# Using Respondent Uncertainty to Mitigate Hypothetical Bias in a Stated Choice Experiment

***Richard C. Ready, Patricia A. Champ, and Jennifer L. Lawton***

ABSTRACT. *In a choice experiment study, willingness to pay for a public good estimated from hypothetical choices was three times as large as willingness to pay estimated from choices requiring actual payment. This hypothetical bias was related to the stated level of certainty of respondents. We develop protocols to measure respondent certainty in the context of a choice experiment, and to calibrate hypothetical choices using these certainty measures. While both the measurement of respondent certainty and the use of certainty measures to calibrate responses are complicated by the multiple-choice nature of choice experiments, calibration successfully mitigated hypothetical bias in this application.* (JEL Q51)

## I. INTRODUCTION

Stated preference (SP) approaches to nonmarket valuation involve asking study participants to make hypothetical trade-off(s) between their wealth and the nonmarket good(s) of interest. Many studies have compared SP responses to actual behaviors involving choices with real money at stake. List and Gallet (2001) conducted a meta-analysis of results from 29 such studies and found that values estimated from hypothetical responses were three times as large (on average) as those estimated from observed behavior in choices involving real money commitments. Little and Berrens (2004) expanded List and Gallet's meta-analysis by adding more studies and including more variables that measured methodological differences among studies. They also found hypothetical willingness to pay (WTP) to be approximately three times that of actual WTP. Murphy et al. (2005) conducted a similar meta-analysis using only WTP studies. They found that hypothetical WTP was, on average, 2.6 times as large as actual WTP. The difference between hypothetical values and actual payment values is referred to as hypothetical bias.

There has been interest in finding ways to modify how SP questions are asked and/or calibrate the values obtained to eliminate or adjust for hypothetical bias. One approach, used primarily with the dichotomous choice (DC) contingent valuation (CV) format, involves identifying SP respondents who are unsure of their preferences. There is evidence that respondents who are unsure whether they would pay a specified amount in return for an increase in a public good tend to say yes to a DC CV valuation question (Ready, Navrud, and Dubourg 2001; Berrens et al. 2002). These unsure respondents can be identified with a follow-up question of the form, "How sure are you that you would choose the option you indicated?" (Li and Mattson 1995). A common approach to mitigating hypothetical bias is to recode DC CV "yes" responses where the respondent reports a low level of certainty to "no" responses, and accepting "yes" responses only if they are given with high confidence (Champ et al. 1997; Johannesson, Liljas, and Johansson 1998; Champ and Bishop 2001). Little and Berrens (2004), in their meta-analysis of validity studies, found that studies that calibrated responses based on the respon-

*Land Economics* • May 2010 • 86 (2): 363–381
ISSN 0023-7639; E-ISSN 1543-8325


The authors are, respectively, associate professor, Pennsylvania State University; economist, U.S. Forest Service, Rocky Mountain Research Center; statistical consultant, Pitney Bowes MapInfo, Ann Arbor. This research was partially funded through agreement 00-JV-11221617-198 with the U.S. Forest Service, Rocky Mountain Research Station.

dent's stated level of certainty had less hypothetical bias.

While this "certainty threshold" approach to calibration has been used primarily for the DC CV format, the choice experiment (CE) format[1] has increasingly been used to value multidimensional nonmarket goods. An advantage of the CE format is that it can be used to value both discrete changes in quantities of public goods and marginal changes in the attributes of those goods. Recent split sample criterion validity studies have found that hypothetical bias can exist for the CE format as well (e.g., Lusk and Schroeder 2004; List, Sinha, and Taylor 2006). Could hypothetical bias in CEs also be related to respondent uncertainty? When more than two options are presented to a respondent, measurement of respondent uncertainty becomes more complicated, as does application of certainty threshold calibration. To see this, consider a respondent who chooses one option out of three available, but states low confidence in that choice. Should that respondent be recoded to one of the other two options? Which one? Should calibration depend on whether the chosen option has a higher cost than the other options?

In this paper we explore these issues. The specific objectives of this study are to (1) construct a choice experiment with hypothetical and real payment treatments and determine whether hypothetical bias exists in the experiment, (2) develop a method for measuring respondent uncertainty in CE surveys, (3) determine whether hypothetical bias is related to respondent uncertainty, and (4) determine whether respondent statements of uncertainty can be used to mitigate hypothetical bias in this application.

## II. PREVIOUS RESEARCH ON SP RESPONDENT UNCERTAINTY

Why might SP respondents be uncertain over their behavior or their preferences? First, there may be details of the valuation scenario that are not completely described, including the exact characteristics of the goods and how they would be provided (Hanemann and Kriström 1995). Second, the respondent may have insufficient time to evaluate his own preferences, insufficient experience with the good, or insufficient motivation to invest the time and effort needed to fully consider the choice task and optimize over his preferences (Alberini, Boyle, and Welsh 2003; Loomis and Ekstrand 1998). It has been argued that some trade-offs are inherently impossible to make with precision, such as trade-offs that involve moral issues, but that even in those cases respondents can place upper and lower bounds on their WTP (Opaluch and Segerson 1989; Kooten, Krcmar, and Bulte 2001).

### *Respondent Uncertainty in DC Valuation*

Traditional SP questions do not accommodate uncertainty. For example, in the DC CV format, the only allowable responses are "yes" and "no." A respondent who is unsure what choice she would make if the decision context were real is forced into one of those two responses. Several studies (Welsh and Poe 1998; Loomis and Ekstrand 1998; Ready, Navrud, and Dubourg 2001; Berrens et al. 2002) have found that DC respondents who self-identify as being unsure over their preferences tend to say "yes" to such a forcing question. The authors speculate that this form of yea-saying could be responsible for much of the hypothetical bias seen in DC values.

If some DC CV "yes" responses are made by respondents who are actually unsure over their preferences, it may be possible to identify those individuals and adjust (calibrate) their responses to account for that uncertainty. Several studies have attempted to do so using information from certainty follow-up questions. One of the first was that of Li and Mattson (1995), who asked all DC respondents how certain they were of their DC response, on a scale from 0% to 100%. They used these responses as state-

[1] This format is also referred to, variously, as the attribute-based method, the stated choice format, and sometimes conjoint analysis, though that term is also used for techniques that rate options.

ments of probability, so that a respondent who said "yes" with a confidence level of 80 was interpreted as having an 80% probability of actually choosing "yes" and a 20% likelihood of actually choosing "no." They adjusted the usual likelihood function to include both possible outcomes and their probabilities for each respondent. They found that accounting for respondent uncertainty decreased estimated WTP values by 50% relative to the standard approach to analyzing DC responses.

Several subsequent studies have explored how to best measure respondent uncertainty and use those measurements to adjust estimated WTP values (Loomis and Ekstrand 1998; Berrens et al. 2002). Others have investigated the role that respondent uncertainty plays in motivating differences among elicitation formats in estimated WTP (Welsh and Poe 1998; Ready, Navrud, and Dubourg 2001). While the findings of these studies generally support the hypothesis that respondent uncertainty can lead to hypothetical bias, they lack external validity criteria against which their WTP estimates can be compared.

Champ et al.'s (1997) was the first study to include certainty follow-up questions in a split sample study with hypothetical and real money treatments. In the context of donations for a public good, they asked DC hypothetical payment respondents how certain they were of their response, with a 1- to 10-point scale from "very uncertain" to "very certain." They found the proportion of respondents who said that they would donate in the hypothetical treatment was larger than the proportion who actually donated. However, if they applied a certainty threshold of 10, that is, counted as positive responses only the "yes" responses with a stated certainty level of 10, the hypothetical WTP estimate was not statistically different from the actual donation WTP estimate. This type of adjustment, where only the highest response level is considered a positive response, is called a "top box" analysis in the marketing literature.

