Contingent Valuation, Hypothetical Bias, and Experimental Economics

James J. Murphy and Thomas H. Stevens

Although the contingent valuation method has been widely used to value a diverse array of non-market environmental and natural resource commodities, recent empirical evidence suggests it may not accurately estimate real economic values. The hypothetical nature of environmental valuation surveys typically results in responses that are significantly greater than actual payments. Economists have had mixed success in developing techniques designed to control for this "hypothetical bias." This paper highlights the role of experimental economics in addressing hypothetical bias, and identifies a gap in the existing literature by focusing on the underlying causes of this bias. Most of the calibration techniques used today lack a theoretical justification, and therefore these procedures need to be used with caution. We argue that future experimental research should investigate the reasons hypothetical bias persists. A better understanding of the causes should enhance the effectiveness of calibration techniques.

Key Words: contingent valuation, experiments, hypothetical bias, stated preference

Consider the challenge faced by a contingent valuation (CV) practitioner who is interested in estimating the economic value of a non-market good, such as visibility at a National Park or the protection of habitat for an endangered species. The CV survey is carefully designed and constructed (e.g., Mitchell and Carson, 1989; Champ, Brown, and Boyle, 2004) and the results are produced with the latest estimation techniques (Haab and McConnell, 2003). We now have an estimate for the economic value of the good—but is this value accurate?

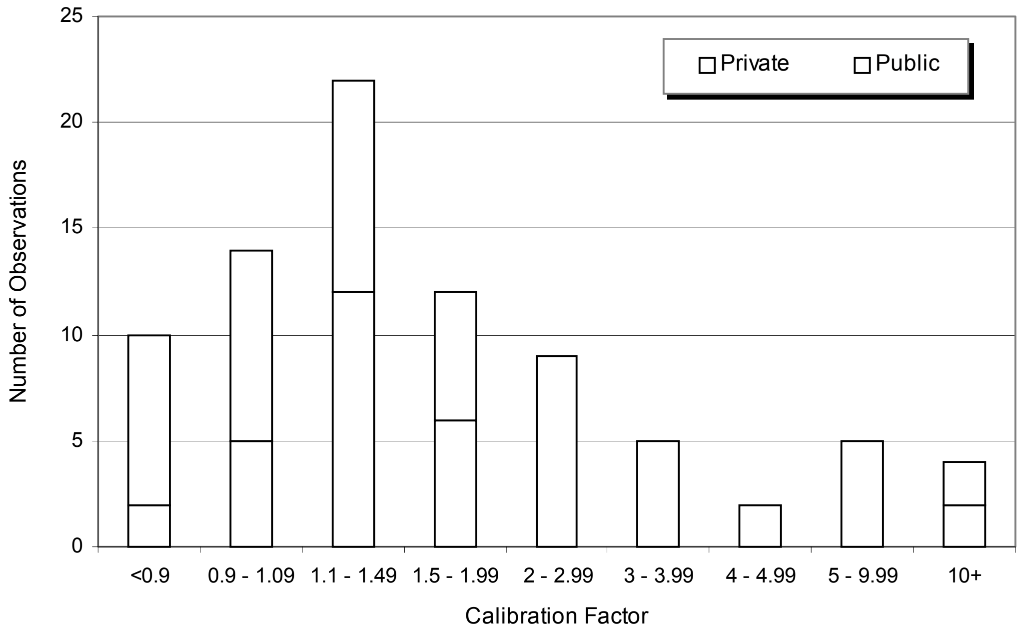
The answer to this question has stirred considerable, and sometimes contentious debate, as highlighted by litigation resulting from the 1989 Exxon Valdez oil spill in Prince William Sound [see Diamond and Hausman (1994); Hanemann (1994); and Portney (1994) for a synthesis of the debate]. Using only field CV data, we cannot be certain that value estimates are accurate. Why? Since CV surveys are hypothetical in both the payment for and provision of the good in question, we do not know whether what an individual says she *would* do in a hypothetical setting matches what she *will* do when actually given the opportunity to do so.[1] And, without the ability to observe the latter, it is difficult to confirm whether the values elicited from a hypothetical survey accurately reflect the real economic value of the good. Some researchers have expressed concern that this lack of a consequential economic commitment in CV surveys often leads to hypothetical bias in which economic values are overstated. For example, as Harrison and Rutström (forthcoming) assert: "As a matter of logic, if you do not have to pay for the good, but a higher *verbal* willingness-to-pay response increases the chance of its provision, then verbalize away to increase your expected utility!"

