Are we doing more harm than good? Hypothetical bias reduction techniques in potentially consequential survey settings

Vasudha Chopra
Christian A. Vossler

November 2024

WORKING PAPER #2024-04

WORKING PAPER SERIES
DEPARTMENT OF ECONOMICS
HASLAM COLLEGE OF BUSINESS
https://haslam.utk.edu/economics/

HASLAM
COLLEGE OF BUSINESS
THE UNIVERSITY OF TENNESSEE, KNOXVILLE

Are we doing more harm than good?
Hypothetical bias reduction techniques in potentially consequential survey settings

Vasudha Chopra[a] and Christian A. Vossler[b,*]
[a] *Department of Data Science, Economics and Business, Plaksha University, India 140306*
[b] *Department of Economics, University of Tennessee, Knoxville, TN 37996*

Abstract: Researchers deploying stated preference surveys to elicit monetary valuations for public goods commonly use techniques devised to reduce bias in hypothetical choice settings. This practice is conceptually at odds with accumulated evidence that most survey respondents instead perceive that their decisions have economic consequences (i.e., affect their future welfare). We examine three bias reduction procedures in both hypothetical choice and incentive compatible, real payment settings: cheap talk, solemn oath, and certainty adjustment. While we find that the oath reduces willingness to pay (WTP) in a hypothetical setting, the oath instead increases WTP by over 30% in a consequential setting. Cheap talk does not alter mean WTP in a consequential setting but leads to a stark difference in WTP across sexes. Applying the common rules for *ex post* adjustment of choices based on stated response certainty leads to significant and large decreases in WTP estimates for both hypothetical and consequential cases. Our results suggest that survey researchers should make use of screening questions to better target hypothetical bias reduction techniques to only those prone to bias.

JEL Classifications: C92, D82, D9, H41, Q51

Keywords: hypothetical bias, consequentiality, stated preferences, experiments, solemn oath, cheap talk, certainty adjustment

---

[*] Direct correspondence to Christian Vossler, Department of Economics, University of Tennessee, Knoxville, TN 37996. E-mail: cvossler@utk.edu. Telephone: 865-974-1699. Fax: 865-974-4601. This paper reports research involving the collection of data on human subjects. Approval from the University of Tennessee Institutional Review Board was obtained, as protocol UTK IRB-21-06225-XM. Funding was provided by the J. Fred Holly Chair of Excellence endowment.

1. Introduction

The traditional view is that using stated preference (SP) surveys to elicit monetary valuations for public goods results in the over-estimation of demand due to "hypothetical bias." Under the traditional view, people view their responses in SP surveys as hypothetical choices, and it is widely documented that people vote "yes" or otherwise indicate a higher demand for public goods in a hypothetical choice setting compared to one where choices are motivated by economic incentives. This bias has important implications for public policy as SP-based value estimates are routinely used in government benefit-cost analyses and in litigations over natural resource damages. As a result, researchers have sought ways to reduce bias, either through *ex ante* interventions designed to mitigate hypothetical bias or *ex post* corrections that use follow-up questions to adjust responses from those suspected to have provided biased signals of demand. Experiments have validated these techniques by applying them to (explicitly) hypothetical choice settings and showing that resulting demand estimates better agree with those obtained from parallel settings with direct financial consequences.

Researchers conducting SP field surveys commonly use one or more hypothetical bias reduction techniques, which is a sensible approach based on the traditional view. However, a somewhat more recent viewpoint, promulgated by the work of Carson and Groves (2007), is that governments use considerable resources to survey people about their preferences, and this information ultimately helps to inform public decision making; therefore, a respondent should perceive that her stated choices are consequential in the sense that they affect her expected (future) welfare. While some researchers have adopted this alternative view, and when asked most survey respondents state they perceive the value elicitation to be consequential, the use of hypothetical bias reduction techniques has nevertheless continued. This begs the question of how approaches designed and validated for hypothetical choice situations perform in environments where many or most respondents perceive the elicitation as consequential. To address this question, this study uses an experiment to examine the effects of using hypothetical bias reduction techniques in an incentive compatible, real payment setting.[1]

Loomis (2014) summarizes hypothetical bias techniques along with accumulated evidence on their performance. The leading *ex post* approach involves recoding choices or otherwise conditioning demand estimates on stated choice certainty.[2] Certainty corrections were popularized by Champ et al. (1997), who demonstrated that they could be used to equate actual and hypothetical donations to a public good. The basic idea is that hypothetical choices from those who are more certain about them are more likely to reflect behavior in settings with direct financial consequences. In a recent meta-analysis, Penn and Hu (2023) find that recoding answers based on follow-up certainty questions can be effective in reducing hypothetical bias, but that adjusted values can be sensitive to recoding rules (and can overcorrect for hypothetical bias). More recently, researchers employing field surveys routinely include follow-up questions designed to gauge whether respondents have consequentiality beliefs and adjust welfare measures based on elicited beliefs (e.g., Herriges et al. 2010; Zawojska, Bartczak, and Czajkowski 2019). While accumulated evidence from these studies supports the view that field surveys are consequential to most respondents, it has further been shown that adjusting welfare estimates based on consequentiality

[1] It is important to point out that consequentiality is a necessary pre-condition for a survey value elicitation mechanism to be incentive compatible (Vossler et al. 2023). More generally, incentive compatibility depends on the value elicitation question(s) used and associated allocation rules, along with respondents' beliefs about how survey choices affect one's future welfare.

[2] Loomis (2014) also discusses "market calibration" and "data screening" techniques, which are uncommon in practice. The former relies on pairing a field survey with a complementary lab experiment where the good described in the survey is offered for real in the lab. This approach has limited appeal given that most field surveys are intended to value large-scale public goods, which cannot be readily funded or implemented without the government. The latter technique is intended to reduce bias due to the use of question formats, such as open-ended questions, which give rise to biases regardless of the presence of economic incentives.

beliefs enhances internal and external validity (e.g., Herriges et al. 2010; Vossler, Doyon, and Rondeau 2012; Vossler and Watson 2013).

Examples of *ex ante* interventions include solemn oaths, cheap talk scripts, consequentiality designs, and social desirability and cognitive dissonance approaches (Loomis 2014). The solemn oath mimics the preamble to sworn testimony made in the court of law and serves as a commitment device to reveal truthful preferences by asking participants to swear upon their honor to give honest answers (Jacquemet et al. 2013). The solemn oath has been shown to reduce hypothetical bias in a variety of contexts, including second-price auctions and referenda (Jacquemet et al. 2011, 2013, 2017). The work of Cummings and Taylor (1999) popularized the use of cheap talk in SP surveys, and in their experiment the cheap talk script informed participants that people tend to vote "yes" in a hypothetical referendum when they would instead vote "no" in a binding referendum. The script further provides potential explanations for the bias, including: (1) the tendency for people to be in favor of pro-social initiatives (e.g., cleaning the environment); and (2) a failure for people to consider their budget constraint. It is clear from the meta-analysis by Penn and Hu (2019) that much evidence has been accumulated on cheap talk scripts.

Consequentiality designs augment the potential link between surveys and public decision-making by, for example, stressing to respondents that information will be shared with relevant authorities, and framing value elicitation questions as advisory referenda. Social desirability reduction approaches ask respondents how they think others would behave, while cognitive dissonance approaches add choice options that allow respondents to signal they are in favor of the policy, but not necessarily at the offered amount (for details, see Loomis 2014).

This study uses an experiment to examine the effects of the solemn oath, a quasi-cheap talk script, and various certainty adjustment approaches, in the context of a consequential decision setting. The value elicitation mechanism we deploy is incentive compatible in the sense that it is a dominant strategy for participants to truthfully reveal preferences. The hypothetical bias reduction approaches were selected based on their use in field SP surveys and because they function in distinct ways. Most cheap talk scripts make explicit the presumed direction of bias and thus directly motivate respondents to reconsider, for example, whether to vote "yes" in a referendum. Solemn oaths can also influence choices but do not explicitly signal a direction of bias or what an "honest" answer entails in a hypothetical referendum. As an *ex post* approach, certainty adjustment does not influence choices. To our knowledge, only one prior study has explored the effects of a hypothetical bias reduction technique (the solemn oath) in the context of a real payment setting: Jacquemet et al. (2017) fail to reject the hypothesis that the oath has no effect on votes in a binding referendum. However, all participants in their study faced the same cost amount, which precludes the estimation of mean or median WTP, and leaves open the possibility that the oath may have altered choices for other cost amounts.

It is pragmatic to explain why we do not evaluate all hypothetical bias reduction methods previously mentioned. Social desirability reduction and dissonance techniques involve value elicitation questions that cannot readily be adapted into incentive compatible elicitation mechanisms. Moreover, researchers rarely use these techniques. Testing consequentiality scripts or the effects of adjusting responses based on follow-up consequentiality questions is clearly problematic in a setting with direct financial consequences. A consequentiality script would cause participants to second guess whether the experiment worked as initially described, and therefore lead to a loss in control. Follow-up consequentiality belief questions would serve to confuse participants in the event the good was funded (perhaps making them think that they now do not have to pay), and otherwise make people feel that the experimenter had deceived them.

