

A Comparison of Approaches to Mitigate Hypothetical Bias

Patricia A. Champ, Rebecca Moore, and Richard C. Bishop

We compare two approaches to mitigating hypothetical bias. The study design includes three treatments: an actual payment treatment, a contingent valuation (CV) treatment with a follow-up certainty question, and a CV treatment with a cheap talk script. Our results suggest that both the follow-up certainty treatment and the cheap talk treatment produce willingness-to-pay (WTP) estimates consistent with the actual payment treatment. However, the follow-up certainty treatment provides response distributions at all offer amounts that are statistically similar to the actual payment treatment, while the cheap talk treatment provides similar responses only at some offer amounts. Furthermore, the cheap talk treatment is effective only for inexperienced individuals. We conclude that the follow-up certainty approach is more consistent than the cheap talk approach for eliminating hypothetical bias.

Key Words: contingent valuation, hypothetical bias, follow-up certainty, cheap talk, nonmarket valuation

While the validity of the contingent valuation (CV) method has been a contentious issue, practitioners have come to recognize that validity is a matter of degree. To address issues of convergent or criterion validity, several studies have compared contingent values to actual payments. These studies generally find that respondents report higher willingness to pay (WTP) in a hypothetical payment situation than in an actual payment situation. Such results provide evidence of the existence of “hypothetical bias” that challenges the validity of the individual study and the CV method in general. The finding of hypothetical bias in many studies has spurred an interest in approaches to mitigate the problem.

Meta-analyses (List and Gallet 2001, Little and Berrens 2003, Murphy et al. 2005) have recently been conducted to investigate study design factors affecting hypothetical bias. List and Gallet (2001) found that private good studies result in less hypothetical bias than studies in which public goods are valued, and that hypothetical bias is

larger in willingness-to-accept studies than willingness-to-pay studies. Elicitation methods were also found to affect hypothetical bias. Little and Berrens (2003) found negative significant coefficient estimates on the certainty correction variables, suggesting that a certainty correction reduces hypothetical bias. The use of a cheap talk script was also found to reduce hypothetical bias in one of the four models estimated in this meta-analysis. Murphy et al. (2005) found that hypothetical bias increased as the hypothetical value increased and that choice-based elicitation (e.g., dichotomous and polychotomous choice, referendum, payment card, and conjoint) mechanisms were associated with less hypothetical bias.

Though the meta-analyses suggest that study design can impact hypothetical bias, there is less insight into the details of how different mitigation approaches actually affect individual respondents. With this in mind, in this article we compare two approaches to mitigating hypothetical bias in order to identify possible differences in how they operate on the study population. Our split sample design includes three treatments: an actual payment treatment that serves as a benchmark for assessing hypothetical bias, a CV treatment with a follow-up certainty question, and a CV treatment with a cheap talk script before the willingness-to-pay question. The follow-up certainty question asks study participants to rate on a 10-point rating

Patricia Champ is an economist at the Rocky Mountain Research Station, U.S. Forest Service, in Fort Collins, Colorado. Rebecca Moore is Assistant Professor in the Warnell School of Forestry and Natural Resources at the University of Georgia in Athens, Georgia. Richard Bishop is Professor Emeritus in the Department of Agricultural and Applied Economics at the University of Wisconsin in Madison, Wisconsin, and a consultant at Stratus Consulting Inc. in Boulder, Colorado, and Washington, D.C.

scale (1=very uncertain and 10=very certain) how certain they are that they would actually pay (or not pay if they said *no* to the contingent valuation question). This self-reported certainty level is used to re-code responses to the willingness-to-pay question to provide an estimate of mean willingness to pay similar to the actual donation treatment. A cheap talk script, which precedes the CV question, describes the issue of hypothetical bias and asks respondents to try to avoid hypothetical bias by responding to the CV question as they would an actual payment question.

Cheap Talk

Cummings and Taylor (1999) were the first to implement cheap talk in a CV setting. They developed a script for laboratory experiments to encourage study participants to respond to the CV question as if they were making an actual financial decision. That script explicitly described the hypothetical bias problem and was read aloud to the study participants prior to the contingent valuation question. The original script was quite long [see Cummings and Taylor (1999) for the substance of the script]. The results of the Cummings and Taylor experiments were promising in that the CV treatment with cheap talk script provided responses that were similar to responses in the actual payment treatment for three different public goods. Since the original Cummings and Taylor study, several others have investigated the use of a cheap talk script.

List (2001) used the Cummings and Taylor (1999) cheap talk script in a laboratory setting with individuals in the market for a private good, sports cards. List found hypothetical bias to exist with both experienced and inexperienced sports card dealers. However, he found that experienced sports card dealers were *not* sensitive to the cheap talk script. In contrast, nondealers did respond to the cheap talk script, and hypothetical bias was eliminated for this group. Brown, Ajzen, and Hrubec (2003) also implemented the long Cummings and Taylor cheap talk script in a laboratory experiment involving a public good. Using a dichotomous-choice format with four offer amounts (\$1, \$3, \$5, \$8), they found that cheap talk worked well at higher offer amounts but had no effect at the lowest offer amounts (\$1 and \$3). Murphy, Stevens, and Weatherhead (2005) also

used the long cheap talk script in a laboratory setting with a public good. They found that overall cheap talk reduced hypothetical bias but did not completely eliminate it. Like Brown, Ajzen, and Hrubec, they found that cheap talk worked well at the higher offer amounts but not at the lower offer amounts. Blumenschein et al. (2008) tested the long cheap talk script in a field experiment with a private good. They found that the cheap talk script failed to eliminate the hypothetical bias.