Champ and Bishop (2001) applied a similar approach and found that a certainty threshold of 8 results in equivalence between the hypothetical and real treatments. They also found that the characteristics of respondents who met this certainty threshold closely matched the characteristics of those who donated in the actual treatment. They take this as evidence that the certainty statements can be used to identify which individual respondents in the hypothetical treatment would actually make a donation if placed into a real choice situation.

Other validity studies have used the certainty threshold approach and have found that it can mitigate hypothetical bias, though there is some variability in the threshold level that provides the closest match between hypothetical and actual WTP values. While many studies have found that the certainty threshold that best mitigates hypothetical bias is at the very the top of the certainty scale (Blomquist, Blumenschein, and Johannesson 2008; Champ et al. 1997), others have found that a somewhat lower threshold works best (Champ and Bishop 2001; Ethier et al. 2000; Vossler et al. 2003; Poe et al. 2002). All studies found that hypothetical bias existed, and that calibration using a fairly strict certainty threshold (7 or higher on a scale of 10) was needed to mitigate that bias. Studies that used verbal descriptions of certainty levels instead of ordinal scales find similar results (Blumenschein et al. 1998, 2008; Johannesson, Liljas, and Johansson 1998).

An alternative to the certainty threshold approach to calibration is to estimate a statistical bias function (Johannesson et al. 1999; Blomquist, Blumenschein, and Johannesson 2008). This can be done only when hypothetical and real choices are observed for the same respondents. For respondents who said "yes" to the hypothetical DC question, a probit regression is estimated that predicts the probability the respondent will actually choose the yes option when faced with a real choice. The response to a certainty follow-up to the hypothetical DC question is used as an explanatory variable in the probit regression. Johannesson et al. (1999) found that the certainty follow-up response was positively related to the probability of an actual "yes" choice. They also found that the probability of an inconsistent response was higher for higher prices. They used the probit regression to predict actual behavior based on the hypothetical responses and the certainty statements and found that such a calibration successfully mitigated the hypothetical bias. They also found that a common statistical bias function could be estimated for two different private goods. Blomquist, Blumenschein, and Johannesson (2008) updated Johannesson et al.'s bias function by including additional data for a third good, a health treatment protocol.

Note that calibration using a certainty threshold or statistical bias function is not made for hypothetical "no" responses. Studies that have compared hypothetical and actual behavior for the same respondent find that hypothetical "no" respondents rarely choose the "yes" option when provided the opportunity in a real payment situation (e.g., Johannesson et al. 1999). There appears to be an asymmetry in how respondent uncertainty biases hypothetical responses.

*Hypothetical Bias in CE Valuation*

The few CE studies that have evaluated hypothetical bias have largely reported results consistent with the DC hypothetical bias studies. These studies can be divided into studies that include an opt-out "no purchase" option and those that do not. Studies valuing attributes of private goods usually include an opt-out option, in which case it is possible to estimate the probability of purchase of the good as well as the marginal WTP for changes in attributes of the goods. When an opt-out option is not provided, it is only possible to estimate marginal WTP for attributes of the good.

Lusk and Schroeder (2004) compare hypothetical and real choices over purchase of steaks with different qualities. They found that the probability of buying a steak was significantly greater in the hypothetical treatment than in the real payment treatment, but that marginal WTP for changes in steak attributes were similar between the two treatments. Similarly, List, Sinha, and Taylor (2006) found that the probability of buying a sports trading card was overstated in hypothetical choices, but that marginal WTP for improvements in card quality were similar in the hypothetical and real treatments.

For a public good, List, Sinha, and Taylor found that hypothetical CE respondents overstated their likelihood of donating for the good by 60%. Likewise, respondents in the hypothetical payment treatment who chose to donate also overstated the probability of choosing the more expensive donation (versus the less expensive donation) by 100%.

Taylor, Morrison, and Boyle (2007) found, for both a private good and a public good, that hypothetical bias existed in the proportion of stated choice respondents who opted in (i.e., chose an option with positive cost), and that the magnitude of the bias was greater when the hypothetical treatment used a provision rule that was not incentive compatible. When the provision rule was not incentive compatible, marginal WTP estimates from hypothetical choices were also biased upward by 30% to 35% relative to actual payment choices. They found mixed results on whether marginal WTP estimates from hypothetical choices were biased when an incentive compatible provision rule was used.

Cameron et al. (2002) found that the proportion of CE respondents who chose to participate in a green energy program was higher in the hypothetical payment treatment than in the actual payment treatment, and that WTP estimates from hypothetical payment responses exceeded those estimated from actual payments, but the differences in WTP were not statistically significant. Because they were limited to only one actual payment scenario, Cameron et al. could not explore whether marginal WTP for attributes differed between the real and hypothetical payment treatments.

Turning to studies that did not include an opt-out option, Carlsson and Martinsson (2001) asked respondents about hypothetical choices over different donations to wildlife programs, followed by identical real choices. In this within-sample comparison, they found no significant differences in marginal WTP for attributes of the donation options. Johansson-Stenman and Svedsäter (2007) essentially repeated Carlsson and Martinsson's experiment (with different money amounts and programs) but used a split-sample design with a real-only treatment and a hypothetical then real treatment. In contrast to Carlsson and Martinsson, Johansson-Stenman and Svedsäter found that marginal WTP for attributes estimated from hypothetical choices were significantly larger than those from the real-only treatment. Alfnes and Rickertsen (2002) compared hypothetical and real choices over steaks with different characteristics. They found that marginal WTP for steak attributes estimated from hypothetical payment choices were 5 to 11 times higher than those estimated from real second-price auctions. Blamey and Bennett (2001) compared hypothetical choices over toilet paper purchases to real purchases obtained from scanner data. They found that parameters of models estimated from hypothetical and real choices differed, but that the hypothetical models did a fairly good job predicting the aggregate market share of products with green attributes. Due to limited variation in the set of real products available, they were unable to compare marginal WTP for product attributes between the hypothetical and real purchase treatments.

To summarize, in almost all studies, hypothetical CE responses generated larger WTP estimates than did choices involving actual payments. There is some evidence that, when the study includes an opt-out option, hypothetical bias may be more of an issue in determining the proportion of respondents who choose an option that costs money and in estimating aggregate WTP for a unit of the good than it is in choices among the costly options and estimates of marginal WTP for attributes of the good, though that evidence is mixed.

*Respondent Uncertainty in CE Valuation*

Only two studies have been conducted that attempted to calibrate CE responses based on certainty follow-up questions. Norwood (2005) compared hypothetical payment CE responses to actual payment DC choices in a public good experiment with undergraduate students, where the contributions to and payoff from the public good were points toward their grade. Each hypothetical choice question had two donation options and an opt-out option. Norwood found that a random utility model estimated from hypothetical CE responses overstated the proportion of students who would contribute to the public good, as compared to the actual payment DC treatment.

