

# 2013 WAEA Keynote Address: Strategies for Overcoming Hypothetical Bias in Stated Preference Surveys

John B. Loomis

In some, but not all, contexts, respondents to stated preference valuation studies state a willingness to pay (WTP) higher than what lab or field experiments indicate is the actual amount they would pay. However, several *ex ante* survey design strategies and *ex post* calibration techniques can be used to minimize or eliminate hypothetical bias. This article reviews and presents evidence on the effectiveness of these *ex ante* and *ex post* approaches. The *ex ante* approaches lead to recommendations to modify survey designs to minimize the bias up front. If the analyst desires, *ex post* calibration of WTP using certainty scales can be used to reduce stated WTP to the point at which it will match actual cash contributions.

*Key words:* choice experiments, conjoint, contingent valuation, hypothetical bias, meta-analysis, nonmarket valuation, stated preference methods

## Introduction

Applied economists are often called upon to value the introduction of new public or private goods for which no markets currently exist. In the case of private goods, agricultural economists are asked to evaluate the market potential of newly configured agricultural products that are not currently marketed. For example, economists assess consumers' willingness to pay a premium, and if so, the size of that premium, for particular organic foods, locally grown foods, food production with less carbon intensity, increased health attributes for foods, etc. (see Lusk and Schroeder, 2004; Norwood and Lusk, 2011; Louviere, Hensher, and Swait, 2000, for a general review).

For decades, environmental economists have been asked to value alternative policy proposals to improve water quality (e.g., how clean is economically justified?), remove existing dams, reintroduce endangered wolves, and so on. This valuation information is often used in policy analyses, environmental impact statements, and resource management plans. The very nature of benefit-cost, policy analysis, and planning is *ex ante*: to guide a decision about what to do in the near future. The action has not been taken yet and the particular good or resource does not exist in that particular location. Hence it is difficult for economists to rely upon actual behavior (revealed preferences). The dams are still in place, so no revealed preference studies can be undertaken. Further, when measuring existence values (e.g., willingness to pay to just know that a self-sustaining wolf population exists in Yellowstone National Park) for proposed actions, there is little behavior to measure (e.g., as there were no wolves in Yellowstone at the time of the Environmental Impact Statement (EIS) and non-use value leaves no behavioral trail).

This article focuses on two stated preference methods commonly used by applied economists: the contingent valuation method (CVM) and choice experiments (CE). For an introduction to CVM,

John B. Loomis is a professor in the Department of Agricultural and Resource Economics at Colorado State University. The author would like to thank the anonymous referees and the Gregmar Galinato for comments that improved the clarity of the manuscript. Funding provided by Colorado Agricultural Experiment Station and W3133 Regional Research Project.

Review coordinated by Gregmar I. Galinato.

see Kling, Phaneuf, and Zhao (2012) and for choice experiments see Louviere, Hensher, and Swait (2000). Readers should bear in mind that this article is not an exhaustive literature review but what the author believes to be studies representative of each of the *ex ante* and *ex post* approaches.

### **Hypotheses about Hypothetical Bias and Resulting Strategies to Overcome**

In stated preference valuation surveys, hypothetical bias can be defined as the difference between what a person indicates they would pay in the survey or interview and what a person would actually pay. Simply put, hypothetical bias in surveys reflects the old saying that “there is a difference between saying and doing.” This concern is pervasive across surveys in any number of social sciences, whether estimating unemployment rates (If you were offered a job this week would you have taken it?) or political polling (If the election were held today, would you vote for or against this proposition?).

However, measuring hypothetical bias is difficult for nonmarketed resources and public goods. In order to measure hypothetical bias, the analyst needs to know the “true” willingness to pay (WTP), but it is hard to deliver and actually force people to pay for a nonexcludable public good. Basic microeconomic theory dating back to Samuelson (1955) suggests that people do not have an incentive to reveal their true WTP and rather hope to “free ride” off others’ donations. Nonetheless, cash validity studies for pure public goods often compare stated WTP with actual donations (Champ et al., 1997; Champ, Moore, and Bishop, 2009). Thus, as Champ et al. (1997) note, *actual* donations may not reflect respondents’ maximum WTP. Compounding this problem, Carson and Groves (2007, p. 188) state that the incentive when using *hypothetical* donations may be to overstate WTP or agree to pay \$Y even if their actual value is less than \$Y, because they wish to see the fund established so that the organization usually described in the survey would undertake fundraising activities to provide public good. Thus the use of a donation vehicle could increase the magnitude of hypothetical bias observed in a survey because the actual donation payment may understate WTP, while the hypothetical donation payment may overstate WTP.

Likewise, in private good experiments, respondents’ maximum WTP may be truncated by their assessment of the price for which a similar good sells in the market (Harrison, 2006; Harrison, Harstad, and Rutstrom, 2004). Furthermore, with private, excludable goods like hunting permits, a respondent’s incentive may be to understate WTP to prevent or minimize any increase in license fee that the respondent feels might arise from their answer to a WTP question. For example, Loomis, Pierce, and Manfredo (2000) found about a 50% understatement of WTP for deer and elk hunting licenses from a contingent valuation method (CVM) survey compared to WTP calculated from a demand curve estimated using the variation in actual hunting license prices for the same species over time.

