Probabilistic Influence and Supplemental Benefits
A Field Test of the Two Key Assumptions Underlying Stated Preferences

Richard Carson
University of California, San Diego

Theodore Groves
University of California, San Diego

John List
University of Maryland

Mark Machina
University of California, San Diego

Draft
February 2004

Abstract

Researchers using stated preference methods typically hold one of two beliefs about the preference information they obtain: respondents always answer questions truthfully or they answer truthfully only when it is in their interest to do so. The second position is consistent with economic theory, but it implies that the interpretation of survey responses depends critically on the incentive structure provided and the success in conveying that incentive structure. We derive the simplest tests capable of distinguishing between the two views. The empirical part of the paper uses a binary discrete choice framework in the context of a vote on a public good—whether a unique baseball memorabilia item should be provided to all or no members of a group at a specified fixed individual price. The tests are cast in terms of the probability of influencing the outcome and the success in creating a true take-it-or-leave-it offer. Empirical results are consistent with the predictions from the theoretical model.

I. Introduction

There is a continuing debate about whether contingent valuation (CV) over or under estimates willingness to pay (WTP) for a good (Carson, Flores, and Meade, 2001). Much of it is driven by conflicting empirical results. For example, take two frequently cited papers. Champ, *et al.,* (1997) find using a split sample approach that respondents say they are willing to contribute more than they actual do when given the opportunity, while Carson *et al.* (1994), in a very large meta-analysis comparing estimates from CV surveys to estimates based upon revealed behavior find CV estimates to be on average somewhat lower than their revealed preference counterparts.[1] These divergent results do not tell us much though, unless the underlying hypothesis is the naïve one that does seem to permeate much of the debate: respondents in a CV survey either always truthfully reveal their preferences irrespective of the incentives they face or CV doesn't work.

Carson, Groves and Machina (2000) have recently put forth a comprehensive neoclassical framework based upon the proposition that respondents should as economic agents behave strategically.[2] Starting from that premise, they derive a rich framework that predicts that in some cases that stated and revealed preference information should be equivalent while in other cases they should be substantially different. For instance, in the Champ *et al.* (1997) setup the Carson, Groves, and Machina framework predicts that the stated preference estimates should be higher than true WTP due to the possibility that strong support in the survey might encourage an actual fundraising effort to be

[1] For the private goods case see Cummings, Harrison, and Rutstrom (1995), Dickie, Gerking, and Fisher (1987) and Neill *et al.* (1994).

[2] Much of the CV literature has taken the position that strategic behavior is a "bad" thing. The Carson, Groves, and Machina framework posits the opposite: incentives for strategic behavior are needed in order to be able to interpret responses in an economic context. The thrust of that paper is to understand what the strategic incentives are, when they correspond with truthful preference revelation, and the direction of the incentive effect when truthful preference revelation is not a dominant strategy for the respondent.

undertaken, while the revealed preference estimates should be lower than true WTP due to possible gains from free riding to the actual fundraising effort. Carson, Groves, and Machina also show that the incentive properties of CV questions differ considerably by whether or not the question is "consequential", by the elicitation format used, by the nature of aggregation rule used, and by whether a coercive payment vehicle is used.[3]

The Carson, Groves, and Machina framework can be useful to help sort out the existing literature on how different types of CV questions perform in a variety of different circumstances. At the heart of that framework are two key propositions. The first is that economic theory has nothing to say about inconsequential or in the jargon of some papers, purely hypothetical questions. The second is that the probability that a CV question is consequential does not influence its incentive properties as long as that probability is positive. This result is in stark contrast to an assertion sometimes made in the CV literature that respondents have an incentive to tell the truth in the inconsequential case. Indeed, because all possible responses to the survey question have the same influence on the respondent's utility level, any response is as good as any other.