For respondents who chose one of the two donation options in a CE question, a follow-up question asked how certain the respondent was that she would actually donate the amount indicated to the public good, if given the opportunity. Unfortunately, this form of certainty follow-up is somewhat ambiguous. Respondents could interpret the question as asking, "If given an actual choice among these three options, how sure are you that you would actually choose the option you indicated?" Alternatively, it could be read as, "If given the opportunity to actually make the donation you indicated, how sure are you that you would actually choose to make the donation?" In the first case, the choice is among all three options, while in the second it is between two. A respondent might be highly certain that he would make a donation, but unsure which of the donation choices he prefers. Such a respondent would answer the first form of the question with a low level of confidence, but answer the second with a high level. It is not clear how respondents interpreted the question.

Given a low level of confidence, how should a CE response be recalibrated? In a three-option choice, the respondent could be reassigned to either of the other two options. Norwood chose to recode all low-certainty CE responses to the opt-out option. Using this calibration procedure, Norwood found that, with a certainty threshold of 6, the estimated random utility model predicted a donation rate that matched the actual donation rate from the real treatment. Note that this calibration protocol assumes away the possibility that unsure respondents who indicate in the hypothetical choice question that they would choose one costly option would actually chose the other costly option when faced with a real payments choice.

Olsson (2005), in a survey valuing fish stocks and water quality in Sweden, used a similar protocol for measuring respondent uncertainty in a CE. He asked only one certainty follow-up question, after all CE questions were answered. He found that WTP estimates based on the CE questions were much larger than those from open-ended or DC valuation questions. Due to the nature of the good, he could not have a real payment treatment.

Olsson found that applying a certainty threshold protocol similar to that used by Norwood (where uncertain respondents are recoded to the opt-out option), decreased estimated WTP for the program, but increased marginal WTP for attributes of the program. This result is counterintuitive. None of the previous CE studies with actual payment treatments found that marginal WTP values from hypothetical choices were smaller than those estimated from real choices. A protocol designed to eliminate hypothetical bias should therefore not increase marginal WTP estimates. These results suggests that recoding all uncertain respondents to the opt-out option gives up important information about the marginal effects of attributes.

In this paper we develop new ways to measure respondent uncertainty in CEs that distinguish between uncertainty over whether the respondent would opt out or opt in from uncertainty over which opt-in option she would choose. We also develop protocols for using that information to calibrate CE responses in ways that allow switching from one costly option to another costly option.

## III. METHODS

### *Measuring Respondent Uncertainty in CEs*

Holmes and Adamowicz (2003) provide a good overview of the theory and practice of choice experiments. In each CE question, respondents are presented with sets of choice options that vary in the levels of several attributes, and are asked which option they prefer from each set. The utility that respondent $i$ receives from option $j$ is typically assumed to be a linear function of the option's attribute levels, $A_j$. Extending the usual model, we consider the possibility that the utility from each choice includes two random components, so that the utility to respondent $i$ from choice $j$ is given by

$$U_{ij} = \sum_{k=1}^{1} \beta_k A_{jk} + \beta_p p_j + \varepsilon_{ij} + \eta_{ij}, \qquad [1]$$

where $\beta_k$ is the marginal utility of attribute $k$, $A_{jk}$ is the level of nonprice attribute $k$ in alternative $j$, and $\beta_p$ is the marginal utility of the price attribute, $p_j$. Since an increase in price decreases income, $-\beta_p$ is the marginal utility of income. The first random component, $\varepsilon_{ij}$, captures individual-specific factors that are known to the respondent but unknown to the researcher. The second random component, $\eta_{ij}$, reflects the respondent's own uncertainty over her preferences.[2]

Prior to making a binding choice, the respondent may not be sure which option she will actually choose. However, she can evaluate her probability of choosing each option. If $\eta_{ij}$ is independent and identically distributed according to a Gumbel distribution, then her subjective probability of choosing option $j$ from choice set $C$ is given by

[2] The reader may be concerned that, with two independent error components, the multinomial logit model will no longer be appropriate. The distribution of a sum of two Gumbel-distributed random terms can be very closely approximated by another Gumbel distribution. If the error terms are normally distributed, then their sum will be normally distributed, and a multinomial probit will still be appropriate.

$$P(j) = \frac{\exp\left(\sum_k \beta_k x_{jk} + \beta_p p_j + \varepsilon_{ij}\right)}{\sum_{m \in C} \exp\left(\beta_k x_{mk} + \beta_p p_m + \varepsilon_{im}\right)}. \qquad [2]$$

Note that the respondent knows the values of $\varepsilon$ when evaluating these probabilities. Ideally, when asked a hypothetical choice question, the respondent would choose the option with the highest subjective probability. However, experimental evidence on respondent behavior in DC settings suggests that respondents who are forced to make a hypothetical choice without knowing their values of $\eta$ tend to choose more costly options than they choose in actual payment situations.

If we wish to calibrate hypothetical choices using information about respondent uncertainty, we need to be able to measure that uncertainty. One approach would be to have the respondent report his subjective probability for each of the options presented. This would likely be a difficult task for the respondent, however. To make the respondent's task easier, we stay with questions of the form, "How sure are you that . . ." Experience with DC formats is that respondents believe that they are able to answer that type of question reliably. A question of the form, "How sure are you of your previous answer?" provides information about *P*(*j*) only for the chosen option. In order to fully recover certainty over all options, though, it is necessary to ask more than one of these questions.

The choice sets used in this study include three options: two costly options, which we call the opt-in options, and a no cost (opt-out) option. These will be respectively referred to as options A, B, and O, though they were not labeled as such in the survey instrument. There are several possible sequences of follow-up certainty questions that should generate similar information about respondent uncertainty. In this study, we employed two different question sequences.

Our first question sequence (sequence S1) is as follows. For respondents who chose option A or B in a hypothetical choice, the respondent was first asked a follow-up question of the form[3]

Q1-S1: How sure are you that you would choose the option you indicated, instead of one of the other two options?

This provides information on P(A) or P(B), depending on which costly option was chosen. We call this the respondent's "unconditional choice certainty." This question is similar to but more specific than those used by Norwood and Olsson. If the respondent indicated a level of certainty in Q1-S1 less than 100%, she was then asked a question of the form

Q2-S1: How sure are you that you would choose either option A or option B, instead of option O?

This provides information on P(A)+P(B). We call this the respondent's "opt-in certainty."

The second sequence of certainty questions (sequence S2) is as follows. For respondents who chose option A or B, the first follow-up question was of the form

Q1-S2: You indicated that you would choose one of the two options that would cost you money. How sure are you that you would choose option A (B), instead of option B (A)?

This provides information on P(A|A∪B) or P(B|A∪B). We call this the respondent's "conditional choice certainty." It is conditional on choosing one of the two costly choices. Regardless of the answer to this question, the respondent was then asked the same opt-in certainty question as in sequence 1, that is, Q2-S2 took the same form as Q2-S1.

In both sequence S1 and sequence S2, respondents who initially chose the opt-out option, O, were asked a follow-up question of the form

Q3: How sure are you that you would choose option O, instead of one of the other two options?

We call this the respondent's "opt-out certainty." It provides information on P(O).

---

[3] Actual wording of the questions and layout of the response screens are available from the authors.

TABLE 1
CERTAINTY TYPES FOR A CHOICE QUESTION WITH TWO COSTLY OPTIONS, A AND B, AND A COSTLESS OPT-OUT OPTION, O

| Option Chosen | Question No. | Certainty Type | Example Question | Probability Measure | No. of Cases Version 1 | No. of Cases Version 2 |
|---|---|---|---|---|---|---|
| A or B | Q1-S1 | Unconditional choice certainty | How sure are you that you would choose A instead of B or O? | P(A) or P(B) | 160 | — |
| A or B | Q1-S2 | Conditional choice certainty | How sure are you that you would choose A instead of B? | $P(A \mid A \cup B)$ or $P(B \mid A \cup B)$ | — | 139 |
| A or B | Q2-S1 and Q2-S2 | Opt-in certainty | How sure are you that you would choose A or B instead of O? | P(A) + P(B) | 153 | 139 |
| O | Q3 | Opt-out certainty | How sure are you that you would choose O instead of A or B? | P(O) | 168 | 193 |

The certainty follow-up questions are summarized in Table 1.

