Consequentiality: A Theoretical and Experimental Exploration of a Single Binary Choice

Richard T. Carson, Theodore Groves, John A. List

Abstract: Researchers, using contingent valuation (CV) to value changes in non-market goods, typically believe respondents always answer questions truthfully or they answer truthfully only when it is in their interest to do so. The second position, while consistent with economic theory, implies that interpreting survey responses depends critically on the incentive structure provided. We derive simple tests capable of distinguishing the two views. Our theoretical model for examining the incentive structure of a single binary choice relaxes the usual expected utility assumption. We test our theory using a field experiment involving voting to provide a public good. Experimental results are consistent theoretical predictions and cast doubt on the relevance of a large experimental literature using inconsequential questions and non-incentive-compatible mechanisms to make inferences about CV. The framework put forth should help in understanding the role played by theoretical conditions for preference elicitation and lend insight into the hypothetical bias literature.

IN THE UNITED STATES, every "economically significant" proposed rule making must undergo a formal benefit-cost analysis. While there is broad agreement on the determinants and major empirical techniques appropriate to estimate the cost side of the equation, a much more difficult task faces the practitioner interested in es-

Richard T. Carson (rcarson@weber.ucsd.edu) and Theodore Groves are both at the University of California, San Diego. John A. List is at the University of Chicago. The authors wish to thank Mark Machina for his help on an earlier version of this paper and the journal's coeditor and referees for constructive comments. We received useful advice from participants, too numerous to name, at seminar and conference presentations of earlier versions of this paper.

timating the total benefits of nonmarketed goods and services. Likewise, when environmental catastrophes occur, such as the oiling of the Prince William Sound in 1989 by the Exxon *Valdez*, credible benefit estimation is necessary to resolve damage disputes. In both cases, policy makers typically turn to a contingent valuation (CV) approach to provide signals of value since they remain the only method that has the ability to obtain the total economic value of the good in question.[1] Nevertheless, a contentious debate about whether a CV approach is able to provide reliable benefit estimation has proliferated among academics, practitioners, and policy makers over the past decade.

The debate is rarely guided by theoretical considerations; rather, it often revolves around a claim that an upward "hypothetical bias" permeates CV estimates. This claim is usually based on evidence from experiments with hypothetical bias typically defined as the difference between purely hypothetical and actual statements related to a good's value (see, e.g., Cummings, Harrison, and Rutström 1995; Cummings, Elliot, and Murphy 1997; List and Gallet 2001; Little and Berrens 2004; Murphy et al. 2005; Harrison 2006; Loomis 2011). Juxtapositioned against the experimental finding of an upward hypothetical bias is the large-scale meta-analysis of Carson et al. (1996), who compare estimates from CV surveys to estimates with their counterparts based on revealed behavior techniques, such as hedonic pricing and the household production/travel cost approach. They found that, on average, CV estimates are somewhat smaller but not statistically different from their revealed behavior counterparts.[2] As the literature has grown large, subsequent meta-analyses have focused on classes of goods and generally find the same result that CV estimates are on average the same as or smaller than estimates based on revealed behavior.[3] There is also a much smaller number of studies (e.g., Johnston 2006; Vossler et al. 2003) that compare predictions of the percentage in favor to the actual vote on a ballot proposition and find that CV results are consistent with the election results.

---

1. Following Carson and Louviere (2011), we use CV to refer to the use of a survey-based approach that elicits information about preferences for a nonmarketed good that are contingent on a constructed market.

2. This meta-analysis looked at 616 comparisons from 83 separate studies. It also found that a study in which the ratio of the CV estimate to the revealed behavior estimate was either close to one or much larger than one was likely to be published. This suggests two distinct positions in the literature.

3. See, for instance, Rosenberger and Loomis (2000), who look at estimates from outdoor recreation studies, and Kochi, Hubbell, and Kramer (2006), who look at value of statistical life estimates. While it is not a formal meta-analysis, Whittington (2010, 209) in his review of results from developing countries notes: "The main conclusion is that households' willingness to pay for a wide range of goods and services offered to respondents in stated preference scenarios is low, in both relative and absolute terms and in comparison to the costs of service provision."

At some level this is not surprising; surveys taken close to a general election are known to be good predictors of the outcome between two competing candidates or positions on a ballot measure. For instance, of over 300 final published polls done for statewide races in 2004, the average error on the margin of victory was 4.4%.[4] The mixed message being sent by these different types of studies is a source of confusion for both applied researchers and policy makers.

To help resolve some of this confusion, we examine the popular single binary choice format recommended by the US National Oceanic and Atmospheric Administration Blue Ribbon Panel (Arrow et al. 1993) from a mechanism design perspective and in the context of a field experiment.[5] Carson and Groves (2007) put forward a mechanism perspective that can help to reconcile some of the empirical results by providing a plausible set of auxiliary conditions that make a question using consequential single binary choice elicitation format incentive compatible for a public good, in the sense that truthful preference revelation is the dominant strategy. We show that a testable implication of their framework is that the fraction of people who favor a policy action in the population of interest should be invariant to the probability that the survey will influence the decision to provide the public good under the specified terms as long as that probability is positive. We also show that violation of one of Carson and Groves's conditions, choices should not have any influence on other decisions, predicts a specific pattern of deviations. We test both of these propositions in a field experiment along with looking at how the inconsequential (purely hypothetical) treatment behaves. The results support the theoretical predictions. In addition, in the case of a purely hypothetical question where Carson and Groves (2007) show that there is no theoretical prediction, we find a moderate size overestimate consistent with the existing literature (Murphy et al. 2005).

Our main theoretical contribution is to show that the conditions put forward by Carson and Groves (2007) for ensuring the incentive compatibility of a single binary choice question do not require agent preferences to conform to expected util-

---

4. For a list of the polling agencies doing statewide general election races and error rates using the standard Mosteller #5 measure, see http://www.surveyusa.com/Scorecards/2004 AllStatewide.xls. Most pollsters would contend that the difficult forecasting problem is not predicting how people will vote if they vote but, rather, in predicting whether they will vote.

5. CV studies are often referred to by the specific format used to collect the stated preference data. A single binary choice is the simplest member of a class of techniques collectively known as discrete choice experiments (Carson and Louviere 2011). Alternative preference elicitation formats may be advantageous from the perspective of gathering more preference information from each respondent and their properties can also be analyzed using mechanism design theory (Carson and Groves 2007). Understanding the properties of a single binary choice under different conditions is usually a good starting point for considering the incentive and informational properties of alternative preference elicitation formats.

ity. In particular, we show that this result holds under the weaker assumption that lies behind cumulative prospect theory (Tversky and Kahneman 1992) and rank-dependent expected utility (Quiggin 1992), two of expected utility's primary competitors. This result is important because the influence of a consequential survey question is through its probabilistic influence on an ultimate, potentially multistage decision of interest, and it has been argued that utility functions that are not linearly additive in probabilities are potentially important in environmental policy analysis (Mason et al. 2005; Shaw and Woodward 2008). Our results also help to emphasize the close relationship between the random lottery approach used in experimental economics that picks one out of k trials for a real monetary payoff and consequential preference questions with probabilistic influence on an agency decision.

1. HYPOTHETICAL BIAS AND CONSEQUENTIALITY

The concept of hypothetical bias has long been problematic (Mitchell and Carson 1989), although magnetically attractive to economists. It has been defined in a number of inconsistent ways in the literature, with the most common current definition focused on the deviation between a survey-based estimate and an estimate from a treatment using "real" money.[6] This might appear to be a reasonable definition until deeper thought is given to operationalizing it. One obvious defect is easily seen by considering voluntary contributions to provide a public good. Using actual voluntary contributions as the "real" benchmark should, due to free riding, grossly underestimate maximum willingness to pay (WTP) for the good, which is the standard measure of economic welfare that an economist wants to estimate.[7] It is useful to go through some of the other issues involved with the studies that populate the meta-analyses (e.g., List and Gallet 2001; Little and Berrens 2004; Murphy et al. 2005) aimed at shedding light on the degree of hypothetical bias likely to be present in CV studies before turning to the consequentiality/inconsequentiality distinction put forward by Carson and Groves (2007).

One major difficulty with the usual operationalization of hypothetical bias arises from the common belief that the estimate from any treatment using real money can be treated as a criterion measure for the purpose of testing validity of CV. This is a folly that many of us have committed. Several decades of experimental economics have convincingly shown that clever but nonintuitive mechanisms, such as the Becker, DeGroot, and Marschak (1964) mechanism and nth-price Vickrey auctions devel-

---

6. Mitchell and Carson (1989) explicitly excluded hypothetical bias from their typology of biases because a survey being hypothetical cannot be the root source of any well-defined directional bias, although they argue that it can influence the amount of effort that respondents put into answering a survey question.

7. Only from a marketing perspective, where the objective is to forecast contributions to a prospective fund-raising campaign, are actual donations an appropriate benchmark.

oped to ensure incentive compatibility, often do not deliver. This is particularly true until the subject has considerable experience with the mechanism. It is also now well known that different mechanisms lead to different estimates, particularly in initial rounds.[8] At a deeper level, preference elicitation approaches like Becker et al. and Vickrey auctions that elicit a continuous WTP response (to be compared to a cost amount to be revealed later) require expected utility to hold in order to be incentive compatible (Horowitz 2006).[9] This is problematic, as expected utility is known to be frequently violated in practice (Starmer 2000).

A further difficulty is that Carson and Groves (2007) show that the same preference revelation mechanism can have different incentive properties in a treatment with real payments than it does in a survey. Two examples vividly illustrate the problem. In a market context, a purchase made under a "take-it-or-leave-it" posted price generally is incentive compatible.[10] In a survey context, a consumer will generally either perceive the survey to be designed to help a firm determine whether a new purchase opportunity should be offered or designed to determine price sensitivity. In the first instance, a consumer who potentially wants to buy the good in the future should strategically indicate that they would purchase if offered. In the second instance, the incentive is to strategically appear to be more price sensitive (underestimation).

---

8. For a recent example and discussion, see figure 1 of Drichoutis, Nayga, and Lazaridis (2011), which shows that WTP in their incentive-compatible lotteries for food items increased over the course of 10 rounds by a factor of 2 to 3, depending on initial training, as subjects overcame the well-known heuristic of trying to buy low. Noussair, Robin, and Ruffieux (2004) compare valuation estimates in repeated rounds for a general population sample that was given induced values and randomly assigned to a Becker et al. or Vickrey auction treatment. They find that in the first round Becker et al. underestimated the induced value by 40% and that the Vickrey auction estimate was 30% lower, a difference that was significant at the $p < .05$ level. They also found that while the Vickrey auction had converged to being unbiased by the fourth round, the Becker et al. estimate had not. Mitani and Flores (2009) using a threshold public good mechanism and induced values found that the "real" treatment significantly underestimates while the "hypothetical" treatment does not.

