Environmental & Resource Economics (2006) 35:137–168
DOI 10.1007/s10640-006-9011-5

Induced-Value Tests of Contingent Valuation Elicitation Mechanisms

CHRISTIAN A. VOSSLER[1,*] and MICHAEL MCKEE[2]
[1]*Department of Economics, University of Tennessee, 534 Stokely Management Center, Knoxville, TN, 37996-0550, USA;* [2]*Department of Economics, University of Calgary, Calgary, Alberta, T2N 1N4, Canada; *Author for correspondence (e-mail: cvossler@utk.edu)*

Accepted 20 June 2006

Abstract. Using an induced-value experimental design that varies whether values for a ''good'' are certain or uncertain and whether payment is real or hypothetical, this study investigates issues of demand revelation, hypothetical bias, and value uncertainty for four elicitation mechanisms used in contingent valuation surveys: dichotomous choice, dichotomous choice with follow-up certainty question, payment card, and multiple-bounded discrete choice. For all elicitation mechanisms, we find no evidence of hypothetical bias: voting decisions do not vary systematically when payment is hypothetical versus when it is real. Under all design conditions we find the fewest deviations between stated and induced values and the strongest evidence of demand revelation with dichotomous choice. Stated uncertainty in dichotomous choice follow-up and multiple-bounded discrete choice questions does correlate with uncertain induced values, but the signal is noisy. We discuss the implications of our findings for the design of contingent valuation surveys.

Key words: contingent valuation, demand revelation, elicitation effects, experiments, hypothetical bias, value uncertainty, willingness to pay

JEL classification: C91, Q51

1. Introduction

Contingent valuation (CV) remains a leading method for eliciting values for public goods and externalities for use in benefit-cost analyses and natural resource damage litigations. This survey-based method asks people to hypothetically make a tradeoff between a change to a good and money. While there is no incentive to reveal true values in a hypothetical setting, there have been theoretical arguments advanced that there is no incentive to lie about one's value in some elicitation settings. Ultimately, whether true values are reported is an empirical question and numerous studies have compared parallel decisions under hypothetical and real payment conditions. While the evidence is mixed, and there are uncontrolled factors that may be driving some results, meta-analyses of studies using field and laboratory data suggest that there exists a positive ''hypothetical bias'', whereby people tend to overstate their values in hypothetical settings (List and Gallett 2001; Little and Berrens 2004; Murphy et al. 2005).[1] The accumulated evidence is that applications that frame the valuation exercise in terms of willingness to pay (WTP), and that use close-ended elicitation formats such as dichotomous choice (DC) referenda, result in the smallest discrepancies between hypothetical and real WTP estimates. However, the underlying causes of hypothetical bias are not yet fully understood and empirical evidence to date has not produced a concrete set of guidelines for mitigating hypothetical bias. This paper offers a fresh perspective on hypothetical bias through the use of induced-value laboratory experiments that compare real and hypothetical payment decisions, across a variety of elicitation formats, with and without value uncertainty.

Consider the question posed to a typical CV survey respondent in a field setting. The individual is asked to respond to a valuation question for a good that is unfamiliar or at least not commonly traded in the marketplace. The individual must first form (construct) her own valuation for the good and next report this value in response to the question posed in the survey. The usual term for the first task is the value formation and the second task the value elicitation. In such a setting, there are two sources of hypothetical bias. The individual may overstate the value formed since the class of goods used in most CV studies is generally socially desirable. As List et al. (2004) show, individuals in such settings may take cues from the behavior of others in their social networks. This may lead to overstating values in hypothetical settings since such behavior would amount to ''cheap talk'' and there is no penalty incentive to tell the truth here.[2] It is this latter source of hypothetical bias that we seek to address through the use of induced-value laboratory experiments. With induced values there is little opportunity for social networks to affect an individual's value since the value formation phase has been eliminated. Thus, the setting allows us to focus on the value elicitation, and the effect of different forms of this elicitation in a setting in which we control for such factors as social interaction. The extent to which our results can inform the elicitation mechanism choice for field surveys will depend on the degree of interaction between the nature of the good being valued and the elicitation mechanism. In principle, we expect such interaction effects to be small. However, we cannot speak to the biases that may arise from such factors as the ''social desirability'' of the good or to the effect of social networks on stated preferences. Finally, since we induce individual values for the ''good'' we know the exact deviation between stated and actual values in hypothetical and real payment settings.

Surprisingly few studies use induced values to investigate CV mechanism design issues. Taylor et al. (2001) investigate hypothetical and real DC referenda and find that decision "errors" (i.e. responses inconsistent with induced values) across the two conditions were quite close: 16.1% errors in real referenda versus 16.8% in hypothetical referenda. Most errors occurred when the difference between induced value and cost was small ($1 or less). Further, errors were not systematic over or under-statements of value, leading the authors to find in favor of demand revelation under either condition, and conclude that there is no evidence of hypothetical bias. Cherry et al. (2004) investigate hypothetical bias in a second-price auction where participants have an outside option to obtain the private good at a fixed price. They find evidence of (positive) hypothetical bias, but that changing the price for the outside option engenders the same *change* in bids for both real and hypothetical payment settings. The observed overbidding (i.e. overvaluation) in the hypothetical setting may stem from the utility gained from the act of winning, where the lack of financial incentives in the hypothetical setting would serve to increase the importance of winning relative to the real setting. This may explain why Taylor et al. (2001) do not find hypothetical bias in their public good setting. In a different vein, Burton et al. (2003) use induced values to investigate responses to follow-up offers in double referenda.

This study focuses on four elicitation formats used in CV surveys: DC, DC with follow-up certainty question, two versions of the payment card, and two versions of the multiple-bounded discrete choice elicitation mechanism. The DC or "take it or leave it" elicitation format is most commonly used, and is appealing on the grounds it mimics a familiar and realistic posted-offer market purchase or yes/no voting situation, and it imposes the lowest cognitive costs on the respondent in keeping with the arguments of Smith and Walker (1993). Gibbard (1973) and Satterthwaite (1975) show that a DC voting mechanism does not induce false responses. Further, if individuals perceive that they are affected by the voting outcome, then DC referenda are incentive compatible under fairly general assumptions.

Alternatives to DC are commonly used in CV surveys, although their theoretical properties are generally not as well understood. These alternative mechanisms differ from DC in terms of including more payment amounts (i.e. bids) and/or eliciting information on respondent uncertainty. DC questions gather rather crude information on the respondent's WTP. Alternative formats that ask the respondent about multiple payment amounts are statistically more efficient, and reduce sample size requirements relative to DC. Our study investigates two such formats, the multiple-bounded discrete choice format of Welsh and Poe (1998) and the payment card. As respondents are unfamiliar with these alternative mechanisms, and the formats are presumably less transparent, an important empirical question is whether this leads to noisier decision-making. Further, it is incumbent on researchers to establish that alternative mechanisms are demand revealing.

Many CV studies involve goods that are not commonly traded in markets, and respondents may thus be unsure of how to place a value on such goods. Accordingly, researchers have engaged in asking respondents to express uncertainty either in qualitative or quantitative terms. Some posit that value uncertainty may be able to at least partially explain evidence on hypothetical bias. For instance, one conjecture – supported by studies that compare DC responses with those from elicitation mechanisms that allow uncertain responses – is that uncertain respondents will state "yes" to a DC question in a hypothetical setting, but say "no" in a real payment situation, leading to the notion that "Yes means Maybe and No means No" (Berrens et al. 2002, p. 165). Much theoretical and empirical work has been devoted to conceptualizing how respondents answer CV questions under uncertainty and developing elicitation formats that capture this uncertainty (e.g., Li and Mattsson 1995). Calibration techniques based on these formats have been somewhat successful at equating hypothetical and real WTP (Vossler et al. 2003). However, Murphy and Stevens (2004), in their survey of the literature, warn, "without a better understanding of the causes of hypothetical bias and why these techniques are effective, they should be used with caution."

This study investigates a DC question coupled with a follow-up question about response uncertainty, and the multiple-bounded discrete choice format which presents respondents with multiple response categories that allow uncertainty. We implement the DC format with follow-up as a separate mechanism than a pure DC format, as it is unknown *a priori* whether the mere presence of the follow-up question affects DC responses. To investigate whether hypothetical bias is at least partially explained by uncertainty and whether assessments of uncertainty correlate with value uncertainty, we include both certain and uncertain value treatments, where respondents in the latter receive a distribution of possible values. This allows an investigation of how uncertainty affects responses to both hypothetical and real payment questions, and represents a significant departure from existing studies that compare uncertain responses to hypothetical questions taken at face value with real payment transactions where the uncertainty underlying purchase decisions is unknown.