Other question sequences are possible, but pretests showed that respondents found the questions used in these two sequences easy to understand and intuitive.

*Study Design*

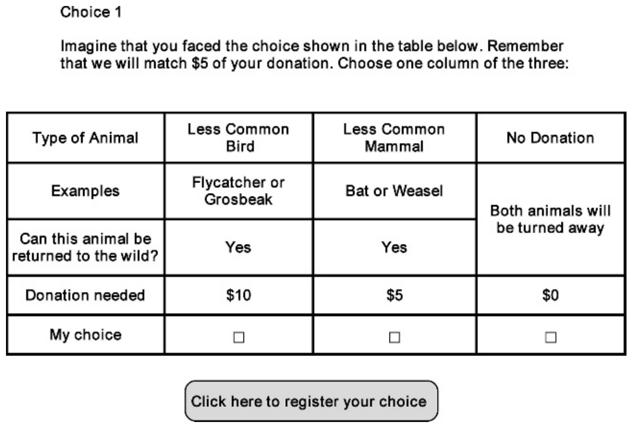
The study was conducted using undergraduate students at Pennsylvania State University. The public good valued was wildlife rehabilitation. A state-licensed wildlife rehabilitator, Centre Wildlife Care (CWC), collaborated on the study. CWC accepts and cares for injured or abandoned wild animals, with the goal of releasing them back into the wild. If this is not possible, the animals are cared for in captivity for life. The operations of CWC are funded entirely through donations of time and money. Not all animals can be accepted for rehabilitation due to time and money constraints.

In the CEs, respondents were offered the opportunity to donate money sufficient to care for one animal. In each choice question, the respondent was offered two donation options and an opt-out option. Respondents were told that CWC must turn away animals because of limited resources. Respondents were told that, if they chose a donation option, the money would be used to rehabilitate one animal of the type they chose. If they chose not to donate, that animal would be turned away and would presumably die. CWC agreed to use the donations for the types of animals indicated, so the respondents choices had real consequences.

The option attributes were type of animal (mammal, bird, or turtle), common versus less common, whether the animal can eventually be returned to the wild, and the donation required of the participant, which took values of 0 for the opt-out option and \$5 or \$10 for the opt-in options.[4] So that the amount donated equaled the actual cost of rehabilitation (cost estimates were obtained from CWC), the study provided a donation match of \$5 for every donation made by a respondent. An example of a hypothetical choice is shown in Figure 1.

The good used in this study—rehabilitation of an injured or abandoned wild animal—is a quasi-public good. It is nonrival in that everyone can benefit from it once it is provided, but it is excludable in that the respondent will benefit from that unit of the good only if she pays the donation. Further, in contrast to quasi-public goods used in previous studies (e.g., Taylor, Morrison, and Boyle 2007; Cum-

[4] Based on pretests, no snakes or amphibians were included (due to revulsion toward those animals), and no endangered species were included (due to overwhelming preference for those animals).


Choice 1

Imagine that you faced the choice shown in the table below. Remember that we will match $5 of your donation. Choose one column of the three:

| Type of Animal | Less Common Bird | Less Common Mammal | No Donation |
|---|---|---|---|
| Examples | Flycatcher or Grosbeak | Bat or Weasel | Both animals will be turned away |
| Can this animal be returned to the wild? | Yes | Yes | |
| Donation needed | $10 | $5 | $0 |
| My choice | ☐ | ☐ | ☐ |

Click here to register your choice

FIGURE 1
EXAMPLE OF A HYPOTHETICAL CHOICE SET

mings and Taylor 1999; Champ et al. 1997), the good used in this study is time sensitive. The provision rule was clearly explained to respondents. If a donation is not made, an animal will be turned away that would otherwise be rehabilitated. The respondent cannot free-ride on the future donations by others for that specific animal. For a respondent who values rehabilitation for that specific animal more than the posted cost, the optimal strategy is to make the donation, and vice versa. Therefore, even though the payment vehicle was described to respondents as a donation, and it would legally be considered a donation for tax purposes, it functioned as a posted price for an excludable good.

After each hypothetical choice, the certainty follow-up questions were asked. For each certainty question, respondents indicated their level of certainty with a slider bar (Figure 2). The choice set was shown with coloring and an arrow or arrows, and the labels next to the slider bar were customized to clearly indicate the option or options to which the certainty question applied.

Respondents were paid $20 to participate. Surveys were conducted individually in a laboratory on a computer. In each choice task, respondent was presented with four choice sets. The same set of choices was used for hypothetical and real tasks. For real choice tasks, respondents were told that one of the four choices would be binding, and would be randomly selected at the end of the survey. At the end of the survey, the administrator determined which choice was binding and subtracted any donation from the $20 incentive.

The study include three survey treatments: hypothetical choices with certainty follow-up sequence S1 followed by real choices (treatment H1), hypothetical choices with certainty follow-up sequence S2 followed by real choices (treatment H2), and real choices only (treatment R).

## IV. RESULTS

A total of 249 surveys were completed, 82 in the H1 treatment, 83 in the H2 treatment, and 84 in the R treatment. The computer-based survey would not allow participants to skip questions, and all who started the survey completed it.

Table 2 shows the proportion of choices where the respondent chose an opt-in option for hypothetical and real choices in each treatment. A Pearson chi-square test was conducted to test the null hypothesis that the frequency of commitments for Animal A, Animal B, or No Donation followed a common distribution across the three real commitment datasets and across the two hypothetical commitment datasets.

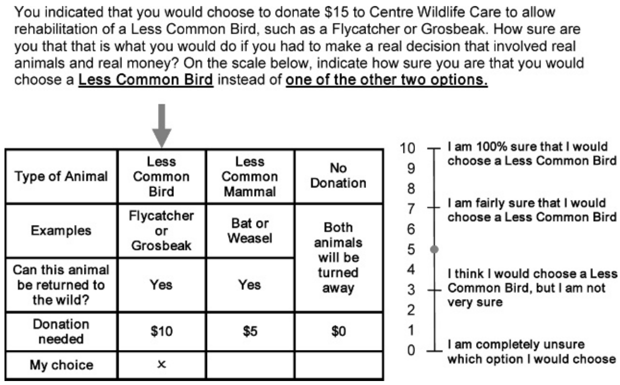



FIGURE 2
Certainty Follow-up Question Q1-S1

For all four choices, for both real and for hypothetical datasets, the test failed to reject the null hypothesis of common choice behavior. Similarly, tests conducted on the opt-in rates (i.e., without distinguishing between Animal A and Animal B) showed no statistical differences across treatments. There is no evidence, therefore, of treatment effects within either the hypothetical or the real treatments.