Thus, in many cash validity studies, what appears to be hypothetical bias may be problems with estimating the “true” WTP in the first place. For the purposes of this article, we define hypothetical bias as the difference between actual cash WTP or actual purchase intentions and stated WTP or stated purchase intentions. However, consistent with others (Little and Berrens, 2004, p. 2), the empirical evidence cited below, and general concern voiced by many skeptical economists (Diamond and Hausman, 1994; Hausman, 2012), this hypothetical bias often manifests itself as overstatement of WTP.

The issue of hypothetical bias in stated preference surveys was first studied not long after the method itself appeared. Knetsch and Davis (1966) made the first comparison of revealed and stated preference estimates and Bohm (1972) performed the first comparison of actual and hypothetical willingness to pay was performed within a decade after Davis (1963) introduced the contingent valuation method (CVM). Since that time several thousand CVM studies been published. There have also been nearly 100 comparisons of actual cash and hypothetical WTP Little and Berrens (2004) and hundreds of comparisons of revealed preference and stated preference (SP) studies (Carson

et al., 1996), including comparisons to real referenda (Vossler and Kerkvliet, 2003; Johnston, 2006; Vossler and Watson, 2013). Many of these studies—especially the more recent ones—do not show hypothetical bias. However, a meta-analysis by Murphy et al. (2005) of just WTP studies shows a median overstatement of 1.35, while the List and Gallet (2001) and Little and Berrens (2004) studies show a median overstatement of about 3, a higher estimate due perhaps in part from including WTA studies in their meta-analysis. How much the magnitude of this overstatement may be due to understatement of actual WTP in the lab versus overstatement of WTP in the SP surveys is not clear at this time and is an important avenue for future research.

Despite all of the empirical work on hypothetical bias, there is still no widely accepted theory of hypothetical bias in stated preference surveys (Murphy et al., 2005, p. 318). Five years later, Mitani and Flores (2010) echoed this same conclusion: “The underlying causes of hypothetical bias are not yet sufficiently understood, and the theoretical or systematic explanation remains as one of the major questions in the stated preference economic analysis.” There are, however, several theories and plausible hypotheses about how a person may respond when asked how much they would hypothetically pay for a particular public good. One of the best-known theories is by Carson and Groves (2007) and is discussed in more detail below.

The hypotheses about how people respond to WTP questions imply different approaches for countering the hypothetical bias. Some of the hypotheses include *ex ante* approaches to reduce hypothetical bias by survey design. Other approaches involve *ex post* recoding or calibration of WTP responses to purge the stated WTP of hypothetical bias. While each method for mitigating hypothetical bias springs from a different hypothesis, these hypotheses are not necessarily inconsistent with one another. Some recent research has combined different *ex ante* methods or combined *ex ante* and *ex post* methods (Whitehead and Cherry, 2007).

### ***Ex Ante* Approaches for Reducing Hypothetical Bias**

There are four primary *ex ante* survey design approaches: (1) emphasize consequentiality of the survey and respondents’ choices; (2) urge respondents to be honest and to act as though they really had to pay here and now; (3) use cheap talk approaches, which explicitly communicate the problem of hypothetical bias to respondents; (4) reduce social desirability bias, the tendency to give answers that the respondent considers to be socially acceptable or what they think the interviewer wants to hear.

#### *Consequentiality Designs*

This *ex ante* approach was developed by Carson and Groves (2007), who suggested including three features in a survey in order for a hypothetical constructed market or simulated voter referendum to be potentially incentive compatible (e.g., truth revealing). First, the survey must be consequential to the respondent; that is, it must have some potential effect on their future utility such as higher taxes or increased probability that the public good will be supplied. One way to do this is to cast the referenda as an “advisory one” to public officials who will be making a decision on the issue. Second, only a binary, dichotomous choice question format is potentially demand revealing. Thus, a valuation question in which the respondent chooses one from among several product profiles or multiple alternative public goods (e.g., some choice experiment (CE) designs) would not meet this condition, but a yes/no vote would. Vossler, Doyon, and Rondeau (2012) formally extended consequentiality theory to binary choice experiments. Finally, the payment mechanism must be compulsory (such as a tax paid by all) if the referendum passes. This last requirement rules out using donations as an incentive-compatible payment vehicle.

The results to date on the performance of consequential CVM and CE surveys are encouraging. A study by Vossler and Kerkvliet (2003) was one of the first to find similarity between an actual referendum (in Oregon) and a CVM study. Johnston (2006) found no statistical difference between the proportion of people who voted yes in a hypothetical CVM referendum on a water supply project in a small town in Massachusetts and the subsequent actual vote four months later. The author stated the survey was consequential to respondents and respondents knew this was an actual water supply project being considered. The results of the CVM survey would influence the town manager's decision of whether to actually hold the water supply referendum or not. If the survey indicated that people would not pay the required cost, no actual referendum would likely have taken place. Vossler and Evans (2009) found that increasing the consequentiality of a vote also resulted in a closer correspondence between SP responses and their criterion in a lab experiment. Landry and List (2007) found that consequentiality resulted in the proportion of respondents willing to pay two different bid amounts not being statistically different than the actual proportion. Vossler and Watson (2013) found that their CVM related to a community referendum in Massachusetts on open space and historic preservation to be paid for by a property tax matched the actual vote (53.5% actual vote versus 50.7% overall hypothetical and 57.4% among those that thought the CVM survey was consequential).