Cheap talk scripts face a similar challenge as consequentiality scripts. That is, by stating that people behave differently in "hypothetical" versus "real" choice settings, those about to participate in an incentivized decision task may second guess whether real money (and provision) is at stake. To overcome this challenge, we use a quasi-cheap talk script that avoids describing the decision task as "hypothetical" and depicts an alternative behavioral mechanism that, as in Cummings and Taylor (1999), works through the budget constraint. Our script describes the tendency for people to act more pro-socially in incentivized experiments compared to parallel non-experimental situations (e.g., Benz and Meier 2008; Laury and Taylor 2008). A potential explanation for this bias is a "house money effect", which implies that people consider their budget constraint differently in an experiment.[3] While there are clear differences between the two, our script retains a fair amount of the language included in the Cummings and Taylor (1999) script.[4]

Under a best-case scenario, a hypothetical bias reduction technique influences the choices of those prone to bias (or adjusts their responses after the fact) while leaving unaltered the choices of everyone else. Investigating techniques in both consequential and hypothetical settings is important for finding hypothetical bias reduction methods that are applicable regardless of whether SP respondents hold consequentiality beliefs. However, it is also possible that these techniques distort demand estimates for respondents who view the elicitation as consequential and thus potentially do more harm than good. An *ex ante* intervention may eliminate consequentiality beliefs or alter a respondent's utility in a way that leads to choices that deviate from the would-be choices made in the direct payment setting the survey endeavors to emulate. An *ex post* correction could mistakenly recode truthful responses.

[3] Of course, this "bias" could also be attributable to an experimenter demand effect.
[4] Our motivation to include a quasi-cheap talk treatment stems from feedback received after presenting preliminary findings, which is perhaps unsurprising given the popularity of this approach.

Consistent with many SP surveys, we elicit values for an environmental public good (tree plantings) with considerable non-use value. The application and incentive compatible value elicitation procedures closely follow Vossler and Zawojska (2020). Parallel procedures for hypothetical choice treatments are comparable to past hypothetical bias experiments that involve hypothetical referenda. Using these procedures, our experiment confirms stylized facts from the literature: WTP is higher in a hypothetical choice setting relative to a consequential setting, i.e., there is an upward hypothetical bias; using a solemn oath reduces but does not eliminate the hypothetical bias; and some certainty adjustment rules are effective at reducing or eliminating hypothetical bias.

Our findings related to the use of hypothetical bias reduction techniques in a consequential setting are as follows. The solemn oath *increases* WTP by 30%. Using evidence from a post-experiment questionnaire, we argue that the oath motivates those with underlying pro-social preferences to increase their WTP. Such behavior is congruent with lab experiment evidence that the oath increases contributions to a public good (Hergueux et al. 2022). A quasi-cheap talk script does not alter mean WTP but does influence preferences: The script drives a significant WTP gap between sexes, and further reduces the variance of the WTP distribution for males. Applying common certainty adjustment rules, which recode or reweight only uncertain "yes" votes, leads to significantly lower WTP estimates. A certainty adjustment procedure that instead adjusts both "yes" and "no" votes, increases WTP. Holding the cost of the project fixed, we find no systematic difference in stated certainty levels associated with "yes" votes between real and hypothetical choice settings. This contrasts with a common conjecture in the literature, specifically that people who vote "yes" in a hypothetical referendum are less certain of their choice than they would be in a consequential setting.

2. Experimental Design

2.1 Treatments and follow-up certainty questions

There are five between-subjects treatments. The consequential treatment (CONS) uses an incentive compatible mechanism to provide a benchmark measure of demand from which to measure hypothetical bias and the effects of using hypothetical bias reduction techniques. The hypothetical treatment (HYPO) uses identical instructions and procedures to the CONS treatment but emphasizes that choices are hypothetical in that choices do not influence outcomes. The HYPO+Oath and CONS+Oath treatments ask participants to sign a solemn oath, which mimics those used in prior hypothetical bias experiments (e.g., Jacquemet et al. 2013). Participants know that signing the oath is voluntary and does not affect the outcome or earnings from the voting experiment.

The last treatment, CONS+CheapTalk, presents participants with a quasi-cheap talk script after introducing the tree-planting project and voting procedures, but prior to the vote. As discussed in the Introduction, the language of the script closely follows the original cheap talk script of Cummings and Taylor (1999), with the important exceptions that "house money effects" are emphasized as a potential source of bias (rather than "hypothetical bias"), and that people are more inclined to contribute to publicly spirited causes when using house money (which parallels the message from Cummings and Taylor that people are more likely to support social causes when no money is at stake). The solemn oath and quasi-cheap talk scripts, as well as experiment instructions for the consequential and hypothetical voting tasks, are in the appendix.

We apply certainty adjustments based on *ex post* procedures. In the experiment, after all votes are submitted, we present respondents with all their voting choices and ask them to state their certainty level, using a scale from 1 ("very uncertain") to 10 ("very certain"), separately for each choice. The instructions emphasize that responses to the certainty questions will not influence the voting outcome in any way. In most related experiments, only the “yes” respondents in hypothetical treatments are asked about response certainty. We asked all participants, regardless of their treatment assignment or voting choices, to answer the certainty questions. Our analysis focuses on the effects of certainty adjustments on CONS and HYPO treatment choices.

2.2 Value elicitation tasks

The valuation scenario and the instructions for the consequential treatments (aside from the cheap talk and oath scripts) are based on Vossler and Zawojska (2020). Participants are asked to consider whether they would fund a project that involves planting and supporting 160 trees in the Appalachian Mountains. Participants are provided with general information on the benefits of tree planting and reforestation, such as improved water quality, soil stabilization and wildlife habitat. Many SP surveys involve environmental public goods, and one advantage of the method is the ability to quantify both use and non-use sources of value. We speculate that our elicitation mainly captures non-use values.

We elicit preferences using a payment card format that has participants vote “yes” or “no” on a set of referenda that vary only in terms of the per-person cost of funding the tree-planting project. The cost amounts are: \$0, \$1, \$2, \$3, \$4, \$5, \$6, \$8, \$10, \$12, and \$15. All referenda are presented simultaneously. After all votes are submitted, one referendum is randomly selected, and a majority vote implementation rule decides whether the randomly selected referendum passes. In the consequential treatments, participants must pay the selected per-person cost if the referendum passes, and the project is implemented; the payment card format along with the allocation rule constitutes an incentive compatible mechanism.[5] In the hypothetical treatments, even if a referendum “passes”, no money is collected and no trees are planted.

The elicitation mechanism was chosen over a single “yes” or “no” vote because it captures more precise information on preferences. Using an induced-value experiment with financial incentives, Vossler and McKee (2006) find that the elicited WTP distribution based on this mechanism is statistically equivalent to the induced WTP distribution. In their experiments that elicited values for similar tree planting projects, Vossler and Zawojska (2020) compare this payment card mechanism with a single binary choice and find that they yield statistically equivalent WTP distributions.

When facing this random selection mechanism, some participants may believe that the “selected” cost depends on the actual cost of the tree plantings, while others may speculate that money collected may exceed the actual cost, in which case money could go towards additional tree plantings. Either belief would result in a loss of incentive compatibility in the consequential treatments. To mitigate these potential biases, instructions make explicit that the actual cost of the project has been negotiated with a conservation organization, the money collected will not exceed the actual cost, and any difference between the amount collected and the actual cost would be covered by the experimenter.

Participants in consequential treatments are informed that if the vote passes, they will be emailed a certificate from the conservation organization One Tree Planted. The certificate acknowledges the donors, the number of trees planted, and the date of payment. This approach enhances the credibility of the elicitation mechanism. Instructions for the hypothetical voting

[5] Azrieli, Chambers, and Healy (2018) prove that, for a set of games or decision problems that, when analyzed individually are incentive compatible, a mechanism that randomly selects one of the games to be binding is also incentive compatible. Each “game” in our mechanism is a single binary referendum with a majority vote rule, which is incentive compatible, and thus this theory proves the incentive compatibility of the payment card mechanism.

treatments also describe this procedure, but instead frame it as what would have happened if instead the vote was binding.

2.3 Testable hypotheses and power analysis

We summarize below the main (null) testable hypotheses:

Hypothesis 1: WTP is equal in the hypothetical (HYPO) and consequential (CONS) treatments.

Hypothesis 2: The solemn oath has no effect on WTP in a hypothetical choice setting.

Hypothesis 3: The solemn oath has no effect on WTP in a consequential setting.

Hypothesis 4: Certainty adjustment has no effect on WTP in a hypothetical choice setting.

Hypothesis 5: Certainty adjustment has no effect on WTP in a consequential setting.

Hypothesis 6: The quasi-cheap talk script has no effect on WTP in a consequential setting.

To help determine sample sizes, we conducted a pilot experiment (n=20) with the HYPO+Oath treatment. Procedures and instructions closely followed those of the actual experiment, and participants were drawn from the same subject pool. Using an interval regression model, the standard deviation of WTP from the pilot is 5.05. The payment card treatment of Vossler and Zawojska (2020) involved the same elicitation procedure and a similar tree planting project, and therefore provides useful data from which to form priors about the WTP distribution for our benchmark consequential treatment. The standard deviation of WTP reported in their paper (Model II) is 3.87.

Assuming the above standard deviations for the respective hypothetical and consequential treatments in our experiment, and accounting for the elicitation format and payment card bid design, we undertook a power analysis using Monte Carlo simulations. We settled on sample sizes of 150 people per treatment. For comparisons of two hypothetical treatments, this allows us to detect a minimum treatment effect of about \$1.60 with 80% power, based on a 5% significance level. For comparisons between two consequential treatments, the minimum detectable effect size is about \$1.25, and for comparisons between real and hypothetical treatments the minimum detectable effect size is about \$1.50. To place some perspective on these numbers, we note that Vossler and Zawojska (2020) estimate a mean WTP of about \$4 for a similar tree-planting project. Meta-analyses suggest that, on average, the ratio of hypothetical to actual WTP is about 3 (List and Gallet 2001; Little and Berrens 2004; Murphy et al. 2005; Penn and Hu 2019).[6] Therefore, the experimental design is well-powered to detect a relatively modest degree of hypothetical bias. Further, the design can detect the effects of interventions that alter valuations as little as 20 to 30% relative to the HYP and CONS baselines, respectively. Penn and Hu (2019, 2023) report that cheap talk and certainty reduction approaches reduce hypothetical to "real" WTP ratios considerably.