Overall, while avoiding purely open-ended questions, the mechanisms we investigate are standard formats that vary in terms of the amount of information (conceptually) elicited on WTP and preference uncertainty. Although CV surveys employing payment cards or multiple-bounded discrete choice questions do not involve an explicit decision rule, to make these mechanisms operational in a real payment setting we use a public goods voting variant of the Becker–DeGroot–Marshack (BDM) mechanism (1964). We later demonstrate that this mechanism is incentive compatible.

2. Experimental Design

Two hundred and sixty-four students were recruited at the University of Tennessee in the summer of 2005. Experiments were conducted primarily in classroom settings, in groups ranging in size from 16 to 29. Three sessions were conducted in a laboratory setting with small groups (4–10). There is no evidence that the experiment size or setting influences behavior. All participants received \$10 at the outset, and earnings for those in real payment treatments ultimately depended on the outcome of a majority rule referendum. These experiments were short, lasting approximately 35 min on average.

With design conditions of hypothetical or real payment crossed with certain or uncertain induced values, there are four "treatments". Within each treatment we examine the performance of DC, DC with follow-up, two versions of the payment card, and two versions of the multiple-bounded discrete choice format. Each participant is given a copy of the instructions pertaining to one treatment, which were read aloud in each session by the same researcher. Consistent with field survey conditions, there are no practice questions. Instructions describe voting decisions on four proposals, with the outcome of each proposal based on a different elicitation mechanism, such that a participant faces all four mechanisms (one version each of the payment card and multiple-bounded discrete choice). The proposal is to provide a public "good", which is simply an amount of money, to everyone in the group at a described cost. Passage of a proposal is determined by a majority-voting rule. To prevent participants from making quick, quasi-simultaneous decisions on all four proposals, as well as to minimize spillover effects, each proposal is read aloud and decisions on a particular proposal are completed before the next proposal is read. To avoid learning effects, participants make decisions for all proposals before a result is determined and are not allowed to revise prior decisions.

After all decisions are made, a volunteer rolls a die to determine which one of the proposals is considered. Votes for the selected proposal are then counted and the outcome announced to the group. Passage of the proposal affects earnings only in real payment treatments. The experiment concludes with a two-page questionnaire, which gathers basic demographics as well as information on number of economics classes taken, previous knowledge of the economics of public goods, academic major, and prior voting in state/national elections.

2.1. FRAMING

There are two important differences between instructions for real and hypothetical payment treatments. First, in the general preamble to the experiments, the following paragraph is included in hypothetical treatment instructions:

> "These voting situations are *hypothetical* in the sense that no money will be collected or payments made as a result of your vote. However, I ask that you make every effort to vote just exactly as you would if this were a real situation – please vote as if money would be collected and you would actually be paid your value if the proposal passed."

Second, directly before each elicitation mechanism is presented, participants are told to "Please remember that this is a *hypothetical* situation and so there will be no money collected or payments made as a result of your vote."

Note that the wording of the valuation questions does not vary between hypothetical and real payment treatments. We do this to avoid potential mischief associated with having the hypothetical vote appear *too hypothetical*. Also, even though there are no direct financial consequences of decisions in hypothetical treatments, votes for the randomly chosen proposal are nevertheless counted and announced to the group.

2.2. INDUCED VALUES

Participants receive a "value card", which indicates their value for the "good", and their payoff outcomes depending on whether or not the proposal passes. The induced value for the participant is constant across the four proposals. Participants are told that not everyone has the same value – so they are to look closely at their personal value. To at least partially avoid responses motivated by fairness or altruism, participants are not told the range or distribution of values for everyone in their group.

For value certainty treatments, induced values across group members are uniformly distributed over the range of $1.50 to $9.50, in $1 increments. For uncertain value treatments, participants are given a $2 range of possible values. These ranges are constructed by adding/subtracting $1 from the set of certain values. This range is wide relative to the value distribution. Participants are instructed that each value in the range (in 25-cent increments) has an equal chance of being selected. After all decisions are made, the exact value for each participant is determined though a die roll.

2.3. COST DISTRIBUTIONS

For the two dichotomous choice formats, the costs (i.e., bids) are \$1, \$3, \$5, \$7, and \$9, with roughly the same number of participants receiving each amount. In order to avoid correlation between induced values and costs, which would produce misleading WTP distributions even if all respondents voted optimally, costs were first associated with each participant number and then induced values were randomly assigned. For a given respondent, costs for the two DC questions differ in about 75% of the cases. We constructed design parameters (DC costs and induced values) for 75 potential participants. For purposes of experimental control, this set of parameters is used for all four treatments. Participants know that DC costs vary across group members but do not know anything about the range or distribution of costs.

For the payment card and multiple-bounded discrete choice formats, participants face possible payment amounts of \$1 to \$10, in \$1 increments. These amounts are the same for all participants. Given the cost and induced-value distributions, the smallest discrepancy between (expected) induced value and cost is 50 cents.

2.4. ELICITATION MECHANISMS AND ASSOCIATED DECISION RULES

The decision rule for a DC referendum is straightforward: if the majority of individuals vote yes, the proposal passes. The design issue confronted in this study is how to determine decision rules based on payment card or multiple-bounded responses. Although surveys generally frame these formats in terms of a voting exercise, an explicit decision rule is not typically described. Indeed, an important empirical question is how respondents believe their decisions will be used. Note that decision rules are necessary, as otherwise there would be no way to implement these mechanisms under real payment conditions, and hence no way to detect hypothetical bias.

For payment card and multiple-bounded discrete choice elicitation mechanisms, we essentially convert decisions into yes and no votes for a randomly determined cost amount and then use a majority-voting rule to determine whether the proposal passes. This mechanism is best thought of as a public goods voting version of the BDM, and our payment card in particular is a discrete-choice version of the Random Price Voting Mechanism of Messer et al. (2006). When framed as a purchase decision, the BDM for private goods has the buyer submit her maximum WTP, the price for the good is then randomly determined, and the buyer purchases the good if WTP exceeds price. The main difference from the BDM and our public goods version is that, instead of triggering a purchase when WTP exceeds the cost, here this triggers a "yes" vote for the proposal and a majority of such votes is needed for implementation.

In a private good setting the BDM is theoretically incentive compatible under the assumption of expected utility maximization.[3] For public goods settings the BDM can be shown to be incentive compatible under similar conditions. Consider the case of a general payment card. The basic logic is as follows. If the randomly selected cost falls below her value, she makes a positive return if the proposal passes. Thus, she is better off by indicating a WTP for all amounts less than her value, as these become yes votes at these amounts, which increases the probability the proposal passes at these amounts (and her expected payoff). If the random cost falls above her value she loses money if the proposal passes. She should not indicate any amount greater than her induced value as that would increase the probability of implementation and decrease her expected earnings at these costs. Hence, the respondent is best off indicating a maximum WTP consistent with her induced value.

Closely following Irwin et al. (1998) and Messer et al. (2006), we construct a more formal proof. Let WTP_i and V_i denote the indication of WTP and value of the "good", respectively, for participant *i*. Further, let C be the randomly chosen cost from the uniform distribution $[C^{\min}, C^{\max}]$, which is paid only if the proposal passes. With Y denoting the participant's initial endowment, the participant earns $Y + V_i - C$ if the proposal passes, and earns Y if the proposal does not pass. Assuming that the participant behaves as if her vote is pivotal, let $[C^{\min}, WTP_i]$ be the range of costs for which the participant determines the proposal passes, and let $[WTP_i, C^{\max}]$ be the range of costs for which the participant determines the proposal fails. Finally, let $U(\cdot)$ denote the utility function. Assuming expected utility maximization, the optimal bid is determined by maximizing the following expression with respect to B_i:

$$EU_i = \int_{C^{\min}}^{WTP_i} p(C)U(Y + V_i - C)dC + \int_{WTP_i}^{C^{\max}} p(C)U(Y)dC$$

Taking the first order condition with respect to B_i yields:

$$dEU_i/dB_i = p(B_i)U(Y + V_i - B_i) - p(B_i)U(Y) = 0.$$

Thus, maximal expected utility has $B_i = V_i$. In words, expected utility is maximal when the participant indicates her WTP is equal to induced value.

This majority vote decision rule is chosen because U.S. residents are generally quite familiar with majority voting rules, and referenda have nice theoretical properties. Specifically, after all decisions have been made, we randomly choose one of the 10 cost amounts by having a volunteer roll a 10-sided die; this chosen amount becomes the cost of the good if the proposal passes. This determination corresponds to a reasonable conjecture actual CV survey respondents may have: multiple amounts are presented because actual costs are not presently known, but once uncertainty about cost is resolved the researchers/policy makers will examine whether the majority of respondents are willing to pay this amount. We now present the value elicitation questions and describe how responses for each determine whether a proposal passes.