Fortunately, most CV surveys are consequential in the sense of Carson, Groves, and Machina in that they potentially can influence policy choices by decision makers, so that observing a difference in behavior between respondent answers when the CV questions are consequential and inconsequential should not be a source of concern unless one holds the view (which clearly clashes with standard economic theory but is popular

[3] One implication of the Carson, Groves, and Machina framework is that private goods typically represent a worse case situation for evaluation how survey responses compare to actual behavior. Indeed, from a strategic standpoint, the incentive structure is very similar to that of voluntary contributions to a public good. As such it is not surprising that experimental papers like Neill et al. (1994) and Cummings, Harrison, and Rutstrom (1995) come to the same (theoretically predicted) conclusion as the voluntary contribution papers (*e.g.,* Seip and Strand, 1992) that surveys tend overestimate demand.

among psychologists) that in general respondents truthfully reveal their preferences in surveys and experiments even if they are inconsequential.[4]

In contrast, finding that the response to a question varies with probability that it influences a policy outcome would be very troubling in a neoclassical economic framework because it is possible to derive, as we will show, a strong invariance result in the case that the CV question is incentive compatible. The only formal test (Cummings and Taylor, 1998) of this proposition in the literature suggests that this invariance principle is violated. We show, however, the invariance result depends crucially on the question being asked being incentive compatible at all the positive probability levels. In many ways this derivation is not surprising since Carson, Groves and Machina argue that three auxiliary conditions are needed in addition to a binary discrete choice question to achieve incentive compatibility: a weakly monotonic influence function in terms of the percent in favor (a condition meet by any plurality voting rule), a coercive payment mechanism, and the good will not be provided in any other way. The first two conditions are met in the Cummings and Taylor (1998) study, although it is not clear that the third, which can be termed the take-it-or-leave-it condition, is.[5]

In this paper, we look at the situation where it is often thought that CV should work the best, at least in a theoretical sense. This situation involves a pure public good with provision based upon a single binary discrete choice question with a clear plurality voting rule as the aggregation mechanism coupled with a coercive payment if the good is

---

[4] The purely hypothetical case has been formally tested against the revealed behavior case by Cummings *et al.* (1997) who reject that the two produced comparable estimates.

[5] The good used in that study was the provision of information booklets on an environmental hazard to poor people in New Mexico by groups recruited for the experiment in Atlanta. Participants in these experiments may have perceived the possibility that either other groups taking part in the experiment could pay for the informational booklets or that there were other avenues of provision if the particular group elected not to provide them. It is also possible that participants in that experiment perceived that provision through some other avenue like mounting an Atlanta fundraising drive was more likely if a “yes” response was provided.

provided, and with no other way to provide the good. We first develop the theoretical framework for three special cases: (a) the probability that the vote is binding is one, the probability that the vote being binding is less than one but greater than zero, and (c) the probability that the vote is binding is zero. Cases (a) and (b) are shown to be equivalent from an incentive perspective, while (c) does not have comparable properties. We then look at what happens when it is possible to obtain the good in some other way and that way is potentially influenced by the response to the original voting opportunity. Here we show that the probability of a "yes" vote is dependent upon the probability that the vote is binding. Under reasonable conditions, it can further be shown that the percent in favor monotonically declines as the probability of the vote being binding increases.

After laying out the theoretical framework and hypotheses associated with it, we experimentally test those hypotheses using groups of people recruited at a well functioning sports memorabilia market. The group gets to vote on whether to purchase a ticket stub for the Kansas City Royal's game where Cal Ripken Jr. broke the world record for consecutive games played. We turn these ticket stubs into a public good by having the group vote under a majority rule provision at a fixed cost. The outcome of the vote is that each individual in the group get either one unit of the good (the ticket stub) or no member of the group gets the good. The Ripken ticket stubs are not generally available in sports memorabilia markets so exclusion of everyone in the group from the good is clearly possible. In some treatments we deliberately induce incentives for non-truthful preference revelation by providing an alternative means to obtain the Ripken ticket stub that is tied to the response to the referendum vote.

II. Theoretical Framework

The basic theoretical model for an incentive compatible binary discrete choice survey question follows that described in Carson, Groves, and Machina. These results can be stated succinctly in the form of five propositions:

Proposition 1: A binding (binary) referendum vote with a plurality aggregation 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 than might be made available to the relevant population.[6]

Proposition 2: changing from a binding referendum to an advisory referendum doesn't change 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, with only an binary weight on the in the aggregation rule) not the binding nature of the referendum that matters.