2.4 Experimental Procedures

Experiment sessions are conducted either online or in the UT Experimental Economics Laboratory using participants drawn from a common subject pool. To minimize mode-of-administration effects, online and in-person sessions involve synchronous participation and are facilitated by an experiment moderator, i.e., the online sessions are conducted like standard laboratory sessions. The experiment is programmed using the software z-Tree (Fischbacher 2007) and online sessions are implemented with z-Tree unleashed (Duch, Grossman, and Lauer 2020). In the online setting, the moderator and participants interact through Zoom.

[6] There is considerable heterogeneity in this ratio, nevertheless. Penn and Hu (2019) find that about two-thirds of the estimated ratios are 1.41 or higher.

A typical session proceeds as follows. An experiment moderator verifies registrations by checking Student ID cards. Written instructions are provided to participants, and a moderator reads instructions aloud while participants follow along. A moderator emphasizes that instructions provide only correct information, and individual decisions will be kept anonymous. Participants enter all decisions on personal computers. The moderator encourages and answers questions.

The experiment has four stages. In the first stage, the participants go through a standard risk elicitation task of the sort popularized by Holt and Laury (2002). In the second stage, participants engage in unrelated, incentivized tasks. These tasks provide a means through which respondents earn money that could be spent on the tree-planting project. Stage three is the voting experiment used to elicit values for the tree-planting project. At the beginning of the session, participants are aware that there are multiple parts to the study but have no details on the voting experiment. In the final stage, participants fill out a questionnaire that includes demographic questions along with various measures intended to provide insight on the drivers of any observed treatment effects. Loomis (2011) points out that the underlying causes for hypothetical bias are not fully understood and apart from these bias reduction approaches, the social psychology literature offers alternative behavioral explanations that may explain the bias.

2.5 Participants

Participants were recruited from an existing database of undergraduate students at the University of Tennessee that had previously registered to receive invitations for economics experiments. Participants could only attend a single session of the experiment. An experiment session lasted approximately 65 minutes and on average participants earned $22. Fifty experiment sessions were conducted between the spring of 2021 and the fall of 2023. Including the pilot, we use data from 743 participants.[7,8]

3. Results

3.1 Data

We begin the analysis by reporting descriptive statistics for the sample (Table 1) and presenting the raw vote percentages by treatment and cost amount (Table 2). The average participant age is approximately 21 years, 55% of participants are female, about 52% are currently employed. Fifty-two percent of the participants have previously taken part in an (unrelated) laboratory experiment. Due to technical issues with some of the online sessions, we are missing complete questionnaire data for about 5% of the sample.

As clear from Table 2, as expected the percentage of "yes" votes decreases as cost increases in all treatments. There are considerable differences between HYPO and CONS treatment votes, with deviations of 28 to 43 percentage points for costs of \$3 to \$10. A Mann-Whitney U test rejects the null hypothesis that that the sum of participant "yes" votes for the two treatments follows the same probability distribution ($p<0.001$). Further, there is evidence that the oath alters votes in the hypothetical choice setting ($p=0.047$). While the data suggest that the oath lowers demand, significant voting differences between HYPO+OATH and CONS ($p<0.001$) remain. These results

[7] We also enrolled 184 students from Appalachian State University in online sessions. After analyzing the data when we were nearing the end of the planned data collection, we uncovered stark differences in the effects of the oath across the university-specific samples. We use explicit University of Tennessee branding on the oath form, and the oath had a negligible effect on the Appalachian State students. In retrospect, this difference was predictable, and it would have been preferrable to use university-specific oath forms. As the power analysis did not account for university-specific differences, we ultimately made the decision to drop the Appalachian State sample and enrolled more Tennessee students.

[8] Due to variations in attendance rates, we fell somewhat short of our sample size goal for the CONS treatment (n=137 instead of n=150). Incorporating the actual versus planned sample sizes in our power analysis yields minimum detectable effect sizes that are virtually identical.

align well with the prior literature that examines the effect of the oath on hypothetical bias (Carlsson et al. 2013; Jacquemet et al. 2013, 2017), and therefore indicate that our combination of methods, public good, and subject pool do not give rise to findings that are singularly out of line with the prior literature. Applying the Mann-Whitney U test reveals that the oath has a significant impact on consequential choices (p=0.018), and there are a higher number of "yes" votes with the oath. The quasi-cheap talk script has no overall effect (p=0.904).

We further apply Fisher's exact tests of vote proportions at specific cost amounts for more granular information on WTP distribution disparities across treatments. Differences between CONS and HYPO occur at every cost (p<0.01), except for $0. The same is true when comparing CONS and HYPO+Oath (p<0.02). HYPO and HYPO+Oath voting choices are statistically different at relatively high cost amounts: costs of $6 (p=0.011), $10 (p=0.038), $12 (p=0.064), and $15 (p=0.040). Further, CONS and CONS+Oath differ at costs that fall towards the middle of the bid distribution: $2 (p=0.018), $4 (p=0.077), $5 (p=0.008), $6 (p=0.087) and $8 (p=0.057). In their related experiment, Jacquemet et al. (2017), find that adding the oath to a binding referendum has no statistical effect. However, this comparison involves a single cost amount, and so our finding is not necessarily at odds with this prior result. There are no differences in the proportion of "yes" votes between CONS and CONS+CheapTalk at any cost level.

3.2 The effects of the solemn-oath and quasi-cheap talk on willingness to pay

We use a censored regression model (Cameron and James 1987) to estimate WTP distributions for each of the between-subjects treatments, and test for treatment effects. In doing so, we interpret a vote y_{it} from respondent i on referendum t as providing a censored signal of (i.e., a bound on) WTP. Let one's valuation be a linear function of covariates such that $WTP_{it}^{*} =$

$\mathbf{x}_i\boldsymbol{\beta} + u_{it}$, where $\mathbf{x}_i$ is a vector of covariates, $\boldsymbol{\beta}$ is a vector of unknown parameters, and u_{it} is a mean-zero error term. Then, a "yes" vote ($y_{it} = 1$) when facing cost c_{it} indicates that $WTP_{it}^* \geq c_{it}$, and a "no" vote ($y_{it} = 0$) signals $WTP_{it}^* < c_{it}$.[9] To facilitate estimation, we assume a normal distribution for the errors, with $u_i \sim Normal(0, \sigma_i^2)$. This gives rise to the following log-likelihood function:

$$[1] \qquad \ln \mathcal{L} = \sum_i \sum_t \left\{ y_{it} \ln\left(1 - \Phi\left(\frac{c_{it} - \mathbf{x}_i\boldsymbol{\beta}}{\sigma_i}\right)\right) + (1 - y_{it}) \ln \Phi\left(\frac{c_{it} - \mathbf{x}_i\boldsymbol{\beta}}{\sigma_i}\right)\right\},$$

where Φ denotes the normal CDF. As illustrated by Haab, Huang, and Whitehead (1999), conclusions drawn from hypothetical bias experiments can be sensitive to the assumption of a common error variance across treatments. We allow the error variance to differ across treatments by specifying a standard deviation function $\sigma_i = \mathbf{z}_i\boldsymbol{\gamma}$, where the $\mathbf{z}_i$ denote treatment-specific indicators. Standard errors are clustered at the participant level to account for within-subject serial correlation.

Table 3 presents estimates from censored regressions that include as covariates indicators for all treatments except for CONS. Estimated coefficients therefore represent differences in WTP relative to the benchmark consequentiality treatment. In models that also include control variables, which are defined in Table 1, the control variables are demeaned which ensures that the intercept can be interpreted directly as mean WTP for the CONS treatment.

Full sample results appear as Model 1 and Model 2, and the remaining four models split the sample based on sex. As expected from random treatment assignment, including control variables has a negligible impact on estimated treatment effects. As such, in the discussion that

[9] Alternatively, we could use the sequence of votes to define a single WTP interval for each participant and apply an interval regression to the resulting cross-section data. Doing so has a negligible effect on estimates. We analyze the voting data instead as a panel to keep consistency with our certainty adjustment analysis, which requires that we apply potentially different weights across the sequence of votes.

follows we will focus on results from models without control variables. We also note that the indicator for mode-of-administration is insignificant in each case. For example, in Model 2 this indicator has a coefficient of -0.075 and is statistically insignificant (p=0.878).

Model 1 produces the following WTP estimates: CONS $3.90, CONS+Oath $5.24, CONS+CheapTalk $3.79, HYPO+Oath $7.77, and HYPO $9.20. We note that the WTP estimate from the benchmark, consequential choice treatment ($3.90) falls in the range of values reported by Vossler and Zawojska (2020) for their related payment card treatment ($3.81 to $4.00, depending on model specification).