9. Horowitz (2006) shows that Becker et al. and similar auction mechanisms may not induce truthful preference revelation in the sense of yielding the desired Hicksian WTP if expected utility does not hold. The reason for this is that the dominant strategy under the Becker et al. mechanism is to reveal the certainty equivalent that is dependent on the perceived distribution of possible cost. Under the expected utility, the Hicksian measure is equal to the certainty equivalence.

10. Even this proposition needs qualification. The number of consumers must be large enough that the actions of one consumer do not influence the actions of the firm (Roberts 1976). Dynamic linkages by a consumer such as going to a new local restaurant, even though the food is bad, in the hope that if the restaurant stays in business it will improve and offer another dining option in the future must also be ruled out.

With a public good and a donation payment vehicle, the only influence that a survey treatment can have is on the likelihood that an actual fund-raising drive is mounted. Optimal behavior in this situation for someone who wants to see the public good provided is to strategically overpledge in the survey and free ride later if the fund-raising drive is mounted. Carson and Groves (2007) further argue that the mechanisms like Becker et al. and Vickrey auctions that are incentive compatible in experiments (under the assumption that consumers are expected utility maximizers) are not incentive compatible in surveys. This is because the "real" experimental variant works because it effectively ensures that the extra preference information contained in a continuous response cannot be used to the agent's disadvantage, but such a guarantee cannot be made for the survey variant since the decision of interest occurs in the future.

The major issue, though, with most of the lab and field experiments examined in the hypothetical bias meta-analyses is that the "survey" treatments were explicitly designed to be "purely hypothetical" in the sense that subjects are told, often repeatedly so, that their responses will not have any influence on anything. Carson and Groves (2007) label such treatments as "inconsequential" and argue that any response in this case has the same influence on utility and, as such, no prediction from economic theory can be made as to how agents should respond.[11] They contrast this with a "consequential survey" question where the agent has a nonzero probability of influencing an outcome that the agent cares about. It is somewhat surprising that inconsequentiality (purely hypothetical) was picked out by those with a nascent interest in experimental economics as the major element of a CV survey that needed testing in order to help assess how such surveys were likely to perform in the field, since most of the major CV surveys done for policy purposes have not been cast as inconsequential.[12] Respondents are typically told that the survey is being done by university researchers to help inform policy decisions and a government agency is the explicit sponsor of the survey in some cases.

Poe and Vossler (2011), in their review of consequentiality, have called the perspective Kuhntian (Kuhn 1962) in nature and argue that it is fundamentally changing how researchers view stated preference data. Under consequentiality, survey respondents are explicitly told that their answers to preference questions will influence

---

11. Creation of a purely hypothetical survey question asking about preferences for a good is a more difficult undertaking than it may first appear, as it is natural for participants to speculate on how the results may be used. A conjecture that leads to overestimates is having participants think that if they indicate an interest in the good then they are more likely to be invited back and paid for participating in future sessions. If true, and encouraged by the use of purely hypothetical preference questions, this would provide an explanation for many experimental results.

12. In fairness, many CV surveys are explicitly cast as purely hypothetical, although this has long (e.g., Mitchell and Carson 1989) not been a recommended practice.

agency decisions concerning the public good presented in the survey. When survey respondents believe that to be the case their responses become revealed economic behavior. A number of studies have started to examine various issues related to consequential survey questions (e.g., Polomé 2003; Landry and List 2007; Carson, Chilton, and Hutchinson 2009; Nepal, Berrens, and Bohara 2009; Vossler and Evans 2009; and Herriges et al. 2010; Vossler, Doyon, and Rondeau 2012; Vossler and Watson 2013; Mitani and Flores, forthcoming) and have found substantial empirical support for predictions concerning economic behavior that follow from it.[13]

Our theoretical work establishes a key invariance result: the probability that a survey question will be binding should not influence its incentive properties provided that probability is positive. In addition to surveys, there are a wide range of decisions that have a similar incentive structure such as citizen or expert advisory committees making an up or down recommendation to a policy maker. If expected utility holds, our probabilistic influence condition is isomorphic to the "random lottery" approach used in experimental economics under which a subject makes k choices with one of them, randomly chosen, receiving a real payoff (e.g., Charness and Rabin 2002; Holt and Laury 2002; Gneezy, Niederle, and Rustichini 2003; Guth et al. 2007).[14]

2. THEORETICAL FRAMEWORK

The approach we take is to show that moving from an incentive-compatible binding referendum to an advisory referendum or an advisory survey does not change the incentive properties of the basic mechanism.[15] This is done by showing that the degree of influence on the decision does not matter as long as it is positive. The case of zero influence is, however, fundamentally different. Finally, we consider the situation where the vote may influence a second outcome with supplemental benefits.

---

13. Landry and List (2007) and this paper are closely related, having shared Carson et al. (2002) as a common ancestor. Landry and List looks at several issues we do not address, such as "cheap talk," sequencing, and the use of different cost amounts. It has only one treatment with probabilistic influence and, as such, does not test the invariance property of interest in this paper. Their basic result on consequentiality is the same as we find; there is no statistical difference between their probabilistic influence and the real cases. This replication should strengthen confidence in the results presented here (Maniadis, Tufano, and List 2014).

14. Holt (1986) and Karni and Safra (1987) show that, in the absence of expected utility holding, subjects should not necessarily treat different trials as independent even though only one of them will be chosen for a real money payoff.

15. An advisory referendum occurs when the public votes on an issue but the vote is advisory in the sense that the government is not legally bound to implement the outcome of the referendum even if the requisite plurality votes in favor. Such referenda are periodically held in places where the government wants to see if there is public support for an action. A well-known example is that some of the public votes on whether to join the European Union were advisory rather than binding referenda.

Our results can be stated succinctly in the form of a set of propositions that can be seen as more formal versions of those put forward in Caron and Groves (2007). It is relatively straightforward to show that they hold if expected utility is assumed.[16] In appendix A, we show that they also hold under a condition, mixture monotonicity, which states that if two lotteries differ only in that more probability is placed on higher-ranked (by the voter's preference) alternatives in one than in the other, then the first lottery is preferred to the second. This condition is stronger than first-order stochastic dominance but does not require the independence axiom that lies behind expected utility.

The propositions are:

Proposition 1: A binding (binary) referendum vote with a plurality voting rule is incentive compatible in the sense that truthful preference revelation is the dominant strategy when the following additional conditions hold: (*a*) the vote is coercive in that all members of the population will be forced to follow the conditions of the referendum if the requisite plurality favors its passage and (*b*) the vote on the referendum does not influence any other offer that might be made available to the relevant population.[17]

Proposition 2: Changing from a binding referendum to an advisory referendum does not alter the incentive structure as long as the decision maker is more likely to undertake the referendum proposed outcome if the specified plurality favors it. This proposition follows from noting that it is the nature of the influence on the decision (the agent is potentially pivotal at one point in the decision space, the requisite plurality) not the binding nature of the referendum that matters.

---

16. The first five propositions were sketched out in an informal sense without proof in Carson and Groves (2007). For a formal proof using expected utility, see Vossler et al. (2012), who also consider additional conditions that need to hold for a sequence of binary choice questions to have similar properties. It is worth noting here that these propositions do not necessarily hold if we further weaken the mixture monotonicity condition to first-order stochastic dominance or if the choice space contains more than two alternatives.

17. This case is well known in the voting literature (Farquharson 1969), although the two auxiliary conditions are often not emphasized. It is also sometimes invoked to demonstrate that the set of incentive-compatible mechanisms allowed by the Gibbard-Satterthwaite theorem is not the null set. That theorem states that only mechanisms with two response alternatives can be incentive compatible, without restrictions on preferences, but it does not imply the opposite, that all mechanisms with two response alternatives are incentive compatible. The condition that no other offers are influenced by the vote effectively rules out side deals and logrolling between voters.

Proposition 3: It is possible to replace the plurality vote aggregation rule used in propositions 2 and 3 with the more general condition that the referendum's proposed outcome increases in a weakly monotonic fashion with the fraction of the population that favors it. This result follows by noting that any plurality voting rule is a special case of this more general influence function and that propositions 1 and 2 hold for any plurality strictly greater than zero and less than one.

Proposition 4: Allowing the decision maker to consider the results of the advisory referendum with a probability less than one but greater than zero does not alter the incentive properties of the mechanism. This is because the probability that the proposed referendum outcome is implemented is still weakly monotonically increasing with the fraction of the population that favors it.

Proposition 5: Replacing a vote of the entire population with an exogenously chosen random sample of the population does not alter the incentive properties of the mechanism. This is an old result of Green and Laffont (1978) that holds for a large class of possible mechanisms, including the ones considered here.

The sample of the population in proposition 5 in the situation we are interested in can be seen as comprising an advisory survey. For it to be incentive compatible the two auxiliary conditions of proposition 1 need to hold. The auxiliary condition 1a can only be invoked with public goods. It is what makes the incentive properties for public and private goods fundamentally. Proposition 2 shows that the binding nature of our referendum in proposition 1 is not the source of its desirable incentive properties. Proposition 3 says that as long as any member of a larger class of influence functions, that includes plurality voting rules as a special case, holds for agents, then a binary choice meeting the other conditions of proposition 1 provides the same incentives for truthful preference revelation.[18] This is important from the perspective of an advisory survey because, while it is straightforward to tell respondents that the more people in the survey who favor providing the good the more likely it is to be provided, it may be difficult or implausible to state a specific plurality voting rule as different policy makers along a decision path are likely to want

---

18. The likelihood that an agent is pivotal with a plurality voting rule has been discussed at length in the voting literature. An advisory survey is less likely to suffer from an agent believing that they have no chance of being pivotal since a survey's sample will almost always be seen as small relative to the population of voters. (Mitani and Flores [2012] have explored whether varying group sizes between 1 and 45 makes a difference in group voting behavior and found that it did not.) It is also likely that respondents may perceive of the decision rule as being strictly monotonic in some regions, such as the probability of implementation increasing as the fraction above 50% in favor increases.

to see different levels of public support before they think that the new public goods should be supplied.[19] Proposition 4 shows that having positive probability that the vote is not considered does not alter incentive properties as long as there is still a positive probability that it will.[20] Proposition 5 says that the incentive properties of an advisory referendum are not altered by implementing a survey version as long as respondents are exogenously chosen.

The next proposition looks at what happens if the condition in proposition 4, of having positive influence on the decision, is changed to having no influence on the decision that Carson and Groves (2007) defined as an inconsequential case.

> **Proposition 6**: If the probability of influencing the decision becomes zero, then any response has the same influence on the agent's utility, so truthful preference revelation is no longer a dominant strategy.[21] This result follows from the influence function with respect to the decision being flat everywhere so that the agent no longer has any chance of being pivotal.

The difficulty here is that neither a yes response nor a no response has any chance of changing the agent's utility level. Not caring about the outcome can also result in the choice being inconsequential but that is typically less of an issue if one of the choices imposes a monetary cost which is typically the case.