Economics experiments offer the potential to shed some light on the accuracy of responses to hypothetical CV questions. Experimental research has a well-established framework which was widely

James J. Murphy is assistant professor, Department of Resource Economics and Center for Public Policy and Administration, and Thomas H. Stevens is professor, Department of Resource Economics, both at the University of Massachusetts-Amherst.

This article is based on an invited paper presented at the annual meetings of the Northeastern Agricultural and Resource Economics Association, held jointly with the Canadian Agricultural Economics Society, Halifax, Nova Scotia, June 20–23, 2004.

[1] The terms "revealed," "real," and "actual" values are used interchangeably, and refer to situations in which an individual makes a consequential economic commitment—in experimental studies, this typically involves payment for a good by the participant. Most studies of hypothetical bias assume these cash-based estimates are unbiased. On the other hand, stated or hypothetical values refer to survey responses that lack any salient economic commitment.



Source: Murphy et al., 2003

Figure 1. Distribution of calibration factors by type of good (where calibration factor = hypothetical value/actual value)

recognized when Vernon Smith became a co-recipient of the Nobel Memorial Prize in Economics "for having established laboratory experiments as a tool in empirical economic analysis." What distinguishes experiments from other empirical techniques are control and replication. The ability to control the environment under which individuals make economic decisions is what gives experiments power. The experimenter can vary treatments to test hypotheses about the effects of different explanatory variables on individual choices. Unlike a typical field CV survey, a carefully designed experiment can include both hypothetical and real payment scenarios. By comparing outcomes in these two settings, one can make some inferences about the existence of hypothetical bias, its causes, and ways to mitigate its effects. Moreover, other researchers can replicate, and perhaps extend, the experiment to test its robustness. Generally, it is the body of experimental evidence, rather than a single study, that allows us to draw more reliable conclusions about what we do and do not know (Roth, 1988).

The existence of hypothetical bias has been well-documented in both laboratory and field settings. In a recent survey of the literature, Harrison and Rutström (forthcoming) found a positive bias in 32 of 39 observations. There are two meta-analyses of the experimental hypothetical bias literature: List and Gallet (2001) and Murphy et al. (2003). The latter study updates the List and Gallet data for some coding differences and conducts a sensitivity analysis of their results. The results of both meta-analyses are consistent with findings reported by Harrison and Rutström, suggesting that mean hypothetical values are about 2.5 to 3 times greater than actual values (but this comes from a highly skewed distribution with a median ratio closer to 1.5). Figure 1 presents the distribution of this bias for 83 observations from 23 studies (Murphy et al., 2003).

Although the presence of this bias is well documented, its underlying causes are not fully understood. Consequently, this paper highlights the need for a better understanding of the causes of this bias, and argues that future experimental research should focus on this issue.

Contingent Valuation Experiments

Running Experiments to Establish Empirical Regularities

Smith (1994) lists a number of reasons why economists run experiments. The evolution of the experimental CV literature clearly encompasses a subset of these. A fundamental reason for running experiments is to establish empirical regularities that guide the development of theory. As noted by Smith, many advances in science occur as a result of well-documented regularities and anomalies that motivated further research. Ultimately, establishing these stylized facts in the lab can help focus the efforts of theorists.

Early experimental CV research focused on establishing the empirical regularity that hypothetical bias exists, and tested the overall validity of contingent valuation. Bohm's seminal paper comparing hypothetical and actual values was published in 1972, but it was not until about a decade later when this literature began to grow. In a series of papers, Bishop and Heberlein found that hypothetical values for hunting permits consistently exceeded actual values (Bishop and Heberlein, 1979, 1986; Heberlein and Bishop, 1986).[2] In contrast, Dickie, Fisher, and Gerking (1987) found that values for pints of strawberries elicited in a hypothetical survey were consistent with those observed when individuals were given an opportunity to actually purchase the good.[3] Subsequent research, however, has consistently suggested that values derived from surveys typically exceed actual values (e.g., Cummings, Harrison, and Rutström, 1995; Fox et al., 1998; List and Shogren, 1998), sometimes by a substantial margin (e.g., Neill et al., 1994). There are exceptions to the conclusion about the existence of hypothetical bias (e.g., Johannesson, 1997; Sinden, 1988; Smith and Mansfield, 1998), but these are in the minority.