Proposition 3: replacing the plurality vote aggregation rule with the more general condition that the referendum proposed outcome is weakly monotonically increasing with the fraction of the population that favors it. This follows from noting that the result in Propositions 1 and 2 holds 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 a 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.

---

[6] Note that this case is well known in the voting literature (Farquharson, 1969). It is also often 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 but does not imply the opposite, that all mechanisms with two response alternatives are not incentive compatible. The condition that no other offers are influenced by the vote effectively rules side-deals and logrolling between the voters.

Propositions 2 through 5 can be combined to yield the result that an "advisory" survey which has the potential to influence a decision maker in the sense that a particular outcome is weakly increasing in the percent in favor from the survey should be incentive compatible if the other auxiliary conditions of Proposition 1 are met.

The next proposition describes what happens if the condition in Proposition 4 of having positive influence on the decision is changed to having no possible influence on the decision, which defined as the inconsequential case in Carson, Groves, and Machina.

Proposition 6: if the probability of influencing the decision becomes zero, then any response has the same influence on the agent's utility so that truthful preference revelation is no longer a dominant strategy.[7] 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. Note that in this case, neither a yes response or a no response has any chance of changing the agent's utility level.

The auxiliary assumptions of Proposition 1 are most likely to be violated in a survey context when the survey has some chance of influencing another decision.[8]

Proposition 7: if there exists the possibility of influencing a second outcome then the response to question is not generally 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.

---

[7] The one philosophical objection that could be put forth here is that in the case of zero influence on the decision, it is not clearly that truthful preference revelation is always a well-defined concept. Treating the Mechanism 1 case, p=1 case as "truth" provides a "practical" way to avoid this issue.

[8] There are any 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 measure constrains the school district to accept either the status quo funding level or 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.

With more specific assumptions about the nature of the second outcome, it is possible to get more specific predictions about the nature of the divergence caused by the possibility of influencing that outcome.

Proposition 8: if a second outcome (a good including its costs) is desired by all agents, the vote on the first good (positively) influences the probability the second good is provided and the plurality used by the decisionmaker is the same with respect to both the first and second goods, then (a) the percent in favor will be higher when it is possible to influence provision of the second good and (b) the divergence between the responses will fall as the probability of the vote being binding increases.[9]

A set of testable hypotheses associated with these propositions is as follows using p as the probability that a particular vote is binding. Mechanism 1 to refer to those treatments that do not involve a second good, and Mechanism 2 to refer to those treatments that involve the possibility of a second good. Hypothesis H1 directly addresses a key implication of the Carson, Groves, and Machina framework—making the influence function stochastic should not alter the optimal response if the mechanism is incentive compatible. This hypothesis 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 comparison to the $0 < p < 1$ Mechanism 1 treatments.[10]

$H1_0$: The percent in favor at $1 > p > 0$ (stochastically binding) is equal to that of p=1 (deterministically binding).

$H1_A$: The percent in favor with $0 < p < 1$ is not equal to that of p=1.[11]

---

[9] The second good can, of course, be the first good a different price (possibly zero). It is also possible to show that a version of Proposition 7 holds in the situation where the shape of the (weakly monotonic) influence functions for the two goods is different if one is prepared to impose some reasonable restrictions on preferences, the relative nature of the goods, and the nature of the probabilistic influence on the decision.

[10] Cummings, *et al*. (1997) and Cummings and Taylor (1998) also rely upon this result in their analysis.

[11] 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 that to expect the $1 > p > 0$ case to result in a higher percent in favor than in the p=1 case.

Hypothesis 2 looks at the Proposition 6 result that the inconsequential and consequential cases need not produce the same response.

> $H2_0$: the percent in favor at p=0 (inconsequential/purely hypothetical case) is equal to that of p > 1 (consequential case).
>
> $H2_A$: The percent in favor at p=0 is not equal to the percent in favor at p=1.[12]

Hypothesis 3 looks at the Proposition 7 result that the deliberate introduction of the possibility of a second good (mechanism 2) may induce a different response that of mechanism 1 with p > 0. The $H3_{A2}$ and $H3_{A3}$ versions of the alternative hypothesis represent increasing degrees of specificity with respect to the theoretical prediction in Proposition 8.