One of the auxiliary assumptions of proposition 1 is violated if the response to the question of interest is perceived as having some chance of influencing another decision.[22]

---

19. Polomé, Van der Veen, and Geurts (2006) compare five provision rules in a binary choice context and show that the three rules consistent with proposition 3 produce statistically indistinguishable fractions in favor while two provision rules, one involving donations and the other being a completely unspecified rule, produce divergent estimates. Vossler and Evans (2009), in an experimental context, look at different provision rules using a binding majority rule vote as the baseline case. They found that two provision rules consistent with subjects having influence on the decision resulted in a statistically equivalent response while a decision rule that insured subject responses were inconsequential and a purely hypothetical treatment were statistically different from the baseline treatment.

20. A variant of proposition 4 can be shown to underlie the random lottery approach where each decision has a positive probability of being implemented. Part *b* of proposition 1 is analogous to ensuring that only one of the choices is given a real payoff.

21. We treat the case where the vote is binding ($p = 1$) as "truth" as it is not clear what truthful preference revelation implies in the $p = 0$ case where any response has the same influence on the agent's utility.

22. There are a number of ways that this can happen. One formal context that has been well analyzed is a school bond funding referendum where a major vote against the funding

Proposition 7: If the possibility of influencing a second outcome exists then the response to a question is generally not incentive compatible with respect to preferences concerning the choice posed. In this case, the optimal response will incorporate the influence on both of the outcomes so that the response may be different than in the case where only the first outcome can be influenced by the vote.

With more specific assumptions about the nature of the second outcome, it is possible to obtain more specific predictions about the nature of the divergence caused by the possibility of influencing that outcome. Proposition 8 looks at the situation where there is the possibility of an unambiguously desired "consolation" prize if the group votes in favor but the vote is not binding.[23] This allows us to address the issue of whether agents always tell the truth even when there is a clear incentive not to do so.

Proposition 8: Let there be a second good (including its costs) that is desired by all agents and let there be a positive probability that this second good will be provided in the event that (*a*) the requisite plurality votes in favor of providing

---

measure constrains the school district either to accept the status quo funding level or to bring back a new funding proposal between that level and the defeated measure. In this case, it may be in the voter's interest to vote against a funding level that is preferred to the status quo but which is in turn less preferred to the level likely to be put forth if the original ballot proposition is defeated. In a CV context, Richer (1995) shows that people expected a different version of the Desert Protection Act to be brought forward if the one currently being considered was defeated. In cases involving donations, the subsequent decision can be whether a (larger) fund-raising drive is mounted. In economic lab experiments, the subsequent decision can be seen as whether the subject is offered the good later or is invited back to participate in future experiments.

23. This proposition was motivated in part by Cummings and Taylor (1998), who found that probabilistic invariance was violated. In their experiment, groups of undergraduate students in Atlanta were asked about paying for information booklets to warn poor Hispanic families about the hazards of groundwater contamination in Albuquerque, New Mexico. We felt that many of the students were unlikely to believe that a nonprofit group in New Mexico was relying solely on a group of Georgia students to provide this good. One plausible belief would be that a yes vote would encourage them to undertake a broader fund-raising effort beyond just their one group. The lower the probability that the group's vote is binding, the lower the cost is of encouraging this larger effort. We do not believe this problem to be specific to Cummings and Taylor's study but rather see it as a generic issue when trying to convert an existing charity into a public good with a coercive payment structure, in which case the natural second good is the ability to free ride. More generally the lack of a clear nexus between subjects and the good offered may prompt consideration of possible supplemental benefits.

the first good but (*b*) the vote turns out to be nonbinding using the random device. Under these conditions the percentage of the sample in favor of provision should be higher than in the case where the vote only influences whether the first good is provided. Further, the divergence between votes without and with the possibility of the second good should decrease as the probability of the vote being binding increases.

A set of testable hypotheses associated with these propositions is as follows, using p to denote the probability that a particular vote is binding. Mechanism 1 is used to refer to those treatments that do not involve the possibility of a second good. Mechanism 2 is used to refer to those treatments that do involve the possibility of a second good.

Hypothesis 1 directly addresses a key implication of the framework—making the subject's influence on the decision stochastic should not alter the optimal response if the mechanism is incentive compatible. It utilizes the proposition 1 result that at $p = 1$ the mechanism is well known to be incentive compatible and tests the empirical validity of proposition 4 by comparing it to the $0 < p < 1$ mechanism 1 treatments.[24] The testable hypothesis is:

H1$_0$: The percentage in favor at $1 > p > 0$ (stochastically binding) is equal to that of $p = 1$ (deterministically binding).[25]

Hypothesis 2 looks at the proposition 6 result that the inconsequential and consequential cases need not produce the same response. The testable hypothesis is:

H2$_0$: The percentage in favor at $p = 0$ (inconsequential case) is equal to that of $p \geq 0$ (consequential case).[26]

Hypothesis 3 with $p > 0$ looks at the proposition 7 result that mechanism 2 may induce a different response than that of mechanism 1. The implementation of mech-

---

24. Cummings et al. (1997) and Cummings and Taylor (1998) also rely upon this result in their analysis.

25. It is possible to formulate this as a one-sided test where, following the results of Cummings and Taylor (1998), the plausible alternative might be to expect the $1 > p > 0$ case to result in a higher percentage in favor than in the $p = 1$ case.

26. This, too, could take the form of a one-sided test. The empirical evidence suggests that the percentage in favor at $p = 0$ is higher (e.g., Cummings et al. 1997). However, we suspect that this result may be quite context dependent, since, as noted earlier, economic theory predicts no particular direction for a divergence, only that there is no reason for the two responses to produce comparable results.

anism 2 under hypothesis 3 is that if a majority votes in favor of providing the good but a random device shows that the vote is not binding, then a second good will be provided. Initially, we make no assumptions about the second good, allowing some people to see it as desirable while others see it as undesirable. Many land use decisions reflect the flavor of this situation, as projects favored by local planning groups often never get higher level approval but competing uses for publicly owned parcels of land emerge once it is put in play. The testable hypothesis is:

$H3_0$: With $p > 0$, the percentage in favor for mechanism 1 is equal to the percentage in favor for mechanism 2 versus $H3_{A1}$ that there is a difference.

Sharper results can be obtained if we make the assumption that the second good is (weakly) desired by all subjects. Hypothesis 3_{A2} is the alternative that the fraction in favor is higher than that of the $p = 1$ case. Underlying it is the simple notion that if one branch of a lottery improves while the other branch stays fixed, then the fraction in favor should increase. Rejecting the null hypothesis that there is no difference in favor of the one-sided alternative would show that our subjects engage in nontruthful preference revelation when there is a clear incentive to do so. Hypothesis 3_{A3} posits a more sophisticated behavior under which, in addition to the 3_{A2} hypothesis that the fraction in favor for $0 < p < 1$ is larger than in the $p = 1$ case, subjects actively trade off probabilities in the two branches. This predicts the fraction in favor should fall as p increases. Table 1 provides a summary of the different hypotheses to be tested.

While our model predicts that the incentive structure of our model is invariant to the value of p conditional on it being greater than zero, it does not say anything

Table 1. Summary of Hypotheses Tested

Hypothesis	Question Addressed	Proposition Tested
1	Does stochastic influence on an incentive compatible binary choice influence the percentage in favor?	Proposition 4
2	Is the percentage in favor the same when the choice is consequential versus inconsequential?	Proposition 6
3 (A1)	Does introduction of influence on a second good alter the choice made?	Proposition 7
3 (A2)	Does introduction of influence on a universally desired good increase the percentage in favor?	Proposition 8
3 (A3)	Does the percentage in favor, when there is influence on a universally desired good, decline as the probability vote is binding increases?	Proposition 8

about the variability of responses. There are two questions with respect to p that are of interest. First, is the variance of the error component decreasing in p (conditional on $p > 0$), which is what one might expect if p influences the level of effort put into making the decision? Second, is the variance of the error component different for the $p = 0$ case versus the $p > 0$ cases? Formal testing of these two hypotheses can be undertaken with the data used to look at hypotheses 1 and 2. These tests take the form of determining whether the error term is heteroscedastic with respect to the value of p for the $p > 0$ for the consequential mechanism 1 observations and an indicator variable for $p = 0$ for all the mechanism 1 observations.

3. EXPERIMENTAL DESIGN

To provide a strict test of our theoretical conjectures, we follow List (2001, 2002) and recruit subjects from a well-functioning marketplace—on the floor of a sports card show in Tucson, Arizona. In the current set of experiments, however, we make use of greater control than what was available in previous field experiments. Naturally, field experiments present a trade-off: they give up some of the controls of a laboratory experiment in exchange for increased realism. Instead of giving up complete control, we use a hybrid approach by recruiting subjects on the floor of a sports card trading show and running the treatments in an adjacent room in the same building. Although this sort of experiment is not common in the literature, it provides a useful middle ground between the tight controls of the laboratory and the vagaries of completely uncontrolled field data.

Each participant's experience typically followed two steps: (1) considering the invitation to participate in an experiment that would take about 30 minutes and (2) actual participation in the experiment. In step 1, the monitor approached potential subjects entering the trading card show and inquired about their interest in participating in an experiment that would take about 30 minutes. If the individual agreed to participate, then the monitor briefly explained that the subject would receive a $10 show-up fee. The monitor explained that at a prespecified time, the subject should enter an adjacent room to take part in the experiment. Directions to the room were provided, and the subject was informed that she would receive instructions for the experiment when she arrived.

Step 2 began when subjects entered the room and signed a consent form in which they acknowledged their voluntary participation in the experiment and agreed to abide by the rules of the experiment. After subjects were comfortably situated in the room, the monitor began the experiment. The exact instruction script read to participants is contained in appendix B.

Participants were told that they would have the opportunity to vote on whether everyone in their group would receive a ticket stub to a prominent Kansas City Royals baseball game where Cal Ripken Jr. broke the world record for the number of consecutive games played. In the baseline "real" treatment, participants were told

that, if more than 50% of the people in the group voted in favor, then all members of the group would get the Cal Ripken Jr. ticket stub via a device referred to as "Mr. Twister" and that all members of the group would pay $10. If 50% or fewer people in the group voted no, then no one would get a ticket stub or pay anything. After the instructions were read aloud, a vote was taken with each subject filling out his or her own decision sheet (see app. C).

In addition to the baseline binding referendum, we ran several "probabilistic referenda" to provide a test of our theoretical predictions from mechanism 1. In these other treatments we set the probability that the referendum would be binding equal to 0%, 20%, 50%, and 80%.[27] In these probabilistic treatments, we followed the above instructions verbatim, except we changed the provisioning rules by adding this excerpt (this example is for the 20% treatment):

Two-Step Referendum Rules:

1: If more than 50% of you vote **YES** on this proposition, then the referendum has **passed**. If the referendum passes, then in Step 2 we will roll a 10-sided die to determine if the referendum is binding. If the referendum does not pass (50% or fewer of you vote **YES**), then no one will pay $10 or receive a Kansas City Royals game ticket stub dated June 14, 1996.