2.4.1. *Dichotomous Choice*

For both DC variants, the WTP question reads as follows:

> What is your vote on the proposal, given that passage of the proposal would cost you $X? (*Please circle ONE response*)
> YES
> NO

If a majority circle "yes", the proposal passes; otherwise, it does not pass. Each participant (hypothetically or actually) pays their particular cost amount and receives the good (i.e., their induced value).

Similar to how follow-up certainty questions come across in CV surveys, we treat the certainty question as a debriefing question that does not affect the outcome of the vote. The particular certainty question we investigate comes from Champ et al. (1997), and has been used in several subsequent studies (e.g., Vossler et al. 2003):

> So you voted YES on this proposal. We would like to know how sure you are of that. On a scale from '1' to '10', where '1' is 'Very *Un*certain' and '10' 'Very Certain', how certain are you about your YES vote? (*Please circle ONE response*)

Very Uncertain									Very Certain
1	2	3	4	5	6	7	8	9	10

Thus, this question is consistent with the common stance in the literature that the "uncertainty" lies with yes DC responses and that it is primarily the yes responses that are poor indicators of actual intentions to pay.

2.4.2. *Payment Card*

We investigate two forms of the payment card present in the literature: one which has the respondent circle the maximum amount they are willing to pay from a set of cost amounts, and another which has the respondent indicate a yes or no response to each amount. The first version, which we label "PC v.1", reads:

> What is the *highest* amount you would be willing to pay and still vote in favor of the proposal? *(Please circle ONE response)*

\$1	\$2	\$3	\$4	\$5	\$6	\$7	\$8	\$9	\$10

After the cost for the good is randomly determined, any amount circled equal to or greater than the cost is considered as a yes vote. All amounts circled below the actual cost became no votes.

The second version, which we label "PC v.2", is:

> What is your vote on the proposal, given that passage of the proposal would cost you these amounts? (*Please indicate a Y or N response for each amount*)

Cost	Y = Yes N = No
\$1	
\$2	
\$3	
\$4	
\$5	
\$6	
\$7	
\$8	
\$9	
\$10	

This decision framework essentially has the respondent answer a series of DC questions. The proposal passes when the majority of participants vote yes to the randomly selected cost.

As described earlier, both payment card versions are best thought of as a public goods version of the BDM mechanism for private goods. For PC v.1, where the participant can only indicate her maximum WTP among the set of costs, the individual is best off by selecting the highest amount lower than her value. For PC v.2, the participant is best off by voting yes to all costs less than her value, and voting no to all other amounts. Given the discrete choice set, these decisions reflect the closest the respondent can get to the theoretically optimal response of indicating a WTP that equals induced value. For respondents with uncertain induced values, these same decisions are optimal under the added assumption of risk-neutrality. Rabin (2000) shows that, within the EU framework and the stakes typically offered in laboratory market settings, only virtual risk neutrality is consistent with realistic levels of risk aversion. Nevertheless, participants may not be EU-maximizers; risk or loss aversion may drive responses.

2.4.3. *Multiple-Bounded Discrete Choice*

The multiple-bounded discrete choice format is best thought of as a payment card (in particular, our PC v.2) that allows respondents to express payment uncertainty.[4] As detailed by Welsh and Poe (1998), the format presents people with a response matrix where they choose among "Definitely Yes", "Probably Yes", "Not Sure", "Probably No" or "Definitely No" responses at each possible payment amount. Researchers tend to interpret and analyze multiple-bounded data in one of two ways. Welsh and Poe (1998), and others, recode categorical choices as yes and no responses to define a WTP interval, and then employ analytical methods used for payment cards. Alternatively, Evans et al. (2003) argue that multiple-bounded responses convey subjective payment probabilities. They assign probability weights to each category and use a generalization of the interval model used for payment cards in data analysis.

An important question is how CV survey respondents facing this format believe their responses will be used. Welsh and Poe (1998) do not provide their respondents with explicit weights that will be applied to the "votes" in the matrix. We are unable to implement their version of the multiple-bounded setting in the lab given our sample sizes and limited ability to collect participant characteristic information. That is, we are unable to map the participant choices into "yes" and "no" votes as did Welsh and Poe. Thus we have chosen to implement a version of this setting with considerable vagueness in the weights that would be applied to the choices the participants indicate in the matrix. We assert that survey respondents may believe that "Definitely Yes" and/or "Probably Yes" responses will be used as indications of WTP at each price level. This interpretation underlies our first version of this format, which we label "MBDC v.1":

> What is your vote on the proposal, given that passage of the proposal would cost you these amounts? (*Circle ONE letter for EACH dollar amount to indicate your vote*)

The decision rule is based on both the random draw for cost and a coin flip. If the coin is "heads", then "Definitely Yes" responses are treated as yes votes and all other responses treated as no votes. A flip of "tails" means both "Definitely Yes" and "Probably Yes" responses to the randomly selected

Cost	Definitely No	Probably No	Not Sure	Probably Yes	Definitely Yes
$1	A	B	C	D	E
$2	A	B	C	D	E
$3	A	B	C	D	E
$4	A	B	C	D	E
$5	A	B	C	D	E
$6	A	B	C	D	E
$7	A	B	C	D	E
$8	A	B	C	D	E
$9	A	B	C	D	E
$10	A	B	C	D	E

cost become yes votes. A volunteer makes the coin flip and the cost determination after all decisions have been collected. With a majority-vote implementation rule, these procedures essentially have a "Definitely Yes" response count as a yes vote, have a "Probably Yes" response count as 0.5 yes votes, and all other categories count as 0 yes votes. While these voting weights for the middle, uncertain response categories do not correspond closely with category labels, recall that our purpose here is to capture a situation where there is considerable vagueness in how responses correspond to votes. The respondent maximizes expected payoffs by selecting "Definitely Yes" for all amounts less than (expected) value, and either a "Definitely No", "Probably No", or "Not Sure" response to all amounts above value.

Consistent with Evans et al. (2003), respondents may conjecture that responses in all categories will be carefully examined, with perceived payment intentions increasing as responses shift from "Definitely No" to "Definitely Yes". This interpretation underlies our other version of this format, which we label "MBDC v.2":

> What is your vote on the proposal, given that passage of the proposal would cost you these amounts? (*Circle ONE letter for EACH dollar amount to indicate your vote*)

The main difference between the two formats is the coupling of probabilities with the response categories. In the experiment, responses become partial yes votes for the randomly selected cost amount. In particular, "Definitely No" responses count as 0 yes votes; "Probably No" responses count as 0.25 yes votes; "Not Sure" responses count as 0.5 yes votes; "Probably Yes" responses count as 0.75 yes votes; and "Definitely Yes" responses count as 1 yes vote. If the sum of the yes votes constitutes a majority, the proposal is implemented.

Optimal responses are similar to those for MBDC v.1. Respondents maximize expected earnings by responding "Definitely Yes" to all amounts

Cost	Definitely No 0% sure of YES vote	Probably No 25% sure of YES vote	Not Sure 50% sure of YES vote	Probably Yes 75% sure of YES vote	Definitely Yes 100% sure of YES vote
\$1	A	B	C	D	E
\$2	A	B	C	D	E
\$3	A	B	C	D	E
\$4	A	B	C	D	E
\$5	A	B	C	D	E
\$6	A	B	C	D	E
\$7	A	B	C	D	E
\$8	A	B	C	D	E
\$9	A	B	C	D	E
\$10	A	B	C	D	E

below (expected) value. However, respondents should respond with "Definitely No" responses to all amounts above value. Thus, the main difference between the two mechanisms is that expected earnings decrease if respondents choose "Probably No" or "Not Sure" for amounts above value with MBDC v.2.

Much of our analysis relies on deviations between stated preferences and induced values (i.e. non-"optimal" responses) and willingness to pay functions. Throughout our analysis, unless otherwise noted, we assume participants are risk-neutral expected utility maximizers. The next section describes three alternative characterizations of deviations from optimal responses, and how we construct willingness to pay functions.

3. Analytical Constructs

3.1. DEVIATIONS

There are at least three reasons why actual responses may deviate from what we characterize above as optimal responses. First, a respondent may simply not be able to decipher what her optimal response is. This could stem from unfamiliarity with the elicitation mechanism or an errant notion that a strategic under- or over-stated valuation actually increases expected earnings.