*Does Hypothetical Bias Exist?*

However, there were large differences in the opt-in rates between the hypothetical treatments and the real treatments (Table 2). The opt-in rate for hypothetical choices was about three times as large, on average, as for choices involving real payments. This difference was statistically significant at the 1% level, regardless of whether the test was conducted using opt-in rates for individual questions or the average opt-in rate for all four questions combined. This is clear evidence of hypothetical bias

TABLE 2
Hypothetical and Real Opt-in Rates

| Treatment | Opt-in Rate: Hypothetical Choices | Opt-in Rate: Real Choices |
|---|---|---|
| R ($n$ = 336) | — | 14.3% |
| H1 ($n$ = 328) | 48.8% | 15.2% |
| H2 ($n$ = 332) | 41.9% | 16.0% |

and demonstrates bias both within sample and between samples.

Behavior in the real choices in treatments H1 and H2 was similar to that in treatment R, so that there does not appear to be any sequencing effect. It is therefore useful to explore within-respondent patterns of behavior. Table 3 shows how many times respondents in the H1 and H2 treatments followed each of five response patterns. Of respondents who chose to opt in in the hypothetical choice, only 26.8% chose to donate for that same animal when presented with a real choice. Most who opted in in the hypothetical choice opted out in the real choice (65.6%), but there were some respondents who chose one animal in the hypothetical choice but then chose the other animal in the real choice (7%). The calibration method used by Norwood and Olsson would have incorrectly reclassified these respondents to the opt-out option.

In contrast, almost all respondents who opted out in the hypothetical choices opted out when faced with real choices (99.5%). This result is similar to that seen in the DC CV studies and implies any calibration protocol should not change hypothetical opt-out responses, regardless of the level of certainty associated with those responses.

Multinomial logit (MNL) models were estimated for the hypothetical and real payment choices, pooled across treatments. When all attributes in the design were included in the model, none were found to

TABLE 3
PATTERNS OF RESPONSES IN HYPOTHETICAL TREATMENTS

| Hypothetical Choice | Real Choice | Category | Proportion |
|---|---|---|---|
| Opts in ($n$ = 299) | | | |
| | Opts in, and chooses same animal | Yes-Yes-Same (YYS) | 26.8% |
| | Opts in, but chooses other animal | Yes-Yes-Different (YYD) | 7.7% |
| | Opts out | Yes-No (YN) | 65.6% |
| Opts out ($n$ = 361) | | | |
| | Opts in | No-Yes (NY) | 0.5% |
| | Opts out | No-No (NN) | 99.5% |

be statistically significant other than donation amount. This may be because of heterogeneity in preferences. For example, some respondents may prefer a bird, while others prefer a mammal. A simple model was estimated that included only the donation amount (COST; \$5 or \$10), a common alternative specific constant for the two options that result in rehabilitating an animal (ANIMAL; =1 for the two opt-in options, 0 for the opt-out option), and the ANIMAL alternative specific constant interacted with a measure of how high a priority the respondent placed on rehabilitating wildlife (PRIORITY; ranging from 0 = low priority to 10 = high priority).

Table 4 presents the MNL results. The estimated model based on real payment choices is presented in the first results column of Table 4, while the estimated model based on hypothetical payment choices is presented in the second column. In both cases, all parameter estimates are statistically significant at the 1% level. The most important difference between the real and hypothetical models is that the marginal disutility of COST estimated from real choices is over twice as large as that estimated from hypothetical choices. A likelihood ratio test showed that the two models differed at the 1% level of significance. This result also held true if only the responses from treatment R (the real-only treatment) were used to estimate the model for real choices.

For both the real payment and hypothetical payment models, utility from rehabilitating an animal increases with the PRIORITY score. Evaluating PRIORITY at its mean (6.43), it is possible to calculate the mean WTP to rehabilitate one animal. If we assume that WTP is bounded from below at 0, then mean WTP estimated from the hypothetical choices was \$5.29, versus \$1.68 from real choices. A Monte Carlo approach was used to simulate sampling variability in the mean WTP estimates and generate 95% confidence intervals. Using Poe, Severance-Lossin, and Welsh's (1994) methods of convolutions, the difference between real and hypothetical WTP estimates was found to be statistically significant at the 1% level. The same results were obtained when the real choices only from treatment R were used.

While the estimated model does not allow us to calculate marginal WTP for changes in the attributes of the good, it does allow us to calculate marginal changes in WTP associated with a change in the respondent's characteristics. Here, a one-unit increase in the PRIORITY score increased WTP estimated from real payment choices by \$0.55, but increased WTP estimated from hypothetical payment choices by \$1.33, a statistically significant difference at the 1% level. We therefore see hypothetical bias in both total WTP and marginal WTP. This result is inconsistent with the findings of Lusk and Schroeder (2004) and of List, Sinha, and Taylor (2006), who found hypothetical bias in total WTP but not in marginal WTP.

*Is Hypothetical Bias Related to Respondent Uncertainty?*

Having established that hypothetical bias is present, the next research objective is to examine the relationship, if any, between respondent uncertainty and hypothetical bias. Analysis of the certainty responses from the H1 treatment showed internal

TABLE 4

MULTINOMIAL LOGIT ESTIMATION RESULTS, DEPENDENT VARIABLE = OPTION CHOSEN

| Variable | Description | All Treatments: Real Choices | Treatments H1 and H2: Hypothetical Choices |
|---|---|---|---|
| ANIMAL | 1 = opt-in choice<br>0 = opt out | −2.713*** (0.462) | −2.023*** (0.355) |
| PRIORITY * ANIMAL | 0 = low priority<br>10 = high priority | 0.388*** (0.051) | 0.332*** (0.043) |
| COST | \$0, \$5, or \$10 | −0.351*** (0.039) | −0.142*** (0.025) |
| Sample size | | 996 | 660 |
| Log likelihood | | −447.059 | −612.510 |
| Mean WTP (95% CI) | | \$1.68 (\$1.35, \$2.07) | \$5.29 (\$4.65, \$6.40) |

*Note:* Standard error in parentheses. WTP, willingness to pay.
*** Significant at $p \leq 0.01$.

inconsistencies. As constructed, a respondent's opt-in certainty should be higher than his unconditional choice certainty. That is, you should be more certain that you will choose some animal than that you will choose a specific animal. However, many H1 respondents expressed higher levels of certainty about which animal they would rehabilitate compared to whether they would donate at all. We conclude that the sequence S1 of certainty follow-up questions used in the H1 treatment does not work well. We believe that respondents were confused about the distinction between opt-in certainty and unconditional choice certainty. We do not use the results from the unconditional choice certainty question in subsequent analysis. Because the opt-in certainty question in treatment H1 was identical to that used in treatment H2, we continue to consider the results from that question in our analysis.

Table 5 shows the average certainty levels for each response pattern for both hypothetical treatments. Respondents who opted in in the hypothetical choice, and then opted in in the real choice (YYS or YYD) had higher opt-in certainty than those who opted out in the real choice (YN). Further, respondents who chose the same donation option in the real treatment (YYS) had higher conditional choice certainty than those who chose the other donation option (YYD). Information from these follow-up questions is likely, therefore, to be useful in predicting which respondents will follow which response pattern. Meanwhile, NN respondents had higher opt-out certainty than NY respondents, but there were very few of the latter.

A probit regression analogous to Johannesson et al.'s (1999) statistical bias function was estimated for choices where the respondent opted in in the hypothetical choice. It predicts the probability of a YYS or YYD pattern versus a YN pattern. The results are shown in Table 6. All estimated parameters are statistically significant at the 1% level. For both hypothetical treatments, the probability of opting in in the real choice was higher as the opt-in certainty increased. A likelihood ratio test showed no difference in the model parameters between the two treatments. This provides strong evidence that stated opt-in certainty can be used to predict which hypothetical respondents will actually opt in when faced with a real choice.