The model reveals clear evidence of hypothetical bias. WTP is more than twice as high in HYPO compared with CONS (ratio=2.4), and the difference is statistically significant ($p<0.001$). This result is consistent with the nonparametric tests. It is also transparent from Table 3 that using an oath in a hypothetical choice setting decreases WTP by $1.43 (a 16% decrease), and this difference is significant (p=0.036). In contrast, and consistent with the Fisher exact test results reported earlier, the oath *increases* WTP in the incentive compatible elicitation setting (p=0.022). The size of the effect is considerable at $1.34 cents, an increase of 34%, which suggests caution in using this approach in a field survey setting where there is presumably a mix of respondents with and without consequentiality beliefs. Our quasi-cheap talk script is estimated to decrease mean WTP by just 11 cents, and this difference is not statistically significant (p=0.829).

The lone control variable that is statistically significant in Model 2 is the indicator for female, which suggests that females are willing to pay $1.75 ($p<0.001$) more for the tree planting project. This large effect motivated us to split the sample by sex, which reveals interesting results.[10]

[10] As we did not originally plan to examine gender effects, this line of investigation should be considered informative but exploratory.

Models 3 and 4 include only females, and Models 5 and 6 only males. For females, the magnitude of hypothetical bias is virtually identical to that of the full sample (ratio=2.4). The oath is ineffective in reducing hypothetical bias for females. However, in a consequential setting, the oath increases WTP by $1.65 (40% increase; p=0.033). The analysis of males reveals contrasting results. For males, the oath reduces WTP in the hypothetical choice setting by $2.32 (p=0.025), which effectively halves the level of hypothetical bias. Although the effect is somewhat large at $1.05 (p=0.247), there is no statistical evidence that the oath impacts WTP in a consequential setting among males. There is evidence that effectiveness of the oath may be heterogenous across participants (Carlsson et al. 2013), although prior studies of the oath have not explored sex-based differences.

There is further suggestive evidence of sex-based elicitation effects as it relates to the quasi-cheap talk script. For females, the script increases WTP by 81 cents while for males the script decreases WTP by 97 cents. Although these estimates not individually statistically significant, and the experimental design is only powered to detect sex-specific effects that are quite large, the difference in WTP for the CONS+CheapTalk treatment across genders is highly significant (p<0.001). Further, evidence from Models 5 and 6 (as well as Model 2) reveal that the quasi-cheap talk script significantly reduces the standard deviation of the WTP distribution for males.

3.3 Certainty adjustment methods

We asked respondents to state their level of certainty for each of their votes separately, using a 10-point Likert scale (1 is “Very Uncertain” and 10 is “Very Certain”). Overall, participants say they are highly certain of their votes. The mean certainty levels are 8.7 and 8.3 for the CONS and HYPO treatments, respectively. These figures are slightly higher, at 9.1 and 8.8, if we restrict the sample to "yes" votes.

Table 4 presents stated certainty levels by vote and cost amount, separately for the HYPO and CONS treatments. Based on two-sample t-tests (allowing for unequal variances), we find no significant differences in the certainty levels, at any cost amount, for "yes" votes across treatments. We do, however, find that certainty levels of "no" respondents are statistically different at all but the lowest three cost amounts. While "no" CONS votes are associated with higher certainty levels, the differences across treatments are nevertheless modest and range from 1.1 to 1.9.

Utilizing the certainty responses, we examine the effects of adjusting the HYPO treatment votes using several approaches that have been applied in the literature. Importantly, we also apply these procedures to the CONS treatment data, which relays the effects of applying these techniques in a setting where respondents view the elicitation as consequential and not hypothetical.

As suggested by the meta-analysis of Penn and Hu (2023), the most widely used certainty correction approaches involve recoding less certain "yes" votes as "no" votes. However, the cutoff levels needed to equate hypothetical and "real" WTP differ across studies, and of course in field survey applications without a real payment benchmark researchers are left to use their own judgement in terms of what recoding rule to use. Here, we consider cutoffs as low as 7 and as high as 10. For example, using a cutoff of 7 (Cert≥7) means that all "yes" votes associated with certainty levels of 7 and higher are retained whereas "yes" votes with lower levels of certainty are recoded as "no".

We further consider two adjustment rules that work by interpreting the certainty levels into probabilities, which is arguably less *ad hoc*. Penn and Hu (2023) refer to the two approaches as the “asymmetric uncertainty model” (ASUM) and the “symmetric uncertainty model” (SUM). For the ASUM, “yes” votes are converted into probabilities using the formula:

[2] $\Pr(\text{yes}|\text{vote}=\text{"yes"}) = \frac{4}{9} + \frac{5}{90} * certainty\ level,$

which recodes a “yes” vote into a probability of voting “yes” between 50% (certainty level of “1”) and 100% (certainty level of “10”). For the SUM, both “yes” and “no” votes are given a probabilistic interpretation. The “yes” votes are converted using formula [2], while the “no” votes are converted to probabilities using the formula:

[3] $\Pr(\text{yes}|\text{vote}=\text{"no"}) = \frac{5}{9} - \frac{5}{90} * certainty\ level,$

which recodes a “no” vote into a probability of voting yes that is between 0% (certainty level of 10) and 50% (certainty level of 1).

For ASUM and SUM, estimation continues to be conducted with the censored regression model [1], with the exception that y_{it} is now a “fractional response” rather than an indicator variable. For ASUM, y_{it} is defined by [2] in the case of a “yes” vote and $y_{it} = 0$ as before for a “no” vote. For SUM, y_{it} is defined by [2] in the case of a “yes” vote and by [3] for a “no” vote.

The top panel of Table 5 presents unadjusted and certainty-adjusted WTP estimates based on the HYPO data. Also reported is the difference in the various WTP estimates and the benchmark CONS estimate. For the threshold recoding approaches, as expected the WTP estimate decreases as the threshold increases. Only for the most extreme recoding scheme, Cert≥10, is the adjusted WTP estimate statistically equal to the benchmark CONS estimate. While this is somewhat dramatic, we note again that certainty levels are high, and so an extreme recoding is needed to decrease WTP estimates by a meaningful amount. The ASUM adjustment procedure reduces WTP slightly while the SUM adjustment procedure increases WTP by $1.45. The latter result reflects the fact that participants are relatively more certain about their “yes” votes.

The bottom panel of Table 5 reports certainty-adjusted WTP estimates using the CONS data. As a point of comparison, reported are differences between the adjusted and unadjusted CONS estimates. As can be gleaned from the table, the adjustment schemes impact CONS and HYPO estimates in qualitatively similar ways. The changes, in percentage terms, induced by each approach track closely when applied to CONS and HYPO data. For all threshold recoding approaches, the adjusted WTP estimate is statistically different, and lower than, the unadjusted estimate. For the most extreme threshold procedure, which equated adjusted hypothetical and consequential WTP, the adjustment procedure when applied to consequential choices decreases WTP by \$1.87, which is a 47% decrease. The ASUM procedure lowers WTP by 34 cents ($p<0.001$) and SUM increases WTP by 88 cents ($p<0.001$).

4. Exploration of behavioral drivers

To better understand the ways in which the oath and quasi-cheap talk script influence WTP, we next analyze data from the post-experiment questionnaire to find potential moderating variables. Included in the questionnaire are: the 10-item Big Five personality instrument of Gosling, Rentfrow, and Swann (2003); two scenarios from the Test of Self-Conscious Affect-3 (Tagney et al. 2000) designed to measure guilt, shame, and blame; and two questions from Lundquist et al. (2009) that measure lying aversion. There are two measures for each of the nine characteristics, which we aggregated (averaged). Our exploration reveals that two of the characteristics – agreeableness and shame – moderate the effect of the solemn oath.[11,12] Both

[11] There are no statistical differences for the agreeableness and shame measures for any pair of treatments, i.e., responses to these elicitations do not appear to be correlated with treatment assignment.

[12] WTP for the two hypothetical choice treatments statistically increases with the 'openness' personality trait and decreases with the 'emotional stability' trait. Those in the quasi-cheap talk treatment with a higher level of 'conscientiousness' have a lower WTP. In these cases, the effects of the oath or cheap talk script do not vary with

variables take on values of 1 to 7, where a higher value indicates a stronger level of the characteristic.

Presented in Table 6 are regressions that allow the WTP for each treatment to vary according to agreeableness and shame. In both oath treatments, WTP increases with the level of agreeableness. This drives a gap between the CONS and CONS+Oath treatments as agreeableness increases, and decreases the effectiveness of the oath to reduce hypothetical bias. In the consequential choice setting, the oath increases WTP by $1.29 for every one-unit increase in agreeableness (p=0.012) relative to the no oath case. At agreeableness levels of 4 or higher, WTP is higher with the oath; otherwise, WTP with and without the oath are statistically equal. A somewhat smaller marginal effect of $1.00 arises when the oath is used (versus not) in the hypothetical choice setting (p=0.059). For all but those with relatively high agreeableness (4.8 or higher), the oath significantly reduces hypothetical bias. As females have a higher agreeableness than males (p<0.001), this provides a possible explanation for why the oath is effective for males but not for females at reducing hypothetical bias.

The psychology literature has linked the agreeableness personality trait with other-regarding behavior. The meta-analysis by Wilmot and Ones (2022) indicates that agreeable individuals are more cooperative, trusting as well as altruistic. Individuals who are generally more agreeable are also pro-social and display pro-environmental attitudes as well as behaviors (Soutter and Mõttus 2021). To the extent that agreeableness captures pro-social attitudes, our results indicate that the oath motivates people to behave more consistently with their underlying social preferences; indeed, in the benchmark treatment WTP does not statistically vary with

the personality trait. Table A1 in the appendix presents regression results associated with our explorations of other Big Five personality traits.

agreeableness. Our finding that the oath increases WTP in the consequential setting is consistent with recent lab experiment evidence that participants give considerably more in a standard voluntary contributions mechanism game after the oath is administered (Hergueux et al. 2022).