Second, other-regarding preferences such as altruism may drive responses. The influence of other-regarding behavior on DC responses is likely to be minimal, however, given that both costs and values varied across group members and it is thus impossible for a participant to know whether deviating from an optimal response would hurt or help the average group member. The influence of other-regarding preferences on payment card and multiple-bounded responses is possible. Although the range or distribution of values is unknown, the presented array of costs might provide a signal (in actuality it is a great indicator). Thus, a respondent with a high (low) value relative to the cost distribution may indicate a lower (higher) value (to the extent possible) if she feels this is in the best interest of the average group member. We looked for such behavior by breaking the sample of participants into those with high and low values, where high (low) values are those greater (less) than half of the cost amounts. Then, for each elicitation format in each treatment, we tested the null hypothesis that the frequency of responses (characterized as optimal, over and under valuations) is equal across high and low value participants using chi-square tests for 2×3 contingency tables. We fail to reject the null hypothesis in 11 of 16 cases.[5] The rejections only occur for PC v.1 and MBDC v.1, and four of the five rejections occur in hypothetical settings. In examining the instances where rejections occur, we note that there are very few over-valuations and there tended to be fewer under-valuations among individuals with low values. This is in contrast to Messer et al. (2006), who find that with symmetric and known value distributions those with low values tend to overstate values by as much as those with high values understate them. This suggests that the differences between high and low value participants are more likely due to attempts at strategy or unfamiliarity with the elicitation mechanism. In sum, while we cannot completely rule out the influence of other-regarding behavior in this experiment, we maintain that its role was likely small and does not confound any of our analysis.

Third, relevant for uncertainty treatments, participants may not be risk-neutral EU-maximizers. As an alternative to considering various forms for the utility function, we consider three measures of deviations from what we characterize as an optimal response, one of which does not rely on the assumption of risk neutrality.

Under the first measure, *Deviation1*, a respondent is said to make a "deviation" when she digresses in any way from the theoretically optimal response. For DC formats a deviation occurs when a respondent votes yes when her (expected) value is below the cost amount or when she votes no when her (expected) value is above the cost. For the payment card, this occurs when the respondent does not circle the highest amount that falls below her (expected) value for PC v.1 or fails to vote yes for all amounts below value, and no for all other amounts for PC v.2. For MBDC formats, deviations occur when the respondent does not indicate a "Definitely Yes" for all amounts below (expected) value and indicate a "Definitely No" (or "Probably No" or "Not Sure" for MBDC v.1) for all amounts above value.

Note that the payment card and multiple-bounded formats require respondents to make a more precise decision than with DC. In our experiment, respondents must give correct indications of value when the (expected) value-cost spread is \$0.50. In order to put DC on equal footing with the payment card and multiple-bounded discrete choice, we construct the measure *Deviation2* for DC formats. Under this measure, a deviation is defined identically to *Deviation1*, but we restrict the DC sample to those who have a value-cost spread that is equal to \$0.50. Approximately one-third of DC respondents had a value-cost spread of \$0.50.

Our third measure, *Deviation3*, allows us to examine whether there is a prevalence of large deviations. *Deviation3* considers only deviations from an optimal response that are larger than \$1. Specifically, DC calculations are based on the subset of respondents that had a (expected) value-cost spread greater than \$1, in absolute value terms. This is approximately two-thirds of respondents. For all PC v.2 and multiple-bounded discrete choice respondents, who make a specific decision for all ten cost amounts, we exclude responses to amounts that fall within a \$1 or less of (expected) value. For example, for a respondent with a value of \$4.50, we simply do not look at decisions at cost amounts of \$4 or \$5. But, *Deviation3* would pick up any deviation from the optimal response to values of \$1 to \$3 and from \$6 to \$10. For PC v.1 respondents, who are asked to circle the highest cost amount they would be willing to pay, a respondent is said to make a deviation when the circled amount exceeds (falls below) \$1 of (expected) value. Thus, a respondent with a value of \$4.50 who circles \$5 is not considered to as having made a deviation under this measure, whereas the same respondent deviates if she circles \$6. While *Deviation3* characterizes more extreme deviations, note that it does not rely on the assumption of risk neutrality. That is, *Deviation3* only considers those deviations that occur when costs fall outside the value range for those in uncertain value treatments. Thus, this serves as the basis of better comparisons between uncertainty and certainty treatment respondents facing the same mechanism.

3.2. WILLINGNESS TO PAY DISTRIBUTIONS

To examine issues of demand revelation and hypothetical bias, we construct theoretical and empirical WTP cumulative distribution functions (cdfs).[6] The empirical distribution is simply the distribution of observed yes responses across cost amounts. The theoretical cdf is the distribution of yes responses that would occur if each individual responds optimally based on her induced value, cost, and elicitation mechanism. An alternative would be to use the cdf of induced values, but this introduces a confounding factor: the (expected) induced value is a single number whereas the elicitation mechanism only provides a bound on WTP. Thus, even if all respondents give the "correct" answer, we can still observe deviations between empirical and theoretical cdfs when constructing the latter from induced values. This is especially important with DC comparisons, as DC elicitation is rather crude and any spurious correlation between induced values and costs can lead to differences between our theoretical cdf and the induced value cdf.

Since PC v.1 and multiple-bounded formats do not simply elicit yes and no responses, we elaborate on how we construct these cdfs. For PC v.1, we make the standard assumption that the respondent is willing to pay all listed amounts below the amount circled, and is unwilling to pay any listed amount above the circled amount. For the multiple-bounded formats, distributions are constructed in a manner consistent with how votes are determined. "Definitely Yes" responses for MBDC v.1 are always considered yes votes. "Probably Yes" responses receive a probability weight of 0.5. All other responses in this version are counted as no votes and receive probability weights of 0.0. MBDC v.2 response probabilities correspond directly with how yes votes are assigned: a "Definitely Yes" response receives a yes probability of 1.0; "Probably Yes" responses receive a probability weight of 0.75; "Not Sure" receives a weight of 0.5; "Probably No" receives a weight of 0.25; and "Definitely No" receives a probability of 0.0.

4. Results

4.1. ANALYSIS OF DEVIATIONS

Table I presents the frequency of respondents who deviate from optimal-decisions, using our three measures, by treatment, for each elicitation mechanism. Table II presents tests for differences in the frequency of deviations *between* elicitation mechanisms. Since each respondent faces the two DC formats and each payment card respondent also faces the multiple-bounded discrete choice format, we compare deviations across these samples using a McNemar Test for two dependent samples. For all other comparisons, e.g. PC v.1 versus PC v.2, the respondent groups differ and so we use the Fisher Exact Test for two independent samples. Using the Fisher Exact Test, we also test the equality of deviation rates across all treatments and selected treatment pairs *within* each elicitation mechanism, and present the results in Table III.

The DC format performs quite well with just an 8.7% deviation rate. For the subset of respondents who face a difference between (expected) value and cost of \$0.50, this rate (*Deviation2*) only rises to 15.3%. Just 5.6% of respondents deviate when the difference between (expected) value and cost is greater than \$1 (*Deviation3*). In comparison, simply adding the certainty question leads to an increase in deviations of almost 5%, and this difference

Table I. Observed deviations from an optimal response

Treatment	n[a]	Deviation 1	Deviation 2	Deviation 3
Dichotomous Choice				
Real, Certainty	72	6.9%	16.7%	2.1%
Hypothetical, Certainty	62	8.1%	10.0%	7.1%
Real, Uncertainty	66	12.1%	25.0%	6.5%
Hypothetical, Uncertainty	64	7.8%	9.5%	7.0%
Total	264	8.7%	15.3%	5.6%
Dichotomous Choice with Follow-up Certainty Question				
Real, Certainty	72	11.1%	20.0%	6.4%
Hypothetical, Certainty	62	9.7%	12.5%	7.9%
Real, Uncertainty	66	19.7%	30.4%	14.0%
Hypothetical, Uncertainty	64	12.5%	18.2%	9.5%
Total	264	13.3%	20.2%	9.4%
Payment Card, Version 1				
Real, Certainty	40	50.0%		40.0%
Hypothetical, Certainty	30	33.3%		33.3%
Real, Uncertainty	37	64.9%		48.7%
Hypothetical, Uncertainty	34	70.6%		47.1%
Total	141	55.3%		42.6%
Payment Card, Version 2				
Real, Certainty	32	34.4%		9.4%
Hypothetical, Certainty	32	21.9%		15.6%
Real, Uncertainty	29	72.4%		20.7%
Hypothetical, Uncertainty	30	53.3%		30.0%
Total	123	44.7%		18.7%
Multiple-Bounded Discrete Choice, Version 1				
Real, Certainty	40	67.5%		50.0%
Hypothetical, Certainty	30	56.7%		46.7%
Real, Uncertainty	37	86.5%		37.8%
Hypothetical, Uncertainty	34	91.2%		50.0%
Total	141	75.9%		46.1%
Multiple-Bounded Discrete Choice, Version 2				
Real, Certainty	32	65.6%		40.6%
Hypothetical, Certainty	32	65.6%		50.0%
Real, Uncertainty	29	96.6%		58.6%
Hypothetical, Uncertainty	30	96.7%		50.0%
Total	123	80.5%		49.6%

[a]Relevant dichotomous choice sample sizes differ for *Deviation2* and *Deviation3*. See Text.

is statistically significant at the 10% level. A closer examination reveals that the divergence between DC and DC with follow-up certainty question stems from more deviations in the latter format in the uncertain value treatments.