Mitani and Flores (2010) hypothesized that hypothetical bias arises because of respondents' uncertainty about the degree of consequentiality, particularly regarding the likelihood of paying. While this hypothesis is similar to Champ et al. (1997, reviewed below), Mitani and Flores' approach focuses on survey design and hence is *ex ante*. In particular, a survey must discuss the provision decision rule and the likelihood of payment. Failure to do so leaves the respondent uncertain about the likelihood of having to pay the full bid amount and the likelihood of the public good being provided, and this uncertainty leads to hypothetical bias. They then demonstrate that if the respondent is told in the survey that the good will be provided based on the results of the survey (probability of provision  $> 0$ ) and the same probability that they will have to pay, then theoretically respondents should reveal their true value. They confirm this effect with an induced valuation experiment in which a series of WTP exercises are undertaken in which the probability pairs of provision and payment are made explicit, along with a provision point type threshold for delivery of the good. Their results have several implications for survey design: "First, it is essential to induce subjective probabilities so that the probability of payment equals the probability of provision. In the experimental or survey designs, it will be important to control both payment and provision sides in the same way" (p. 38).

### *Honesty and Realism Approaches*

In this approach, respondents are exhorted to report what they would honestly pay, here and now, not what they think the good would sell for in a market. Loomis et al. (1996) made one of the first attempts at this approach and were able to reduce but not eliminate hypothetical bias. The honesty approach was recently formalized by administering an oath to respondents, who were asked to swear to tell the truth (Jacquemet et al., 2013). This approach asked subjects in the lab to sign an oath stating, "I the undersigned (name) swear upon my honour that, during the whole experiment, I will tell the truth and always provide honest answers." The first test of the method showed it to be a promising approach for reducing or eliminating hypothetical bias, although it appears to be more powerful when combined with "cheap talk" (discussed below). A recent replication of the oath also found that the oath was able to eliminate hypothetical bias among students (Stevens, Tabatabaei, and Lass, 2013).

**Table 1. Results of Cheap Talk Tests of Reducing Hypothetical Bias**

Study	Eliminated	Reduced	Over-Corrected	No Effect
Cummings and Taylor (1999)	1			
Aadland and Caplan (2003)		1		
Brown, Ajzen, and Hrubes (2003)		1		
Morrison and Brown (2009)	1		1	
Champ, Moore, and Bishop (2009)		1		
Blumenschein et al. (2008)				1
Landry and List (2007)	1			
<b>Total</b>	<b>3</b>	<b>3</b>	<b>1</b>	<b>1</b>

### *Cheap Talk*

Cummings and Taylor (1999) used a “cheap talk” script that explicitly confronts the problem of hypothetical bias by telling respondents that participants in past surveys have been shown to overstate their WTP. Respondents are instructed not to do this, but instead to respond with what they would actually do if this were a real decision with their own money. Cummings and Taylor found this approach successful in the lab (to obtain hypothetical WTP equal to actual) as have some others (Aadland and Caplan, 2003; Landry and List, 2007). However, other studies have had less success (Blumenschein et al., 2008; Aadland and Caplan, 2006, who found it increased WTP). Some studies have found that “cheap talk” is primarily effective for just certain types of respondents, such as individuals unfamiliar with the good (List, 2001; Champ, Moore, and Bishop, 2009). The tabulation of the available studies shown in table 1 indicates that out of eight estimates, “cheap talk” eliminated hypothetical bias in three study estimates, reduced but did not eliminate hypothetical bias in three other estimates, had no effect in one study, and over-corrected for hypothetical bias (leading to stated WTP below actual WTP) in one study.

### *Reducing Social Desirability Bias and Cognitive Dissonance*

Another approach attempts to minimize the tendency a respondent might have to give a socially acceptable answer or one that they think the interviewer or the survey sponsor wants to hear (Babbie, 1992). This tendency stems from the potential for the respondent to feel some emotional discomfort—or what psychologists call “cognitive dissonance”—revealing their actual value. One approach for reducing social desirability bias and cognitive dissonance is called “inferred valuation,” as it asks respondents what they think others would pay for the good rather than asking what they themselves would pay (Lusk and Norwood, 2009; Norwood and Lusk, 2011). This method is based on the hypothesis that respondents gain utility from giving answers that are consistent with social norms rather than their own personal values. Specifically, the respondent gains utility from pleasing the interviewer or maintaining a positive self-image. However, when they are faced with an actual cash situation, respondents behave more selfishly. Respondents also have a more negative view of what they think others would pay. Thus, when asked what they think others would pay, they provide answers that are stripped of social desirability bias and are more realistic indications of what they themselves would actually pay in a real situation. In laboratory experiments, this method has worked well to eliminate hypothetical bias in two different experiments (Lusk and Norwood, 2009; Norwood and Lusk, 2011).