2: Contingent on the referendum passing (more than 50% of you vote YES), I will roll this 10-sided die [roll die on the table]. If I roll a 1 or 2 the referendum will be binding and all of you will pay $10 and Mr. Twister's crank will be turned, providing a Ripken ticket stub to everyone. If I roll a 3–10, the referendum is not binding. In this case, no one pays $10 and Mr. Twister fails to be funded.

In the 50% (80%) treatment, we replaced 1 or 2 with 1–5 (1–8), and 3–10 with 6–10 (9 or 10). Since these rules were necessarily new for many subjects we carefully described the two-stage process using several examples.

In the 0% treatment, we used language in the spirit of the literature (see, e.g., List 2001) that emphasized to the participants that the referendum was purely hypothetical. For example, we used passive language where appropriate: "suppose you have the opportunity to vote on whether 'Mr. Twister' will be funded," and we

---

27. Our design allocated subjects equally between the different treatments with the exception that the baseline treatment received a double allocation to enhance power in comparing it to other treatments.

reminded subjects that "regardless of the vote outcome, no one will pay $10 or receive a Kansas City Royals game ticket stub."[28]

Besides these treatments designed to examine the predictive power of mechanism 1, we also explicitly test mechanism 2's major conjectures. The second good is defined to be the first good (Ripken ticket stub) but is provided for free if a majority of the group votes in favor of the provision and the random device shows the vote not to be binding. This design is rather stark with respect to its incentives for some participants to say "yes" even though that would not be the optimal response under mechanism 1. Using a free version of the first good as the second outcome, coupled with having that outcome occur only in the case the group's vote is not binding, avoids the issue of substitution effects. Provision for free ensures that all participants desire the second good. The mechanism 2 treatments are a direct extension of the mechanism 1 20% and 80% treatments. For example, the mechanism 2 20% treatment is identical to that of the mechanism 1 20% treatment with one important deviation: a roll of 3–10 provides the good for free if the referendum passes. Hence, if 50% or more vote yes, the public good is provided with certainty. The roll of the die determines whether each subject pays for the public good. The language read to subjects was the same as for mechanism 1 except that the second step of the two-step referendum rules shown earlier for the 20% case was replaced with:

> Contingent on the referendum passing (more than 50% of you vote YES), I will roll this 10-sided die [roll die on the table]. If I roll a 1 or 2 the referendum

---

28. We chose 20% as the lowest positive probability. At first blush, it may seem that we should have used a very small probability since this is consistent with how many economists view surveys. However, in reality such manipulation, while theoretically interesting due to well-known difficulties people have with dealing with small probabilities, is of little practical relevance. It is analogous to commissioning a survey and then telling respondents that the results would be completely ignored with very high probability. We are not sure how this could be credibly conveyed in a survey. Such a statement would likely to be seen by respondents as offensive in the sense of wasting their time or it would convey to respondents that something was being hidden from them. Mitani and Flores (2010) do formally test it, using students and induced values, a treatment that was binding with $p = .01$ in addition to treatments with $p = .25$, $.50$, $.75$, and 1.0. Their treatments with $p \geq .25$ all resulted in the correct fraction (.6) of those in the treatment voting yes, while the .01 treatment substantially underestimated with only 42%. Mitani and Flores (2012) ran a later experiment at p .25, $p = .1$, and $p = .01$ and show that the $p = .25$ and $p = .1$ treatments behave in a similar manner while the $p = .01$ treatment does not. They also show that errors given the induced values are concentrated in the $p = .01$ treatment with too many subjects having small positive gains voting no.

will be binding and all of you will pay $10 and Mr. Twister's crank will be turned, providing a Ripken ticket stub to everyone. If I roll a 3–10, **no one pays $10**, but Mr. Twister is <u>still</u> provided (thus each of you will receive a Ripken ticket stub free).

To provide variation over the probability of receiving the good for free, we ran a second treatment that replaces 20% with 80%. Hence, in this second treatment, if the referendum passes, then subjects have a 20% chance of receiving the good for free and an 80% chance of paying for the good.

Before moving to the experimental results, we should mention a few noteworthy aspects of our experimental treatments. First, we are eliciting homegrown rather than induced values and sports memorabilia fans are known to have very heterogeneous tastes. Second, no subjects participated in more than one treatment. Third, we randomized assigned subjects to treatments. In a small number of cases, interested subjects could not make their allocated treatment for personal reasons (e.g., "I have to babysit at 4 p.m.") and were reassigned to times when they could attend.[29] Fourth, we were careful to design a public good ("Mr. Twister") that would share important characteristics of public goods in the field yet remain deliverable within our experiment in the sense of being a publicly provided private good. Fifth, we chose sports memorabilia as the public good output since it has an abstract quality relative to normal marketed goods that are directly consumed. The uniqueness of the deliverable good (of the more than 400 subjects, only one had previously seen the Ripken ticket stub) essentially guaranteed that the subjects were unfamiliar with the good in the sense of having never previously owned or dealt with this piece of baseball memorabilia. Sixth, there was no established price for the good we were offering and it was difficult to obtain elsewhere, both features more typical of public goods than private goods.[30]

---

29. We collected data on subject-specific characteristics (see app. C) to provide for controls for possible overrepresentation in some groups, but their use did not substantively alter results.

30. Unlike baseball trading cards and some other types of sports memorabilia where there are fairly well established prices and availability, ticket stubs for particular games are very thinly traded. At the time of this experiment, we knew of one stub, similar to those we used, that was for sale for $40 by a dealer two thousand miles away in Baltimore. Subsequently, we have seen a similar stub fail to sell on eBay for $5 and in most time periods have not seen any listed for sale. We were fortunate enough to obtain the unique piece of sports memorabilia in quantity because one of the coauthors personally attended the sporting event and collected the stubs from in and around the ballpark.

4. EXPERIMENTAL RESULTS

Voting distributions for the various treatments are reported in table 2. The top panel of table 2 contains the mechanism 1 data, whereas the bottom panel contains the mechanism 2 data summary. For convenience, we label the baseline plurality referendum that is binding with certainty "Real," the probabilistic referenda $P(Z$ %), where Z represents the probability of the YES vote being binding, and we term the mechanism where $Z = 0$ as "inconsequential or purely hypothetical." Labels in the lower panel, $PF(20\%)$ and $PF(80\%)$, represent probabilistic referenda mechanisms where if the referendum passes, then subjects pay \$10 for provision with 20% ($PF(20\%)$) or 80% ($PF(80\%)$) probability; or likewise receive the good for free with 80% ($PF(20\%)$) and 20% ($PF(80\%)$) probability.

Cursory examination of the voting distribution summary in the top panel of table 2 indicates that there are only small differences in voting behavior across the different mechanism 1 consequential treatments and that there is no particular pattern as p changes. However, we do observe a much more sizable difference between the percentage of yes votes between the actual and purely hypothetical referenda. Results of the binding treatment show that 45.8% (44/96) of the total participants voted yes compared to the 60.3% (35/58) that voted yes when the vote was inconse-

Table 2. Experimental Data

| Regime | Treatment | Individual Vote Summary | |
		% Yes	Total Subjects
Binding majority	Real mechanism 1 (100% chance of binding)	45.8	96
Binding majority	$P(80\%)$ mechanism 1 (80% chance of binding)	41.3	46
Binding majority	$P(50\%)$ mechanism 1 (50% chance of binding)	48.1	52
Binding majority	$P(20\%)$ mechanism 1 (20% chance of binding)	44.0	50
Binding majority	Inconsequential (0% chance of binding)	60.3	58
Binding majority with free chance	$PF(80\%)$ mechanism 2 (for passed proposition: 80% chance of payment, 20% chance of free provision)	58.2	55
Binding majority with free chance	$PF(20\%)$ mechanism 2 (for passed proposition: 20% chance of payment, 80% chance of free provision)	71.4	49

quential (0% chance of the vote being binding).[31] The voting pattern in the actual referendum is similar to the voting patterns in the probabilistic referenda: $P(80\%)$: 41% yes votes (19/46); $P(50\%)$: 48% yes votes (25/52); and $P(20\%)$: 44% yes votes (22/50).[32] While these figures are roughly consistent with one another, we find that the percentages of yes votes in the treatments with a possibility of free provision appear to be larger than comparable mechanism 1 proportions: $PF(80\%)$: 58.2% (32/55) and $PF(20\%)$: 71.4% (35/49).

There are several statistical ways to test our hypothesis 1 that there is no difference between the $p = 1$ binding treatment and the three consequential $0 < p < 1$ treatments. Perhaps the most straightforward is with a simple contingency table. The $\chi^2(3)$ statistic for this table's test of independence between the vote [yes/no] and the probability of the vote being binding is 0.4991, which has a p-value of 0.919. Two test statistics that take into account the possibility of a monotonic response to changes in the probability of being binding are the gamma statistic, which is 0.0054 with an asymptotic standard error (ASE) of 0.0984, indicating a very insignificant result, and the Spearman rank order correlation coefficient, which also is very close to zero and not significantly different from it, 0.0035 (ASE = 0.0641).

A test with more power, in the sense of taking into account the specific probability of being binding, is to run a simple probit regression of the probability of a yes response against the probability that the vote was binding ($P(\text{BINDING})$). This model, shown in table 3, produces an estimated coefficient on $P(\text{BINDING})$ which is zero out to six decimal places, a result clearly in accord with our theoretical prediction that the probability of the vote being binding should not matter as long as it was positive.

Having accepted hypothesis 1_0, then it is possible to collapse all of the mechanism 1 $p > 0$ treatments ($N = 302$). The test of hypothesis 2 is that the consequential cases behave differently from the inconsequential case, and it takes on the straightforward form of a 2×2 contingency table of vote by consequential (yes/no). The $\chi^2(1)$ test statistic from this table is 4.374, which has a p-value of .037. An alternative test is to estimate a probit model with a single indicator variable for coming from the inconsequential treatment. Table 4 reports the estimated coefficients and standard errors for this model.

---

31. This 1.31 ratio between our real and hypothetical treatments is consistent with what would have been expected from Murphy et al.'s (2005) meta-analysis suggesting that there is nothing unusual about the difference between the two treatments observed in this study.

32. Using a ticket stub to a less prominent baseball game and a similar setup to the one we use, Landry and List (2007, table 2, p. 425) find there is no statistical difference between their real and probabilistically binding ($p = .5$) treatments. At a $5 cost, 33% were in favor in the real treatment and 32% were in favor in the $p = .5$ treatment. At $10, 19% were in favor in the $p = .1$ treatment, while 20% were in favor in the $p = .5$ treatment.