Common Elements of Valuation Experiments and Their Effects on Hypothetical Bias

Valuation experiments have a number of common elements. A group of subjects is recruited for the experiment. While these subjects are often university students, a substantial number of valuation experiments are conducted in the field with non-students [e.g., Champ et al. (1997) surveyed Wisconsin residents about their willingness to pay for road removal in the Grand Canyon; and List (2001) conducted his experiments using people who attended a sportscard show]. Cummings, Harrison, and Rutström (1995) recruited subjects from both undergraduate business classes and from various church groups; they observed similar patterns of hypothetical bias in both groups. The meta-analysis results reported by Murphy et al. (2003) provide some evidence that, although hypothetical bias exists with both students and non-students, the bias may be greater with student subjects. However, because most student experiments are conducted in group settings, Murphy et al. note it is unclear whether the effect is due to the subject pool or the experiment setting.

Subjects participating in a CV experiment are asked about their value for a specified good. Most experiments focus on an individual's willingness to pay (WTP), although a few studies address willingness to accept (e.g., Bishop and Heberlein, 1979; Coursey, Hovis, and Schulze, 1987; Smith and Mansfield, 1998; List and Shogren, 2002). An important distinction among experiments is the type of good valued—some studies use private goods, others public goods. Ultimately, the goal of CV is to value public goods. However, using private goods may be convenient because respondents are often familiar with the good and its substitutes, and may have considered or engaged in a market transaction for the good at some point. Even if the good is unfamiliar, subjects may be more comfortable with valuing private goods, whereas the subject may have never considered placing an economic value on a public good like atmospheric quality. If subjects are more comfortable valuing goods they commonly purchase, then they may be less prone to error (List and Gallet, 2001). Private goods also avoid any biases due to free-riding. Another argument in support of using private goods is that if CV cannot accurately estimate economic value in these relatively familiar settings, it is probably unlikely to do so with public goods. As documented by List and Gallet (2001), hypothetical bias exists with both types of goods, but the bias is lower with private goods. On the other hand, Murphy et al. (2003) found the evidence mixed regarding any differences in bias for public and private goods. Later, we will discuss some potentially important distinctions between these two types of goods which future research may need to address.

Nearly all valuation experiments attempt to elicit what are known as "homegrown" values.[4] This feature distinguishes valuation experiments from the prototypical economics experiment. An experimenter can attempt to control preferences, or can try to

[2] Hanemann (1984) highlights the sensitivity of this conclusion.

[3] Harrison and Rutström (forthcoming) argue that a more detailed examination of their data yields mixed results and, on average, hypothetical values exceed actual values by 58%.

[4] Some notable exceptions include Taylor (1998); Burton et al. (2003); and Polomme (2003).

measure them. In many experiments, such as most market or public goods games, it is important to be able to state that individual values vary in particular ways. Smith (1976) outlines the conditions under which the use of salient rewards based on experiment performance allows the experimenter to exert control over individual preferences. Induced value theory basically states that under the right conditions, the reward medium of the experiment—typically cash—dominates all other considerations, and individuals are motivated to maximize their experimental earnings. In a simple market experiment, for example, sellers may be given a reservation price that represents the cost of production. These subjects are told they will be paid, in cash, the difference between the price at which they trade and this reservation price. Since higher earnings in the lab translate into more cash outside the lab, sellers have the incentive to maximize their experimental earnings by seeking the highest possible trading price. Similarly, buyers can be given reservation prices that represent redemption values, which create incentives to maximize consumer surplus. Because the experimenter assigned the reservation prices, he or she can construct the market supply function and market demand function and compare these with observed outcomes to test hypotheses. The experiments presented by Anderson (2004) are an application of induced-value theory.

Most valuation experiments, on the other hand, do not use induced values. Rather than attempting to control preferences using induced values, the researcher tries to measure subjective, homegrown values [Harrison, Harstad, and Rutström (2004) discuss some methodological issues that need to be considered when eliciting subjective values]. As with an actual field CV study, these homegrown values cannot be known with certainty. However, by carefully manipulating the conditions under which values are elicited, the experimenter can test whether changes in explanatory variables influence responses. Most hypothetical bias experiments include both a hypothetical CV survey and a real payment scenario. For example, some subjects in the Champ et al. (1997) Grand Canyon road removal study were asked a hypothetical contingent donation question. Other subjects participated in a similar survey in which they were asked for actual contributions for road removal. Both surveys made clear that the project was the only source of funding for the program. The estimated mean WTP in the hypothetical treatment ($46–$89) was significantly greater than mean actual contributions ($9).