Table II. Between elicitation mechanism tests of deviations (p-value)[a]

Hypothesis	Deviation 1	Deviation 2	Deviation 3
DC = DCF	0.096	0.375	0.227
PC v.1 = PC v.2	0.109		0.000
MBDC v.1 = MBDC v.2	0.377		0.622
All DC = All PC	0.000	0.000	0.000
All DC = MBDC	0.000	0.000	0.000
All PC = All MBDC	0.000		0.000

DC = Dichotomous Choice; DCF = Dichotomous Choice with Follow-up Certainty Question; PC v.1 = Payment Card, Version 1; PC v.2 = Payment Card, Version 2; MBDC v.1 = Multiple-Bounded Discrete Choice, Version 1; MBDC v.2 = Multiple-Bounded Discrete Choice, Version 2.
[a]Fisher exact test used for all hypothesis with the exception of DC = DCF and All PC = All MBDC, where a McNemar test is used.

There is also a higher propensity for those with the follow-up question to make deviant yes votes.

In sharp contrast to the DC formats, roughly half of respondents make payment card deviations and over three-quarters make deviations with the multiple-bounded discrete choice format. Focusing on large deviations, PC v.2 performs reasonable well, with a deviation rate of 18.7%. Version 1 of the payment card, as well as the multiple-bounded formats, yield large deviations from about 45% of respondents. Overall, statistical tests indicate that DC formats have the lowest frequency of deviations, followed by PC v.2, with PC v.1 and the two multiple-bounded formats bringing up the rear of the pack.

Turning to deviation rate comparisons across treatments, Table III reveals very clear patterns. Deviation rates do not differ across treatments for either DC format. Deviation rates differ across treatments for the payment card and multiple-bounded formats, with the difference arising between certain and uncertain value treatments. These differences vanish if only large deviations are considered, and so may be at least partially explained by risk preferences. There is never a difference in deviation rates between corresponding real and hypothetical payment treatments. Thus, there is no evidence of hypothetical bias in the sense that asking a hypothetical payment question does not lead to a higher (or lower) frequency of deviations.

Finally, we formally model individual-level deviations using a probit model, with explanatory variables that control for design conditions and characteristics of the respondent. Descriptions for explanatory variables appear in Table IV. To examine whether deviations are mostly in one direction, we constructed the variable "Winner" that equals 1 if (expected) value exceeds cost. We interact this with design variables. We initially

Table III. Within elicitation mechanism tests of deviations (Fisher exact *p*-value)

Hypothesis	DC	DCF	PC v.1	PC v.2	MBDC v.1	MBDC v.2
Basis of analysis: Deviation1						
All equal	0.732	0.370	0.013	0.000	0.002	0.000
RC = HC	1.000	1.000	0.223	0.405	0.455	1.000
RU = HU	0.561	0.342	0.623	0.180	0.712	1.000
RC = RU	0.386	0.235	0.250	0.005	0.062	0.003
HC = HU	0.608	0.778	0.005	0.017	0.003	0.003
Basis of analysis: Deviation2						
All equal	0.558	0.517				
RC = HC	0.673	0.702				
RU = HU	0.238	0.491				
RC = RU	0.710	0.511				
HC = HU	1.000	0.694				
Basis of analysis: Deviation3						
All equal	0.644	0.688	0.580	0.220	0.696	0.586
RC = HC	0.336	1.000	0.624	0.708	0.813	0.616
RU = HU	1.000	0.738	1.000	0.552	0.345	0.604
RC = RU	0.356	0.301	0.496	0.287	0.360	0.204
HC = HU	1.000	1.000	0.314	0.230	0.808	1.000

DC = Dichotomous Choice; DCF = Dichotomous Choice with Follow-up Certainty Question; PC v.1 = Payment Card, Version 1; PC v.2 = Payment Card, Version 2; MBDC v.1 = Multiple-Bounded Discrete Choice, Version 1; MBDC v.2 = Multiple-Bounded Discrete Choice, Version 2; RC = Real Payment, Certainty; HC = Hypothetical Payment, Certainty; RU = Real Payment, Uncertainty; HU = Hypothetical Payment, Uncertainty.

included income and an indicator variable for whether the respondent has studied the economics of public goods. Coefficients on these variables were jointly insignificant in all models, and we exclude these from our final analysis in lieu of fairly high item nonresponse rates.

We estimate mechanism-specific probit models explaining individual deviations; *Deviation1* is the dependent variable. We pool the two dichotomous choice data sets, allowing for the mean deviation rate to differ between the two by including the dummy variable "DCF", which equals 1 if the observation corresponds to the DC with follow-up treatment. This pooled model is justified by a likelihood ratio test ($\chi^2_{(12)} = 5.858$, $p = 0.923$). For the payment card and multiple-bounded formats, we analyze deviations for each cost amount, rather than simply model whether some deviation occurred during the decision exercise. Since participants make decisions on ten different amounts for these formats, we construct ten observations per participant in a straightforward manner. We prefer this modeling strategy as it allows us to investigate the effect of the value-cost spread on deviations and

Table IV. Description of explanatory variables for probit models

Variable name	Description	Sample mean[a] (SD)
Hypothetical	= 1 if payment is hypothetical	0.477 (0.500)
Uncertainty	= 1 if induced value is uncertain	0.492 (0.500)
Hypo*Uncertainty	= 1 if payment is hypothetical and induced value is uncertain	0.244 (0.430)
Distance	Absolute value of (expected) value-cost difference	2.403 (2.173)
Winner*Uncertainty	= 1 if (expected) induced value exceeds cost *and* Uncertainty = 1	0.265 (0.442)
Winner *Certainty	= 1 if (expected) induced value exceeds cost *and* Uncertainty = 1	0.279 (0.449)
Group size	Number of participants in the participant's group	19.156 (6.272)
Lab	= 1 if experiment conducted in designated experimental lab	0.080 (0.272)
Age	Participant's age, in years	23.565 (6.480)
Male	= 1 if participant is male	0.573 (0.495)
Voted	= 1 if participant voted in any prior state or national election	0.840 (0.367)
Math	= 1 if participant has a math-intensive major (e.g., engineering, statistics)	0.302 (0.459)
DCF	= 1 if elicitation format is dichotomous choice with follow-up certainty question	0.500 (0.500)

[a]Sample means are for the dichotomous choice model sample.

results in comparably specified models across elicitations. For all models, we use robust standard errors adjusted for clustering at the individual-level.

Table V presents estimated models. We present estimated marginal effects, as they are easier than probit coefficients to interpret. The DC model estimates the probability of deviating from an optimal response increases by 3.5% when a follow-up certainty question is present, *ceteris paribus*, and this effect is significant at the 10% level. The coefficient on the variable "Hypothetical" is insignificant in all models, suggesting that participants in hypothetical payment treatments are no more likely to make deviations than those in real payment treatments. The significant "DCF" coefficient and insignificant "Hypothetical" coefficients are consistent with nonparametric test results.

The only variable significant in all models is "Distance". The marginal effect for this variable ranges from -0.026 (PC v.2) to -0.071 (MBDC v.2), indicating the probability of making a deviation decreases by 2.6% to 7.1% for every \$1 increase in the difference between (expected) induced value and cost. Uncertainty leads to a positive and significant increase in deviations in all formats except for PC v.2 and MBDC v.2. The effects of "Distance" and "Uncertainty" jointly suggest deviations are universally related to the difficulty of the decision problem. In value certainty treatments, "Winners" are generally more likely to make deviations. This effect is quite large for PC v.1 and MBDC v.1, where winners in certainty treatments are 18.6% and 36.7% more likely to make deviations. Group size and participating in a lab setting (versus a classroom) are statistically insignificant factors. Note that no "Lab" coefficient is estimated for MBDC v.1 and PC v.1 models since all these respondents participated in a classroom setting.