Few researchers have experimented with shorter cheap talk scripts using mail or telephone surveys. Aadland and Caplan (2003) used a short cheap talk script in a telephone survey that elicited households' willingness to pay for curbside recycling, an impure public good. They found that cheap talk was more effective at eliminating hypothetical bias for households with higher willingness to pay. Lusk (2003) also used a shorter version with a mail survey and found that cheap talk had no effect on individuals who had experience or knowledge of goods similar to the good valued in his study. The Lusk study did not include an actual payment treatment so it was not possible to evaluate the extent of the hypothetical bias. Aadland and Caplan (2005) developed a short "neutral" cheap talk script for use in a telephone survey. The script did not mention the usual direction of hypothetical bias (hypothetical > actual). They found that the neutral cheap talk script treatment provided willingness-to-pay estimates that were *greater* than the standard contingent valuation treatment. In other words, the neutral script exacerbated the hypothetical bias.

The previous research on cheap talk provides a few insights: the effectiveness of cheap talk can vary across offer amounts (Brown, Ajzen, and Hrubec 2003, Murphy, Stevens, and Weatherhead 2005), cheap talk may be more effective with less experienced or less knowledgeable individuals (List 2001, Lusk 2003), and short cheap talk scripts can reduce hypothetical bias for some types of respondents (Aadland and Caplan 2003, Lusk 2003). We built on these studies to develop a "medium" length cheap talk script included in a mail survey.

Certainty Scale

Several approaches have been taken to allow for expressions of uncertainty associated with will-

ingness-to-pay questions, such as using follow-up certainty questions (Li and Mattsson 1995, Champ et al. 1997, Blumenschein et al. 1998, Champ and Bishop 2001), polychotomous response formats (Ready, Whitehead, and Blomquist 1995), and multiple-bounded discrete choice formats (Welsh and Poe 1998). The approach taken in the current study was to have the survey respondent go through the cognitive challenge of responding to a dichotomous choice willingness-to-pay question and then follow up with a question about how certain the individual was that she would actually pay if asked to do so. Li and Mattsson (1995) followed up a dichotomous-choice willingness-to-pay question with "How certain are you of your answer to the previous question?" Respondents circled a point on an interval scale from 0 to 100 percent (0 = absolutely uncertain, 100 = absolutely certain). They found that the certainty information improves the efficiency of parameter estimates, but could not comment on how such information could be used to mitigate hypothetical bias as they did not include an actual payment treatment.

Champ et al. (1997) investigated the use of a follow-up certainty question to mitigate hypothetical bias. After a dichotomous choice CV question for a public good, respondents who said *yes* to the CV question were asked how certain they were that they would actually pay if asked to do so. Responses were given on a 10-point scale, with 1 labeled "very uncertain" and 10 labeled "very certain." Champ et al. (1997) found that recoding the *yes* responses to the CV question so that those who said they were very certain (circled 10) were considered positive responses to the CV question and all others were coded as negative responses provided an estimate of mean WTP that was similar to the actual payment treatment. As the data for that study were collected with mail surveys, it was possible for respondents to read ahead, and so the presence of the follow-up certainty question could influence responses to the CV question. Therefore, a standard CV treatment without a follow-up certainty question was also included to assess whether the inclusion of the certainty question affected the distribution of response to the CV question. The distributions of responses to the standard CV question and the CV question with the follow-up certainty question were similar.

Blumenschein et al. (1998) implemented a similar approach in a laboratory setting. However, instead of using a 10-point scale they asked whether individuals were probably sure or definitely sure about their response to the dichotomous choice CV question. They recoded the CV response in such a way that "definitely sure" was coded as a *yes* and "probably sure" was coded as a *no*. This recoding scheme provided results similar to the actual payment treatment. Johannesson et al. (1999) conducted a very similar study in a laboratory setting with a private good. However, they used the words "fairly sure" and "absolutely sure" in the follow-up question. They found that coding only the "absolutely sure" responses as a *yes* to the CV question resulted in an underestimate of responses in the actual payment treatment.

Ethier et al. (2000) used the follow-up certainty scale developed by Champ et al. (1997) in a study with a public good. They found that using a cutoff of 7 on the 10-point scale to recode the CV data provided results similar to the actual payment treatment. Champ and Bishop (2001) replicated their earlier study with a different public good across a wider range of offer amounts. In this study they found that using a cutoff of 8 on the certainty scale provided an estimate of mean WTP that was indistinguishable from the actual payment treatment. More importantly, they found that the CV respondents who circled 8, 9, or 10 on the certainty scale were similar across a range of measures (i.e., attitudes, experience, demographics) to the individuals who had actually paid for the good. Norwood (2005) implemented the Champ et al. approach in a choice experiment that allowed students to donate class participation points. He found that calibrating responses at a certainty level of 8 underestimated actual participation.

Blumenschein et al. (2008) conducted a study with a private good that included a follow-up certainty question that asked participants if they were probably sure or definitely sure about their response to a dichotomous choice CV question. When the CV data were recoded in such a way that only the "definitely yes" responses were counted as positive responses to the CV question, results were similar to the actual payment treatment. Samnaliev, Stevens, and More (2006) implemented a split sample general population mail survey that allowed for a comparison of the fol-

low-up certainty 10-point scale to a *yes, no, or not sure* response format. The CV questions included three offer amounts (\$3, \$5, and \$10). They found that certainty levels were not affected by the offer amounts; however, the number of “not sure” responses decreased with the higher offer amounts. Adjusting for uncertainty with the 10-point scale approach substantially reduced the WTP estimates, while the “not sure” treatment did not reduce the WTP estimates.

The bottom line with the follow-up certainty studies is that the approach has been repeatedly shown to mitigate hypothetical bias. However, it is not clear how certain individuals need to be for their response to be considered equivalent to an actual payment situation. The studies that used the 10-point scale (1 = very uncertain, 10 = very certain) suggest that respondents should be fairly certain (i.e., certainty level ≥ 7) for hypothetical payment responses to correspond to actual payment decisions.