TABLE 5

MEAN CERTAINTY LEVELS BY TREATMENT AND RESPONSE PATTERN

| Response Pattern | Treatment H1: Opt-in Certainty | Treatment H2: Opt-in Certainty | Treatment H2: Conditional Choice Certainty |
|---|---|---|---|
| YYS | 6.7 | 8.1 | 7.5 |
| YYD | 7.4 | 6.1 | 5.5 |
| YN | 4.7 | 4.6 | 5.2 |
| NN | 6.8 | 7.3 | — |
| NY | 4.9 | 1.0 | — |

*Note:* YYS, yes-yes-same; YYD, yes-yes-different; YN, yes-no; NN, no-no; NY, no-yes.

TABLE 4
*(Extended)*

| Treatment H2: Hypothetical Choices | Treatment H2: Calibrated Hypothetical Choices | |
|---|---|---|
| | Without Option Switching | With Option Switching |
| −2.249*** (0.488) | −4.306*** (0.733) | −3.412*** (0.714) |
| 0.344*** (0.059) | 0.380*** (0.077) | 0.384*** (0.078) |
| −0.132*** (0.036) | −0.085 (0.056) | −0.210*** (0.057) |
| 332 | 332 | 332 |
| −296.154 | −172.273 | −166.402 |
| $5.12 ($4.24, $7.61) | $1.71 (undefined) | $1.56 ($1.16, $2.23) |

The second issue to address is whether we can identify which respondents will follow the YYD pattern versus the YYS pattern. Here we use the conditional choice certainty measures from the H2 treatment. Initial analysis of the pattern of responses showed that the YYD pattern was more likely in situations where the option chosen in the hypothetical choice was more expensive than the other opt-in option. To capture this effect, a relative cost dummy variable was constructed equal to 1 if the option chosen in the hypothetical choice was more expensive than the other opt-in option, and 0 if it was less expensive or the same cost.

A probit regression predicting the probability of a YYS pattern versus a YYD pattern was estimated for the H2 treatment. Results are shown in Table 7. The probability that the respondent will stick with the same animal in the real choice, versus switching to the other animal, was higher if the respondent stated higher conditional choice certainty, and was lower if the initial animal chosen was more expensive. Both effects were statistically significant at the 1% level.

Based on these results, there is strong evidence that respondent uncertainty is related to inconsistent choices, in this case YN and YYD. Respondents who opted in in the hypothetical choice but had lower stated opt-in certainty were more likely to switch to the opt-out option in the real choice. Respondents who opted in in the hypothetical choice but had lower conditional choice certainty and/or originally chose an animal with higher cost were more likely to donate for the other animal when presented with the real choice.

These results, which are consistent with the findings of Johannesson et al. (1999) and Blomquist, Blumenschein, and Johannesson (2008), are useful in that they resolve a criticism of using respondent certainty measures to calibrate hypothetical responses. This criticism can be best described in the context of DC CV. If the proportion of respondents who say "yes" in a hypothetical treatment is larger than the proportion who actually purchase the good in a treatment with real payments, then for any continuous measurable attribute of the respondent, there is some threshold value

TABLE 6
Probit Regression Model: Dependent Variable = 1 if YYS or YYD Response Pattern, = 0 if YN Response Pattern

| Variable | Description | Treatment H1 | Treatment H2 | Pooled |
|---|---|---|---|---|
| Intercept | =1 | −1.812*** (0.299) | −1.910*** (0.318) | −1.874*** (0.217) |
| Opt-in certainty | 0 = completely unsure<br>10 = 100% sure | 0.23*** (0.05) | 0.26*** (.046) | 0.25*** (0.03) |
| Sample size | | 160 | 139 | 299 |
| Log likelihood | | −85.156 | −72.966 | −158.481 |

*Note:* Standard error in parentheses. YYS, yes-yes-same; YYD, yes-yes-different; YN, yes-no.
*** Significant at $p \leq 0.01$.

TABLE 7

PROBIT REGRESSION MODEL: DEPENDENT VARIABLE = 1 IF YYS RESPONSE PATTERN, = 0 IF YYD RESPONSE PATTERN

| Variable | Description | Parameter Estimate |
|---|---|---|
| Intercept | =1 | −0.068 (0.580) |
| Conditional choice certainty | 0 = completely unsure; 10 = 100% sure | 0.17** (0.08) |
| Relative cost dummy | =1 if option chosen is more expensive than other option | −0.923** (0.419) |
| Sample size | | 52 |
| Log likelihood | | −23.761 |

*Note:* Standard error in parentheses. YYS, yes-yes-same; YYD, yes-yes-different.
** Significant at $p \leq 0.05$.

that, if used to calibrate the hypothetical responses, would give the "correct" proportion of yes responses. That this can be done using respondent certainty as the calibrating metric does not prove that hypothetical bias is related to respondent uncertainty. However, the within-sample results in Tables 6 and 7 show that respondent uncertainty is indeed related to hypothetical bias, affirming the use of respondent certainty as a calibration metric.

*Can Hypothetical Choices Be Calibrated Using Respondent Uncertainty Measures?*

Respondents who give hypothetical answers that are inconsistent with their real behavior tend to state lower levels of certainty in their choices. The certainty follow-up measures might therefore be useful in calibrating hypothetical choices. We consider two different calibration protocols. The first is similar to that used by Norwood and by Olsson. In this protocol, hypothetical choices where the respondent opted in but had low opt-in certainty are recoded to the opt-out option, but choices are never switched from one opt-in option to the other opt-in option. A certainty threshold of 7 or higher was used, because it best matched the actual donation rate.[5]

The second calibration rule allows for the possibility of switching from one opt-in option in the hypothetical choice to the other opt-in option in the real choice. The calibration rule used here is shown diagrammatically in Figure 3. For hypothetical opt-in choices, there is first a decision whether the respondent would opt in in the real choice situation. This rule is the same as in the first protocol; respondents with low opt-in certainty are recoded to the opt-out option. For hypothetical opt-in choices with high opt-in certainty, however, a second determination must be made whether the respondent would switch to the other opt-in option (i.e., follow a YYD pattern). Information on the conditional choice certainty is used to identify these respondents, but the switching rule varies depending on the relative cost of the two opt-in options.

Three cases exist. The first case is where the opt-in option chosen in the hypothetical choice is less expensive than the other opt-in option. In treatments H1 and H2, such respondents almost never followed a YYD pattern (this occurred in only 1 response out of 158). Accordingly, the calibration rule in this case does not switch any respondents to the more costly opt-in option. The second case is where the two opt-in options have the same cost. In this case, most respondents who opted-in in both the hypothetical and the actual choice stayed with the same option (i.e., followed the YYS pattern), though some (38%) did switch. The calibration rule in this case switches a respondent from one opt-in option to the other only if her conditional choice certainty is very low (less than 5). The third case is where the opt-in option chosen in the hypothetical choice

[5] Continuous measures of certainty from the slider bars were rounded to the nearest 0.1 before the threshold was applied, so the actual threshold used was 6.95.