Turning to respondent characteristics, age is statistically significant in the two payment card models, although the marginal effects are quite small and the direction of the effect is inconsistent. For some formats, prior voting experience and facility in mathematics decrease the probability of making a deviation. In particular, voting has a statistically significant effect for DC and the seemingly more transparent versions of the payment card (PC v.2) and multiple-bounded discrete choice (MBDC v.2) formats. Math-intensive majors are 7.4% and 10.4% less likely to make deviations to dichotomous choice and MBDC v.2 questions, respectively.

4.2. ANALYSIS OF WILLINGNESS TO PAY DISTRIBUTIONS

We compare cumulative empirical and theoretical, "true" WTP distributions using one-sample Kolmogorov–Smirnov tests and present results in Table VI. The null hypothesis of equal distributions is not rejected in all cases, with the exception of PC v.1 and MBDC v.1, where we reject equality for the real and hypothetical payment treatments with certain values at the

Table V. Probit analysis of deviations (entries in table are marginal effects)

	Dichotomous Choice	Payment Card, v.1	Payment Card, v.2	MBDC, Version 1	MBDC, Version 2
Dependent variable = 1 if respondent deviates from an optimal response; = 0 otherwise					
Hypothetical	0.012 (0.34)	−0.002 (−0.06)	−0.001 (−0.02)	−0.040 (−0.68)	0.070 (0.62)
Uncertainty	0.092** (2.16)	0.112*** (2.95)	0.042 (0.93)	0.203*** (2.49)	0.140 (1.09)
Hypo*Uncertainty	−0.033 (−0.65)	0.012 (0.27)	0.082 (0.99)	0.017 (0.23)	−0.054 (−0.24)
Distance	−0.033*** (−4.60)	−0.035*** (−8.77)	−0.026*** (−6.04)	−0.053*** (−9.43)	−0.071*** (−8.37)
Winner*Uncertainty	−0.017 (−0.55)	0.047 (1.47)	−0.010 (−0.43)	0.135*** (2.78)	−0.047 (−0.76)
Winner*Certainty	0.076** (2.15)	0.186*** (5.13)	0.032 (1.05)	0.367*** (6.46)	0.051 (0.75)
Group size	−0.003 (−1.03)	−0.001 (−0.53)	−0.013** (−2.05)	0.001 (0.13)	−0.005 (−0.33)
Lab	−0.055 (−0.64)	–	−0.149 (−1.62)	–	−0.044 (−0.19)
Age	0.001 (0.65)	−0.002* (−1.69)	0.007*** (3.00)	−0.002 (−1.34)	0.005 (0.63)
Male	−0.003 (−0.13)	−0.007 (−0.33)	0.003 (0.14)	−0.068* (−1.75)	−0.072 (−1.25)
Voted	−0.050* (−1.88)	−0.025 (−0.85)	−0.077*** (−2.92)	−0.082 (−1.44)	−0.177** (−2.45)
Math	−0.074** (−2.45)	−0.005 (−0.23)	−0.003 (−0.11)	−0.026 (−0.71)	−0.104* (−1.81)
DCF	0.035* (1.75)	–	–	–	–
Constant	−0.095 (−1.24)	−0.094* (−1.75)	0.000 (0.00)	−0.129 (−1.32)	0.128 (0.32)
Log-L	−158.538	−416.151	−320.946	−565.316	−654.516
χ^2	47.55	94.55	67.68	204.18	90.07
Pseudo-R^2	0.131	0.203	0.160	0.213	0.119
n	524	1390	1230	1390	1230

Asymptotic t-ratios in parenthesis.
*, **, and *** denote parameter is statistically different from zero at the 10%, 5%, and 1% significance levels, respectively.

5% level. Thus, even for formats such as PC v.2 and MBDC v.2 with a high incidence of deviations, these deviations occur for both those with a positive or negative (expected) value-cost spread such that they cancel each other out. Looking back to the probit results for PC v.1 and MBDC v.1, the frequency of deviations is quite high for "Winners", which leads to a systematic under-revelation of the true WTP distribution. Further, the largest discrepancy between observed and true WTP distributions occurs at fairly large amounts, either $5 or $6, which presumably leads to underestimates of median WTP. Drawing more from probit model results, the lack of statistically significant effects on deviation rates for voters and mathematically inclined individuals suggests that these elicitations confront respondents with an unfamiliar decision framework that is relatively more difficult to decipher.

Table VII presents Kolmogorov–Smirnov distribution tests for analogous empirical real and hypothetical WTP distributions. In all cases, even for mechanisms that produce biased WTP distributions, we fail to reject the null hypothesis of equality. This result, coupled with the earlier finding of no systematic differences between deviation rates, suggests there is no evidence whatsoever of hypothetical bias.

4.3. ANALYSIS OF STATED PREFERENCE UNCERTAINTY

In our final stage of analysis, we investigate whether there is a correspondence between stated uncertainty in the DC follow-up certainty question and multiple-bounded formats and uncertain induced values. Table VIII presents mean certainty level responses for selected subgroups as well as correlation coefficients (Spearman's ρ) between the (expected) value-cost spread and certainty level, by treatment, for DC follow-up certainty questions. The mean certainty level is about 8 out of 10 across the board. The mean certainty levels for value certainty treatments were surprising at first glance. However, even participants with certain induced values may have difficulty with the decision task and convey this decision uncertainty through responses to this question. Consistent with this conjecture, we generally see lower certainty levels for those facing a small difference between (expected) value and cost, including those respondents who had a negative (expected) value-cost spread.

Estimated correlation coefficients between the (expected) value-cost spread and certainty level show a statistically significant relationship between the two in all treatments. The correlation coefficients are 0.29 and 0.31, respectively, for the real and hypothetical payment treatments with certain values. These correlation coefficients are significant at the 10% level. The correlation coefficients for uncertainty treatments are larger, 0.56 for real payment and 0.52 for hypothetical payment, and are statistically different

Table VI. Theoretical versus empirical willingness to pay distributions (Kolmogorov–Smirnov test)

Treatment	*n*	*D*-statistic	Reject Null?
Dichotomous Choice			
Real, Certainty	72	0.0769	No
Hypothetical, Certainty	62	0.1429	No
Real, Uncertainty	66	0.0833	No
Hypothetical, Uncertainty	64	0.1429	No
Dichotomous Choice with Follow-up Certainty Question			
Real, Certainty	72	0.0667	No
Hypothetical, Certainty	62	0.1429	No
Real, Uncertainty	66	0.1250	No
Hypothetical, Uncertainty	64	0.1333	No
Payment Card, Version 1			
Real, Certainty	40	0.2500	Yes, at 5% level
Hypothetical, Certainty	30	0.2667	Yes, at 5% level
Real, Uncertainty	37	0.0811	No
Hypothetical, Uncertainty	34	0.1471	No
Payment Card, Version 2			
Real, Certainty	32	0.0625	No
Hypothetical, Certainty	30	0.1333	No
Real, Uncertainty	26	0.1154	No
Hypothetical, Uncertainty	30	0.1000	No
Multiple-Bounded Discrete Choice, Version 1			
Real, Certainty	40	0.2375	Yes, at 5% level
Hypothetical, Certainty	30	0.2667	Yes, at 5% level
Real, Uncertainty	34	0.1618	No
Hypothetical, Uncertainty	34	0.1765	No
c*Multiple-Bounded Discrete Choice, Version 2*			
Real, Certainty	32	0.0469	No
Hypothetical, Certainty	29	0.1034	No
Real, Uncertainty	26	0.0865	No
Hypothetical, Uncertainty	30	0.0583	No

from zero at well beyond the 1% level. Overall, the certainty question captures value uncertainty as well as uncertainty regarding the decision task.

Recall that there is weak evidence that the mere presence of the follow-up certainty question leads to more deviations for the DC question. In particular, those with uncertain values and a small (expected) value-cost spread are relatively more likely to deviate when the follow-up question is present. Since these participants do indicate a low level of certainty, it is likely they are conveying that they *could* be made better off with a yes vote, but the chances

Table VII. Tests of hypothetical bias: real versus hypothetical willingness to pay distributions (Kolmogorov–Smirnov test, *p*-value)

Elicitation mechanism	*n*	Certainty	*n*	Uncertainty
Dichotomous Choice	62	0.552	62	0.552
Dichotomous Choice with Follow-up Certainty Question	62	0.116	62	0.916
Payment Card, Version 1	29	0.951	33	0.851
Payment Card, Version 2	30	0.594	26	0.926
Multiple-Bounded, Version 1	29	0.572	31	0.615
Multiple-Bounded, Version 2	31	0.951	26	0.733

Table VIII. Follow-up certainty question responses

Treatment	*n*	Mean	Mean, |value-cost| = \$0.50	Mean, deviation voters	Spearman's ρ (value-cost), certainty
Real, Certainty	39	7.95	7.38	6.75	0.2946 (p = 0.069)
Real, Uncertainty	34	7.65	5.67	5.00	0.5561 (p = 0.001)
Hypothetical, Certainty	31	8.19	7.09	5.50	0.3215 (p = 0.078)
Hypothetical, Uncertainty	35	7.71	6.91	5.25	0.5181 (p = 0.001)

of it are low. Without a means of qualifying her answer, a respondent facing a cost that falls at the upper end of her value distribution is less likely to deviate from an optimal response.