While the literature on using respondent uncertainty to mitigate hypothetical bias in CV studies continues to grow, a consensus on an underlying mechanism that describes the relationship of uncertainty to hypothetical bias has not yet emerged. Norwood, Lusk, and Boyer (2008) make this point and offer up two possible explanations for the relationship between uncertainty and hypothetical bias: risk aversion and commitment cost. While neither explanation has been rigorously tested, they both appear to be intuitively plausible. The intuition behind risk aversion hypothesis is that when individuals are risk adverse and uncertain about the utility they will derive from a good, the resultant expected utility will be depressed. If individuals are more risk adverse when they have to actually pay for a good relative to the hypothetical payment setting of CV, actual payment values will be lower than contingent values. The commitment cost hypothesis put forth by Norwood, Lusk, and Boyer argues that if individuals are uncertain about the value of a good and expect to learn more about that value in the future but are forced to make a decision about purchasing the good today, they will state a lower WTP than they would if they had no uncertainty about the value of the good. Thus the difference between WTP with certainty and WTP without uncertainty, referred to as “commitment cost,” is positive. If commitment cost is considered only in actual payment situations, the difference between

actual and hypothetical payments (hypothetical bias) will be more pronounced as uncertainty increases. The data collected for the current study do not allow for an analysis of the relationship between uncertainty and risk aversion or commitment cost. However, future studies could be designed to develop and test conceptual models of the relationship between respondent certainty and hypothetical bias.

Certainty Scale and Cheap Talk

Blumenschein et al. (2008) conducted an interesting field experiment in which a cheap talk treatment and a follow-up certainty treatment were compared to actual payments to better understand which approach was more successful at mitigating hypothetical bias. The study involved a private good, participation in a diabetes management program. Study participants were recruited from a group of individuals known to have Type-2 diabetes. Participants convened at a central location where they were divided into one of three treatment groups: actual payment, hypothetical payment with a cheap talk script preceding the willingness-to-pay question, or hypothetical payment with a follow-up question about whether respondents were probably sure or definitely sure that they would buy the diabetes management service. All three treatments implemented a dichotomous choice format with three offer amounts. Blumenschein et al. (2008) found evidence of hypothetical bias in both the cheap talk treatment and the treatment with the follow-up certainty question. However, when study participants who said they were “probably sure” that they would buy the good were re-coded as “no” responses, the hypothetical bias was removed. The authors concluded that the cheap talk treatment was not effective and that the follow-up certainty treatment was effective at removing hypothetical bias.

Study Design

In this study we implemented a split sample field study similar to the Blumenschein et al. (2008) study. The study included three treatments: CV with follow-up certainty question, CV with medium-length cheap talk script preceding the CV question, and an actual payment treatment. All treatments were administered via a mail survey

and were similar in all aspects other than the treatments previously mentioned. However, our study differs from the Blumenschein et al. study in four important aspects. First, we used a mail survey to administer the survey, rather than the group setting of Blumenschein et al. Second, the good used in our study was a public good. Third, the follow-up certainty question asked respondents on a 10-point scale (from 1 = very uncertain to 10 = very certain) how certain they were that they would actually pay the offer amount from the willingness-to-pay question. Finally, the study was administered to a general population sample.

The Good

Though once widespread, whooping cranes were on the brink of extinction, with only 16 remaining in 1941. Since the 1950s only one flock of whooping cranes has survived. Whooping cranes are threatened primarily by the conversion of their wetlands habitat into agricultural lands or urban development areas. The International Crane Foundation has been orchestrating efforts to ensure the survival of the species. As part of these efforts, a second flock of whooping cranes is being bred and introduced into the wild. Each year, whooping crane chicks are hatched in captivity and taught behaviors crucial to their survival in the wild. As whooping cranes are migratory birds, one important aspect of this program is teaching the young cranes how to make the 1,250-mile migration journey from central Wisconsin to Florida. After being led to Florida by an ultralight aircraft their first year, the cranes are then able to make the return trip to Wisconsin unassisted the next spring. They also continue the migration annually as a flock, without the assistance of an aircraft. However, to ensure the success of the program, radio transmitters are placed on the leg of each crane to monitor the birds during migration and throughout the year. If a bird is in danger or sick, scientists will intervene and rescue the bird. The first class of cranes, 18 birds, was hatched in the spring of 2001. The project will continue until the flock has grown to 125 cranes (approximately 10–25 years). At the time of our study, funding was needed to purchase radio transmitters for whooping crane chicks who were to be hatched in the spring of 2004. The transmitters cost around \$300 each, and while survey respondents were

not told the cost of the transmitters, they were told that the transmitters could be provided only if there were sufficient support in the form of donations.

The Survey

In January 2004, a mail survey was administered to a random sample of Madison, Wisconsin, residents. The sample was randomly split into three treatments: (i) an actual donation treatment, (ii) a contingent donation with follow-up certainty questions treatment, and (iii) a contingent donation treatment with a cheap talk script. For all three treatments, each person received a cover letter on University of Wisconsin, Department of Agricultural and Applied Economics letterhead, a question and answer sheet, and a survey booklet. The beginning of the survey booklet described the endangered nature of the whooping cranes as well as the ongoing project to establish a second flock of whooping cranes and the role of radio transmitters in this project. The willingness-to-donate question came after the description, and was followed by questions concerning previous knowledge of the reintroduction project, environmental attitudes, and socio-demographic information. Ten days after the survey packet was mailed, a reminder/thank you postcard was sent to all respondents. A second survey packet was sent to all nonrespondents two weeks after the postcard.

Five hundred and five surveys were sent in each of the three treatments. All three treatments used a dichotomous choice question to ask for donations of a specific dollar amount to purchase radio transmitters for the whooping cranes. The five offer amounts used in each treatment were \$10, \$15, \$25, \$50, and \$100. Each treatment included a similar distribution of offer amounts and the assignment of offer amounts was random. Those who said *yes* in the actual payment treatment were asked to include a check payable to the International Crane Foundation for the stated amount with their completed survey.¹ For all three treatments, the surveys were returned to the University of Wisconsin. Checks from the actual payment treatment were then mailed by the re-

¹ Respondents who sent checks for amounts greater than the offer amount were coded as saying *yes* to the offer amount. Likewise, respondents who sent checks for less than the offer amount were coded as saying *no* to the offer amount.

search team to the International Crane Foundation. The certainty treatment included a contingent donation question that was immediately followed by a certainty question (see Appendix A). This question asked the respondent to state on a 10-point scale how certain they were that they would actually send a donation if they were asked to do so; or, if they said *no* to the donation question, how certain they were that they would not make a donation. The endpoints of the scale were labeled as 1 being “very uncertain” and 10 being “very certain.”