> $H3_0$: The percent in favor for Mechanism 1 with p > 0 is equal to the percent in favor for Mechanism 2 with 0 < p < 1.
>
> $H3_{A1}$: The percent in favor for Mechanism 2 is different than that using Mechanism 1.
>
> $H3_{A2}$: The percent in favor for Mechanism 2 is greater than using Mechanism 1.
>
> $H3_{A3}$: The percent in favor for mechanism 2 is greater than that using Mechanism 1 and this percentage is decreasing as p for Mechanism 2 increases.

To summarize, these three hypotheses, H1, H2, and H3, test whether introduction of stochastically influencing the decision alter the response relative to the base binding binary discrete choice cases where the auxiliary conditions for incentive compatibility are met, whether moving from the consequential to inconsequential purely hypothetical case makes a difference, and whether one can deliberately induce non-truthful preference

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

revelation. Finding that the degree of influence did not matter as long as there was influence would provide support for the neoclassical model. Finding that the move from consequential to inconsequential makes a difference and that non-truthful preference revelation can be rejected would demonstrate that respondents do not always tell the truth from two different directions and again would provide support for the neoclassical model. The opposite results would provide evidence against the neoclassical model.

While the prediction that the incentive structure of the neoclassical model is invariant to the value of p conditional on its being greater than zero, it is possible to put forth a model which postulates some type of random component where the variance of that term is decreasing in p because the value of p influences the level of effort put into making the decision. Formal testing of this model in terms of the experiment underlying H1 and H2 take the form of a test of whether there is heteroskedasticity in the error term relative to the value of p.

III. 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 sportscard show in Tucson, AZ in April 2001. 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 tradeoff: 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 sportscard 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 does provide 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 pre-specified time (1pm, 2pm, 3pm, 4pm Saturday or 1pm, 2pm, 3pm Sunday) 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. Excerpts of the instructions for the benchmark referendum read as follows (Appendix A contains the instructions):

> 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.

After the instructions were read aloud, a vote to fund the public good, Mr. Twister, was taken. Each subject filled out his/her own decision sheet, where a sample decision sheet is provided in Appendix B.

In addition to this baseline plurality referendum that represents a mechanism that is binding with certainty, we ran several "probabilistic referenda" to provide strict tests of our theoretical predictions from Mechanism 1. In these other treatments, we set the probability that the referendum will be binding equal to 0%, 20%, 50%, and 80%. 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 subjunctive language, emphasizing to the participants that the referendum was pure hypothetical, or inconsequential using the more precise terminology of Carson, Groves and Machina.

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 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 only occur in the case the group's vote is not binding avoids the issue of any substitution or complementarity effects and provision for free ensures that all participants desire the second good.[13] The Mechanism 2 treatments are direct extension of the Mechanism 1 20% and 80% treatments. For example, the Mechanism 2 20% treatment is identical to

[13] This is what happens for instance in a survey involving a voluntary payment. If the survey can influence whether an actual fundraising drive is undertaken, it may be optimal for the respondent to say "yes" to the survey and then say "no" to the actual fundraising drive in order to engage in classic free riding behavior. A similar situation can occur with private goods whereby a "yes" response makes it more likely that the good will be provide so that a respondent with any positive probability of wanting to purchase the good in the future may find it optimal to state they would purchase the good.

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 following excerpt represents the two-step referendum rules for this 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, **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 run 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, no subjects participated in more than one treatment. Second, we randomized subjects into treatments to assure an equal representation across referenda. But, in certain cases interested subjects could not make their allocated treatment for personal reasons (*e.g.,* "I have to babysit at 4pm"). To provide a control for an overrepresentation of certain groups in treatments, we also collected data on subject specific characteristics (see Appendix B). Third, we were careful to design a public good ("Mr. Twister") that would share some of the important characteristics of public goods in the field, yet remain deliverable within our experiment. Fourth, we choose sports memorabilia as the public good output since it has an abstract quality beyond the normal market good. Furthermore, the uniqueness of the deliverable (of the more than 400 subjects, only 1 had previously seen the Ripken ticket stub) essentially guaranteed that the subject had not previously dealt with the piece of memorabilia. 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. Finally, although we have rarely seen the Ripken ticket stub sold in markets, there are reports that the good has sold for more than $40 in a market in the Baltimore region.