At high levels of shame, the effect of the oath is neutralized in both consequential and hypothetical cases. In the consequential setting, the oath decreases relative WTP by $1.14 for every one-unit increase in shame (p=0.038), leading to no statistical treatment effect at levels of shame of 3.1 or higher. In the hypothetical setting, the oath increases WTP by 88 cents compared to the HYPO baseline (p=0.041); as a result, the effectiveness of the oath in mitigating hypothetical bias decreases with shame. For shame levels of 3.0 or higher, the oath has no statistical effect on hypothetical bias. The psychology literature has found that shame constrains immoral behavior and is positively associated with honesty (e.g., Cohen et al. 2011). We therefore interpret the findings related to shame as showing that the oath, intuitively, has no discernable effect among those predisposed to honest behavior.

While the analysis of agreeableness and shame sheds some light on the behavioral response to the oath, none of the moderating variables we considered are correlated with the quasi-cheap talk script. However, some insight can be gleaned from the follow-up questions directly related to participant votes for the tree planting projects. The questionnaire asks participants whether they agree with the statement "I had thought about how others in the room might feel about the project before I voted." Those in the quasi-cheap talk treatment had a statistically different, and lower, level of agreement relative to those in the consequentiality benchmark treatment ($p = 0.049$). This may be explained by the fact that the script frames a "yes" vote as consistent with socially desirable behavior. Therefore, the script may have diverted the attention of some to the greater good rather than the good of the voting group.

Following Jacquemet et al. (2017), we also asked follow-up questions to accumulate evidence on whether the oath treatments are correlated with self-reports of honest behavior in the experiment, and whether participants are happy following the experiment. Using Mann-Whitney U tests, there are no statistical differences (even at the 10% level) in self-reported honesty, perceived honesty of others, or happiness, between any pair of treatments. Jacquemet et al. (2017) find that people are less honest in a hypothetical treatment relative to a real payment, but this difference goes away when one introduces the oath. Further, the same authors find that people in the hypothetical plus oath treatment are less happy than in the hypothetical and real payment treatments. Our results indicate that these earlier findings are not robust to other subject pools and/or applications.

5. Discussion

Belief elicitation questions reveal that most respondents perceive that high-quality stated preference (SP) surveys have the potential to influence policy decisions, and that increasing the provision of public goods would come at a cost to households. Moreover, accumulated evidence shows that controlling for respondent beliefs about the consequentiality of an SP survey can impact welfare estimates and improve internal and external validity (Rondeau and Vossler 2024). These findings are at odds with the traditional view that SP survey respondents interpret the elicitation as hypothetical, and therefore question whether hypothetical bias reduction techniques are needed to provide accurate welfare estimates. In this study, we ask whether the commonly employed hypothetical bias reduction techniques have unintended effects on valuations in settings where respondents hold consequentiality beliefs. To provide an answer to this question, we conducted an experiment that examines the performance of hypothetical bias reduction techniques in an incentive compatible, real payment decision setting. The evidence is univocal: *ex ante* reduction techniques alter preferences, and *ex post* certainty corrections distort revealed preferences.

There is clear evidence that using a solemn oath, an *ex ante* approach, increases willingness to pay (WTP) in our consequential choice setting by 30%. Therefore, while the oath may reduce bias in hypothetical choice settings (as documented here and elsewhere), in a consequential setting the evidence shows that the oath can change underlying preferences. For both hypothetical and consequential choice settings, we find that the oath increases WTP for those with a stronger "agreeableness" personality trait, which has been linked to stronger pro-social behavior. In the hypothetical choice case, the oath is effective only for those with a moderate to low agreeableness level. In the consequential choice setting, the oath increases WTP compared to the no oath case for those with moderate to high agreeableness levels. That the oath can motivate pro-social behavior in an incentivized experiment is congruent with Hergueux et al. (2022), who find the oath increases contributions in a voluntary contributions mechanism game. The oath has differential effects on WTP across sexes, and the fact that females tend to be more agreeable is a possible driver of this result. We also find that the oath has a null effect, in both hypothetical and consequential settings, among those prone to feelings of shame. As those prone to shame tend to behave more honestly, it may be the case that these people are unphased by an appeal to 'tell the truth.'

The quasi-cheap talk script we examine did not alter average WTP, but nevertheless has detectable effects in a real payment setting: The script leads to a statistically significant difference in WTP between men and women, and decreases the variance of the WTP distribution for males. The original cheap talk script of Cummings and Taylor (1998) explicitly mentions hypothetical bias, and we had to alter the script in the real payment setting to avoid confusing participants.

Nevertheless, the fact that the script altered preferences suggests, like the solemn oath, it may have unintended effects in survey settings that respondents perceive as consequential. It is intuitive that a cheap talk script that indicates the elicitation is “hypothetical” will make those who otherwise perceived the elicitation to be consequential to second-guess this belief, resulting in a clear loss of incentive compatibility.

In the case of using a follow-up certainty question to make *ex post* adjustments of choices, a common belief among researchers is that those who say they are uncertain about their “yes” vote in a hypothetical setting would have instead voted “no” in a consequential setting. The common procedure is to then recode uncertain “yes” votes as “no.” Of course, in a consequential setting, people can still be uncertain, and this common method of certainty adjustment serves to bias WTP estimates downward, which we show by applying common adjustment techniques to data from real payment elicitations. Of potential interest, “yes” respondents in our consequential treatment are just as certain about their responses as are “yes” respondents in the hypothetical setting.

The evidence on consequentiality beliefs suggests that for most field applications we should expect a mix of respondents, some who view the survey as consequential and others who view the survey as hypothetical. Our evidence suggests that while hypothetical bias techniques are likely to reduce hypothetical bias for the latter group, applying the same techniques to the former group may either alter preferences or recode valid indicators of preferences. When applying an *ex ante* technique, ‘treating’ respondents with and without consequentiality beliefs is unavoidable unless one elicits beliefs prior to the voting scenarios. Lloyd-Smith, Adamowicz, and Dupont (2019) first proposed the possibility of asking about consequential beliefs beforehand and find that doing so increases the fraction holding this belief. An online survey (or interviewer) could then display the oath (or a cheap talk script) only to those indicating the survey is hypothetical. In the case of *ex post* techniques such as adjustments based on response certainty, information on beliefs (regardless of whether they were asked before or after the value elicitation scenario(s)) can be used to delineate which subset of respondents the certainty-based adjustments should be applied to.

An important and related issue is how to best identify those prone to hypothetical bias. One possibility is to ask a "policy consequentiality" question (e.g., 'to what extent do you think the survey will be considered by authorities?'). However, a common but not universal finding is that conditioning welfare estimates on beliefs over whether the survey could influence policy *increases* WTP estimates, which goes in the opposite direction of hypothetical bias. Instead, or in addition to, a "payment consequentiality" question (e.g., 'if the policy is undertaken do you believe you will have to pay?') is a logical candidate. Those who perceive policy consequentiality but not payment consequentiality should in theory have a strategic incentive to over-reveal demand. Further, given those in typical "hypothetical bias" experiments should neither perceive policy nor payment consequentiality, those with similar beliefs in the SP setting should also be prone to a positive bias. Vossler et al. (2023) speculate that consequentiality belief questions may also be identifying "protest" respondents, such as those that do not trust the government to provide the good if funded. Protesters are not necessarily prone to hypothetical bias, and in fact are inclined to vote "no." Sorting the protesters from others who do not hold payment consequentiality beliefs may therefore be important when targeting hypothetical bias reduction techniques.

References

Azrieli, Yaron, Christopher P. Chambers, and Paul J. Healy. 2018. "Incentives in Experiments: A Theoretical Analysis." *Journal of Political Economy* 126(4): 1472-1503.

Benz, Matthias, and Stephan Meier. 2008. "Do people behave in experiments as in the field? Evidence from donations." *Experimental Economics* 11: 268-281.

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: 269-276.

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

Carlsson, Fredrik, Mitesh Kataria, Alan Krupnick, Elina Lampi, Åsa Löfgren, Ping Qin, and Thomas Sterner. 2013. "The truth, the whole truth, and nothing but the truth—A multiple country test of an oath script." *Journal of Economic Behavior & Organization* 89: 105-121.

Champ, Patricia A., Richard C. Bishop, Thomas C. Brown, and Daniel W. McCollum. 1997. "Using donation mechanisms to value nonuse benefits from public goods." *Journal of Environmental Economics and Management* 33: 151-162.

Cohen, Taya R., Scott T. Wolf, A.T. Panter, and Chester A. Insko. 2011. "Introducing the GASP scale: A new measure of guilt and shame proneness." *Journal of Personality and Social Psychology* 100(5): 947–966

Cummings, Ronald G., and Laura O. Taylor. 1999. "Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method." *American Economic Review* 89(3): 649-665.

Duch, Matthias L., Max R.P. Grossmann, and Thomas Lauer. 2020. "z-Tree unleashed: A novel client-integrating architecture for conducting z-Tree experiments over the Internet." *Journal of Behavioral and Experimental Finance* 28: 100400.

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

Gosling, Samuel D., Peter J. Rentfrow, and William B. Swann. 2003. "A very brief measure of the Big-Five personality domains." *Journal of Research in Personality* 37(6): 504-528.

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.

Hergueux, Jérôme, Nicholas Jacquemet, Stéphane Luchini, and Jason F. Shogren. 2022. "Leveraging the honor code: Public goods contributions under oath." *Environmental and*

Resource Economics 81: 591-616.