Because the experimenter is usually eliciting unknown homegrown values, it is important to emphasize the inferences which can be drawn from the results. If hypothetical values exceed actual values, then the data clearly support the argument that these values differ. However, without knowing the true economic value of the good, we make the reasonable *assumption* that the responses in the real settings represent the true economic value, and therefore the hypothetical values must be overstated. Based solely on what can be inferred from the data, it is entirely possible the converse is true: i.e., the hypothetical values are accurate and the real values misstated (Harrison, 2002). This point is revisited later.

Running Experiments to Develop and Test New Calibration Techniques

Experiments can also provide a testing ground for new institutions (Smith, 1994) or, in the case of CV, new techniques to calibrate for hypothetical bias. Because experiments offer a high degree of control, economists can test the properties of new institutions or new calibration techniques in a relatively low-cost setting before they are actually implemented. In the case of CV calibration, if the techniques fail to adequately account for the bias in a carefully controlled experiment, then it is unlikely these instruments will be effective in more complex field settings.

After the Exxon Valdez spill, the National Oceanic and Atmospheric Administration (NOAA) convened a panel of experts led by Nobel laureates Kenneth Arrow and Robert Solow to review CV. The resulting report (Arrow et al., 1993) influenced the design of CV studies and the evolution of the experimental literature. The NOAA panel noted the "unfortunate" lack of data that could be used to calibrate CV responses. This inevitably helped spawn a growing literature in the development and testing of calibration techniques. These techniques rest on the assumption that, although responses to hypothetical CV questions may be biased, these responses provide useful information about true economic values.

In response to the NOAA report, Blackburn, Harrison, and Rutström (1994) developed a simple statistical bias function using socioeconomic characteristics to estimate the extent to which people overstate their value.[5] Taking experimental data from

[5] The idea of estimating a bias function for public goods was first proposed by Kurz (1974).

Cummings, Harrison, and Rutström (1995), they used the bias function estimated using the data from one experiment to calibrate responses in another experiment. These calibrated hypothetical responses are statistically indistinguishable from those in the real treatment. Mansfield (1998) expands on this concept by estimating the influence of these individual characteristics on the extent of bias, as opposed to the influence on individual preferences. More recently, Hofler and List (2004) adopted a stochastic frontier regression model to successfully calibrate hypothetical responses to those from an actual auction for a private good (a baseball card with a \$350 market value). These statistical bias functions are potentially valuable because they calibrate for hypothetical bias using data normally collected in a CV survey. However, apart from recognizing that individual characteristics might play an important role in an individual's valuation decision, there is no reason a priori to explain why these particular characteristics matter, or, if they do, the appropriate econometric model specification. Using laboratory experiments to calibrate field data, Fox et al. (1998) note that these calibration functions may be commodity- and context-specific. Moreover, results reported by both Mansfield (1998) and List and Shogren (2002) indicate calibration functions may also be individual-specific. Thus, while the approach developed in these papers is promising, calibration functions may be unique to a particular situation, and the absence of a theoretical motivation makes it difficult to generalize the results.

Another ex post technique for mitigating hypothetical bias is the use of an uncertainty adjustment. The NOAA panel recommended that values be elicited using a referendum format that included "no answer" or "don't know" as an explicit option. A series of papers focused on the implications of a "don't know" response, and more broadly on the notion of respondent uncertainty—i.e., respondents may be unsure of their actual value for the good (Opaluch and Segerson, 1989; Gregory et al., 1995; Li and Mattsson, 1995; Gregory and Slovic, 1997; Wang, 1997). This uncertainty might occur because subjects are unfamiliar with the good described or how to value it, or perhaps they have not put explicit thought into how much they value the good. Moreover, if an individual's preferences for the good are formed during the survey, then the one-shot nature of CV surveys may not provide an individual with the experience necessary for these values to stabilize (Loomis and Ekstrand, 1998).