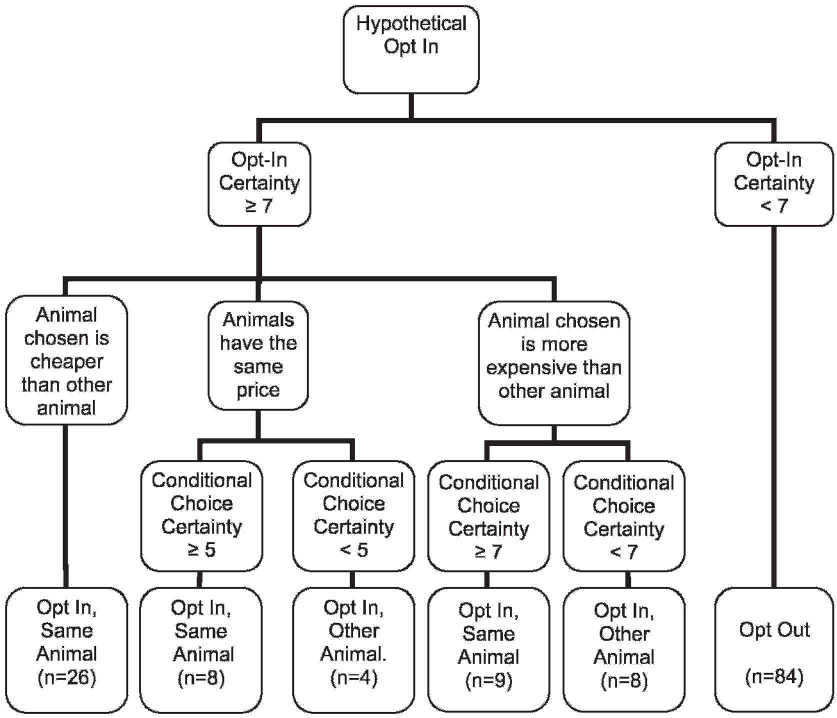



FIGURE 3
Calibration Rule Allowing Switching between Opt-in Options

is more expensive than the other opt-in option. In this case, more respondents followed the YYD pattern (i.e., switched to the cheaper opt-in option) than followed the YYS pattern (57% vs. 43%). Accordingly, the calibration rule uses a higher conditional choice certainty threshold than in Case 2. Respondents are assumed to switch to the cheaper opt-in option if their conditional choice certainty is less than 7, the same threshold level used to identify YN respondents. Figure 3 shows the number of cases where each type of recalibration was applied for the responses from treatment H2. The resulting proportions of respondents following the YYS, YYD, and YN patterns closely match the actual proportions measured in treatments H1 and H2.

For both calibration protocols, hypothetical opt-out choices were left unchanged, regardless of the stated opt-out certainty.

The hypothetical choices from treatment H2 were calibrated using both protocols, and MNL models were estimated. Results for calibrated and uncalibrated data are shown in the third through fifth results columns of Table 4, along with calculated mean WTP and 95% confidence interval on mean WTP for a respondent with average PRIORITY score. Both calibration protocols resulted in estimates of mean WTP that were similar to, and not significantly different from, mean WTP estimated from real choices. Similarly, calibration removed the impact of hypothetical bias on marginal WTP estimates. The marginal change in WTP associated with a one-unit increase in PRIORITY was $0.70 when estimated from the data calibrated without option switching, and $0.59 when estimated from data calibrated with option switching. Neither calibrated marginal WTP differed significantly from the marginal WTP estimated from real payment choices, which was $0.55.

While both calibration rules appear to remove hypothetical bias from both total and marginal WTP estimates, there are important differences between them. In particular, the estimated parameter for the COST attribute is negative but not significantly different from zero for the calibration rule without switching. As a consequence, the simulated 95% confidence interval for the estimated mean WTP is undefined. In contrast, the estimated parameter for the COST attribute is significantly different from zero for the calibration rule with switching, and for the real choices. Likelihood ratio tests show that the MNL model estimated from the hypothetical choices calibrated using the protocol without option switching is significantly different from the model estimated from real choices, but that the model for the data calibrated using the protocol with option switching is not significantly different from the model estimated from real choices. Allowing for option switching appears to be important in preserving price sensitivity in the choices. If calibration without allowing for option switching results in lower estimates of the marginal disutility of expenditures, this could help explain why Olsson found that calibration resulted in increases in the marginal WTP estimates.

## V. DISCUSSION

The main conclusions of this research are (1) hypothetical bias does exist in CE data, (2) hypothetical bias in CE data is related to respondent uncertainty, and (3) certainty follow-up questions can be used to calibrate hypothetical CE data. The pattern of results suggests that hypothetical bias exists not only in the respondents' propensity to choose a costly option, versus the opt-out option, but also when respondents choose among costly options. Calibration protocols should therefore include the possibility of the respondent switching from one costly option to another costly option, instead of always assuming that unsure respondents would switch to the opt-out option.

When the number of options presented to the respondent is larger than two, measurement of respondent uncertainty becomes more complicated. There exists more than one sequence of possible certainty questions that could be used. We found that not all sequences perform equally well. More research is needed on how respondents interpret certainty follow-up questions in a multiple-choice context, and how the questions can be asked in ways that enhance their reliability and usefulness. Berrens et al. (2002), in the context of DC CV, propose using direct questions about the probability of each choice. We believe that this would be difficult for respondents in a CE setting, but it is worth exploring.

The results show that calibration must take into account the possibility of switching between opt-in options. More work remains to be done in developing such calibration rules. In particular, the certainty thresholds used here for option switching should be researched further. We chose a certainty threshold for the opt-in decision that matched the calibrated hypothetical opt-in rate to the actual opt-in rate. However, we had too few observations to be able to do such frequency matching when setting the thresholds for option switching. Studies with larger samples would be needed to find option switching thresholds that are more firmly rooted in empirical measurement of behavior. Work is also needed to determine whether the certainty thresholds that mitigate hypothetical bias are consistent across studies and across goods. Studies that calibrate CV data have consistently concluded that the certainty threshold that performs best lies between 7 and 10. Several CE studies will be needed valuing different goods before we can determine whether CE calibration protocols are transferable across goods.

Although studies, including this one, have consistently found an empirical relationship between respondent uncertainty and hypothetical bias, there has not emerged a consensus on the behavioral mechanism underlying that relationship. Norwood, Lusk, and Boyer (2008) offer up two possible utility-theoretic explanations why respondent uncertainty might lead to hypothetical bias: risk aversion and commitment cost. According to the risk aversion hypothesis, individuals who are uncertain about the utility they will derive from a good may be more risk averse when they have to actually pay for a good than when they are in a hypothetical payment setting, so that actual payment values are lower than contingent values. The commitment cost hypothesis argues that if individuals are uncertain about the value of a good, expect to learn more about that value in the future, but are forced to make a purchase decision today, they will state lower WTP than they would if they had no uncertainty about the value of the good. Commitment cost is defined as the difference between WTP with certainty and WTP with uncertainty. If commitment cost is only considered in actual payment situations, the difference between actual and hypothetical payments (hypothetical bias) will be more pronounced as uncertainty increases.

Our results cannot directly confirm or refute either hypothesis. However, respondents in our experiment were explicitly told that they would not be receiving any new information prior to making their decision. Still they expressed uncertainty over their behavior, which was related to hypothetical bias. This suggests that commitment costs were not an important mechanism in this case. More work is needed to develop and test alternative conceptual models of the relationship between respondent certainty and hypothetical bias.

## References


Alberini, Anna, Kevin Boyle, and Michael Welsh. 2003. "Analysis of Contingent Valuation Data with Multiple Bids and Response Options Allowing Respondents to Express Uncertainty." *Journal of Environmental Economics and Management* 45 (1): 40–62.