Table 3. Probit Regression Test of Hypothesis 1

Parameter	Estimate	Standard Error
Intercept	−.1237	.1960
P(BINDING)	.0000	.0026

Table 4. Probit Regression Test of Hypothesis 2

Parameter	Estimate	Standard Error
Intercept	−.1236	.0805
TREATMENT ($P = 0$)	.3858	.1850

The p-value on the two-sided test that the coefficient on the TREATMENT($P =$ 0) indicator variable is zero is 0.037. Thus, we reject hypothesis 2_0 at the .05 level. The one-sided test suggested by the hypothetical bias literature yields a p-value of 0.019.[33]

Next we turn to a test of hypothesis 3 that offering the possibility of getting the good for free induces nontruthful preference revelation. Again, accepting hypothesis 1_0, a simple test of hypothesis 3_0 against hypothesis 3_{A1} comes from a 2×2 contingency table of vote by mechanism [1 or 2 | $p > 0$] ($N = 348$), which yields a $\chi^2(1)$ statistic of 10.941 which has a p-value of 0.001, suggesting a clear rejection of the hypothesis that nontruthful preference revelation cannot be induced.

Hypothesis 3_{A2} can be tested using a simple probit model that regresses the vote on an indicator variable for the observation coming from mechanism 2 rather than mechanism 1 (with $p > 0$). This model is shown in table 5. The one-sided test suggested by hypothesis 3_{A2}, which takes into account the predicted direction of the mechanism 2 inducement, rejects the null hypothesis at $p < .0001$.

We can test hypothesis 3_{A3} by running a probit model that regresses the probability of a yes vote from the $p > 0$ mechanisms 1 and 2 observations ($N = 348$) on an intercept term and variable [BINDING2], which is the probability of being binding for the mechanism 2 observations and zero otherwise. The results of this

---

33. While we did not design this experiment to have power for comparing the $p = 0$ treatment to each of the other mechanism 1 treatments individually, a referee suggested that it would be useful to provide these tests. They can be examined using a one-sided Fisher's exact test where the test's direction is taken from the hypothetical bias literature which suggests that the estimate from the purely hypothetical case is larger. For $p = 0$ versus $p = .2$, $p = .5$, $p = .8$, and $p = 1.0$, the p-values are 0.066, 0.136, 0.041, and 0.057, respectively.

Table 5. Probit Regression Test of Hypothesis 3_0 versus Hypothesis 3_{A2}

Parameter	Estimate	Standard Error
Intercept	−.1236	.0805
Mechanism 2 indicator	.4934	.1495

Table 6. Probit Regression Test of Hypothesis 3_0 versus Hypothesis 3_{A3}

Parameter	Estimate	Standard Error
Intercept	−.0489	.0764
BINDING2	.0046	.0024

Table 7. Alternative Probit Regression Test of Hypothesis 3_0 versus Hypothesis 3_{A3}

Parameter	Estimate	Standard Error
Intercept	−.1838	.1287
$PF(80\%)$	.3904	.2135
$PF(20\%)$	.7498	.2294

estimation are given in table 6, where the p-value on the one-sided test suggested by hypothesis 3_{A3} is .026.

One possible objection to the estimates in tables 4–6 is that they use observations from mechanism 1 for $p = .5$ while there is no comparable mechanism 2 treatment. In addition, observations for the $p = 1$ case exist only in mechanism 1 but not in mechanism 2 since there is no chance to get the good for free. To examine this issue, table 7 provides a probit model using only the 80% and 20% treatments from mechanisms 1 and 2 ($N = 200$) with dummy variables for mechanism 2 ($p = .80$ [$PF(80\%)$] and $p = .20$ [$PF(20\%)$]).

The one-sided tests suggested by hypothesis 3_{A3} are that the coefficient on $PF(80\%)$ should be greater than zero, which yields a p-value of .034; and that the coefficient on $PF(20\%)$ should be greater than 0, which yields a p-value of .001. A one-sided test of whether $PF(20\%) - PF(80\%) > 0$ yields a p-value of .080.

While the results reported in tables 2–7 are consistent with our theoretical expectations, they do not control for respondent characteristics. Doing so should not be necessary due to random assignment of respondents to treatments, and implementing such controls does not substantively change any of our results. The best

predictor variables from the participants' characteristics are income, which is, as expected, positively related to voting "yes" and is highly significant. Being a dealer, the number of years of experience trading sports memorabilia, and age are also typically positively related to the probability of a yes vote, although the effects are much smaller. Gender and education are never significant predictors. We did not find any significant interactions between the treatments and respondent characteristics with the exception of an interaction between income and the $p = 0$ treatment, which suggests that higher income subjects were somewhat less likely to say "yes" than lower income subjects when faced with this treatment.[34]

The literature has also discussed the importance of considering variance with the choices (see, e.g., Haab, Huang, and Whitehead 1999). Exploring the issue of whether there is heteroscedasticity with respect to the value of p in the experimental setup for testing hypotheses 1 and 2 requires the inclusion of covariates for identification. The set of covariates we use here are the significant ones: Age, Dealer, and Income. Other covariates that are not at least marginally significant (i.e., sex, years of schooling, and years of experience with sports memorabilia) are not included.

For hypothesis 1 (mechanism 1, $p > 0$; $N = 244$), one can look at the log likelihoods (LL) from probit models for three specifications: only covariates entered linearly (LL $= -131.772$), covariates plus the probability that the vote is binding (LL $= -131.668$), and the covariates plus the probability that the vote is binding plus allowing the error term using the standard heteroscedastic probit formulation to be a function of the probability that the vote is binding (LL $= -131.638$). This yields three possible likelihood ratio tests. Adding the probability of being binding results in a $\chi^2(1)$ test statistic equal to .2084 (p-value $= .648$), which is consistent with our earlier result on hypothesis 1. Adding the heteroscedastic variance parameter results in a $\chi^2(1)$ test statistic equal to .0600 (p-value $= .806$). The third likelihood ratio statistic considers both the addition of prob(BINDING) to both the location and scale parts of the heteroscedastic probit model. Here we get a $\chi^2(2)$ test statistic equal to .2690 (p-value $= .874$), suggesting that one cannot reject the hypothesis that the prob(BINDING) has no effect on either the mean or the variance of the underlying willingness to pay distribution.

For hypothesis 2 (mechanism 1, $p = 0$ and $p > 0$; $N = 302$), the log-likelihood of the model using only covariates is -175.058, the covariates and a dummy for $p = 0$ model has a LL of -172.838, and adding a heteroscedasticity parameter for the $p =$

---

34. The negative interaction with income appears to be a fairly mechanical functional form correction that occurs because high-income subjects already have a fairly high proclivity to vote for receiving the ticket stub. The interaction is only marginally significant when log income is used instead.

0 indicator yields a LL of −169.274. The first likelihood ratio test here yields a $\chi^2(1)$ test statistic equal to 4.439 (p-value = .0351) consistent with our earlier results on hypothesis 2 that the $p = 0$ treatment results in a different percentage in favor than the $p > 0$ treatments. The second likelihood ratio test looks at letting the error component be potentially heteroscedastic by letting it be a function of an indicator for the $p = 0$ treatment. Doing so yields a $\chi^2(1)$ test statistic equal to 7.127 (p-value = 0.008), suggesting, if anything, that the impact of the $p = 0$ treatment on the variance is even more significant than its influence on the percentage in favor, a finding consistent with Haab et al. (1999). The third likelihood ratio test looks at both the location and scale effect and yields a $\chi^2(2)$ of 11.561 (p-value = .003), suggesting that a more appropriate way to look at hypothesis 2 is that the $p = 0$ treatment appears to influence both parameters of the underlying willingness to pay distribution.[35]

At some level, the lack of a variance effect in the $p > 0$ treatments is not surprising given the straightforward nature of the choice task faced by agents. This result, however, may not carry over to CV studies that require respondents to fill in too many missing details and that do not motivate them to put sufficient effort into the task. The addition of a large increase in the variance for adding the $p = 0$ observations provides another reason for viewing the results obtained under such a condition skeptically.

5. CONCLUDING REMARKS

Our work here provides one explanation for why the existing experimental evidence that suggests a substantial upward hypothetical bias is so at odds with other evidence. These experimental comparisons are typically based on treatments in which subjects are told that they are engaged in a purely hypothetical exercise, a condition Carson and Groves (2007) refer to as inconsequential. From a theoretical perspective, there is no reason for estimates from inconsequential questions to bear any resemblance to those from consequential questions that have long characterized state-of-the-art CV surveys. The "real" treatments from studies summarized in these meta-analyses often use private goods, voluntary donations, and complex incentive compatible mechanisms, all of which are problematic if these "real" estimates are the "truth" that survey estimates strive to hit. We find the meta-analyses of hypothetical bias interesting from an academic standpoint and, even more so, as they add addi-

---

35. If one assumes that the $p = 0$ indicator should be used first in modeling the error component, then the $\chi^2(1)$ test statistic for the addition of that indicator to the regression component is 2.960 (p-value = .084), emphasizing the importance of the influence of the purely hypothetical treatment on the error component.

tional studies and code for additional factors (e.g., Little, Broadbent, and Berrens 2012) since they can shed light on the properties of different mechanisms. However, we do not believe that they are anywhere near the point where much credence should be given to statements such as CV studies using a particular approach usually over (or under) estimates by an average of X %.[36]

The results from our experiment provide strong support for two key predictions from neoclassical mechanism design theory concerning the incentive structure of survey questions. First, for an incentive-compatible consequential single binary choice question, the probability that the responses are taken into account in making a decision does not matter as long as it is bounded away from zero. Second, respondents will take advantage of transparent incentives that encourage misrepresentation. Further, in tests comparing consequential treatments ($p > 0$) to the inconsequential treatment ($p = 0$), we find that a different response is obtained at $p = 0$, in terms of both the mean and variance of the response. This suggests that results obtained for the inconsequential purely hypothetical case should not be used to make inferences about how the standard consequential $p > 0$ case behaves. One way of summarizing our results is that people report truthfully when it is in their interest to do so, that they do not do so when it is in their interest not to, and that respondents reporting when there are no incentives can diverge from situations where there are incentives for truthful preference revelation.

Our results suggest that understanding how to ensure consequentiality in CV surveys should be a major focus for survey designers. This message is in concert with the long-standing advice (e.g., Mitchell and Carson 1989) that emphasizes the need for realism in the design of such surveys. Consequentiality will not in general be as easy as telling respondents that the survey's results may influence some vague policy, and it brings on a set of difficult challenges.[37] When survey responses rep-

---

36. A further issue with the meta-analyses that look at hypothetical bias is that a large fraction of the estimates used in those studies are taken from experiments using students. While lab experiments with students are useful for generating qualitative insights, they are not well suited for providing quantitative estimates of key parameters in real world settings (Levitt and List 2007).

37. In some studies, consequentiality can be invoked by using an official government sponsor. In others, university researchers will need to explore the effectiveness of different ways of making their survey instruments consequential from the perspective of respondents. When the consequentiality of the survey is in doubt, it may be useful to try to identify those respondents who did not believe the exercise was consequential. These attempts may suffer from some of the same endogeneity issues that plague certainty scales, so experimentation with different question formats and wording may be useful. It may also be worth looking at other measures that point toward potential root sources of seeing the survey as inconsequential, such as identifying respondents who face a payment vehicle that for them is not coercive.

resent real economic commitments, respondents care about program details.[38] Consequentiality and realism need to permeate all parts of a valuation survey. Respondents instinctively know when a survey instrument is a well-thought-out presentation of a substantive policy decision that seriously asks them for their input.