A number of approaches have emerged for incorporating respondent uncertainty. These typically involve a post-decision question asking the individual how certain she is of her response to the CV question. The format of this follow-up question can be on a qualitative scale, such as "fairly sure" or "absolutely sure" (as in Blumenschein et al., 1998; and Johannesson, Liljas, and Johansson, 1998), or a numeric certainty scale, such as a 10-point Likert scale in which 1 = very uncertain and 10 = very certain (e.g., Champ et al., 1997; Champ and Bishop, 2001; Ethier et al., 2000; Poe et al., 2002). For responses to a hypothetical question in which the respondent reports a high degree of uncertainty, these studies generally find that recoding uncertain "yes" responses to "no" yields estimates of WTP which are a good approximation of actual payments.[6] This recoding scheme has effectively calibrated hypothetical responses in a number of studies, but the cutoff point at which this technique works varies. For example, Champ et al. (1997) concluded that 10 was an appropriate cutoff, Champ and Bishop (2001) used a cutoff of 8, and Poe et al. (2002) chose 7. In contrast, Wang (1997) found that treating uncertain responses as "no" clearly underestimates mean WTP.

This approach assumes those who are uncertain about their "yes" response in the hypothetical setting are likely to respond "no" when confronted with a real payment situation. While this assumption is quite reasonable, it is also somewhat arbitrary, because little is known about the underlying source of uncertainty or hypothetical bias. Although a growing body of experimental evidence is establishing the stylized fact that this approach may be effective, it is not entirely clear why this interpretation of uncertainty is preferable to alternative interpretations and coding schemes. After all, if the bias is caused by too many people responding "yes," then by construction, *any* procedure that reduces the number of "yes" responses while holding the "no" responses constant must reduce the difference between "yes" and "no" responses.

Recent results regarding another calibration technique, cheap talk (first introduced by Cummings and Taylor, 1999), illustrate the importance of not only establishing that a calibration technique works, but also developing an understanding of *why* it works. Cheap talk is an ex ante calibration technique in which the researcher attempts to elicit unbiased

[6] However, Johannesson, Liljas, and Johansson (1998) found this approach resulted in a significant understatement of economic value.

responses by reading a script which draws respondents' attention to the hypothetical bias problem. As observed by Cummings and Taylor (1999, p. 663), "the cheap talk design was successful in eliciting responses that were indistinguishable from responses to valuation questions for actual payments." This approach is appealing because, like uncertainty adjustments, it is intuitive and easily implemented. According to the premise behind this technique, simply making respondents aware of hypothetical bias, regardless of its underlying causes, is sufficient to eliminate it.

The evidence about cheap talk's robustness is mixed, however. In an unpublished manuscript, Cummings, Harrison, and Osborne (1995) found that a shortened version of the script was not effective, but a lengthier script similar to the one used by Cummings and Taylor (1999) was successful. Similarly, Poe et al. (2002) concluded a short script did not influence decisions. Conversely, Aadland and Caplan (2003) reported their use of a short cheap talk script was effective. Based on findings by List (2001), the long script did not reduce hypothetical bias with experienced card dealers, but was effective with inexperienced participants; both Lusk (2003) and Aadland and Caplan (2003) report similar results. Brown, Ajzen, and Hrubes (2003) found that the long cheap talk script was successful, but only for high payment amounts.

Our point is not that the calibration techniques such as uncertainty adjustments or cheap talk are inherently flawed or inappropriate. On the contrary, there is a growing body of experimental evidence suggesting these techniques do have an effect on hypothetical WTP responses. Nevertheless, as the cheap talk results indicate, these techniques may not work in all situations. Clearly, without a better understanding of the causes of hypothetical bias and why these techniques are effective, they should be used with caution. The next section provides a discussion of some approaches that have been used to interpret CV responses.

Toward a Better Understanding of the Causes of Hypothetical Bias

Interestingly, calibration techniques such as statistical bias functions and uncertainty adjustments have been quite successful at mitigating hypothetical bias without a formal theory to explain the fundamental causes of the bias. This lack of theoretical support may ultimately limit the generality of these methods. A number of plausible reasons have been offered to explain why hypothetical bias exists, but many of these hypotheses have not yet been rigorously tested in an experimental setting. This is a surprising contrast to the experimental literature in other contexts, such as public goods or bargaining experiments. Once the experimental evidence clearly established stylized facts not predicted by neoclassical theory (e.g., people do contribute to the public good, and the modal offer in an ultimatum game is an even split of the pie), subsequent research began to explore possible explanations for these results. Conversely, the experimental valuation literature seems to have focused more on developing better CV instruments and calibration techniques without directly addressing the underlying sources of the bias. This section discusses some of the possible explanations for hypothetical bias that have been posited and explores the evidence (experimental and nonexperimental) about these hypotheses.