Alfnes, Frode, and Kyrre Rickertsen. 2002. "SC-X: Calibrating Stated Choice Surveys with Experimental Auction Markets." Discussion Paper #D-5/2002, Department of Economics and Social Sciences, Norwegian University of Life Sciences, Aas, Norway.

Berrens, Robert P., Hank Jenkins-Smith, Alok K. Bohara, and Carol 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 (1): 144–68.

Blamey, Russell, and Jeff Bennett. 2001. "Yea-saying and Validation of a Choice Model of Green Product Choice." In *The Choice Modelling Approach to Environmental Valuation*, ed. Jeff Bennett and Russell Blamey. Cheltenham, UK: Edward Elgar.

Blomquist, Glenn C., Karen Blumenschein, and Magnus Johannesson. 2008. "Eliciting Willingness to Pay without Bias using Follow-up Certainty Statements: Comparisons between Probably/Definitely and a 10-Point Certainty Scale." Working paper, Department of Economics, University of Kentucky, Lexington.

Blumenschein, Karen, Magnus Johannesson, Glenn C. Blomquist, Bengt Liljas, and Richard M. O'Conor. 1998. "Experimental Results on Expressed Certainty and Hypothetical Bias in Contingent Valuation." *Southern Economic Journal* 65 (1): 169–77.

Blumenschein, Karen, Glenn C. Blomquist, Magnus Johannesson, Nancy Horn, and Patricia Freeman. 2008. "Eliciting Willingness to Pay without Bias: Evidence from a Field Experiment." *Economic Journal* 118 (525): 114–37.

Cameron, Trudy A., Gregory L. Poe, Robert G. Ethier, and William D. Shulze. 2002. "Alternative Non-market Value-Elicitation Methods: Are the Underlying Preferences the Same?" *Journal of Environmental Economics and Management* 44 (3): 391–425.

Carlsson, Fredrik, and Peter Martinsson. 2001. "Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments?" *Journal of Environmental Economics and Management* 41 (2): 179–92.

Champ, Patricia A., and Richard C. Bishop. 2001. "Donation Payment Mechanisms and Contingent Valuation: An Empirical Study of Hypothetical Bias." *Environmental and Resource Economics* 19 (4): 383–402.

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): 151–62.

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

Ethier, Robert G., Gregory L. Poe, William D. Schulze, and Jeremy Clark. 2000. "A Comparison of Hypothetical Phone and Mail Contingent Valuation Responses for Green-Pricing Electricity Programs." *Land Economics* 76 (1): 54–67.

Hanemann, W. Michael, and Bengt Kriström. 1995. "Preference Uncertainty, Optimal Design, and Spikes." In *Current Issues in Environmental Economics*, ed. Per Olov Johansson, Bengt Kriström, and Karl-Göran Maler. Manchester, UK: Manchester University Press.

Holmes, Thomas P., and Wiktor L. Adamowicz. 2003. "Attribute Based Methods." In *A Primer on Nonmarket Valuation*, ed. Patricia A. Champ, Thomas Brown, and Kevin Boyle, 171–219. Dordrecht: Kluwer Academic Publishers.

Johannesson, Magnus, Glenn C. Blomquist, Karen Blumenschein, Per-Olov Johansson, Bengt Liljas, and Richard M. O'Conor. 1999. "Calibrating Hypothetical Willingness to Pay Responses." *Journal of Risk and Uncertainty* 18 (1): 21–32.

Johannesson, Magnus, Bengt Liljas, and Per-Olov Johansson. 1998. "An Experimental Comparison of Dichotomous Choice Contingent Valuation Questions and Real Purchase Decisions." *Applied Economics* 30 (5): 643–47.

Johansson-Stenman, Olof, and Henrik Svedsäter. 2007. "Hypothetical Bias in Choice Experiments: Within versus Between Subject Tests." Working Papers in Economics 252. Göteborg: Department of Economics, Göteborg University.

Kooten, G. Cornelis van, Emina Krcmar, and Erwin H. Bulte. 2001. "Preference Uncertainty in Non-market Valuation: A Fuzzy Approach." *American Journal of Agricultural Economics* 83 (3): 487–500.

Li, Chuan-Zhong, and Leif Mattsson. 1995. "Discrete Choice under Preference Uncertainty: An Improved Structural Model for Contingent Valuation." *Journal of Environmental Economics and Management* 28 (2): 256–69.

List, John A., and Craig A. Gallet. 2001. "What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? Evidence From a Meta-Analysis." *Environmental and Resource Economics* 20 (3): 241–54.

List, John A., Paramita Sinha, and Michael H. Taylor. 2006. "Using Choice Experiments to Value Non-market Goods and Services: Evidence from Field Experiments." *Advances in Economic Analysis and Policy* 6 (2): Article 2.

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

Loomis, John, and Earl Ekstrand. 1998. "Alternative Approaches for Incorporating Respondent Uncertainty When Estimating Willingness to Pay: The Case of the Mexican Spotted Owl." *Ecological Economics* 27 (1): 29–41.

Lusk, Jayson L., and T. C. Schroeder. 2004. "Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks." *American Journal of Agricultural Economics* 86 (2): 467–82.

Murphy, James J., P. Geoffrey Allen, Thomas H. Stevens, and Darryl Weatherhead. 2005. "A Meta-analysis of Hypothetical Bias in Stated Preference Valuation." *Environmental and Resource Economics* 30 (3): 313–25.

Norwood, F. Bailey. 2005. "Can Calibration Reconcile Stated and Observed Preferences?" *Journal of Agricultural and Applied Economics* 37 (1): 237–48.

Norwood, F. Bailey, Jayson L. Lusk, and Tracy Boyer. 2008. "Forecasting Hypothetical Bias: A Tale of Two Calibrations." In *Environmental Economics, Experimental Methods*, ed. Todd L. Cherry, Stephan Kroll, and Jason F. Shogren. New York: Routledge.

Olsson, Björn. 2005. "Accounting for Response Uncertainty in Stated Preference Methods." Presented at the EAERE Congress, Bremen, Germany, June 23–26.

Opaluch, James J., and Kathleen Segerson. 1989. "Rational Roots of 'Irrational' Behavior: New Theories of Economic Decision-Making." *Northeastern Journal of Agriculture and Resource Economics* 18 (2): 81–95.

Poe, Gregory L., Jeremy E. Clark, Daniel Rondeau, and William D. Schulze. 2002. "Provision Point Mechanisms and Field Validity Tests of Contingent Valuation." *Environmental and Resource Economics* 23 (1): 105–31.

Poe, Gregory L., Eric K. Severance-Lossin, and Michael P. Welsh. 1994. "Measuring the Difference ($X - Y$) of Simulated Distributions: A Convolutions Approach." *American Journal of Agricultural Economics* 76 (4): 904–15.

Ready, Richard C., Ståle Navrud, and W. Richard Dubourg. 2001. "How Do Respondents with Uncertain Willingness to Pay Answer Contingent Valuation Questions?" *Land Economics* 77 (3): 315–26.

Taylor, Laura O., Mark D. Morrison, and Kevin J. Boyle. 2007. "Exchange Rules and the Incentive Compatibility of Choice Experiments." Working paper, Center for Environmental and Resource Economic Policy, North Carolina State University.

Vossler, Christian A., Robert G. Ethier, Gregory L. Poe, and Michael 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, Michael P., and Gregory L. Poe. 1998. "Elicitation Effects in Contingent Valuation: Comparisons to a Multiple Bounded Discrete Choice Approach." *Journal of Environmental Economics Management* 36 (2): 170–85.