Also included in this category are approaches that reduce respondents’ potential cognitive dissonance in dichotomous choice or referendum format situations. In these formats, different respondents are asked whether they would pay a dollar (bid) amount that varies systematically across the sample in order to reveal maximum WTP. Brown et al. (1996) first suggested that respondents confronted with this question face a cognitive dissonance between two objectives, especially at high bid amounts: (a) say “No,” they would not pay that high amount, which creates the potential worry

that they are signaling the interviewer (or the institution the interviewer is representing) that they do not value the good at all, or (b) engage in “yea-saying” in order to indicate they do value the good, even if they do not value the good as high as the dollar amount presented to them in their survey version. To alleviate this cognitive dissonance, Blamey, Bennett, and Morrison (1999) and Loomis, Traynor, and Brown (1996) provided a third option, or what Loomis et al. call “trichotomous choice.” As the name implies, a third choice is offered, allowing the respondent to vote for the program, but not at the original bid amount. This provides the respondent a way out of the cognitive dissonance and signals the researcher that the respondent’s value is positive but lower than the original bid amount. Both Blamey, Bennett, and Morrison (1999) and Loomis, Traynor, and Brown (1996) showed that this method reduced WTP, but Morrison and Brown (2009) performed the first formal validity test. They found that having six payment options for the respondents resulted in hypothetical WTP being slightly less than actual WTP. More replication is needed to determine whether this approach is always conservative.

### ***Ex Post Approaches to Reducing Hypothetical Bias***

#### *Data Screening*

*Ex post* screening of survey responses is particularly useful if the survey uses open-ended WTP questions, as there is no upper bound on the monetary amount a respondent can offer in these cases. This problem is somewhat minimized with the use of payment cards to elicit WTP. Nonetheless, even with payment cards, some respondents may circle monetary amounts that may appear to the researcher to be an “implausibly high” fraction of their income. One approach to minimize the influence of these potentially questionable observations is to report the median WTP and what Mitchell and Carson (1989, p. 227) referred to as an “alpha trimmed mean.” If multiple regression techniques are applied to the data, with WTP as the dependent variable, another possibility is to identify outliers (e.g., more than three standard deviations from the regression line) and remove them. Another approach applicable to dichotomous choice response data is to use the Turnbull lower-bound approach (Haab and McConnell, 2002). This nonparametric approach truncates the upper end of the WTP integral to the highest bid amount asked in the survey so as to not extrapolate the WTP distribution beyond the highest bid used in the survey.

A recent and more sophisticated *ex post* calibration method is called the Orbit model (Davies and Loomis, 2010). This technique combines an ordered probit model and a tobit model to address the fact that stated open-ended amounts beyond some point might be realistically treated as ordinal indicators rather than cardinal amounts. The technique anchors the combined ordered probit and tobit model at two points: (a) usually at zero, as those respondents that state zero are believed to reflect valid responses, and (b) the median or mean of the data. Beyond the median, reported values are treated as ordinal responses (hence the ordered probit part of the likelihood function). One advantage of this method is that it allows differential calibration of stated quantities between zero and the median and beyond the median. The approach has been applied to stated trips, but has yet to be applied to WTP.

#### *Related Market Calibration*

Another *ex post* approach involves calibrating stated WTP based on “actual behavior” responses in a market or lab experiment. Fox et al. (1998) proposed a calibration approach called CVM-X. Using this approach, the researcher conducts a CVM survey; then, a subset of those respondents are asked to participate in a lab validity experiment with the same good as used in the survey. Based on the lab validity test, the analyst then calculates the ratio of actual WTP to hypothetical WTP elicited from the same respondents in an experiment. This is then applied to the CVM WTP estimates from the survey. The CVM-X cannot generate “universal” calibration factors because the authors warn that

**Table 2. Summary of Effect of *ex post* Uncertainty Calibration Recoding on Hypothetical Bias**

Study	Eliminated	Reduced	Over-Corrected
Champ et al. (1997)	1		
Morrison and Brown (2009)	1		1
Champ, Moore, and Bishop (2009)	1	1	
Ethier et al. (2000)	1		1
Champ, Moore, and Bishop (2009)	1		
Macha, Donfouet, and Mahieu (2012)		1	
Blumenschein et al. (2008)	1		
<b>Total</b>	<b>6</b>	<b>2</b>	<b>2</b>

calibration factors developed for deliverable goods typically used in the lab may not be transferrable to some public goods needing to be valued (Fox et al., 1998, p. 464). The method is also limited to goods that have some deliverability in the lab, making its application to pure public goods difficult.

### *Uncertainty Recoding*

Champ et al. (1997) suggested that hypothetical bias originates in respondents' uncertainty about various dimensions of the CVM survey (e.g., what the good is worth to them). They argue that hypothetical bias in a dichotomous choice WTP response can be reduced by recoding responses that vote "Yes" to a "No" if the respondent expresses a high degree of uncertainty about their affirmative WTP response. This approach is particularly relevant in valuing public goods that the respondent may have not thought about (e.g., removing roads on the north rim of the Grand Canyon, Champ et al., 1997) or for which they had never thought of in monetary values (enhancing whooping crane populations, Moore et al., 2010). In this *ex post* approach, a respondent answers the dichotomous choice valuation question and then indicates how certain they are of their previous answer (often on a 1–10 scale). Based on the public opinion polling literature and several past calibrations of actual donations and hypothetical WTP, uncertain respondents that give "Yes" responses are coded as "No" in the WTP statistical analyses (Champ, Moore, and Bishop, 2009, see). The limited evidence suggests that recoding respondents with a certainty level of less than seven (or definitely sure in the narrative scale) matches actual cash donations reasonably well (Ethier et al., 2000; Morrison and Brown, 2009; Blumenschein et al., 2008). However, (as noted earlier) actual cash donations may understate a respondent's maximum WTP, so it is possible that the approach to uncertainty calibration taken by Champ and colleagues is over-correcting, leading to underestimates of maximum WTP.