The cheap talk treatment included the same dichotomous choice question as the contingent donation with certainty treatment, but did not include the follow-up certainty question. Instead the donation question was preceded by a medium-length cheap talk script (see Appendix B). This script explained to participants that studies have shown that it is difficult for many people to answer questions about a hypothetical situation as if it were real, and that “often more individuals say they will make a donation in the hypothetical situation than when the situation is real.” The script then asked respondents to try to avoid this problem by making sure they answered the contingent donation question as if they were being asked to make an actual donation. The three treatments differed only by the presence or absence of the cheap talk script, the certainty question, or the request for an actual donation. The rest of the survey materials, including the project description and other survey questions, were identical across treatments. In June 2004, additional surveys were sent to a separate sample to increase the number of responses. The sample used for the January and June mailings were drawn from the same sample frame, and a comparison of the data does not suggest any difference between the two respondent pools. Additionally, there were not statistically significant differences between those who responded to the first, rather than the second, January mailing. In total, 975 surveys were mailed in the actual donation treatment, 730 surveys were mailed in the CV with follow-up certainty treatment, and 760 in the CV with cheap talk treatment.

Contingent valuation practitioners often struggle with what is the most appropriate provision mechanism for a particular good or program. Trade-offs in terms of credibility must be weighed against other issues such as incentive compatibil-

ity (Champ et al. 2002). While donation mechanisms lack incentive compatibility, there are some very practical reasons for using such mechanisms to provide observations of actual behavior that closely mimic a state-of-the-art contingent valuation question. Comparing incentive-compatible mechanisms for public goods in a field setting to similar actual behaviors has been problematic. Furthermore, it is not common for provision of small-scale public goods to be based on referendum results, and many public goods are provided via donation mechanisms. Champ et al. (1997) argue that donations are a lower bound to Hicksian surplus values, and in many cases this lower bound is enough to establish that the benefits of a policy exceed the costs in the context of a benefit-cost analysis. As one goal of this study was to implement actual and hypothetical payment treatments in a field setting, a donation mechanism was deemed appropriate.

Results and Discussion

Overall the response rates for the two hypothetical treatments were significantly higher than those for the actual donation treatment (Table 1). Responses to other questions in the survey were compared across treatments to test for sample selection bias. We examined differences in prior knowledge, environmental interest, and demographic characteristics. Based on these comparisons, we conclude that respondents in the three treatments represent the same population. Table 2 summarizes the characteristics of the survey respondents.

Actual and CV Treatment Groups

We found that, similar to past studies, more individuals said *yes* to the donation question in the hypothetical treatment relative to the actual donation treatment. As shown in Table 3, 26 percent of the respondents in the actual donation treatment said they would donate and send a check. In comparison, 55 percent of respondents in the CV treatment said they would donate if actually asked to do so. The difference in the percentage answering *yes* to the willingness-to-donate (WTD) question between the actual and CV treatments was significant at all offer amounts. Table 4 presents parametric and non-parametric estimates of the WTD for the different treatment groups. The

Table 1. Response Rates by Offer Amount and Treatment

Offer Amount	Usable/Deliverable Surveys (Response Rate)		
	Actual	CV	Cheap Talk
\$10	38/139 (27%)	39/114 (34%)	45/96 (47%)
\$15	52/220 (24%)	52/160 (32%)	54/165 (33%)
\$25	45/229 (20%)	53/156 (34%)	43/177 (24%)
\$50	57/188 (30%)	55/167 (33%)	55/190 (29%)
\$100	33/157 (21%)	36/110 (33%)	34/100 (34%)
Total	225/933 (24%)	235/707 (33%)	231/728 (32%)

Note: "Usable" surveys are those that were returned with the donation question answered. "Deliverable" is the number of surveys mailed minus those that were returned and marked "undeliverable" or "deceased."

Table 2. Survey Respondent Characteristics by Treatment

	Actual	CV	Cheap Talk
Mean offer amount	37	37	37
Mean age	51	54	51
Mean number of years in Wisconsin	38	41	41
Percent female	42% ^a	32% ^a	39%
Mean education	college graduate	college graduate	college graduate
Mean income	\$65,640	\$63,230	\$61,440
Had prior knowledge of the reintroduction project	72%	72%	69%
Had knowledge of the International Crane Foundation	93%	91%	89%
Had visited the International Crane Foundation	32%	31%	28%

^a Significant difference between actual treatment and CV treatment at $\alpha = .05$.

results clearly indicate that WTD is higher in the CV treatment than in the actual donation treatment. The parametric estimate of mean WTD is \$68 for the CV treatment and only \$21 for the actual donation treatment. Similarly, the non-parametric estimate of WTD is \$45 for the CV treatment and only \$17 for the actual donation treatment. If we take the responses from the actual treatment group as a baseline, these results strongly suggest the existence of hypothetical bias in the CV treatment.

Cheap Talk Treatment

The cheap talk treatment shows potential for reducing hypothetical bias. Thirty-eight percent of the respondents in the cheap talk treatment said they would make a donation, compared to 55 percent of those in the CV treatment (Table 3). Thus the cheap talk script reduced the number of respondents answering *yes* to the donation question compared to the standard CV approach. Relative to the actual payment treatment, significantly

Table 3. Percentage of Respondents Answering Yes to Willingness-to-Donate Question

	Actual (n=225)	CV (n=235)	Cheap Talk (n=231)	Certainty (7) (n=233)	Certainty (8) (n=233)	Certainty (9) (n=233)
\$10	47%	77% ^a	44% ^b	59%	51%	31%
\$15	31%	67% ^a	54% ^a	52% ^a	46%	29%
\$25	33%	57% ^a	51% ^c	37%	33%	24%
\$50	14%	40% ^a	22% ^b	24%	14%	7%
\$100	6%	36% ^a	15% ^b	22% ^c	19%	11%
Total	26%	55% ^a	38% ^{a, b}	39% ^a	33%	20% ^c

^a Significantly different from actual donation treatment at $\alpha = .05$

^b Significantly different from CV treatment at $\alpha = .05$

^c Significantly different from actual donation treatment at $\alpha = .10$

Note: The n for the Certainty (7), Certainty (8), and Certainty (9) treatment is 2 less (233 compared to 235) than CV because 2 respondents answered the CV question but not the follow-up certainty question.