APPENDIX A
THEORETICAL MODEL

We consider n agents/voters who are asked to vote (formally in a referendum election or, perhaps, informally in a survey of households) for or against a specific collective project. A *project* might typically consist of a fixed amount of a public good (e.g., a park, a dam, a highway, etc.) and a specified cost to each voter, but it is sufficient to treat a project completely generally and denote it by the generic notation α. An agent's *preferences*, $\succeq_i$, are defined over alternative projects α and a status quo, z, alternative, which is the default outcome if the project under consideration is not adopted. Although generally a project will involve a certain quantity of a public good and, hence, be desirable to any voter, a project would typically also come with a cost to the voter and, as such, depending on the desirability of the public good and the amount of the cost, a voter may be better off, worse off, or indifferent with any given project α than at the status quo z, that is, preferences may be either $\alpha \succ_i z$ or $z \succeq_i \alpha$ or both.[39]

An agent/voter is asked to vote either for or against a specific project—α. We consider rules for deciding on adopting projects that depend on the number (or frequency) of affirmative votes for the project under consideration. The family of such rules is *Ballot Rules* and may be defined in the following manner: Let $m_i = 0$ denote a vote by i against the proposed project and $m_i = 1$ denote a vote by i in favor of the project. Define $\mathbf{m} = \{m_1, \ldots m_n\}$ to be the *full vote* of all n agents and define

---

Further, once a survey instrument becomes consequential for most respondents, those seeing the survey as inconsequential may well have lower values than those who did not, as a recent study by Vossler and Watson (2013) found.

38. Substantial underestimation may occur in consequential CV studies because a nontrivial fraction of the public thinks that the government is unlikely to deliver a promised program (and, hence, discount the probability of provision from the certainty level the survey designer would like) and a nontrivial fraction of the public believes that government programs suffer from cost overruns that the public ends up having to pay for (Carson et al. 2004). Strong and Flores (2007) and Mitani and Flores (forthcoming) provide an interesting initial look at this issue.

39. The weak preference order $\succeq$ is defined in the usual manner as encompassing both the strong preference order $\succ_i$ and the indifference order $\sim_i$.

$r = r(\mathbf{m}) = \sum_i m_i$ to be the total vote in favor of the project—$(n - r$ is then the vote against the project).

A Ballot Rule $\rho : \{1, \ldots, n\} \to [0, 1]$ is a rule that specifies the *probability* the project is undertaken given the vote, r, in favor of the project. A voting rule is *deterministic* if and only if $\rho(r) = 0$; or 1; is *monotonic* if and only if $\rho(r) \leq \rho(r')$ whenever $r < r'$, and is *consequential* if and only if it is monotonic and also satisfies $\rho(0) < \rho(n)$. An *inconsequential* Ballot Rule is thus one that always ignores every vote.

Since a ballot rule defines a probability of the project being adopted for every full vote $\mathbf{m}$, the outcomes facing a voter at the time he or she is asked to express an opinion on a project are *lotteries* over the two alternatives: the project a under consideration and the status quo z outcome. Casting his or her vote potentially will affect the probability of getting project a, and, hence, the voter must weigh the effect his or her vote has on the resulting lotteries. We assume only that the voter's preferences over project lotteries satisfy a *mixture monotonicity* property, which states that if two lotteries differ only in that more probability is placed on higher ranked (by the voter's preference $\succsim_i$) alternatives in one than in the other, then the first lottery is preferred to the second.[40]

Given a ballot rule, a voter's preferences over project lotteries (which we also denote $\succsim_i$) will induce preferences over all full votes $m_{\sim}$ as follows: $\mathbf{m} \succsim_i \mathbf{m}'$ if and only if

$$L(\mathbf{m}) = \{[\rho(r(\mathbf{m})), a]; [(1 - \rho(r(\mathbf{m}))), z]\} \succsim L(\mathbf{m}') = \{[\rho(r(\mathbf{m}')), a]; [(1 - \rho(r(\mathbf{m}'))), z]\}.$$

Thus, given any particular Ballot Rule, an n-person game is defined in which $I = \{1, \ldots, n\}$ is the set of players, $M_i = \{0, 1\}$ is the strategy space available to the ith player (i.e., i can only vote for or against the project under consideration), and $\succsim_i$ over full votes $\mathbf{m}$ is the ith player's preferences over joint strategies (full votes).

Given the game defined by a particular Ballot Rule, and the project a under consideration, it is easy to characterize the induced preferences over joint strategies (full votes) under the mixture monotonicity property assumed to hold for the agent's preferences over project lotteries. Simply stated, full vote $\mathbf{m}$ is weakly preferred to full vote $\mathbf{m}'$ if and only if the number of votes in favor of the project is higher for $\mathbf{m}$ than for $\mathbf{m}'$ whenever the voter prefers project a to the status quo z, and vice versa when z is preferred to a. It is now easy to see that a voter would

---

40. The property that preferences over lotteries satisfy the mixture monotonicity property is stronger than the property that they satisfy first-order stochastic (vector) dominance but is weaker than the independence axiom required for expected utility. This property ensures that cumulative prospect theory conforms to first-order stochastic dominance that was perceived as a defect with the initial version of prospect theory (Kahneman and Tversky 1979).

always vote in favor of the project a whenever $a \succ_i z$ under any consequential ballot rule, since voting in favor when project a is preferred to the status quo z could never lower the likelihood that the project would be adopted and would in some situations strictly increase the probability that the favored project is adopted, and thus make the voter better off.

We summarize in the following theorem:

Theorem: If voters' preferences satisfy *mixture monotonicity* then, for any consequential Ballot Rule considered for deciding on a public project, it is the *unique* weakly dominant strategy for every voter to vote in favor of the project if $a \succ_i z$ and to vote against the project if $z \succ_i a$, that is, to vote his or her true preferences.

Proof: Let a denote the public project and z denote the status quo

Case 1: $a \succ_i z$.

For every $\mathbf{m}$, $r(m_\sim/1) \geq r(m_\sim/0)$ and for some $\mathbf{m}'$, $r(\mathbf{m}'/1) > r(\mathbf{m}/0)$.

Thus, $\rho(r(\mathbf{m}/1)) \geq \rho(r(\mathbf{m}/0))$ and $\rho(r(\mathbf{m}'/1)) > \rho(r(\mathbf{m}'/0))$, where $\mathbf{m}'/m_i = (m'_{i1}, \ldots, m'_{i-1}, m_i, m'_{i+1}, \ldots, m'_n)$, and, by mixture monotonicity, lottery $\{[\rho(r(\mathbf{m}/1)), a]; [(1 - \rho(r(\mathbf{m}/1))), z]\}$ is preferred to lottery $\{[\rho(r(\mathbf{m}/0)), a]; [(1 - \rho(r(\mathbf{m}/10))), z]\}$ for all $\mathbf{m}$ and strictly preferred for some $\mathbf{m}'$.

Hence, voter i prefers joint strategy $\mathbf{m}/1$ to joint strategy $\mathbf{m}/0$ for every $\mathbf{m}$ and strictly prefers joint strategy $\mathbf{m}'/1$ to joint strategy $\mathbf{m}'/0$ for some $\mathbf{m}'$. Thus, player i's strategy to vote in favor of the project, $m_i = 1$, is at least as good as the alternative strategy of voting against the project, $m_i = 0$, for every possible set of votes by all other voters and strictly better for at least some other set of votes by all the others. Thus, $m_i = 1$ is the unique weakly dominant strategy for voter i.

Case 2: $z \succ_i a$. By an identical argument, it is a weakly dominant strategy to vote against the project in this case. QED

Corollary 1: Both strategies (voting for and voting against) are dominant for the voting game of the theorem when $a \sim_i z$.

Proof: Immediate since the voter does not care if the project is adopted or not and hence both strategies have the same effect (namely, none) on his or her utility.

Corollary 2: Both strategies (voting for and voting against) are dominant for the voting game defined by an inconsequential Ballot Rule.

Proof: Immediate since voting for or against the project will never have an effect on the outcome, and, thus, the induced preferences over strategies derived from

the preferences over projects is the trivial preference defined by the property that all joint strategies (full votes) are indifferent.

Deterministic Ballot Rules and Advisory Referenda/Surveys

Consider a public decision process consisting of two stages. First, a survey or referendum is employed to determine the support for a public project—as measured by the number of votes in favor of its adoption. However, the actual decision to implement the project depends on some other agency, for example, a legislature or an administrative body. In other words, the survey or referendum is merely advisory to the actual decision maker. We may suppose that the stronger the majority is in favor of the project the more likely it will be that the project will be adopted by the ultimate decision maker.

If the decision maker always follows the outcome of the advisory referendum or survey and adopts the project if it has a majority (or some other specific plurality cut-off number), then the resulting ballot rule would be deterministic. Otherwise, the decision process of first conducting an advisory referendum/survey and then deciding on whether or not to adopt the project is just a monotonic ballot rule that is consequential, unless the decision body completely ignores the referendum/survey.

For procedures based on advisory referenda/surveys of this type, the theorem establishes that the voter/respondent's unique dominant strategy is always to vote his or her true preference on the project, no matter what the precise weight is that the decision maker places on the outcome of the referendum/survey—as long as the decision maker is responsive to some degree to the referendum/survey's outcome.

Supplemental Benefits: The Model with a "Backup" or Consolation Alternative

We now explore the situation in which a vote (whether expressed in a referendum or merely on an advisory survey) for or against the project under consideration may also influence the adoption of an alternative project, called here the *consolation* project, if the main project is not adopted. Such a situation may arise if a particular project that has strong support is nonetheless not adopted but the decision maker then chooses to adopt a "similar" but less costly project. Or, perhaps another agency or entity decides to adopt an unimplemented project that had expressions of wide support in a referendum or survey. If a is the *main* project, let b denote the *consolation* project.

Formally, we consider a choice process consisting of two parts. First is an advisory referendum (or survey) for or against the main project a. Each agent registers a vote, m_i, for or against the project. A Ballot Rule, $\rho_a(\cdot)$, then determines the probability project a is adopted, $\rho_a(r(\mathbf{m}))$, based on the total number of affirmative votes, $r(\mathbf{m})$. The uncertainty is resolved and project a is either adopted or rejected. If the project a is rejected, the second stage is reached, and another Ballot Rule, $\rho_b(\cdot)$, is applied to determine if the consolation project b is to be adopted ac-

cording to the probability $\rho_b(r(\mathbf{m}))$, which is also based on the total number of affirmative votes originally registered by all the voters for project a. The uncertainty is then resolved and project b is either adopted or rejected, leaving the status quo, z, intact.

Under such procedures, the (a priori) alternatives facing agents at the time of their vote are lotteries involving three outcomes: (1) the main project a, (2) the consolation project b, and (3) the status quo z. Let $L = [(p_a, a), (p_b, b), (1 - p_a - p_b, z)]$ denote such a lottery and $\mathbf{L}$ the class of all such lotteries.