This review of the available *ex post* validity studies suggests that the uncertainty recoding has been able to eliminate hypothetical bias in six out of ten WTP estimates, reduced it in two estimates, and over-corrected in two estimates. Table 2 provides a summary of these studies.

Of course, the ability to eliminate bias hinges on knowing the uncertainty level at which to convert "Yes" responses into "No" responses. This is difficult to know without a parallel actual donation, but Champ, Moore, and Bishop (2009, p. 179) stated, "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 (sample) mean level of certainty may be an appropriate cutoff point." This empirical fix begs for a choice theoretic or utility theoretic model of what determines how certain different respondents might be and why a particular level of certainty appears to yield valid statements of actual WTP.

Akter and Bennett (2012) offered a conceptual framework for incorporating respondent uncertainty about the scenario respondents are given, the policy itself, and their preferences. Recent advances in the uncertainty calibration method (Moore et al., 2010, p. 396) have moved away from simply recoding the data to developing an "uncertain respondent model" equation that the authors believe "... provides the theoretical foundation for directly incorporating information on respondent

choice uncertainty—as expressed on the certainty scale...—into an econometric model of choice behavior.” The authors indicate their method provides a lower bound on actual WTP.

### What Meta-Analyses Tell Us about the Effectiveness of *ex post* and *ex ante* Methods

We review three meta-analyses that evaluate hypothetical bias in stated preference surveys and experiments. Hypothetical bias is primarily assessed either through lab experiments or field studies. In the case of field studies, a split sample is used, in which one treatment receives a hypothetical WTP question as in the standard CVM and another treatment is provided the same information about the good but is asked to actually pay. All three of the meta-analyses reviewed focused on hypothetical bias as revealed in a mix of experiments and field studies.

The List and Gallet (2001, hereafter LG) meta-analysis utilized twenty-nine cash validity studies, split roughly equally between lab and field experiments, providing a total of 174 observations. They suggested that their analysis provides important insights into typical CVM studies, such as various elicitation methods including dichotomous choice, as well as other factors influencing calibration factors between hypothetical and actual WTP (List and Gallet, 2001, p. 242). Murphy et al. (2005) utilized twenty-eight studies with eighty-three observations. Murphy et al. did not include WTA studies in their analysis and had a few other differences in minor selection criteria, leading to the difference in the number of studies analyzed. Little and Berrens’ (Little and Berrens, 2004, hereafter LB) meta-analysis expanded on LG’s dataset both in the number of observations and the number of variables examined (adding three variables, one each for certainty correction, cheap talk, and referendum protocols). Further, they used a “clustering” procedure to control for repeated observations from the same study.

With respect to hypothetical bias, List and Gallet (2001, p. 246) found a median calibration factor of 3 (hypothetical payment divided by actual payment), Little and Berrens (2004, p. 5) found a median calibration factor of 3.13 in their expanded dataset, while Murphy et al. (2005), found a raw median ratio of hypothetical to actual WTP of just 1.35. The reduced degree of hypothetical bias in Murphy et al. is largely a result of not including WTA studies, which have a higher degree of hypothetical bias. This finding of added hypothetical bias with WTA is part of the reason why WTP is usually recommended as the welfare measure in CVM studies (Arrow et al., 1993; National Oceanic and Atmospheric Administration, 1994, 1996), and in fact is the most common welfare measure elicited.

There is some difficulty in directly comparing the meta-regressions of LG/LB to Murphy et al., since LG/LB used the calibration factor (hypothetical divided by actual WTP) as their dependent variable and Murphy et al. used the actual value as a dependent and include hypothetical value as an independent variable. Thus, the three studies can only be compared qualitatively:

- Type of good: Public versus Private. The evidence from LG’s meta-analyses indicates that using a private good reduces hypothetical bias (i.e., reduced the calibration factor in LG; increased actual value in Murphy, which—holding hypothetical value constant—also reduced the calibration factor). LB did not find this effect.
- Value Elicitation Question Format: LG’s and LB’s meta-analyses indicated that using a dichotomous choice question format did not reduce the degree of hypothetical bias and in fact has no statistical effect on the calibration factor. However, LB found that the referendum form of dichotomous choice reduced hypothetical bias. Murphy et al. included the dichotomous choice format with conjoint and payment card. They found that the closed-ended WTP question formats increased actual value in two out of three of their meta-analyses, hence reducing the calibration factor. With respect to commonly used laboratory elicitation methods (e.g., Smith auction, random price auction, etc.), LG found no statistically significant effect on the calibration factor, while LB found that the first price auction reduced hypothetical bias, but the Smith auction did not.