Table 4. Mean Willingness to Donate (WTD) by Treatment

	Actual (n = 225)	CV (n = 235)	Cheap Talk (n = 231)	Certainty (7) (n = 233)	Certainty (8) (n = 233)	Certainty (9) (n = 233)
Mean WTD ^a	\$21	\$68	\$36	\$42	\$33	\$20
95% CI ^b	[16, 34]	[51, 113]	[28, 57]	[31, 73]	[25, 60]	[14, 48]
Turnbull lower bound	\$17	\$45	\$22	\$28	\$24	\$13
95% CI	[11, 22]	[36, 53]	[15, 28]	[20, 37]	[17, 31]	[6, 20]

^a The mean WTD estimate is based on a logit model with response to the WTD question as the dependent variable and offer as the independent variable. The mean is the area above logistic probability density function.

^b "CI" stands for confidence interval and was estimated with the Krinsky-Robb procedure with 10,000 draws of β .

more respondents said *yes* in the cheap talk treatment for two of the five offer amounts. Likewise, the overall percentage saying *yes* to the willingness-to-donate question is significantly higher for the cheap talk treatment than for the actual payment treatment (Table 3). The parametric estimates of mean willingness to donate are significantly different between the cheap talk and actual payment treatments (Table 4). The nonparametric estimates of mean willingness to donate for the cheap talk and actual payment treatments are not significantly different from each other (Table 4).

Certainty Treatment

We also consider the effect of recoding the CV treatment according to respondents' certainty of their response. In the certainty treatment, the fol-

low-up certainty question is used to convert some of the *yes* responses to the CV question to *no* responses. The idea is that individuals who say *yes* but are very uncertain are likely to say *no* in an actual payment question. In our study, the mean certainty level is 7.7 and the median is 8 on the 10-point scale, with 60 percent of the respondents circling an 8, 9 or 10 (Table 5). Based on this, we use 7 on the certainty scale as our starting point for recoding the less certain *yes* responses to *no* responses. We also look at cutoff points of 8 and 9. Tables 3 and 4 report the results of the recoding at three different levels of certainty cutoff. Certainty(7), Certainty(8), and Certainty(9) use cutoff levels of 7, 8, and 9, respectively. These results show that using 8 as a cutoff gives a distribution of responses to the CV question that is statistically similar to that of the actual dona-

Table 5. Distribution of Responses to Follow-Up Certainty Question (n = 128)

Certainty Level	Percent	Cum. Percent
1 = very uncertain	2%	2%
2	1%	3%
3	1%	4%
4	1%	5%
5	5%	10%
6	19%	29%
7	11%	40%
8	23%	63%
9	8%	71%
10 = very certain	29%	100%

Mean certainty level = 7.7
Median certainty level = 8

tion treatment at every offer amount (Table 3). Likewise the estimated mean willingness to donate is similar for the CV treatment with the data recoded to that of the actual payment treatment (Table 4). However, in this study, using a cutoff point of 9 on the 10-point scale also provides an estimate of mean willingness to donate that is similar to the actual payment treatment. The consistency between the actual payment data and the CV data recoded at certainty level 8 shown in Table 3 suggests that 8 is the more appropriate cutoff point.

In summary, considering the hypothetical bias mitigation treatments individually, our results are similar to those of previous studies. The cheap talk treatment reduced the hypothetical bias, though results are inconclusive regarding the elimination of the bias. We were able to identify a certainty level that could be used as a cutoff for recoding the CV data in a manner that produces results similar to the “actual” treatment. The remainder of this paper compares the two hypothetical bias mitigation treatments.

Comparing the Cheap Talk and Certainty Treatments

The results in Tables 3 and 4 suggest that for our study, the certainty treatment more consistently

eliminates the hypothetical bias. Perhaps the strongest test is the simple result shown in Table 3. The certainty treatment provided a percentage of positive responses that is statistically similar to the actual payment treatment at every offer amount and overall. While the cheap talk treatment lowered the estimated WTD compared to the standard CV treatment, it did not definitively produce an estimate that matched that of the actual donation treatment. It should be noted that the effectiveness of the cheap talk approach could depend on the length of the cheap talk script used. As noted above, we used a medium-length script. We do not know if a longer script would have been more effective.

One advantage of the certainty treatment is that it provides information in addition to the response to the CV question. To provide insight into differences or similarities between individuals who we suggest are certain of their response to the donation question (i.e., those who circled 8–10 on the certainty scale) and those who are uncertain (i.e., those who circled 1–7), we estimated a logistic regression with “certainty” as the dependent variable. The results are shown in Table 6. This model included only those respondents who were in the CV treatment *and said they would donate*. The regression results suggest that the offer amount was not related in a statistically significant manner to certainty responses. Individuals with higher incomes and those with a strong interest in the environment were more likely to be very certain that they would actually pay if asked to do so. Responses to three questions that were asked only of the *yes* respondents about why they said *yes* were also included in the regression. Having said *yes* because they thought the reintroduction program would be worth the assigned offer amount was associated with higher certainty levels. Respondents who said *yes* because they knew they would not really have to come up with the money were less certain. Finally, those who said *yes* to show their support for the crane reintroduction were more certain of their positive response. It is not clear that any one of these measures alone could be used like the certainty scale response to predict which *yes* respondents to the CV question would actually donate. The results suggest that certainty is related to a combination of factors: ability to pay (*income*), strength of support for the issue (*environment*,

Table 6. Logistic Regression (dependent variable 1 = certainty level 8–10, 0 = certainty level 1–7, n = 117)