We begin with the simplest possible explanation for hypothetical bias—the lack of any consequences associated with an individual's response. While this may explain why an individual might *misstate* her value, it does not account for why she systematically *overstates* her value. The results of some induced-value experiments conducted by Taylor et al. (2001) using an incentive-compatible referendum support this rather straightforward observation. They observed similar rates of misvotes (about 16%) in both the real and hypothetical referenda. Clearly, the inconsequential nature of hypothetical CV questions must be interacting with other factor(s).

Perhaps one of the most intuitively appealing explanations for hypothetical bias is that if the respondent has a positive value for the good, and if her response to the valuation question may increase the likelihood of the good's provision at little or no cost to her, then it makes sense for her to report an inflated value (Bohm, 1984; Harrison and Rutström, forthcoming). Along these lines, it is possible that individuals are merely expressing a positive attitude for the good without necessarily agreeing to contribute toward its provision (Champ and Bishop, 2001). While this theory may help explain the bias in some public good settings, it fails to explain why the bias is also prevalent in experiments with private goods. Roughly half the experimental studies include private goods, and hypothetical values consistently exceed actual values even though the outcome of the hypothetical survey has no bearing on the provision of the good (Murphy, et al. 2003). As shown by figure 1, private goods, in fact, have a larger share of calibration factors which exceed 1.5 (54% vs. 36%).[7] Clearly, hypothetical bias is not unique to public goods, and the underlying causes of the bias must be more complex than simply an attempt to increase the likelihood of a good's provision at no cost to the respondent.

The notion of respondent uncertainty as a potential source of bias has garnered much attention recently. We have already discussed how this information has been used to calibrate responses. Now we consider some of the justifications that have been proposed. Much of the research testing hypotheses about interpreting respondent uncertainty has been conducted in a nonexperimental setting (e.g., Opaluch and Segerson, 1989; Ready, Whitehead, and Blomquist, 1995; Li and Mattsson, 1995; Wang, 1997; Loomis and Ekstrand, 1998), highlighting the need for more direct tests using experiments.

Opaluch and Segerson (1989) present one of the earliest attempts to explain inconsistencies in valuation responses using ambivalence theory. They posit that individuals become ambivalent when presented with difficult tradeoffs—e.g., between money and environmental amenities—because they may be unsure where the indifference curves lie. Essentially, the value may not be a constant known to the respondent, but rather a random variable. So, although an individual may not know her value precisely, she may be able to place it within a range. The implications of this scenario are perhaps best illustrated with an example. Consider a respondent who estimates her willingness to pay for a non-market good is "about $25 or so." If asked about her willingness to pay a rather large sum (e.g., $500), this respondent may be able to unambiguously state, "no," she is absolutely certain she would not pay this much under any situation. Similarly, she may express a high degree of confidence about her willingness to pay extremely small amounts (say $1). Uncertainty sets in as the amount asked approaches $25.

Thus, ambivalence theory would suggest higher levels of uncertainty correspond with amounts that are approaching the respondent's actual value, whereas highly certain responses reflect values that deviate noticeably from the true value. If ambivalence theory is correct, then the recoding of uncertain "yes" responses to "no" may not be fully capitalizing on the information provided by these uncertain responses. The data from Loomis and Ekstrand (1998) are consistent with ambivalence. Figure 2 is adapted from their study, and illustrates the quadratic relationship between amount asked and certainty. Respondents are relatively certain for both low and high amounts asked, but for intermediate values, certainty decreases. Contrasting results were found by Champ and Bishop (2001), who report that uncertainty is invariant to the amount asked.