- WTP vs WTA: LG found using WTP reduced the calibration factor; surprisingly, LB did not find such an effect.
- Sample Pool: Murphy et al. found that using students in lab experiments significantly reduced actual value, hence increasing the hypothetical bias.
- *Ex ante* and *ex post* calibration efforts; LB included a separate variable for the *ex post* certainty correction in their meta-analysis. They found this correction resulted in a statistically significant reduction in hypothetical bias. Cheap talk did not have any significant effect. Murphy et al. combined the cheap talk and uncertainty recoding into a variable they called “Calibrate.” Their variable had a significant effect on increasing actual WTP in two out of three of Murphy et al.’s meta-regressions. By increasing actual WTP, *ceteris paribus*, hypothetical bias and the calibration factor were reduced.

There seems to be only limited agreement (i.e., in two out of the three author’s meta-models) for several factors reducing hypothetical bias. Tentatively, it appears that using uncertainty recoding and the referendum form of dichotomous choice CVM question format to elicit WTP may contribute to reducing hypothetical bias in CVM survey derived estimates of WTP.

As more studies comparing stated and actual valuation behavior are performed, meta-analyses of the calibration factors will hopefully begin to show consistent results regarding the extent to which *ex ante* survey designs and *ex post* correction adjustments influence the magnitude of the calibration factor. If and when such a state of affairs is attained, an intriguing possibility exists for applying the results of these meta-analyses to *ex post* calibrate WTP estimates from stated preference valuation studies. In particular, there is a potential to use a solid meta-analysis equation to calculate calibration factors to account for the specific details of a stated preference study. For example, a stated preference study valuing a public good with a payment card WTP question format might have a larger calibration factor than one for a private good (e.g., recreation) valued using a dichotomous choice survey with a cheap talk survey design.

## Discussion

What does a preponderance of empirical evidence suggest to an applied economist? It seems fairly clear that the uncertainty calibration correction is a promising approach to better align stated WTP estimates with actual willingness to pay (or donate). However, since donations may understate the correct welfare measure—maximum WTP—caution is warranted when applying this method. Further, the main stumbling block here is determining what level of uncertainty to use to recode “yes will pay responses” to “no” responses. The recommendation from Champ, Moore, and Bishop (2009) to use sensitivity analysis in order to report WTP at several reasonable cutoff levels of uncertainty (e.g., 6–9 on their ten point scale) is helpful when performing a benefit-cost or policy analysis. If the nonmarket benefits are the bulk of a project or policy’s benefits and these benefits are greater than the cost (even when using a high level of certainty as the cutoff), then it is likely the alternative or policy option is economically efficient. Thus, stated preference surveys would be wise to include a question asking respondents to indicate their level of certainty for their response.

The results from several empirical studies—especially those of Vossler and colleagues—suggest that making the survey as consequential as possible also seems to improve the consistency between actual behavior and stated behavior. Following Carson and Groves’ (2007) guidance, casting the choice in a binary voter referendum format, with a compulsory payment vehicle (e.g., some form of taxes or mandatory fee) appears to be an important *ex ante* design feature.

## Conclusion

The wide variety of explanations for hypothetical biases and resulting corrections indicates that there is no universal agreement on a single theory of behavior for stated preference survey respondents. Thus, there is no consensus regarding the best method to correct for hypothetical bias. While each theory has a plausible underlying explanation, some of these correction methods are more practical than others. *Ex ante* survey design methods are one approach. These methods include: (a) consequential survey designs, in which the survey uses a realistic advisory referendum with a compulsory payment vehicle such as taxes; (b) cheap talk, which explicitly informs respondents about hypothetical bias and how respondents often report higher WTP values than what they are actually willing to pay; (c) reducing social desirability bias by asking what the respondent thinks others will pay or providing response options that allows the respondent to register that they value the good without having to vote “Yes” for a bid amount higher than their “true” maximum WTP; (d) honesty priming methods, which ask respondents to give an answer that they would honestly pay, even going as far as having them sign an oath.

*Ex post* methods of adjusting WTP responses already obtained in a survey include: (a) uncertainty recoding of “yes would pay” responses into “no won’t pay” responses when the respondent is uncertain about their positive WTP response; (b) relying on median WTP responses rather than mean WTP responses, especially when using an open-ended WTP response question format or treating responses several times the mean response as an ordinal rather than a cardinal indicator when analyzing the data; and (c) relying on the degree of hypothetical bias uncovered in an experiment with a deliverable good to scale the WTP from a stated preference survey.

In addition, two or more of these approaches can be combined. For example, an *ex ante* method such as “cheap talk” can be used with another *ex ante* method, such as the oath in Jacquemet et al. (2013). Even combining *ex ante* and *ex post* approaches is feasible (Whitehead and Cherry, 2007). However, researchers should be cautious about combining too many *ex ante* methods or combining *ex ante* and *ex post* methods. It is possible to over-correct for hypothetical bias and hence underestimate WTP.