IV. Experimental Results

Voting distributions for the various treatments are reported in Table 1. The top panel of Table 1 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 as "Actual", the probabilistic referenda as 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.

Table 1: Experimental Data

Regime	Treatment	Individual Vote Summary %Yes	 Total Subjects
Binding Majority	Actual (100% chance of binding)	45.8	96
Binding Majority	P(80%) (80% chance of binding)	41.3	46
Binding Majority	P(50%) (50% chance of binding)	48.1	52
Binding Majority	P(20%) (20% chance of binding)	44.0	50
Binding Majority	Inconsequential (0% chance of binding)	60.3	58
Binding Majority With Free Chance	PF(80%) (for passed proposition: 80% chance of payment, 20% chance of free provision)	58.2	55
Binding Majority With Free Chance	PF(20%) (for passed proposition: 20% chance of payment, 80% chance of free provision)	71.4	49

A cursory examination of the voting distribution summary in the top panel indicates that there are little differences in voting behavior across treatments. But, we do observe a deviation in 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 inconsequential (0% chance of the vote being binding). 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). 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 first hypothesis H1. Perhaps the most straightforward is with a simple contingency table, which is displayed in Table 2. The chi-square (df=3) statistic for the 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 .9191. Two test statistics, which take into account that one might expect to see a monotonic response with respect to the change 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 close very close to zero and not significantly different from it, 0.0035 (ASE=0.0641).

Table 2: Vote by Probability of Being Binding

Frequency/ Column %					
VOTE	P(20%)	P(50%)	P(80%)	P(100%)	Total
No	28 56.0%	27 51.92%	27 58.70%	52 54.17%	134
Yes	22 44.00%	25 48.08%	19 41.30%	44 45.83%	110
Total	50	52	46	96	244

A test with more power, in the sense of taking into account the specific probability of being binding is to run a simple logistic regression equation of the probability of a yes response against the probability that the vote was binding [P(BINDING)]. This model shown in Table 3 produces a coefficient on P(BINDING) which is zero out to four decimal places, a result clearly in accord with the neoclassical theoretical prediction.

Table 3: Test of Hypothesis H1

Parameter	Estimate	Standard Error
Intercept	-0.1974	0.31340
P(BINDING)	0.0000	0.00413

If we accept the $H1_0$, then it is possible to collapse all of the Mechanism 1 $p > 0$ treatments, so that the test of H2 takes on a straight forward form where the 2 x 2 contingency table of vote by consequential [yes/no] which yields a chi-square(1) test statistic of 4.374 which has a p-value=.0365. An alternative test is to estimate a logistic regression model with a single dummy variable for coming from the inconsequential treatment are almost identical. Table 4 reports that logistic regression model:

Table 4: Test of Hypothesis H2

Parameter	Estimate	Standard Error
Intercept	-0.1974	0.1287
TREATMENT(P=0)	0.6172	0.2977

The p-value on a two-sided test that the coefficient on the TREATMENT(P=0) dummy variable is for this model is 0.0381. Thus, we reject $H2_0$ at the .05 level. A one-sided test against TREATMENT(P=0) > 0 would have resulted in an even stronger rejection.

Next, we turn to a test of H3. Again, accepting $H1_0$, a simple test of $H3_0$ against $H3_{A1}$ comes from a 2 x 2 contingency table of vote by Mechanism [1 or 2 | p > 0] which gives yields a chi-square(df=1) statistic of 10.941 which has a p-value .0010. An alternative test comes from estimating a simple logistic model that regresses vote on a dummy variable for the observation coming from Mechanism 2 rather than Mechanism 1 (with p > 0). The results of this logistic regression are provided in Table 5.