We now turn to analysis of stated response uncertainty in the multiple-bounded format. We focus our attention on MBDC v.2, given the systematic bias in MBDC v.1 certainty treatments confounds comparisons between MBDC v.1 uncertain and certain value treatments. Table IX presents the aggregate usage of uncertainty response categories across respondents, while Table X show the percentage selecting each category in relation to the difference between (expected) value and cost under the four treatment conditions. Table IX reveals that over half of respondents with certain induced values choose at least one uncertain response category, with approximately 90% of uncertain value respondents making an uncertain response choice. Further, roughly 17% and 33% of certain and uncertain value respondents, respectively, choose all three uncertain response categories during the decision task.

Surprisingly, a noticeable fraction of respondents (about 10–20%) choose an uncertain response category even when the (absolute) difference between value and cost is \$2.50 or higher. However, fortunately, we see that most uncertain responses are given when value is close to cost. Further, for respondents with induced-value uncertainty, we see a relatively higher percentage of respondents with uncertain responses when the cost falls within their value range. Indeed, in real payment treatments with an (expected) value-cost difference of −\$0.50, 31% of respondents with certain induced values choose an uncertain response category as compared to 66% with uncertain induced values. Using a Kolmogorov–Smirnov test, we find that the response distributions for certain and uncertain value respondents facing a value-cost spread of −\$0.50 is statistically different at the 5% level under either payment condition. Response distributions are statistically different at the 10% level when one considers a value-cost difference of \$0.50, but only when payment is hypothetical. Overall, there is a higher propensity for respondents with uncertain induced values to select response categories associated with a lower level of uncertainty, but the pervasive use of uncertain response categories for rather salient differences between value and cost is troublesome.

5. Conclusion

This study reports the results from a set of induced-value laboratory experiments designed to investigate demand revelation, hypothetical bias and value uncertainty for several value elicitation formats commonly used in contingent valuation surveys. We evaluate the performance of DC, DC with follow-up certainty question, two versions of the payment card and two versions of the multiple-bounded discrete choice format. Although field applications do not typically involve an explicit implementation rule, especially for payment cards and multiple-bounded formats, our laboratory elicitations use majority-vote implementation rules that are theoretically incentive-compatible under real payment conditions. The experimental design allows us to explore basic mechanism design issues by (largely) controlling factors such as context and other-regarding behavior. At a minimum, responses under real payment conditions capture whether the incentives underlying elicitation formats are well-understood and immune to (perceived) strategic responses. Inducing value uncertainty allows a direct comparison of stated and *actual* uncertainty, which would be quite difficult to undertake with "homegrown" values. Further, inducing uncertainty allows us to test whether uncertainty leads to hypothetical bias, a conjecture that is commonly put forth in the literature. Responses to hypothetical payment scenarios serve as a more stringent indicator of mechanism transparency, and allow observance of whether respondents systematically state high or low valuations when there is no real money on the line.

Table IX. Multiple-bounded discrete choice (Version 2) uncertain response category usage across respondents

Treatment	% Using all three categories	% Using at least two categories	% Using at least one categories
Real Payment, Certainty	12.5	40.6	62.5
Real Payment, Uncertainty	38.5	73.1	88.5
Hypothetical Payment, Certainty	20.7	27.6	52.4
Hypothetical Payment, Uncertainty	26.7	70.0	93.3

Our analysis of deviations from optimal responses, WTP distributions, and stated indications of uncertainty lead to four important results. First, we find no clear evidence of hypothetical bias. The frequency of deviations, as well as empirical WTP distributions, never differs between otherwise identical real and hypothetical payment treatments for any mechanism. This result should be refreshing for contingent valuation practitioners, as it suggests the method is not doomed simply because we ask people to make decisions without financial commitments.

Induced-value tests (Taylor et al. 2001; Burton et al. 2003; and ours reported here) provide evidence on the empirical performance of elicitation mechanisms under the premise that the value for the good has been formed.[7] If a mechanism performs comparably under both real and hypothetical payment conditions, this suggests that hypothetical bias does not simply arise because payment is hypothetical. Instead, it may suggest there is a value formation problem; people form systematically different values in a contingent versus a real market. That is, the control afforded by our experimental design rules out factors that may be driving the oft-observed hypothetical bias in field and laboratory studies, such as yea-saying, altruism, social context, and value formation.[8] While no studies have tested whether a value formation problem exists whereby individuals form values differently in the context of a hypothetical payment scenario, it seems reasonable that people spend relatively less time, or do not draw from the same information set, when making a hypothetical payment decision. Studies investigating the effects of "time to think" on stated values find that providing this time *lowers* stated WTP. For example, Whittington et al. (1992) use a split-sample design to test whether allowing respondents additional time to think (and encouraging them to do so) before completing a CV survey and report that this results in lower stated WTP values. Thus, it is possible that better value formation may reduce the gap between stated and actual WTP. We feel future research is warranted on entreaties that motivate individuals to think

Table X. Multiple-bounded discrete choice (Version 2) real and hypothetical payments

Value-Cost	Certainty					Uncertainty					*p*-value^
	DY	PY	NS	PN	DN	DY	PY	NS	PN	DN	
Real payment treatments											
> $ 4.50	97	3	0	0	0	100	0	0	0	0	
$ 4.50	75	19	6	0	0	88	6	6	0	0	
$ 3.50	76	14	5	5	0	83	11	6	0	0	
$ 2.50	81	12	4	0	4	81	5	10	5	0	
$ 1.50	80	7	3	7	3	67	13	13	8	0	
$ 0.50	59	19	9	9	3	35	38	8	12	8	0.282
– $ 0.50	6	19	3	9	63	15	12	19	35	19	0.006
– $ 1.50	3	3	10	10	72	0	4	13	9	74	
– $ 2.50	0	4	12	8	77	0	0	10	15	75	
– $ 3.50	0	4	8	8	79	0	0	6	17	78	
– $ 4.50	0	0	10	5	85	0	0	0	14	86	
< – $ 4.50	0	0	9	9	83	0	0	0	8	92	
Hypothetical payment treatments											
> $ 4.50	85	15	0	0	0	86	7	0	7	0	
$ 4.50	93	7	0	0	0	75	13	6	6	0	
$ 3.50	84	11	0	0	5	80	10	0	10	0	
$ 2.50	83	13	0	0	4	71	17	4	4	4	
$ 1.50	61	29	4	0	7	68	7	14	4	7	
$ 0.50	55	14	21	3	7	23	47	7	10	13	0.082
– $ 0.50	0	0	17	17	66	3	13	13	40	30	0.037
– $ 1.50	0	0	8	8	85	0	15	7	4	74	
– $ 2.50	0	0	0	9	91	0	0	8	21	71	
– $ 3.50	0	0	0	5	95	0	0	5	18	77	
– $ 4.50	0	0	0	0	100	0	0	6	6	89	
< – $ 4.50	0	0	0	0	100	0	0	0	16	84	

^*p*-value corresponds with Kolmogorov-Smirnov Test of "Certainty" v. "Uncertainty"

about the decision exercise in a hypothetical payment situation as they would in a real payment situation.

Second, there is strong evidence of demand revelation for both real and hypothetical DC referenda, under value certainty and uncertainty. Deviations occur for less than 10% of respondents and there is a close correspondence between stated and theoretical WTP distributions. While one version each of the payment card and multiple-bounded discrete choice format elicit unbiased WTP distributions, the very large number of deviations observed in this very simple experimental setting should give practitioners pause. Nevertheless, assuming these mechanisms elicit unbiased value estimates in more general settings, an important empirical question is whether there are still efficiency gains over DC despite such noisy decision-making.