If an analyst is concerned about hypothetical bias, he or she can design surveys that minimize the bias and, if necessary, correct for any remaining bias after the data is collected. These options should give the applied economist some confidence he or she can provide conservative estimates of WTP using stated preference methods. These options are in the spirit of what Arrow et al. (1993) suggested as best practice in their recommendations. Whether for *ex ante* benefit-cost analysis of potential projects or policies or evaluating EIS alternatives of potential management actions, relying on what people say they would do can provide economically valuable information for decision makers.

[Received August 2013; final revision received December 2013.]

## References

- Aadland, D., and A. J. Caplan. “Willingness to Pay for Curbside Recycling with Detection and Mitigation of Hypothetical Bias.” *American Journal of Agricultural Economics* 85(2003):492–502.
- . “Cheap Talk Reconsidered: New Evidence from CVM.” *Journal of Economic Behavior & Organization* 60(2006):562–578.
- Akter, S., and J. Bennett. *Valuing Climate Change Mitigation: Applying Stated Preferences in the Presence of Uncertainty*. Northampton, MA: Edward Elgar, 2012.
- Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner, and H. Schuman. “Report of the NOAA Panel on Contingent Valuation.” *Federal Register* 58(1993):4601–4614.
- Babbie, E. R. *The Practice of Social Research*. Belmont, CA: Wadsworth, 1992, 6th ed.

- Blamey, R. K., J. W. Bennett, and M. D. Morrison. "Yea-Saying in Contingent Valuation Surveys." *Land Economics* 75(1999):126–141.
- Blumenschein, K., G. C. Blomquist, M. Johannesson, N. Horn, and P. Freeman. "Eliciting Willingness to Pay Without Bias: Evidence from a Field Experiment." *Economic Journal* 118(2008):114–137.
- Bohm, P. "Estimating Demand for Public Goods: An Experiment." *European Economic Review* 3(1972):111–130.
- Brown, T. C., I. Ajzen, and D. Hrubes. "Further Tests of Entreaties to Avoid Hypothetical Bias in Referendum Contingent Valuation." *Journal of Environmental Economics and Management* 46(2003):353–361.
- Brown, T. C., P. A. Champ, R. C. Bishop, and D. W. McCollum. "Which Response Format Reveals the Truth about Donations to a Public Good?" *Land Economics* 72(1996):152.
- Carson, R. T., N. E. Flores, K. M. Martin, and J. L. Wright. "Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods." *Land Economics* 72(1996):80–99.
- Carson, R. T., and T. Groves. "Incentive and Informational Properties of Preference Questions." *Environmental and Resource Economics* 37(2007):181–210.
- Champ, P. A., R. C. Bishop, T. C. Brown, and D. W. McCollum. "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods." *Journal of Environmental Economics and Management* 33(1997):151–162.
- Champ, P. A., R. Moore, and R. C. Bishop. "A Comparison of Approaches to Mitigate Hypothetical Bias." *Agricultural and Resource Economics Review* 38(2009):166–180.
- Cummings, R. G., and L. O. Taylor. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review* 89(1999):649–665.
- Davies, S. P., and J. B. Loomis. "An Improved Method for Calibrating Purchase Intentions in Stated Preference Demand Models." *Journal of Agricultural and Applied Economics* 42(2010):679–693.
- Davis, R. K. "Recreation Planning as an Economic Problem." *Natural Resources Journal* 3(1963):239–249.
- Diamond, P. A., and J. A. Hausman. "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives* 8(1994):45–64.
- Ethier, R. G., G. L. Poe, W. D. Schulze, and J. Clark. "A Comparison of Hypothetical Phone and Mail Contingent Valuation Responses for Green-Pricing Electricity Programs." *Land Economics* 76(2000):54–67.
- Fox, J. A., J. F. Shogren, D. J. Hayes, and J. B. Kliebenstein. "CVM-X: Calibrating Contingent Values with Experimental Auction Markets." *American Journal of Agricultural Economics* 80(1998):455–465.
- Haab, T. C., and K. E. McConnell. *Valuing Environmental and Natural Resources: The Econometrics of Non Market Valuation*. Northampton, MA: Edward Elgar, 2002.
- Harrison, G. W. "Experimental Evidence on Alternative Environmental Valuation Methods." *Environmental & Resource Economics* 34(2006):125–162.
- Harrison, G. W., R. M. Harstad, and E. E. Rutstrom. "Experimental Methods and Elicitation of Values." *Experimental Economics* 7(2004):123–140.
- Hausman, J. "Contingent Valuation: From Dubious to Hopeless." *Journal of Economic Perspectives* 26(2012):43–56.
- Jacquemet, N., R. Joule, S. Luchini, and J. F. Shogren. "Preference Elicitation under Oath." *Journal of Environmental Economics and Management* 65(2013):110–132.
- Johnston, R. "Is Hypothetical Bias Universal? Validating Contingent Valuation Responses Using a Binding Public Referendum." *Journal of Environmental Economics and Management* 52(2006):469–481.

