN 37996-0550
Email: cvossler@utk.edu

GreenTrees, LLC
ATTN: Reed Haynie, Director of Business Development
4243 Jackson Street, Box 250
The Plains, VA 20198

Dear Mr. Haynie,

Enclosed you will find a check made payable to GreenTrees, LLC, to support the project we discussed to plant 160 trees in the Mississippi River Valley. This payment is made possible by a research study conducted through the University of Tennessee Experimental Economics Laboratory. Participants in the study are students currently enrolled at the University of Tennessee. As a part of the study, the group of participants voted to use some of their earnings from the experiment to help pay for the tree plantings. On behalf of the students, it is my honor to present you with this support.

In return, I have a favor to ask. Upon receiving this payment, I would be very thankful to receive from you a formal acknowledgement of this purchase that I can share with the participants of this study.

Sincerely,

Christian A. Vossler
Professor of Economics
Director, UT Experimental Economics Laboratory
http://web.utk.edu/~cvossler/