For lotteries defined by the two-stage voting process and Ballot Rules, ρ_a and ρ_b, if $\rho_a(r)$ is the probability the main project is chosen, given the total number r votes in favor, and $\rho_b(r)$ is the probability the consolation outcome is chosen, given that the main project was rejected, then $p_a = \rho_a(r)$ and $p_b = [1 - \rho_a(r)] \cdot \rho_b(r)$.

We assume that agents' preferences satisfy the mixture monotonicity property.[41] Thus, for example, if an agent prefers the consolation project b to the main project a and both to the status quo alternative z, then raising the probability of the consolation project in any lottery and decreasing (weakly) the probability of the main project a and the status quo alternative z will improve the lottery for the agent.

Given the Ballot Rules that determine the probabilities of the main and consolation projects based on the number of affirmative votes for the main project, an agent's induced preferences over the full votes of all agents is easily stated: full vote $\mathbf{m}$ is weakly preferred to full vote $\mathbf{m}'$ if and only if full vote $\mathbf{m}$ defines a lottery that is weakly preferred to the lottery defined by the full vote $\mathbf{m}'$. We consider only monotonic Ballot Rules, and thus, both rules $\rho_a(r)$ and $\rho_b(r)$ are weakly increasing in the number of affirmative votes, r, cast.

While a single vote in favor of the pair of projects would weakly increase the probability of getting one or the other project, it is not necessarily an optimal decision for an agent to vote in favor of the pair even when he or she prefers either project to the status quo alternative which occurs if both projects are rejected. Also, even if an agent prefers the status quo to both projects, it nonetheless could be optimal to vote in favor of the project pair. In considering his or her optimal vote, an agent must make the calculation of which vote, for or against, leads to the most preferred lottery.

The critical comparison that determines an agent's optimal vote, given the votes of the other agents, $\mathbf{m}_{-i}$, is between the two lotteries that are defined for each of the agent's possible votes, $m_i = 1$ and $m_i = 0$, or for and against the project. These two lotteries are:

---

41. As introduced earlier, preferences satisfy the mixture monotonicity property if an agent prefers any lottery A to another lottery B if lottery A differs from lottery B only in that lottery A places greater probability on higher ranked (by the agent's utility) alternatives than lottery B.

$$L(\mathbf{m}/1) = \{[\rho_a(r(\mathbf{m}/1)), a]; [[1 - \rho_a(r(\mathbf{m}/1))] \cdot \rho_b(r(\mathbf{m}/1)), b];$$
$$[1 - \rho_a(r(\mathbf{m}/1)) - [1 - \rho_a(r(\mathbf{m}/1))] \cdot \rho_b(r(\mathbf{m}/1))], z\}$$

and

$$L(\mathbf{m}/0) = \{[\rho_a(r(\mathbf{m}/0)), a]; [[1 - \rho_a(r(\mathbf{m}/0))] \cdot \rho_b(r(\mathbf{m}/0)), b];$$
$$[1 - \rho_a(r(\mathbf{m}/0)) - [1 - \rho_a(r(\mathbf{m}/0))] \cdot \rho_b(r(\mathbf{m}/0))], z\}.$$

For the basic model with only one project under consideration (the main project *a*), we showed that it is a dominant strategy of an agent to vote either for the project or against it, irrespective of the probability $p_a = \rho_a(r)$, depending on whether the agent preferred the project *a* to the status quo *z*. In the situation under consideration here, where a consolation project may be adopted in case the main project is rejected, it is no longer necessarily optimal for an agent to vote for the project even if he or she prefers both the main and consolation projects to the status quo, nor is it necessarily optimal for an agent to vote against the project if he or she prefers the status quo to both projects. This can be easily seen by comparing the lotteries, $L(\mathbf{m}/1)$ and $L(\mathbf{m}/0)$ in the circumstance in which, given the votes of all the other agents, $\mathbf{m}_{-i}$, agent *i* is pivotal for the main project *a* and the consolation project *b* will be undertaken for certain if the main project is rejected. That is, when

$$\rho_a(r(\mathbf{m}/1)) = 1 \textbf{ but } \rho_a(r(\mathbf{m}/0)) = 0 \textbf{ and } \rho_b(r(\mathbf{m}/1)) = \rho_b(r(\mathbf{m}/0)) = 1.$$

In this case lottery $L(\mathbf{m}/1) = \{(1, a); (0, b); (0, z)\}$ and $L(\mathbf{m}/0) = \{(0, a); (1, b); (0, z)\}$. It is easy to see that agent *i* will vote for the main project *a* if and only if he prefers the main project to the consolation project, $a \succ_i b$, even if both are worse than the status quo.[42] The situation where there is an implicit consolation prize can be important in experiments, surveys, and market transactions because it can influence revealed behavior.

A Special Case and Some Experimental Testable Hypotheses

A particular special case of interest, because it can be used to generate testable hypotheses, is defined by ballot rules based on majority voting. Specifically, let $I_M(r)$ denote the *indicator* function of total votes *r* when the fraction of votes r/n exceeds 1/2. Thus,

---

42. In this particular case, the status quo alternative is no longer available since the consolation project will be adopted for certain if the main project is rejected.

$$I_M(r) = \begin{cases} 1, & \text{if } r \geq \dfrac{n}{2}; \\ 0, & \text{otherwise.} \end{cases}$$

The function $I_M(r)$ is just the standard *majority rule* criterion for adoption of a project. The ballot rules of the special case under consideration are defined to be the product of the majority rule indicator function and a fixed probability, p. Thus,

$$\rho_a(r) = p_a \cdot I_M(r) \quad \text{and} \quad \rho_b(r) = p_b \cdot I_M(r).$$

For notational ease, define the function $I_M(\mathbf{m}) = I_M(r(\mathbf{m}))$, which also is the majority indicator function of the vector of all n agents' votes (i.e., the full vote) instead of just their sum. Given the ballot rules and votes of all other agents, $\mathbf{m}_{-i}$; agent i's optimal choice is to vote in favor (i.e., set $m_i = 1$) if and only if

$$L(\mathbf{m}/1) = \{[\,p_a \cdot I_M(\mathbf{m}/1)), a]; [(1 - p_a) \cdot p_b \cdot I_M(\mathbf{m}/1)), b]; [1 - [\,p_a + (1 - p_a) \cdot p_b]I_M(\mathbf{m}/1))], z\}$$

$$\succ_i$$

$$L(\mathbf{m}/0) = \{[\,p_a \cdot I_M(\mathbf{m}/0)), a]; [(1 - p_a) \cdot p_b \cdot I_M(\mathbf{m}/0)), b]; [1 - [\,p_a + (1 - p_a) \cdot p_b]I_M(\mathbf{m}/0))], z\}.$$

Note that either $L(\mathbf{m}/1) \succ_i L(\mathbf{m}/0)$ or $L(m_\sim/0) \succ_i L(m_\sim/1)$, so that it would make a difference how agent i votes, only when agent i is pivotal, that is, $I_M(\mathbf{m}/1) = 1$ and $I_M(\mathbf{m}/0) = 0$, in which case,

$$L(\mathbf{m}/1) = \{(p_a, a); ((1 - p_a) \cdot p_b, b); [1 - p_a - (1 - p_a) \cdot p_b], z\}$$

and

$$L(\mathbf{m}/0) = \{(0, a); (0, b); (1, z)\} = \{1, z\}.$$

Thus, we can conclude the following.

Results

(1): In the case in which both projects a and b are preferred to the status quo z, $L(\mathbf{m}/1) \succ_i L(\mathbf{m}/0)$ for all (nonzero) p_a and p_b, and thus, it is a dominant strategy for agent i to always vote in favor of the project.

(2): In the case in which the status quo z is preferred to both projects a and b, $L(\mathbf{m}/0) \succ_i L(\mathbf{m}/1)$ for all (nonzero) p_a and p_b, and thus, it is a dominant strategy for agent i to always vote against the project.

But,

(3): In the case in which project *a* is preferred to project *b*, $L(\mathbf{m}/1) \succsim_i L(\mathbf{m}/0)$ only for p_a sufficiently close to unity (1); otherwise, $L(\mathbf{m}/0) \succsim_i L(\mathbf{m}/1)$. Thus, for any (sub)population of agents of this type, the fraction voting in favor of the project would increase as p_a goes from (close to) zero to (close to) unity.

Also, similarly,

(4): In the case in which project *b* is preferred to project *a*, $L(\mathbf{m}/1) \succsim_i L(\mathbf{m}/0)$ only for p_a sufficiently close to zero (0); otherwise, $L(\mathbf{m}/0) \succsim_i L(\mathbf{m}/1)$. Thus, for any (sub)population of agents of this type, the fraction voting in favor of the project would decrease as p_a goes from (close to) zero to (close to) unity.

Based on result (4), it is possible to construct an experiment to examine if the fraction of agents voting in favor of a project decreases as the probability p_a increases and that it will be adopted if favored by a majority of the voters, as predicted by result (4). Specifying the main project *a* to be sufficiently costly for a group of voters will ensure that they would prefer the status quo to the project and, in a referendum involving only project *a*, would always vote against it, no matter how small p_a. But, offering them the public good of the main project *a* as a consolation project at, say, no cost, ensures that all these voters would prefer the consolation project *b* to the status quo. Then, fixing the probability p_b of the consolation project being adopted, should the main project be rejected, and varying the probability p_a of the main project, would (according to result [4]) show an increasing fraction of voters accepting the proposal (i.e., voting in favor) as the probability p_a decreases from (close to) unity to (close to) zero.

APPENDIX B
EXPERIMENTAL INSTRUCTIONS FOR BASELINE PLURALITY REFERENDUM

Hello! Thanks for attending the experiment. You should have all been given the $10 show-up fee—did everyone receive $10 from the monitor when they entered the room? In this experiment it is very important that you understand the rules, so if you have any questions please do not hesitate to raise your hand and a monitor will come by and answer your question(s). It is important that when we begin <u>no one</u> talks to anyone but a monitor. Are we ready?

Welcome to Lister's Referendum. Today you have the opportunity to vote on whether "Mr. Twister," this small metal box, will be "funded." If "Mr. Twister" is funded, I will turn the handle and N [the amount of people in the room] Kansas

City Royals game ticket stubs dated June 14, 1996, which were issued for admission to the baseball game in which Cal Ripken Jr. broke the world record for consecutive games played, will be distributed—one to each participant [illustrate]. To fund Mr. Twister, **all** of you will have to pay $10. Below please find the proposition and referendum rules.

Proposition: Everyone in the room will contribute $10 to the fund. The contribution will be used for the purpose of funding Mr. Twister, a mechanism that if funded will distribute one Kansas City Royals game ticket stub dated June 14, 1996, to each participant [illustrate].