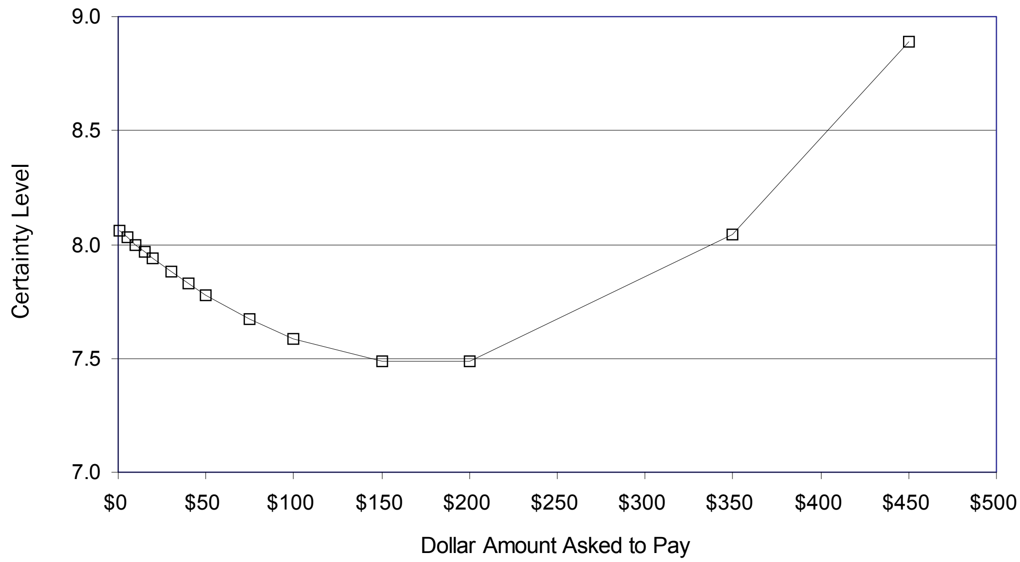
Ambivalence theory alone offers no guidance as to how an individual will respond to an amount that lies within her ambivalence region in either a hypothetical or a real setting. If the individual is unable to resolve the uncertainty through introspection and is forced to make difficult choices, it does not necessarily follow that responses will be biased upwards. In fact, there are competing theories regarding how an individual will respond. A model of respondent behavior was developed by Hoehn and Randall (1987) in which there is uncertainty about the good described and the subject has insufficient time to determine a value. Their model suggests the CV respondent would adopt a conservative approach and reject deviations from the status quo.[8] On the other hand, Crocker and Shogren (1991) present a theoretical model which predicts respondents will overstate WTP when preferences are uncertain. A third possibility is that the individual follows a lexicographic rule which gives dominance to one of the alternatives (Ready, Whitehead, and Blomquist, 1995).

Using the data from two field CV studies, Ready, Whitehead, and Blomquist (1995) found a tendency by those individuals asked a dichotomous choice question to express a lower WTP than those who are given an opportunity to express uncertainty through a polychotomous choice instrument.[9] They conclude that the dichotomous choice respondents demonstrate some degree of conservatism (bias toward the status quo, i.e., a "no" response), although they acknowledge the results are also consistent with lexicographic preferences which give dominance to money over the environmental amenity. If uncertain individuals systematically adopt a conservative strategy (as proposed by Hoehn and Randall, 1987), then this might lend some support to the recoding of uncertain "yes" responses to "no."

[7] However, it is worth noting that most of the very large calibration factors (> 5) are from public goods experiments.

[8] Samuelson and Zeckhauser (1988) discuss a status quo bias in consequential settings with no uncertainty.

[9] Rather than only allowing a simple yes/no response, the polychotomous choice questionnaire allowed for six possible responses: "definitely yes," "probably yes," "maybe yes," "maybe no," "probably no," and "definitely no."



Source: Loomis and Ekstrand, 1998

Figure 2. Relationship between amount asked and certainty level

However, Hoehn and Randall focus on uncertainty about the good, not uncertain preferences, and whether uncertainty about the latter also implies a conservative approach needs to be established.

More importantly, preference uncertainty is not necessarily unique to hypothetical scenarios, and therefore ought also to be present in real payment settings. While consequential decisions might induce an individual to engage in more intense introspection which reduces the ambivalence region, this does not necessarily imply that responses would differ between hypothetical and real situations. In fact, a greater tendency toward "no" responses in hypothetical settings would suggest actual values should *exceed* hypothetical ones—which brings us back to the original question of why individual behavior in a hypothetical setting deviates from that in a real payment setting in a consistent, systematic way.

Loomis and Ekstrand (1998) note that the recoding of uncertain "yes" responses to "no" may overlook the potential for fuller uses of the uncertainty data, and they explore the implications of different interpretations of respondent uncertainty. They observed that roughly one-third of the "no" respondents were uncertain (< 8) about their response. Similarly, more than half of the "no" respondents in Champ and Bishop (2001) indicated they were uncertain about their hypothetical decision. These findings beget the question of how to interpret the uncertain "no" responses.[10] Using field CV data which include a certainty question, Loomis and Ekstrand (1998) estimate six different models. Their standard dichotomous choice model does not incorporate the uncertainty response. Similar to Champ and Bishop (2001), models YES10, YES910, and YES810 convert "yes" responses to "no" if the level of certainty is < 10, < 9, and < 8, respectively. Following Li and Mattsson (1995), their last two models interpret the numerical certainty categories as the probability of actually paying.[11] Their asymmetric uncertainty model (ASUM) replaces the "yes" responses with a probability but keeps all "no" responses unchanged. The symmetric uncertainty model (SUM) converts both "yes" and "no" responses to probabilities.