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.

Holt, Charles A., and Susan K. Laury. 2002. “Risk aversion and incentive effects.” *American Economic Review* 92(5): 1644-1655.

Jacquemet, Nicolas, Alexander G. James, Stéphane Luchini, and Jason F. Shogren. 2011. “Social psychology and environmental economics: a new look at ex ante corrections of biased preference evaluation.” *Environmental and Resource Economics* 48(3): 413-433.

Jacquemet, Nicolas, Robert-Vincent Joule, Stéphane Luchini, and Jason F. Shogren. 2013. “Preference elicitation under oath.” *Journal of Environmental Economics and Management* 65(1): 110-132.

Jacquemet, Nicolas, Alexander James, Stéphane Luchini, and Jason F. Shogren. 2017. “Referenda under oath.” *Environmental and Resource Economics* 67(3): 479-504.

Laury, Susan K., and Laura O. Taylor. 2008. “Altruism Spillovers: Are Behaviors in Context-free Experiments Predictive of Altruism Toward a Naturally Occurring Public Good?” *Journal of Economic Behavior and Organization* **65**(1): 9-29.

List, John A., and Craig A. Gallet. 2001. “What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? Evidence from a Meta-Analysis.” *Environmental and Resource Economics* 20: 241-254.

Little, Joseph, Robert Berrens. 2004. “Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta-Analysis.” *Economics Bulletin* 3(6): 1-13.

Lloyd-Smith, Patrick, Wiktor Adamowicz, and Diane Dupont. 2019. Incorporating stated consequentiality questions in stated preference research. Land Economics 95(3): 293-306.

Loomis, John. 2011. “What's to know about hypothetical bias in stated preference valuation studies?.” *Journal of Economic Surveys* 25(2): 363-370.

Loomis, John B. 2014. “Strategies for overcoming hypothetical bias in stated preference surveys.” *Journal of Agricultural and Resource Economics* 39(1): 34-46.

Lundquist, Tobias, Tore Ellingsen, Erik Gribbe, and Magnus Johannesson. 2009. “The aversion to lying.” *Journal of Economic Behavior & Organization* 70: 81-92.

Murphy, James J., P. Geoffrey Allen, Thomas H. Stevens, and Darryl Weatherhead. 2005. “A meta-analysis of hypothetical bias in stated preference valuation.” *Environmental and Resource Economics* 30(3): 313-325.

Penn, Jerrod, and Wuyang Hu. 2019. “Cheap talk efficacy under potential and actual hypothetical bias: A meta-analysis.” *Journal of Environmental Economics and Management* 96: 22-35.

Penn, Jerrod, and Wuyang Hu. 2023. “Adjusting and calibrating elicited values based on follow-up certainty questions: A Meta-analysis.” *Environmental and Resource Economics* 84: 919-946.

Rondeau, Daniel, and Christian A. Vossler. 2024. “Incentive compatibility and respondent beliefs: Consequentiality and game form.” Working Papers 2024-02, University of Tennessee, Department of Economics.

Soutter, Alistair Raymond Bryce, and René Mõttus. 2021. “Big Five facets' associations with pro-environmental attitudes and behaviors.” *Journal of Personality* 89(2): 203-215.

Tangney, June P., Ronda L. Dearing, Patricia E. Wagner, and Richard Gramzow. 2000. *Test of Self-Conscious Affect–3 (TOSCA-3)*. Fairfax, VA: George Mason University.

Vossler, Christian A., Stéphane Bergeron, Maurice Doyon, and Daniel Rondeau. 2023. “Revisiting the gap between the willingness to pay and willingness to accept for public goods.” *Journal of the Association of Environmental and Resource Economists* 10(2): 413-435.

Vossler, Christian A., Maurice Doyon, and Daniel Rondeau. 2012. "Truth in consequentiality: theory and field evidence on discrete choice experiments." *American Economic Journal: Microeconomics* 4(4): 145-71.

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

Vossler, Christian A., and Sharon B. Watson. 2013. “Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field.” *Journal of Economic Behavior and Organization* 86: 137-147.

Vossler, Christian A., and Ewa Zawojska. 2020. “Behavioral drivers or economic incentives? Toward a better understanding of elicitation effects in stated preference studies.” *Journal of the Association of Environmental and Resource Economists* 7(2): 279-303.

Wilmot, Michael P., and Deniz S. Ones. 2022. “Agreeableness and its consequences: A quantitative review of meta-analytic findings.” *Personality and Social Psychology Review* 26 (3): 242-280.

Zawojska, Ewa, Anna Bartczak, and Mikołaj Czajkowski. 2019. “Disentangling the effects of policy and payment consequentiality and risk attitudes on stated preferences.” *Journal of Environmental Economics and Management* 93: 63-84.

Table 1. Description of Data

Variable name	Description	Mean (Std dev)
Vote	=1 if voted 'yes'; =0 if 'no'	0.504 (0.500)
Cost	Cost of the referendum, \$0 to \$15	6.000 (4.553)
CONS	=1 if baseline consequential treatment	0.184 (0.388)
HYPO	=1 if baseline hypothetical treatment	0.202 (0.401)
HYPO+Oath	=1 if hypothetical with oath treatment	0.205 (0.403)
CONS+Oath	=1 if consequential with oath treatment	0.203 (0.402)
CONS+CheapTalk	=1 if consequential with quasi-cheap talk treatment	0.206 (0.404)
Certainty	Certainty level of vote, 1 to 10	8.577 (2.352)
Control Variables		
Female	=1 if participant is female	0.545 (0.498)
Risk Averse	=1 if selected > 5 safe choices in lottery task	0.489 (0.500)
Experience	=1 if person took part in prior paid experiment	0.518 (0.500)
Employed	=1 if employed; else=0	0.524 (0.499)
Age	Participant's age, in years	20.582 (2.908)
Earnings	Earnings prior to voting, in \$	25.771 (4.157)
Online	=1 if participated in an online session	0.701 (0.458)
Other Variables		
Honesty-Self	Self-reported honesty rating, scale 1-10	9.246 (1.236)
Honesty-Others	Peer-rated honesty of other participants, scale 1-10.	7.459 (1.906)
Happiness	Self-reported happiness rating, scale 1-10	7.626 (1.873)
Agreeableness	Measure of 'agreeableness' personality trait; 1 (“Strongly Disagree”) to 7 (“Strongly Agree)	4.518 (1.206)
Openness	Measure of 'openness' personality trait; 1 (“Strongly Disagree”) to 7 (“Strongly Agree)	5.123 (1.165)
Extraversion	Measure of 'extraversion' personality trait; 1 (“Strongly Disagree”) to 7 (“Strongly Agree)	4.251 (1.604)
Emotional Stability	Measure of 'emotional stability' personality trait; 1 (“Strongly Disagree”) to 7 (“Strongly Agree)	4.467 (1.432)
Conscientiousness	Measure of 'conscientiousness' personality trait; 1 (“Strongly Disagree”) to 7 (“Strongly Agree)	5.469 (1.232)
Shame	Measure of proneness to guilt; 1 (“Strongly Disagree”) to 7 (“Strongly Agree)	2.521 (1.299)

Table 2. Percentage of "yes" votes by cost of the proposal

Cost	CONS	HYPO	HYPO+Oath	CONS+Oath	CONS+CheapTalk
\$0	93.43	94.00	96.05	93.38	94.77
\$1	78.83	92.67	93.42	84.77	78.43
\$2	64.96	90.67	89.47	78.15	66.01
\$3	53.28	86.67	86.18	59.60	52.94
\$4	41.61	83.33	78.95	52.32	39.87
\$5	31.39	74.67	71.71	47.02	32.68
\$6	22.63	62.67	48.03	31.79	20.26
\$8	15.33	48.00	39.47	24.50	13.73
\$10	10.95	39.33	27.63	16.56	9.15
\$12	9.49	30.00	20.39	13.91	5.88
\$15	8.03	28.00	17.76	11.92	4.58
Overall	39.08	66.36	60.83	46.72	38.03

Table 3. Willingness to pay regressions

	(1) Full sample	(2) Full sample	(3) Females only	(4) Females only	(5) Males only	(6) Males only
HYPO	5.294***	5.314***	5.569***	5.505***	4.891***	5.111***
	(0.666)	(0.693)	(0.826)	(0.822)	(1.090)	(1.116)
HYPO+Oath	3.867***	3.864***	5.005***	4.853***	2.567***	2.661***
	(0.580)	(0.591)	(0.791)	(0.781)	(0.844)	(0.864)
CONS+Oath	1.342**	1.403**	1.648**	1.524*	1.047	0.910
	(0.586)	(0.613)	(0.774)	(0.778)	(0.904)	(0.916)
CONS+CheapTalk	-0.112	-0.188	0.814	0.560	-0.973	-0.978
	(0.519)	(0.546)	(0.724)	(0.729)	(0.714)	(0.790)
Intercept	3.902***	3.883***	4.113***	4.208***	3.616***	3.596***
	(0.398)	(0.425)	(0.503)	(0.507)	(0.644)	(0.655)
Standard deviation function (σ):						
HYPO	1.062	0.709	1.015	0.957	1.051	1.231
	(0.663)	(0.713)	(0.800)	(0.799)	(1.113)	(1.169)
HYPO+Oath	0.013	-0.531	0.692	0.599	-1.178	-1.086
	(0.611)	(0.655)	(0.814)	(0.781)	(0.937)	(1.003)
CONS+Oath	0.391	0.141	0.364	0.404	0.385	0.291
	(0.668)	(0.725)	(0.821)	(0.806)	(1.107)	(1.111)
CONS+CheapTalk	-0.708	-1.345**	0.242	0.297	-3.067***	-3.132***
	(0.639)	(0.638)	(0.799)	(0.765)	(0.855)	(0.870)
Intercept	5.137***	5.429***	4.739***	4.657***	5.644***	5.609***
	(0.488)	(0.551)	(0.608)	(0.580)	(0.801)	(0.826)
Control variables?	No	Yes	No	Yes	No	Yes
Log-Likelihood	-4,080.6494	-3,993.6432	-2,165.7827	-2,129.9284	-1,806.7718	-1,798.2634
Observations	8,173	8,151	4,444	4,444	3,707	3,707

Notes: Cluster-robust standard errors in parentheses (clustered by participant). *** p<0.01, **p<0.05, * p<0.1. Control variables are defined in Table 1, and are demeaned so that the intercept has a consistent interpretation across specifications.