- Kling, C. L., D. J. Phaneuf, and J. Zhao. "From Exxon to BP: Has Some Number Become Better than No Number?" *Journal of Economic Perspectives* 26(2012):3–26.
- Knetsch, J., and R. Davis. "Comparison of Methods for Recreation Valuation." In A. V. Kneese and S. C. Smith, eds., *Water Research*, Baltimore, MD: Johns Hopkins University Press, 1966.
- Landry, C. E., and J. A. List. "Using Ex Ante Approaches to Obtain Credible Signals for Value in Contingent Markets: Evidence from the Field." *American Journal of Agricultural Economics* 89(2007):420–429.
- List, J. A. "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards." *American Economic Review* 91(2001):1498–1507.
- List, J. A., and C. A. Gallet. "What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values?" *Environmental & Resource Economics* 20(2001):241–254.
- Little, J., and R. Berrens. "Explaining Disparities between Actual and Hypothetical Stated Values: Further Investigation Using Meta-Analysis." *Economics Bulletin* 3(2004):1–13.
- Loomis, J., T. Brown, B. Lucero, and G. Peterson. "Improving Validity Experiments of Contingent Valuation Methods: Results of Efforts to Reduce the Disparity of Hypothetical and Actual Willingness to Pay." *Land Economics* 72(1996):450–461.
- Loomis, J., C. Pierce, and M. Manfredo. "Using the Demand for Hunting Licenses to Evaluate Contingent Valuation Estimates of Willingness to Pay." *Applied Economics Letters* 7(2000):435–438.
- Loomis, J., K. Traynor, and T. Brown. "Trichotomous Choice: A Possible Solution to Dual Response Objectives in Dichotomous Choice Contingent Valuation Questions." *Journal of Agricultural and Resource Economics* 24(1996):572–583.
- Louviere, J. J., D. A. Hensher, and J. D. Swait. *Stated Choice Methods: Analysis and Application*. Cambridge: Cambridge University Press, 2000.
- Lusk, J. L., and F. B. Norwood. "An Inferred Valuation Method." *Land Economics* 85(2009):500–514.
- Lusk, J. L., and T. C. Schroeder. "Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks." *American Journal of Agricultural Economics* 86(2004):467–482.
- Macha, R. R., H. Donfouet, and P. Mahieu. "A Comparison of Oath and Certainty Calibration in Contingent Valuation Method: An Application to Community Health Fund." Working paper, Department of Economics and Statistics, University of Dodoma, Dodoma, Tanzania, 2012.
- Mitani, Y., and N. E. Flores. "Hypothetical Bias Reconsidered: Payment and Provision Uncertainties in a Threshold Provision Mechanism." Paper presented at the World Congress on Environmental and Resource Economics, Montreal, Canada, July 1, 2010.
- Mitchell, R. C., and R. T. Carson. *Using Surveys to Value Public Goods: The Contingent Valuation Method*. Washington, DC: Resources for the Future, 1989.
- Moore, R., R. C. Bishop, B. Provencher, and P. A. Champ. "Accounting for Respondent Uncertainty to Improve Willingness-to-Pay Estimates." *Canadian Journal of Agricultural Economics* 58(2010):381–401.
- Morrison, M., and T. C. Brown. "Testing the Effectiveness of Certainty Scales, Cheap Talk, and Dissonance-Minimization in Reducing Hypothetical Bias in Contingent Valuation Studies." *Environmental and Resource Economics* 44(2009):307–326.
- Murphy, J. J., P. G. Allen, T. H. Stevens, and D. Weatherhead. "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation." *Environmental & Resource Economics* 30(2005):313–325.
- National Oceanic and Atmospheric Administration. "Natural Resource Damage Assessment: Proposed Rules." *Federal Register* 59(1994):23,098–23,111.

- . “Natural Resource Damage Assessment: Final Rules.” *Federal Register* 61(1996):439–510.
- Norwood, F. B., and J. L. Lusk. “A Calibrated Auction-Conjoint Valuation Method: Valuing Pork and Eggs Produced under Differing Animal Welfare Conditions.” *Journal of Environmental Economics and Management* 62(2011):80–94.
- Samuelson, P. A. “Diagrammatic Exposition of a Theory of Public Expenditure.” *Review of Economics and Statistics* 37(1955):350–356.
- Stevens, T. H., M. Tabatabaei, and D. Lass. “Oaths and Hypothetical Bias.” *Journal of Environmental Management* 127(2013):135–141.
- Vossler, C. A., M. Doyon, and D. Rondeau. “Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments.” *American Economic Journal: Microeconomics* 4(2012):145–171.
- Vossler, C. A., and M. F. Evans. “Bridging the Gap between the Field and the Lab: Environmental Goods, Policy Maker Input, and Consequentiality.” *Journal of Environmental Economics and Management* 58(2009):338–345.
- Vossler, C. A., and J. Kerkvliet. “A Criterion Validity Test of the Contingent Valuation Method: Comparing Hypothetical and Actual Voting Behavior for a Public Referendum.” *Journal of Environmental Economics and Management* 45(2003):631–649.
- Vossler, C. A., and S. B. Watson. “Understanding the Consequences of Consequentiality: Testing the Validity of Stated Preferences in the Field.” *Journal of Economic Behavior & Organization* 86(2013):137–147.
- Whitehead, J. C., and T. L. Cherry. “Willingness to Pay for a Green Energy program: A Comparison of Ex-Ante and Ex-Post Hypothetical Bias Mitigation Approaches.” *Resource and Energy Economics* 29(2007):247–261.