Third, we find that versions of the payment card and multiple-bounded discrete choice format that force respondents to think more about the decision task work better. In terms of deviations and bias, a payment card which has respondents make a yes or no choice for each possible cost amount is better at eliciting WTP than a payment card where respondents circle the amount corresponding with maximum WTP. In terms of bias, a multiple-bounded discrete choice mechanism where each response category corresponds with a fractional yes vote is superior to a version where incentives are less transparent. Through post-experiment discussions, it is apparent that some respondents believed they could somehow make more money by strategically under-reporting their WTP in the poorer-performing versions of the two mechanisms.

Fourth, we find that stated uncertainty in the DC follow-up certainty question and multiple-bounded discrete choice format statistically correlates with uncertain induced values and basic uncertainty regarding the decision task. However, the mere presence of the follow-up question results in a slightly higher frequency of DC deviations. Further, there is a high propensity for multiple-bounded respondents to choose an uncertain response, even with certain induced values and a sizeable difference between value and cost. This suggests that by increasing the action space we are sending a signal to respondents that they should exercise these additional options. This suggests that stated response uncertainty in field studies should be interpreted with care as not all stated uncertainty appears to be genuine. The overall message is that formats that endeavor to elicit information on uncertainty have merit, but at the expense of noisy decision-making.

Our results under induced value uncertainty are in contrast to studies in the literature that assert that hypothetical bias may stem from preference uncertainty, and be resolved through eliciting information on uncertainty. We only investigate one of many possible sources of uncertainty, however. For instance, there may also be uncertainty related to cost and underlying value distributions may be asymmetric. Further, CV surveys and naturally occurring public goods purchase decisions generally involve larger amounts of money than in our experiments. Holt and Laury (2002) provide evidence that people exhibit more risk aversion as payoffs are increased. Eliciting information on preference uncertainty could thus prove to be more valuable in settings involving higher stakes. Future research in this direction is warranted.

Acknowledgements

We thank Mary Evans, John List, Greg Poe and seminar participants at Georgia State University for their helpful comments and suggestions. We thank David Bruner and Zach Richards for their research assistance. Research supported by the J. Fred Holly Endowment.

Notes

1. These uncontrolled factors include the existence of an outside option in laboratory studies involving private goods (Cherry et al. 2004), social context (List et al. 2004), and the use of a different elicitation mechanism in real versus hypothetical settings.
2. Since the decision is often viewed as having only a remote consequence to the individual one may ask the level of cognitive resources that will be invested in the decision to respond to the valuation question. As Smith and Walker (1993) show, decision costs must be balanced by decision rewards if we are to observe accurate statements of values. If the respondent views the CV survey as having no consequence, we may observe considerable variance in the responses or what we could label "hypothetical noise".
3. There is some debate concerning the efficacy of the BDM mechanism in private goods settings. Irwin et al. (1998) demonstrate that the mechanism is cognitively transparent and that experimental participants can be motivated even with very small payoffs. Horowitz (2005) argues that the BDM is not incentive compatible for some behavioral models; however, for EU maximizers and for many non-EU behaviors (such as rank dependent expected utility) Horowitz recognizes that the BDM is incentive compatible.
4. Arguments of theoretical incentive compatibility are similar to those given for the payment card.
5. For these tests, the results of which are available from the authors upon request, we exclude those with values of $1.50 and $9.50 as these respondents could really only err in one direction.
6. We initially compared parametric and nonparametric estimates of mean/median WTP across formats, and to mean/median induced values. We found that such comparisons would be rather misleading. In particular, the crudeness of the DC format coupled with the small sample sizes led to a scenario where, even if all participants behaved according to theory, estimated mean/median WTP differed substantially from mean/median induced values.
7. This is also true for the extensive literature investigating the performance of auction mechanisms (see Cherry et al. 2004) and public good provision mechanisms.
8. Examples of such interaction among respondents are found in the work of List et al. (2004) as discussed in the introduction section.

References

Becker, G. M., M. H. DeGroot and J. Marschak (1964), 'Measuring Utility by a Single-Response Sequential Method', *Behavioral Science* **9**, 226–232.

Berrens, R. P., H. Jenkins-Smith, A. K. Bohara and C. L. Silva (2002), 'Further Investigation of Voluntary Contribution Contingent Valuation: Fair Share, Time of Contribution, and Respondent Uncertainty', *Journal of Environmental Economics and Management* **44**, 144–168.

Burton, A. C., K. S. Carson, S. M. Chilton and W. George Hutchinson (2003). 'An Experimental Investigation of Explanations for Inconsistencies in Responses to Second Offers in Double Referenda', *Journal of Environmental Economics and Management* **46**(3), 472–489.

Champ, P. A., R. C. Bishop, T. C. Brown and D. W. McCollum (1997), 'Using Donation Mechanisms to Value Nonuse Benefits from Public Goods', *Journal of Environmental Economics and Management* **33**(2), 151–162.

Cherry, T. L., P. Frykblom, J. F. Shogren, J. A. List and M. B. Sullivan (2004), 'Laboratory Testbeds and Non-Market Valuation: The Case of Bidding Behavior in a Second-Price Auction with an Outside Option', *Environmental and Resources Economics* **29**(3), 285–294.

Evans, M. F., N. E. Flores and K. J. Boyle (2003), 'Multiple Bounded Uncertainty Choice Data as Probabilistic Intentions', *Land Economics* **79**(4), 549–560.

Gibbard, A. (1973), 'Manipulation of Voting Schemes: A General Result', *Econometrica* **41**(3), 587–602.

Holt, C. A. and S. K. Laury (2002), 'Risk Aversion and Incentive Effects', *American Economic Review* **92**(5), 1644–1655.

Horowitz, J. K. (2005), The Becker–DeGroot–Marshak Mechanism is Not Always Incentive Compatible, Even for Non-Random Goods. Working paper, Department of Agricultural and Resource Economics, University of Maryland.

Irwin, J. R., G. H. McClelland, M. McKee, W. D. Schulze and N. E. Norden (1998), 'Payoff Dominance vs Cognitive Transparency in Decision Making', *Economic Inquiry* **36**(2), 272–285.

Li, C.-Z. and L. Mattsson (1995), 'Discrete Choice under Preference Uncertainty: An Improved Structural Model for Contingent Valuation', *Journal of Environmental Economics and Management* **28**, 256–269.

List, J. A. and C. A. Gallet (2001), 'What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? Evidence from a Meta-Analysis', *Environmental and Resource Economics* **20**(3), 241–254.

List, J. A., R. Berrens, A. Bohara and J. Kerkvliet (2004), 'Examining the Role of Social Isolation in Stated Preferences', *American Economic Review* **94**(3), 741–752.

Little, J. and R. Berrens (2004), 'Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta-Analysis', *Economics Bulletin* **3**(6), 1–13.

Messer, K. D., L. P. Gregory, W. D. Schulze, D. Rondeau and C. A. Vossler (2006), Exploring Voting Anomalies Using a Demand Revealing Random Price Voting Mechanism. Working Paper, Department of Applied Economics and Management, Cornell University, Ithaca, New York.

Murphy, J. J., P. Geoffrey Allen, T. H. Stevens and D. Weatherhead (2005), 'A Meta-Analysis of Hypothetical Bias in Contingent Valuation', *Environmental and Resource Economics* **30**(3), 313–325.

Murphy, J. J. and T. H. Stevens (2004), 'Contingent Valuation, Hypothetical Bias, and Experimental Economics', *Agricultural and Resource Economics Review* **33**, 182–192.

Rabin, M. (2000), 'Risk Aversion and Expected-Utility Theory: A Calibration Theorem', *Econometrica* **68**(5), 1281–1293.

Satterthwaite, M. (1975), 'Strategy-Proofness and Arrow Conditions: Existence and Correspondence Theorems for Voting Procedures and Welfare Functions', *Journal of Economic Theory* **10**(2), 187–217.

Smith, V. L. and J. M. Walker (1993), 'Monetary Rewards and Decision Costs in Experimental Economics', *Economic Inquiry* **31**(April), 245–261.

Taylor, L. O., M. McKee, S. K. Laury and R. G. Cummings (2001), 'Induced-Value Tests of the Referendum Voting Mechanism', *Economics Letters* **71**, 61–65.

Vossler, C. A., R. G. Ethier, G. L. Poe and M. P. Welsh (2003), 'Payment Certainty in Discrete Choice Contingent Valuation Responses: Results from a Field Validity Test', *Southern Economic Journal* **69**(4), 886–902.

Welsh, M. P. and G. L. Poe (1998), 'Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach', *Journal of Environmental Economics and Management* **36**, 170–185.

Whittington, D., V. K. Smith, A. Okorafor, A. Okore, J. L. Liu and A. McPhail (1992), 'Giving Respondents Time to Think in Contingent Valuation Studies: A Developing Country Application', *Journal of Environmental Economics and Management* **22**(3), 205–225.