Variable	Description	Coefficient	Std. Error	p-value
<i>Constant</i>		-2.434	0.816	0.003
<i>Offer</i>	\$10, \$15, \$25, \$50, \$100	-0.013	0.008	0.128
<i>Income</i>	In thousands of dollars	0.015	0.009	0.081
<i>Gender</i>	1 = male, 0 = female	0.771	0.523	0.141
<i>Environment</i>	Interest in the environment (1 = strongly or extremely interested, 0 = otherwise)	1.101	0.506	0.030
<i>WhyYes1</i>	The Whooping Crane Reintroduction Program would be worth that much to me (1 = circled, 0 = not circled)	1.177	0.484	0.015
<i>WhyYes2</i>	I knew I would not really have to come up with the money (1 = circled, 0 = not circled)	-3.115	1.242	0.012
<i>WhyYes3</i>	I wanted to show my support for the whooping crane reintroduction (1 = circled, 0 = not circled)	1.104	0.816	0.003
-2 log likelihood		115.328		
% predicted correctly		76.90		

WhyYes1, and *WhyYes3*), and how seriously the respondent took the WTD question (*WhyYes2*). A text box appearing before the WTD question in the certainty treatment told respondents the following: “The Whooping Crane Reintroduction Program is a real program. However, as this is a research project, we are not asking you to make an actual donation. Nonetheless, we would like you to answer the following question as you would for an actual donation solicitation.” Clearly, some respondents did not follow this request, and even told us so when asked (*WhyYes2*). It is not obvious how a standardized cheap talk script could be developed to move a diverse population of respondents to consider all the possible causes of hypothetical bias and to avoid them.

Looking now at the cheap talk technique, an interesting result that has been observed in previous cheap talk studies is the role of experience on the effectiveness of cheap talk and the variation in results across offer amounts. To examine the role of experience, we break our respondent pool into those who have visited the International Crane Foundation and those who have not (Table 7). We define those who have visited the International Crane Foundation as “experienced” and those who have not visited as “inexperienced.”

Consistent with the List (2001) study, we find that hypothetical bias exists for both experienced and inexperienced respondents. For both groups, the percentage responding *yes* to the hypothetical donation question was higher than the percentage answering *yes* to the actual donation question. Also consistent with the List study, comparing the cheap talk and CV treatments we find that cheap talk effectively eliminates hypothetical bias for the inexperienced group but not for the experienced group. This result is somewhat counterintuitive as it is not apparent why experienced respondents would be subject to hypothetical bias and insensitive to cheap talk. One might conjecture that experience would make the contingent valuation exercise easier, but that does not appear to be the case, as we find experienced individuals to be as susceptible to hypothetical bias as are inexperienced individuals. This result is consistent with the Blumenschein et al. (2008) study, as the participants were individuals with Type-2 diabetes and the good was participation in a diabetes management program. It would appear that the study population was “experienced,” susceptible to hypothetical bias, and insensitive to cheap talk. While the cheap talk technique seems to operate differently for experienced and inexperi-

Table 7. Percentage Responding Yes to Willingness-to-Donate Question by Treatment and Visitation to International Crane Foundation

Treatment	Visited International Crane Foundation	Never Visited International Crane Foundation
Actual (n = 223)	28%	26%
CV (n = 235)	67% ^a	50% ^a
Cheap talk (n = 231)	58% ^a	30%
Certainty (8) (n = 233)	42%	29%

^aSignificantly different from actual treatment at $\alpha = .05$.

enced respondents, the effectiveness of the follow-up certainty treatment does not appear to be as sensitive to experience. We further explore this result when we include the certainty and cheap talk data in the regression model described below.

To better understand how the certainty and cheap talk treatments compare to the actual payment treatment with respect to providing a positive response to the willingness-to-donate question, we estimated a multivariate model in which the probability of answering *yes* to the donation question is conditioned in part on treatment. Table 8 describes the independent variables and Table 9 summarizes the results of the model.² The results suggest that individuals asked about higher offer amounts (*offer*), those who felt they could not afford to make a donation (*afford*), and those who thought the radio transmitters would be purchased regardless of whether they made a donation (*regardless*) were less likely to answer *yes* to the donation question. Individuals who perceived a relationship between making a donation and the likelihood of seeing a whooping crane (*donate*), those who thought the whooping cranes were worth the cost of reintroduction (*worth*), and those who thought the radio transmitters would have a positive impact on saving the cranes (*save*), as well as those who had a strong interest in the environment (*environment*), were more likely to answer *yes* to the donation question. The estimated coefficients on the demographic vari-

ables (*gender* and *income*) were not significantly different from zero. Two dummy variables were created to indicate the certainty correction at 8 (*cert*) or cheap talk (*CT*) treatment with the actual payment treatment as the baseline. The coefficients on these variables were not significant. Given the potential role experience plays in the effectiveness of the treatment, interaction terms between the treatment and having visited the International Crane Foundation were included in the model. For the cheap talk treatment, the coefficient on the interaction term *CT*Visit* is positive and statistically significant, confirming the result shown in Table 7 that those who visited the International Crane Foundation were less sensitive to the cheap talk treatment and were more likely to say *yes* to the willingness-to-donate question. The parameter on the interaction term *Cert*Visit* is not significantly different from zero. The model was also run without recoding the CV data, and in that case the CV treatment variable was significant and positive, confirming the existence of hypothetical bias if no correction is made.³

Conclusions

The goal of this study was to compare two approaches to mitigating hypothetical bias associated with a dichotomous choice contingent valua-

² The model in Table 9 had different independent variables than the model in Table 7 because we are modeling different phenomena. In the Table 7 model, only respondents who said *yes* to the willingness-to-donate question are included. Therefore, we include some variables that are available for only those respondents. In Table 9, the model includes all respondents in the study (i.e., both *yes* and *no* respondents to the willingness-to-donate question).

³ An alternative approach to including treatment variables in the model is to run separate models for each treatment and joint models that combine two of the treatments. Likelihood ratio tests can be conducted to test whether the coefficients are simultaneously equal between treatments. Individual coefficients can also be tested for equality across treatments. We estimated the necessary models for this approach and found the results to be similar to running the model in Table 9 with the treatment variables. We decided to go with the model in Table 9 because the exposition was easier.