Table 5: Test of $H3_0$ Versus $H3_{A1}$

Parameter	Estimate	Standard Error
Intercept	-0.1974	0.1287
MECHANISM 2 Dummy	0.7911	0.2419

Here the p-value on the two-sided test on the Mechanism 2 coefficient suggested by $H3_{A1}$ is .0011, which is highly significant. The one-sided test suggested by $H3_{A2}$ is even more significant.

We can test $H3_{A3}$ by running a simple logistic model that regresses the probability of a yes vote from the p > 0 Mechanism 1 and 2 observations are an intercept term and for the Mechanism 2 observations, the probability of being binding [BINDING2]. The results of this estimation are given in Table 6:

Table 6: Test of $H3_0$ Versus $H3_{A3}$

Parameter	Estimate	Standard Error
Intercept	-0.0782	0.1218
BINDING2	0.0074	0.0038

where the p-value on the one-sided test suggested by $H3_{A3}$ is .0262.

One possible objection to the estimates in Table 4 though 6 is that it uses information from Mechanism 1 for p=.5 while there is no comparable Mechanism 2 treatment and use information from the p=1 case where Mechanism 1 and 2 should coincide since there is no chance to get the good for free. A logistic regression model using only the 80% and 20% treatments from Mechanisms 1 and 2 with dummy variables for Mechanism 2 p=.80 [PF(80%)] and p=.20 [PF(20%)] is given in Table 7.

Table 7: Alternative Test of $H3_0$ Versus $H3_{A3}$

Parameter	Estimate	Standard Error
Intercept	-0.2938	0.2063
PF(80%)	0.6240	0.3425
PF(20%)	1.2101	0.3776

The one-sided tests suggested by $H3_{A3}$ are that PF(80%) > 0 for which the model yields a p-value of .0343, PF(20%) > 0 for which the model yields a p-value of .001, and that PF(20%) - PF(80%) > 0 which yields a p-value of .0805.

While the results reported in Tables 2 through 7 are consistent with our theoretical expectations if the neoclassical model is applicable, they do not control for respondent characteristics. While doing so should not be necessary due to random assignment of respondents to treatments, implementing such controls does not substantively change any of the results. The best and most consistent predictor variable from the participants characteristics is income, which is, as expected, positively related to voting "yes", and 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 a small effect involving income and the p=0 treatment that suggested higher income subjects were somewhat less likely to say "yes" than lower income subjects when faced with this treatment.

V. Effect on Variance

Looking at the issue of whether there is heteroscedasticity with respect to the value of p in the experimental set up for testing H1 and H2 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 H1 (mechanism 1, $p > 0$), one can look at three the log likelihood from a probit model for three specifications: only covariates entered linearly (LL= -131.77199), covariates plus the probability that the vote is binding (LL=-131.66752), 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=-131.6375). This yields three possible likelihood ratio tests. Adding the probability of being binding results in a $\chi(1)$ test statistic equal to .2084 (p-value=.648), which is consistent with our earlier result on H1. Adding the heteroscedastic variance parameter results in a $\chi(1)$ 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 $\chi(2)$ 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 H2 (mechanism 1, p=0 & p > 0), the LL of the covariates only model is –175.0549, the covariates and a dummy for p=0 model has a LL of -172.8787, and adding a heteroskedasticity parameter for the p=0 dummy yields a LL of -169.2742. The first likelihood ratio test here yields a $\chi(1)$ test statistic equal to 4.439 (p-value=.0351) consistent with our earlier results on H2 that the p=0 treatment results in a different percent in favor than the p > 0 treatments. The second likelihood ratio test looks at adding the heteroskedasticity term and yields a $\chi(1)$ 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 the effect on the actual percent in favor, a finding consistent with Haab, Huang and Whitehead (1999). The third likelihood ratio test looks at both the location and scale effect and yields a $\chi(2)$ of 11.561 (p-value=.003) suggesting that a more appropriate way to look at the H2 hypothesis is the p=0 treatment appears to influence both parameters of the underlying willingness to pay distribution.

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 poorly designed contingent valuation studies which require agents to fill in too many missing details and do not motivate them to put sufficient effort into the task. The addition of a large increase in the variance for the p=0 case, provides another reason for viewing the results obtained under such a condition skeptically.