Table 4. Follow-up certainty question responses, by vote and referendum cost

	\$0	**\$1**	**\$2**	**\$3**	**\$4**	**\$5**	**\$6**	**\$8**	**\$10**	**\$12**	**\$15**
						Consequential choice treatment (CONS)					
Yes	9.81	9.50	9.31	8.90	8.60	8.30	8.03	8.55	8.29	8.08	7.27
No	7.44	8.14	8.33	7.98	8.16	8.18	8.30	8.37	8.60	8.79	9.02
						Hypothetical choice treatment (HYPO)					
Yes	9.74	9.62	9.46	9.22	8.84	8.64	7.94	7.89	7.86	7.44	7.26
No	5.62	6.60	7.00	6.83	6.74	6.32	6.67	6.91	7.22	7.67	7.76
						Difference (CONS - HYPO)					
Yes	0.08	-0.12	-0.16	-0.32	-0.24	-0.34	0.10	0.66	0.42	0.63	0.01
No	1.82	1.54	1.33	1.15*	1.43**	1.86***	1.62***	1.46***	1.37***	1.12***	1.26***

Notes: *** p<0.01, **p<0.05, and * p<0.1 indicate a statistically significant difference between the CONS and HYPO treatments based on a two-sample t-test (unequal variances).

Table 5. Certainty-adjusted willingness to pay estimates

	Hypothetical choice treatment (HYPO)						
	Unadjusted	Cert. ≥7	Cert. ≥8	Cert. ≥9	Cert. ≥10	ASUM	SUM
HYPO	9.196***	7.326***	6.539***	5.600***	4.405***	8.288***	10.643***
	(0.534)	(0.458)	(0.448)	(0.449)	(0.475)	(0.483)	(0.552)
HYPO – CONS (unadjusted)	5.294***	3.425***	2.637***	1.698***	0.503	4.386***	6.742***
	(0.666)	(0.608)	(0.600)	(0.600)	(0.620)	(0.626)	(0.681)
	Consequential choice treatment (CONS)						
	Unadjusted	Cert. ≥7	Cert. ≥8	Cert. ≥9	Cert. ≥10	ASUM	SUM
CONS	3.902***	3.209***	2.911***	2.476***	2.031***	3.561***	4.778***
	(0.399)	(0.349)	(0.334)	(0.329)	(0.336)	(0.372)	(0.449)
CONS – CONS (unadjusted)	n/a	-0.693***	-0.991***	-1.426***	-1.871***	-0.340***	0.876***
		(0.174)	(0.195)	(0.227)	(0.271)	(0.060)	(0.127)

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Cluster-robust standard errors in parentheses. Column headings "Cert. ≥ X" refer to the rule used to recode "yes" votes into "no" votes, e.g., "Cert ≥7" means that all "yes" votes are recoded as "no" except for those associated with a certainty level of 7 or higher. ASUM refers to "asymmetric uncertainty model" and SUM refers to "symmetric uncertainty model." See text for estimation details.

Table 6. 'Agreeableness', 'shame', and willingness to pay

	(7)	(8)	(9)	(10)
HYPO	4.191	4.040	7.811***	7.945***
	(2.699)	(2.743)	(1.513)	(1.596)
HYPO+Oath	-1.748	-1.565	4.113***	4.140***
	(2.294)	(2.325)	(1.312)	(1.364)
CONS+Oath	-3.916	-3.836	4.600***	4.824***
	(2.394)	(2.471)	(1.500)	(1.571)
CONS+CheapTalk	-3.599	-3.436	-0.194	-0.168
	(2.425)	(2.466)	(1.311)	(1.359)
CONS × Agreeableness	-0.371	-0.473		
	(0.388)	(0.411)		
HYPO × Agreeableness	-0.059	-0.126		
	(0.429)	(0.419)		
HYPO+Oath × Agreeableness	0.938***	0.793***		
	(0.309)	(0.293)		
CONS+Oath × Agreeableness	0.917***	0.801**		
	(0.339)	(0.338)		
CONS+CheapTalk × Agreeableness	0.463	0.302		
	(0.346)	(0.336)		
CONS × Shame			0.318	0.343
			(0.413)	(0.442)
HYPO × Shame			-0.577*	-0.608*
			(0.347)	(0.355)
HYPO+Oath × Shame			0.303	0.299
			(0.257)	(0.241)
CONS+Oath × Shame			-0.827**	-0.878**
			(0.368)	(0.368)
CONS+CheapTalk × Shame			0.438	0.425
			(0.286)	(0.269)
Intercept	5.267***	5.732***	2.863***	2.795**
	(1.799)	(1.914)	(1.037)	(1.125)
Control variables?	No	Yes	No	Yes
Treatment-specific std deviations?	Yes	Yes	Yes	Yes
Log-Likelihood	-3,818.0936	-3,758.0291	-3,833.8573	-3,760.2646
Observations	7,766	7,766	7,766	7,766

Notes: Cluster-robust standard errors in parentheses (clustered by participant). *** p<0.01, **p<0.05, * p<0.1. Control variables are defined in Table 1 and are demeaned so that the intercept has a consistent interpretation across specifications.

Supplemental Appendix

Title: Are we doing more harm than good? Hypothetical bias reduction techniques in potentially consequential survey settings

Authors: Vasudha Chopra and Christian A. Vossler

Date: November 2024

This appendix includes:

- Supplemental Analysis
- Solemn oath
- Quasi-cheap talk script
- Instructions for consequential voting treatments
- Instructions for hypothetical voting treatments
- Post-experiment questionnaire

Table A1. 'Openness,' 'extraversion,' 'emotional stability,' 'conscientiousness,' and willingness to pay

	Openness	Extraversion	Emotional Stability	Conscientiousness
HYPO	0.349	4.713**	8.348***	4.343
	(2.871)	(2.065)	(2.365)	(3.373)
HYPO+Oath	0.807	2.994*	6.718***	0.604
	(2.687)	(1.694)	(2.059)	(2.787)
CONS+Oath	2.263	2.622	2.862	-0.144
	(3.031)	(1.913)	(2.422)	(2.968)
CONS+CheapTalk	0.809	0.324	0.773	0.419
	(2.536)	(1.695)	(1.972)	(2.680)
CONS × Big-Five measure	0.142	-0.105	-0.041	-0.581
	(0.390)	(0.302)	(0.359)	(0.386)
HYPO × Big-Five measure	1.164***	0.102	-0.641*	-0.354
	(0.393)	(0.326)	(0.342)	(0.447)
HYPO+Oath × Big-Five measure	0.787**	0.164	-0.638**	0.062
	(0.330)	(0.226)	(0.263)	(0.309)
CONS+Oath × Big-Five measure	0.071	-0.284	-0.246	-0.199
	(0.415)	(0.307)	(0.373)	(0.372)
CONS+CheapTalk × Big-Five measure	0.007	-0.146	-0.204	-0.647**
	(0.284)	(0.207)	(0.251)	(0.276)
Intercept	2.893	4.041***	3.805**	6.802***
	(2.062)	(1.347)	(1.647)	(2.195)
Control variables?	Yes	Yes	Yes	Yes
Treatment-specific std deviations?	Yes	Yes	Yes	Yes
Log-Likelihood	-3,756.8219	-3,785.2408	-3,767.2581	-3,766.3316
Observations	7,766	7,766	7,766	7,766

Notes: Cluster-robust standard errors in parentheses (clustered by participant). *** p<0.01, **p<0.05, * p<0.1. Control variables are defined in Table 1 and are demeaned so that the intercept has a consistent interpretation across specifications. For a concise table display, please note that each 'Big-Five measure' in the interaction term aligns with the characteristic shown at the top of each column. The dependent variable in all regressions is (latent) willingness to pay.

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Topic: “A study of survey voting procedures”; UTK IRB-23-07837-XM

I undersigned ……………………………. swear upon my honor that, during the whole experiment, I will:

Tell the truth and always provide honest answers.

Signature…………………………… Date………………

Quasi-cheap talk script

Research suggests that people in experiments are more inclined to spend money in ways that benefit others. In a voting setting like ours, this means that more people would vote “yes” in an experiment than would if this were not an experiment. This may be due to a “house money” effect. The question is then how we can get people to think about their votes in this experiment as they would if voting outside an experiment?

Let me tell you why I think that we see differences in behavior. I think that when we hear about a referendum that involves something that is basically good – helping people in need, improving environmental quality, or anything else – our basic reaction is to think: sure, I want to do this. I really want to vote “yes” to spend some of the money I have earned from the experiment.