Table 8. Explanatory Variables in Logistic Regression

Variable Name	Description	Mean (Std. Dev.)		
		Actual	Cert	Cheap Talk
<i>Offer</i>	\$10, \$15, \$25, \$50, \$100	.38 (.30)	.38 (.30)	.37 (.30)
<i>Visit</i>	Ever visited International Crane Foundation (1 = yes, 0 = no)	.32 (.47)	.31 (.46)	.28 (.45)
<i>Afford</i>	Can't afford to make a donation (1 = agree, 0 = otherwise)	.42 (.49)	.28 (.45)	.29 (.46)
<i>Donate</i>	I felt that if I said I would donate, I would more likely see a whooping crane in Wisconsin (1 = agree, 0 = otherwise)	.26 (.44)	.37 (.48)	.29 (.45)
<i>Worth</i>	Having whooping cranes in Wisconsin is worth the cost of reintroduction (1 = agree, 0 = otherwise)	.76 (.43)	.84 (.37)	.80 (.40)
<i>Save</i>	Fitting the whooping cranes with radio transmitters will have a positive impact on the ability of researchers to save the whooping cranes (1 = agree, 0 = otherwise)	.77 (.42)	.78 (.41)	.83 (.38)
<i>Regardless</i>	The radio transmitters will be purchased regardless of whether or not I make a donation (1 = agree, 0 = otherwise)	.64 (.48)	.64 (.48)	.67 (.47)
<i>Environment</i>	Interest in the environment (1 = strongly or extremely interested, 0 = otherwise)	.48 (.50)	.59 (.49)	.57 (.50)
<i>Gender</i>	1 = male, 0 = female	.58 (.49)	.68 (.47)	.61 (.49)
<i>Income</i>	In thousands of dollars	.65 (.29)	.63 (.29)	.61 (.26)

tion study. One approach explicitly told respondents about the hypothetical bias problem and asked them to try to avoid it. The other involved a correction after the CV question based on how certain respondents said they were about their response to the willingness-to-donate question. Using an actual payment treatment as our benchmark, we found evidence of hypothetical bias. Considering the overall percentage of respondents who said *yes* to the contingent donation question, the cheap talk treatment did not eliminate the hypothetical bias. When the follow-up certainty question was used to recode the CV data at a cut-off point of 8 on the 10-point scale, the percentage that said *yes* at every offer amount and overall was similar to that in the actual payment treatment. Nonparametric estimates of mean willingness to donate based on the actual payment, cheap talk, and certainty treatments are statistically indistinguishable. Likewise, a multivariate model suggests that both the cheap talk and the

certainty cutoff at 8 treatments is similar to the actual payment treatment. The ability of the follow-up certainty correction to eliminate hypothetical bias was not affected by whether or not the respondent had visited the International Crane Foundation. However, individuals who had previously visited the International Crane Foundation (our measure of “experience”) were less sensitive to the cheap talk treatment. This result was similar to the List (2001) and Blumenschein et al. (2008) studies with private goods. This unintuitive result is worthy of additional exploration. Conventional wisdom suggests that individuals who have experience with a particular good should find the CV exercise less cognitively burdensome. If that is really the case, we would expect to find less hypothetical bias with experienced respondents relative to those without experience. This result is cause for concern when applying the cheap talk treatment to reduce hypothetical bias, as most study populations will

Table 9. Logistic Regression of Willingness-to-Donate Function (dependent variable: donation question, 0 = no, 1 = yes; n = 561)

Explanatory Variable	Coefficient	Std. Error	p-value
Constant	-1.494	0.668	0.025
<i>Offer</i>	-0.026	0.005	0.000
<i>Visit</i>	-0.786	0.492	0.110
<i>Afford</i>	-4.042	0.625	0.000
<i>Donate</i>	0.710	0.261	0.007
<i>Worth</i>	1.625	0.426	0.000
<i>Save</i>	1.048	0.374	0.005
<i>Regardless</i>	-0.993	0.260	0.000
<i>Environment</i>	0.871	0.253	0.001
<i>Gender</i>	-0.325	0.264	0.217
<i>Income</i>	0.005	0.005	0.267
<i>Cert</i>	-0.548	0.376	0.144
<i>CT</i>	-0.334	0.375	0.370
<i>Cert*Visit</i>	0.981	0.633	0.121
<i>CT*Visit</i>	1.445	0.670	0.031
-2 Log likelihood	431.773		
Percentage predicted correctly	80.90		

include individuals who are experienced with the good, and the cheap talk treatment may not be effective for those individuals.

There are a couple of caveats to generalizing the results of this study. While the vast majority of surveys are self-administered, one limitation of the approach is the relative lack of control the researcher has over how the survey is administered, making it impossible to know what was and was not read by the respondent and the order in which the respondent read the material and answered the questions. This lack of control could affect our results in two ways. First, if some in the cheap talk group did not read the cheap talk script, they effectively did not receive the treatment assigned to them. This would likely dilute the effectiveness of the treatment relative to a cheap talk script that is read aloud in an in-person

setting. Second, respondents in the certainty group may have read the follow-up certainty question before they answered the donation question, and this could have influenced their donation response. We do not have a “treatment-free” contingent donation group, but such an effect was not found in Champ et al. (1997). Likewise, the significantly higher *yes* rate for the certainty treatment relative to the actual donation treatment is consistent with numerous other studies that compared hypothetical and actual payment decisions that did not include a follow-up certainty question.

Based on the results of this study and similar research (Little and Berrens 2003, Blumenschein et al. 2008), we suggest that the follow-up certainty scale has an advantage over a medium-length cheap talk script in that it appears to work

more predictably across offer amounts and respondent experience. If the researcher is not sure what cutoff point on the scale is most appropriate, she can report willingness-to-pay estimates associated with several cutoff points. While more research is needed, the evidence to date suggests that the mean level of certainty may be an appropriate cutoff point. Another benefit of the follow-up certainty approach over the cheap talk approach is that the uncertainty information can be incorporated into the willingness-to-pay estimates (Shaikh, Sun, and van Kooten 2007, Moore 2006) if a more rigorous approach is desired.