VI. Concluding Remarks

The experimental results presented here strongly support the standard neoclassical economic framework as it has been applied to CV surveys in the case of incentive compatible binary discrete choice questions. The key underlying assumption of the Carson, Groves, and Machina framework, that the probability of influencing the decision does not matter, as long as that probability is not zero, is clearly accepted.

In contrast, the competing hypothesis that respondents always tell the truth is rejected in two different tests. The first of these involves comparing the case of the probability of the referendum being binding being zero versus that probability being positive. Here 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 p=0, inconsequential/purely hypothetical case, should not be used make inferences about how CV works in the standard consequential p > 0 case.

The second involves deliberately providing incentives for non-truthful preference revelation. Here we find that the response in this case is different from the incentive compatible case in the theoretically predicted manner. This suggests that agents will take advantage of a clear incentive for preference misrepresentation. Comparing our results to those of Cummings and Taylor (1998), we find that their results closely resemble our Mechanism 2 (incentive compatible does not hold) case. This suggests that it may be difficult to create the conditions necessary for an incentive compatible CV question in some situations, particular if one does not have a coercive tax payment vehicle available.

References

Carson, Richard T., Nicholas E. Flores, and Norman F. Meade (2001), "Contingent Valuation: Controversies and Evidence," *Environmental and Resource Economics* 19(2), 173-210.

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(1), 80-99.

Carson, Richard T., Theodore Groves, and Mark J. Machina (2000), "Incentive and Informational Properties Preference Questions," Paper presented at the Kobe Conference on Theory and Application of Environmental Valuation. Kobe: Kobe University, January.

Champ, Patricia A., Richard C. Bishop, Thomas C. Brown, and Daniel W. McCollum (1997), "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods," *Journal of Environmental Economics and Management* 33(2), 151-162.

Cummings, Ronald G. Glenn W. Harrison, and E. Elisabeth Rutstrom (1995), "Homegrown Values and Hypothetical Surveys: Is the Dichotomous Choice Approach Incentive-Compatible?," *American Economic Review*, 85(1): 260-266.

Cummings, Ronald G., Steven Elliot, Glenn W. Harrison, and James Murphy (1997), "Are Hypothetical Referenda Incentive Compatible?" *Journal of Political Economy*, 105(3): 609-621.

Cummings, Ronald G. and Laura Osborne Taylor (1998), "Does Realism Matter in Contingent Valuation Surveys," *Land Economics*, 74(2), 203-215.

Cummings, Ronald and Laura Osborne Taylor (1999), "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review*, 89(3), 649-65.

Dickie, Mark T., Ann N. Fisher, and Shelby Gerking (1987), "Market Transactions and Hypothetical Demand Data: A Comparative Study," *Journal of the American Statistical Association* 82(397), 69-75.

Farquharson, R. (1969), *Theory of Voting.* (New Haven: Yale University Press).

Haab, Timothy C., Ju-Chin Huang and John C. Whitehead (1999), "Are Hypothetical Referenda Incentive Compatible? A Comment," *Journal of Political Economy* 107, 186-196.

Green, J.R. and J.J. Laffont (1978), "A Sampling Approach to the Free Rider Problem," in Agnar Sandmo, ed., *Essays in Public Economics* (Lexington, MA: Lexington Books).

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.

List, John A. (2002) “Preference Reversals of a Different Kind: The More is Less Phenomenon," *American Economic Review*, 92, 1636-1643.

Neill, Helen R., Ronald G. Cummings, Philip T. Ganderton, Glenn W. Harrison, and Thomas McGuckin (1994), “Hypothetical Surveys and Real Economic Commitments,” *Land Economics* 70(2), 145-154.

Richer, Jerrell R. “Willingness to Pay for Desert Protection,” *Contemporary Economic Policy* 13, no. 4 (1995), pp. 93-104.

Seip, K. and J. Strand (1992), ‘Willingness to Pay for Environmental Goods in Norway: A CV Study with Real Payment’, *Environmental and Resource Economics* **2**, 91-106.

Appendix A 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 no one 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 the 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 B 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