But when making choices outside an experiment, we think differently about how we spend our own money to pay for something. We basically still would like to see good things happen, but when we are faced with the possibility of having to spend money that we didn’t earn from an experiment, we think about our options differently: if I spend money on this, that’s money I don’t have to spend on other things…we are more likely to vote in a way that takes into account the limited amount of money we have – That is just my opinion, of course, but it’s what I think may be going on in experiments like this.

So if I were in your shoes … I would ask myself: if I were voting outside the experiment, and I had to pay money if the referendum passed: do I really want to spend my money this way? If I really did, I would vote yes; if I didn’t, I would vote no. In any case, I ask you to vote just exactly as you would if you were making decisions outside of an experiment. Please keep this in mind in our referendum.

<u>Experiment Instructions (tree-planting: incentive compatible treatments)</u>

In this experiment, you will be asked to vote in a referendum on whether all participants in the room will collectively fund an **actual** tree-planting project. If the referendum passes, you and the other participations will pay for the tree plantings using some of the money you have earned in the prior experiments. If the referendum does not pass, no money will be collected from you and no trees will be planted.

About the project …

The project involves planting and maintaining 160 trees in the Appalachian Mountains, a region that stretches from southern New York state to northern Alabama and Georgia. To carry out the tree planting, we will partner with the non-profit organization One Tree Planted. This organization plants trees to return formerly unproductive mining, logging, and agricultural land to a natural state.

The benefits of planting trees include …

Improved water quality. The intricate root systems of trees act like filters, removing pollutants and slowing down the water's absorption into the soil. This natural water filtration can lower costs associated with drinking water treatment.
Improved air quality. Trees help to clean the air we breathe by absorbing harmful pollutants. Healthy, strong trees act as carbon sinks, reducing our carbon footprint and reducing the effects of climate change.
Flood control. Trees play a key role in capturing rainwater and reducing the risk of natural disasters like floods and landslides.
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.

Payment procedures

If the referendum passes, we will subtract a specified amount from your prior earnings in today's session and set this aside. We will use this money to purchase the tree plantings while you are completing the post-experiment questionnaire. Since the price of the tree plantings is more than the amount we would collect from you, we will use money from a research grant to pay the difference.

We will forward the confirmation email we receive from One Tree Planted. Attached to this email will be a certificate. The certificate will acknowledge students at the University of Tennessee for funding 160 trees.

If the referendum does not pass, no money will be subtracted from your earnings. No money will be given to One Tree Planted and the tree planting project will not be funded.

Budget reminder: Please keep your budget in mind when voting and think about whether funding the project is worth it to you, and other things you can spend your money on.

The voting process

In this experiment we will ask you to vote YES or NO separately for several possible cost amounts, which will range from $0 to $15.

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

Your YES or NO vote to 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, are there any questions?

Experiment Instructions (hypothetical treatments)

In this experiment, you will be asked to vote in a **hypothetical** referendum on whether all participants in the room would collectively fund a tree-planting project. This referendum is **hypothetical** in the sense that, regardless of how everyone votes, **no money** will be subtracted from your earnings, and **no trees** will be planted. To be as clear as possible, this is not a real referendum, but we want you to imagine how you would vote if given the opportunity to fund a tree-planting project.

About the project …

The project would involve planting and maintaining 160 trees in the Appalachian Mountains, a region that stretches from southern New York state to northern Alabama and Georgia. To carry out the tree planting, we would have partnered with the non-profit organization One Tree Planted, if this were a real referendum. This organization plants trees to return formerly unproductive mining, logging, and agricultural land to a natural state.

The benefits of planting trees include …

Improved water quality. The intricate root systems of trees act like filters, removing pollutants and slowing down the water's absorption into the soil. This natural water filtration can lower costs associated with drinking water treatment.
Improved air quality. Trees help to clean the air we breathe by absorbing harmful pollutants. Healthy, strong trees act as carbon sinks, reducing our carbon footprint and reducing the effects of climate change.
Flood control. Trees play a key role in capturing rainwater and reducing the risk of natural disasters like floods and landslides.
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.

Payment procedures

If this were a real referendum, and it passed, we would have subtracted a specified amount from your prior earnings in today's session and set this aside. We would have used this money to purchase the tree plantings while you were completing the post-experiment questionnaire. Since the price of the tree plantings would have been more than the amount we would have collected from you, we would have used money from a research grant to pay the difference.

We would have forwarded the confirmation email we would have received from One Tree Planted. Attached to this email there would have been a certificate. The certificate would have acknowledged students at the University of Tennessee for funding 160 trees.

If this were a real referendum, and it did not pass, no money would have been subtracted from your earnings. No money would have been given to One Tree Planted and the tree planting project would not have been funded.

Budget reminder: Please keep your budget in mind when voting and think about whether funding the project would have been worth it to you, and other things you could have spent your money on.

The voting process

In this experiment we will ask you to vote YES or NO separately for several possible cost amounts, which will range from $0 to $15.

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

Your YES or NO vote to 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.

Please keep in mind that this referendum is **hypothetical**. Regardless of whether the referendum passes, **no money** will be subtracted from your earnings and **no trees** will be planted. To be as clear as possible, this is not a real referendum, but we want you to imagine how you would vote if given the opportunity to fund a tree-planting project.

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

Post-experiment questionnaire

Part 1: About the Experiment

We would now like you to complete a short questionnaire. Please answer the following questions to the best of your knowledge. All information is completely anonymous and confidential. The first questions relate to your experience in today's experiment.

1. Have you previously participated in a paid study that took place in an experimental economics laboratory?
a. Yes b. No

2. Please indicate your level of agreement with the following statement: “I understood well the instructions for the voting experiment.”
1 - Strongly Disagree; 2 – Disagree; 3 – Neutral; 4 – Agree; 5 - Strongly Agree

3. Please indicate your level of agreement with the following statement: "I was well compensated for my participation in this study.”
1 - Strongly Disagree; 2 – Disagree; 3 – Neutral; 4 – Agree; 5 - Strongly Agree

4. Did you sign the oath statement to “Tell the truth and always provide honest answers”?.
Yes; No

5. Please indicate whether you agree or disagree with the following statements related to the Tree-planting project.
1 - Strongly Disagree; 2 – Disagree; 3 – Neutral; 4 – Agree; 5 - Strongly Agree

Generally speaking, I am happy to give at least some money to pay for tree plantings.

I had thought about how others in the room might feel about the project before I voted.

I did not have enough information to make a comfortable decision in the referendum.

I am in favor of more trees being planted, but cannot afford to pay for it.

I was confused about the procedure used to determine whether the referendum passed.

The responsibility to protect the environment may be shared between the general public and the government.

If I knew other participants were in favor of the proposal, this would have increased the chances that I voted yes.

Part 2: Honesty questionnaire

Using the scale below, please indicate how happy you are at the moment:
Scale of 1 to 10; 1 denotes Very unhappy and 10 denotes Very happy

Using the scale below, please indicate how honest you were during the experiment:
Scale of 1 to 10; 1 denotes Very dishonest and 10 denotes Very honest

Using the scale below, please indicate how honest you think the other participants were during the experiment
Scale of 1 to 10; 1 denotes Very dishonest and 10 denotes Very honest

Please rate the following statements on a scale of 1-7 where 1 denotes Strongly Disagree and 7 denotes Strongly Agree.

If I promise someone to tell the truth, that makes it very difficult for me to lie to that person.

I am more inclined to lie the more I have to gain from that lie.

Suppose you break something at work and find out a co-worker is blamed for the error. Respond to the following the level of agreement from 1 - Strongly Disagree to 7 - Strongly Agree.
You would think "This is making me anxious. I need to either fix it or get someone else to fix it".
You would think about quitting.
You would think "a lot of things aren't made well these days".

Suppose you make a mistake at work and find out a co-worker is blamed for the error. Respond to the following the level of agreement from 1 - Strongly Disagree to 7 - Strongly Agree.
You would think the company did not like the co-worker.
You would keep quiet and avoid the co-worker.
You would feel unhappy and eager to correct the situation.

Part 3: Personality traits

Here are a number of personality traits that may or may not apply to you. Please write a number next to each statement to indicate the extent to which you agree or disagree with that statement. You should rate the extent to which the pair of traits applies to you, even if one characteristic applies more strongly than the other. All questions below are to be rated from 1-7. 1 represents strongly disagree and 7 represents strongly agree.

I see myself as:
a. Extroverted, enthusiastic
b. Critical, quarrelsome
c. Dependable, self-disciplined
d. Anxious, easily upset
e. Open to new experiences, complex

f. Reserved, quiet
g. Sympathetic, warm
h. Disorganized, careless
i. Calm, emotionally stable
j. Conventional, uncreative

Part 4: About yourself

The next questions tell us something about you.
1. What is your age?
2. How do you describe yourself?
a. Male b. Female c. Transgender d. Do not identify myself as female, male, or transgender
3. What is your academic major?
4. What is your current student classification?
a. Freshman b. Sophomore c. Junior d. Senior e. Master's Student f. Law Student g. Doctoral Student h. Other
5. What was your student status for the Spring 2019 semester?
a. Full-time student b. Part-time student c. Not a student
6. In what range is your cumulative GPA?
a. 0 to 2.0 b. 2.1 to 2. c. 2.6 to 3.0 d. 3.1 to 3.5 e. 3.6 to 4.0
7. How many economics courses have you completed at the university level?

8. How would you best describe your current employment status?
a. Employed Full-Time b. Employed Part-Time c. Self-Employed Full-Time d. Self-Employed Part-Time e. Unemployed

Please use the following space to write any comments (positive or negative) you may have about the experiment.
____