References

- Aadland, D., and A.J. Caplan. 2003. "Willingness to Pay for Curbside Recycling with Detection and Mitigation of Hypothetical Bias." *American Journal of Agricultural Economics* 85(2): 492–502.
- _____. 2005. "Cheap Talk Reconsidered: New Evidence from CVM." *Journal of Economic Behavior and Organization* 60(4): 562–578.
- Blumenschein, K., G.C. Blomquist, M. Johannesson, N. Horn, and P. Freeman. 2008. "Eliciting Willingness to Pay without Bias: Evidence from a Field Experiment." *The Economic Journal* 118(1): 114–137.
- Blumenschein, K., M. Johannesson, G.C. Blomquist, B. Liljas, and R.M. O'Conor. 1998. "Experimental Results on Expressed Certainty and Hypothetical Bias in Contingent Valuation." *Southern Economic Journal* 65(1): 169–177.
- Brown, T.C., I. Ajzen, and D. Hrubes. 2003. "Further Tests of Entreaties to Avoid Hypothetical Bias in Referendum Contingent Valuation." *Journal of Environmental Economics and Management* 46(2): 353–361.
- Champ, P.A., and R.C. Bishop. 2001. "Donation Payment Mechanisms and Contingent Valuation: An Empirical Study of Hypothetical Bias." *Environmental and Resource Economics* 19(4): 383–402.
- 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.
- Champ, P.A., N.E. Flores, T.C. Brown, and J. Chivers. 2002. "Contingent Valuation and Incentives." *Land Economics* 78(4): 591–604.
- Cummings, R.G., and L.O. Taylor. 1999. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review* 89(3): 649–665.
- Ethier, R.G., G.L. Poe, W.D. Schulze, and J. Clark. 2000. "A Comparison of Hypothetical Phone and Mail Contingent Valuation Responses for Green-Pricing Electricity Programs." *Land Economics* 76(1): 54–67.
- Johannesson, M., G.C. Blomquist, K. Blumenschein, P.-O. Johansson, B. Liljas, and R.M. O'Conor. 1999. "Calibrating Hypothetical Willingness to Pay Responses." *Journal of Risk and Uncertainty* 18(1): 21–32.
- 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(2): 256–269.
- List, J.A. 2001. "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards." *The American Economic Review* 91(5): 1498–1507.
- List, J.A., and C.A. Gallet. 2001. "What Experimental Protocol Influences Disparities Between Actual and Hypothetical Stated Values?" *Environmental and Resource Economics* 20(3): 241–254.
- Little, J., and R. Berrens. 2003. "Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta-Analysis." *Economics Bulletin* 3(6): 1–13.
- Lusk, J.L. 2003. "Effects of Cheap Talk on Consumer Willingness-to-Pay for Golden Rice." *American Journal of Agricultural Economics* 85(4): 840–856.
- Moore, R. 2006. "Essays on Non-market Valuation: Innovations in Theory and Methods." Ph.D. dissertation, University of Wisconsin-Madison, WI.
- Murphy, J.J., P.G. Allen, T.H. Stevens, and D. Weatherhead. 2005. "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation." *Environmental and Resource Economics* 30(3): 313–325.
- Murphy, J.J., T.H. Stevens, and D. Weatherhead. 2005. "Is Cheap Talk Effective at Eliminating Hypothetical Bias in a Provision Point Mechanism?" *Environmental and Resource Economics* 30(3): 327–343.
- Norwood, F.B. 2005. "Can Calibration Reconcile Stated and Observed Preferences?" *Journal of Agricultural and Applied Economics* 37(1): 237–248.
- Norwood, F.B., J.L. Lusk, and T. Boyer. 2008. "Forecasting Hypothetical Bias: A Tale of Two Calibrations." In T.L. Cherry, S. Kroll, and J.F. Shogren, eds., *Environmental Economics, Experimental Methods*. New York: Routledge.
- Ready, R.C., J.C. Whitehead, and G.C. Blomquist. 1995. "Contingent Valuation When Respondents Are Ambivalent." *Journal of Environmental Economics and Management* 29(2): 181–196.
- Sammaliev, M., T.H. Stevens, and T. More. 2006. "A Comparison of Alternative Certainty Calibration Techniques in Contingent Valuation." *Ecological Economics* 57(3): 507–519.
- Shaikh, S.L., L. Sun, and G.C. van Kooten. 2007. "Treating Respondent Uncertainty in Contingent Valuation: A Comparison of Empirical Treatments." *Ecological Economics* 62(1): 115–125.
- 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(2): 170–185.

Appendix A. CV Question with Follow-Up Certainty Scale

1. If you were asked to make a donation to purchase radio transmitters today, would you be willing to donate \$_____? (CIRCLE ONE NUMBER)

1 No → Skip to Question 4

2 Yes

2. If you answered YES to question 1, on a scale of 1 to 10, where 1 means “very uncertain” and 10 means “very certain,” how certain are you that you would donate \$___ if you had an opportunity to actually do so? (CIRCLE ONE NUMBER)

1 2 3 4 5 6 7 8 9 10
VERY UNCERTAIN VERY CERTAIN

Appendix B. Cheap Talk Script

Please consider this information as you answer the following questions.

The Whooping Crane Reintroduction Program is a real program. However, as this is a research project, we are not asking you to make an actual donation. Nonetheless, we would like you to answer the following question as if this were an actual donation solicitation. Studies have shown that answering a question about a hypothetical donation as if the donation is for real is difficult for many people. We consistently find that more individuals say they will make a donation in the hypothetical situation than when the situation is real. One reason we think this happens is because when the donation is hypothetical, respondents might be thinking “Sure I’ll donate, this is a worthy cause.” But when the decision involves actually making a donation, respondents might think “Do I *really* want to spend my money on this cause?” We ask that you try to avoid this problem and answer the following question as you would a solicitation for an actual donation.