Referendum Rules:

• If more than 50% of you vote **YES** on this proposition, **all** of you will pay $10. In return, "Mr. Twister" will be funded and I will crank the handle, providing one Kansas City Royals game ticket stub dated June 14, 1996, to each participant [illustrate].
• If 50% or fewer of you vote **YES** on this proposition, **no one** will pay $10 and "Mr. Twister" will **not be funded**. Hence, no one will receive a Kansas City Royals game ticket stub dated June 14, 1996.

Are there any questions? Please turn over to your decision sheet.

[After the instructions were read aloud and all questions answered, the vote to fund Mr. Twister was taken. Each subject filled out his or her decision sheet.]

APPENDIX C
DECISION SHEET

My vote for the proposition is (please circle one response):

YES   **NO**

Please sign the line below to verify your vote.

Signature_____

We now want to ask you a few more questions. These questions will be used for statistical purposes only. THIS INFORMATION WILL BE KEPT STRICTLY CONFIDENTIAL AND WILL BE DESTROYED UPON COMPLETION OF THE STUDY.

1. How long have you been involved with sportscards and memorabilia? _____yrs

2. If you are a dealer, how long have you been an active dealer? _____yrs

3. Gender: 1) Male   2) Female

4. Age _____   Date of Birth _____

5. What is the highest grade of education that you have completed. (Circle one)

    1) Eighth grade   3) 2-Year College   5) 4-Year College

    2) High School   4) Other Post-High School   6) Graduate School Education

6. What is your approximate yearly income from all sources, before taxes?

    1) Less than $10,000   5) $40,000 to $49,999

    2) $10,000 to $19,999   6) $50,000 to $74,999

    3) $20,000 to $29,999   7) $75,000 to $99,999

    4) $30,000 to $39,999   8) $100,000 or over

REFERENCES

Arrow, Kenneth J., Robert M. Solow, Paul R. Portney, Edward E. Learner, Roy Radner, and Howard Schuman. 1993. Report of the NOAA Panel on contingent valuation. *Federal Register* 58:4601–14.

Becker, Gordon M., Morris H. DeGroot, and Joseph Marschak. 1964. Measuring utility by a single response sequential method. *Behavioral Science* 9:226–32.

Carson, Katherine S., Susan M. Chilton, and W. George Hutchinson. 2009. Necessary conditions for demand revelation in double referenda. *Journal of Environmental Economics and Management* 57:219–25.

Carson, Richard T., Michael B. Conaway, W. Michael Hanemann, Jon A. Krosnick, Robert C. Mitchell, and Stanley Presser. 2004. *Valuing oil spill prevention: A case study of California's central coast.* Dordrecht: Kluwer.

Carson, Richard T., Nicholas E. Flores, Kerry Martin, and Jennifer L. Wright. 1996. Contingent valuation and revealed preference methodologies: Comparing the estimates for quasi-public goods. *Land Economics* 72:80–99.

Carson, Richard T., and Theodore Groves. 2007. Incentive and informational properties preference questions. *Environmental and Resource Economics* 37:181–210.

Carson, Richard T., Theodore Groves, John A. List, and Mark J. Machina. 2002. Probabilistic influence and supplemental benefits: A field test of the two key assumptions underlying stated preferences. Unpublished manuscript, Department of Economics, University of California, San Diego.

Carson, Richard T., and Jordan J. Louviere. 2011. A common nomenclature for stated Preference Elicitation Approaches. *Environmental and Resource Economics* 49:539–59.

Charness, Gary B., and Matthew Rabin. 2002. Understanding social preferences with simple tests. *Quarterly Journal of Economics* 117:817–69.

Cummings, Ronald G., Steven Elliot, and James Murphy. 1997. Are hypothetical referenda incentive compatible? *Journal of Political Economy* 105:609–21.

Cummings, Ronald G., Glenn W. Harrison, and E. Elisabeth Rutström. 1995. Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive-compatible? *American Economic Review* 85:260–66.

Cummings, Ronald G., and Laura Osborne Taylor. 1998. Does realism matter in contingent valuation surveys? *Land Economics* 74:203–15.

Drichoutis, Andreas C., Rodolfo M. Nayga, and Panagiotis Lazaridis. 2011. The role of training in experimental auctions. *American Journal of Agricultural Economics* 93:521–27.

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

Flores, Nicholas E., and Aaron Strong. 2007. Cost credibility and the stated preference analysis of public goods. *Resource and Energy Economics* 29:195–205.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. Performance in competitive environments: Gender differences. *Quarterly Journal of Economics* 118:1049–74.

Green, Jerry R., and Jean-Jacques Laffont. 1978. A sampling approach to the free rider problem. In *Essays in public economics*, ed. Agnar Sandmo, 95–114. Lexington, MA: Lexington Books.

Guth, Werner, M. Vittoria Levati, Matthias Sutter, and Eline van der Heijden. 2007. Leading by example with and without exclusion power in voluntary contributions experiments. *Journal of Public Economics* 91:1023–42.

Haab, Timothy C., Ju-Chin Huang, and John C. Whitehead. 1999. Are hypothetical referenda incentive compatible? A comment. *Journal of Political Economy* 107:186–96.

Harrison, Glenn W. 2006. Experimental evidence on alternative environmental valuation techniques. *Environmental and Resource Economics* 34:125–62.

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

Holt, Charles A. 1986. Preference reversals and the independence axiom. *American Economic Review* 76:508–15.

Holt, Charles A., and Susan K. Laury. 2002. Risk aversion and incentive effects. *American Economic Review* 92:1644–55.

Horowitz, John K. 2006. The Becker-DeGroot-Marschak mechanism is not necessarily incentive compatible, even for non-random goods. *Economics Letters* 93:6–11.

Johnston, Robert J. 2006. Is hypothetical bias universal? Validating contingent valuation responses using a binding public referendum. *Journal of Environmental Economics and Management* 52:469–81.

Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47:263–92.

Karni, Edi, and Zvi Safra. 1987. Preference reversal and the observability of preferences by experimental methods. *Econometrica* 55:675–85.

Kochi, Ikuho, Bryan Hubbell, and Randall Kramer. 2006. An empirical Bayes approach to combining and comparing estimates of the value of a statistical life for environmental policy analysis. *Environmental and Resource Economics* 34:385–406.

Kuhn, Thomas S. 1962. *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Landry, Craig E., and John A. List. 2007. Using *ex ante* approaches to obtain credible signals for value in contingent markets: Evidence from the field. *American Journal of Agricultural Economics* 89:420–29.

Levitt, Steven D., and John A. List. 2007. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives* 21:153–74.

List, John A. 2001. Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards. *American Economic Review* 91:1498–1507.

———. 2002. Preference reversals of a different kind: The more is less phenomenon. *American Economics Review* 92:1636–43.

List, John A., and Craig A. Gallet. 2001. What experimental protocol influence disparities between hypothetical and actual values. *Environmental and Resource Economics* 20:241–54.

Little, Joseph M., and Robert P. Berrens. 2004. Explaining disparities between actual and hypothetical stated values: Further investigations using meta-analysis. *Economics Bulletin* 3:1–13.

Little, Joseph M., Craig D. Broadbent, and Robert P. Berrens. 2012. Meta-analysis of the probability of disparity between actual and hypothetical valuation responses: Extension and preliminary new results. *Western Economic Forum* 11:1–12.

Loomis, John B. 2011. What's to know about hypothetical bias in stated preference studies. *Journal of Economic Surveys* 25:363–70.

Maniadis, Zacharias, Fabio Tufano, and John A. List. 2014. One swallow doesn't make a summer: New evidence on anchoring effects. *American Economic Review* 104:277–90.

Mason, Charles F., Jason F. Shogren, Chad Settle, and John A. List. 2005. Investigating risky choices over losses using experimental data. *Journal of Risk and Uncertainty* 31:187–215.

Mitani, Yohei, and Nicholas E. Flores. 2009. Demand revelation, hypothetical bias, and threshold public goods provision. *Environmental and Resource Economics* 44:231–43.

———. 2010. Public goods referenda without perfectly correlated prices and quantities. Paper presented at the Allied Social Sciences Meeting, Atlanta.

———. 2012. Robustness tests of incentive compatible referenda: Consequential probability, group size, and value-cost difference. Unpublished manuscript, Norwegian University of Life Sciences.

———. Forthcoming. Hypothetical bias reconsidered: Payment and provision uncertainties in a threshold provision mechanism. *Environmental and Resource Economics*.

Mitchell, Robert M., and Richard T. Carson. 1989. *Using surveys to value public goods: The contingent valuation method*. Washington, DC: Resources for the Future.

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* 30:313–25.

Nepal, Mani, Robert P. Berrens, and Alok K. Bohara. 2009. Assessing perceived consequentiality: Evidence from a contingent valuation survey on global climate change. *International Journal of Ecological Economics and Statistics* 14:14–29.

Noussair, Charles, Stephane Robin, and Bernard Ruffieux. 2004. Revealing consumers' willingness-to-pay: A comparison of the BDM mechanism and the Vickrey auction. *Journal of Economic Psychology* 25:725–41.

Poe, Gregory L., and Christian A. Vossler. 2011. Consequentiality and contingent values: An emerging paradigm. In *The international handbook on non-market environmental valuation*, ed. Jeffery W. Bennett, 122–41. Cheltenham: Edward Elgar.

Polomé, Philippe. 2003. Experimental evidence on deliberate misrepresentation in referendum contingent valuation. *Journal of Economic Behavior and Organization* 52:387–401.

Polomé, Philippe, Anne Van der Veen, and Peter Geurts. 2006. Is referendum the same as dichotomous choice contingent valuation? *Land Economics* 82:174–88.

Quiggin, John. 1992. *Generalized expected utility theory: The rank dependent model*. Dordrecht: Kluwer.

Richer, Jerrell R. 1995. Willingness to pay for desert protection. *Contemporary Economic Policy* 13:93–104.

Roberts, John. 1976. The incentives for correct revelation of preferences and the number of consumers. *Journal of Public Economics* 6:359–74.

Rosenberger, Randall S., and John B. Loomis. 2000. Using meta-analysis for benefit transfer: In-sample convergent validity tests of an outdoor recreation database. *Water Resources Research* 36:1097–1107.

Shaw, W. Douglass, and Richard T. Woodward. 2008. Why environmental and resource economists should care about non-expected utility models. *Resource and Energy Economics* 30:66–89.

Starmer, Chris. 2000. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature* 38:332–82.

Tversky, Amos, and Daniel Kahneman 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty* 5:297–323.

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

Vossler, Christian A., and Mary Evans. 2009. Bridging the gap between the field and the lab: Environmental goods, policy maker input and consequentiality. *Journal of Environmental Economics and Management* 58:338–45.

Vossler, Christian A., Joe Kerkvliet, Stephen Polasky, and Olesya Gainutdinova. 2003. Externally validating contingent valuation: An open-space survey and referendum in Corvallis, Oregon. *Journal of Economic Behavior and Organization* 51:261–77.

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* 86:137–47.

Whittington, Dale. 2010. What have we learned from 20 years of stated preference research in less-developed countries. *Annual Review of Resource Economics* 2:209–36.