Because Loomis and Ekstrand (1998) did not include a treatment in which actual payments were collected, we cannot tell which model most accurately predicts the true WTP. Presumably the standard dichotomous model suffers from hypothetical bias. Since the WTP estimates from the two probabilistic

[10] It should be noted that experimental evidence suggests people who respond "no" in a hypothetical setting rarely change their minds when confronted with real payment opportunities. For example, neither Blackburn, Harrison, and Rutström (1994) nor Johannesson et al. (1999) observed any subjects doing so.

[11] These uncertainty adjustments assume (*a*) although an individual's valuation response may be biased, her certainty response is not; and (*b*) all respondents interpret the uncertainty scale similarly (Loomis and Ekstrand, 1998).

models (SUM and ASUM) are not statistically different from the uncalibrated standard dichotomous choice estimate, it appears these models failed to adequately correct for the bias, lending some empirical support to the recoding of uncertain "yes" responses to "no." However, Wang (1997) found the opposite—that recoding uncertain responses to "no" yields an underestimate of economic value. Furthermore, theory provides no guidance about the relative merits of two equally plausible and intuitive interpretations: converting uncertain "yes" to "no" (e.g., Champ and Bishop, 2001; Poe et al., 2002) or interpreting these responses as probabilities (e.g., Li and Mattsson, 1995; Wang, 1997; Loomis and Ekstrand, 1998). As acknowledged by Champ and Bishop (2001), this sensitivity of the calibration results to the interpretation of the uncertainty response highlights the need for a better understanding of the causes of hypothetical bias and respondent uncertainty.

A Comment on Private Goods Experiments

Earlier, we noted that hypothetical bias experiments have been conducted using both public and private goods. The basic rationale for using private goods is that if CV cannot produce accurate estimates in this more familiar setting, then it is less likely to do so with public goods. The results from experiments with each type of good are consistent: hypothetical values tend to exceed actual payments. Since the direction of the bias is the same across good types, it may seem reasonable to pool the data and use inferences drawn from private goods experiments to shed light on the hypothetical bias problem with public goods. However, we are becoming increasingly convinced that the sources of the difference between hypothetical and actual values for public and private goods may be due to entirely different sets of factors. Most of the aforementioned hypotheses about causes of hypothetical bias discussed in the previous section make sense primarily in the context of public goods. For example, a higher hypothetical WTP for a private good does not increase the likelihood of the good's provision, and it is unlikely that people are expressing a positive sentiment for an electric juicer or baseball card.

With homegrown values, we can never be sure whether the hypothetical or the actual response reflects the true value. With public goods, some of the factors discussed in the previous section may potentially lead to overstated hypothetical values. However, with private goods, we hypothesize that perhaps it is the hypothetical values which are truthfully revealed, but responses to real payment questions are biased because the actual payments are bounded by the market price for the good (net of transaction costs). A subject is unlikely to agree to actually purchase (or sell) a good in an experiment at a price exceeding what she could pay for it outside the lab. This would suggest that values elicited in the lab may be censored by the prevailing market price (Harrison, Harstad, and Rutström, 2004). Without careful controls in the experimental design, it is possible that an individual truthfully reports her maximum WTP in a hypothetical setting but censors actual economic commitments at a perceived market price. Future experimental research may want to focus on testing this conjecture specifically, and more generally on the applicability of the results from private goods experiments to public good valuation techniques.

Concluding Remarks

Let's take stock of what we do and do not know. It has been well-established that hypothetical values exceed actual values, and calibration techniques have had some success in aligning these values. Less is known about the underlying causes of these differences in values. The inconsequential nature of the CV survey cannot alone explain these differences; it might explain greater variability in hypothetical responses, but not higher mean and median responses. Therefore, there must be an interaction of the hypothetical decision with other factor(s). Expressing positive sentiment toward the good, or an attempt to increase the likelihood of the good's provision, is certainly possible in a public goods setting, but this fails to explain why hypothetical bias also exists with private goods since these factors are not applicable. There is some evidence that respondents may be uncertain, but the causes of this uncertainty and its implications for valuation are not well understood. Ultimately, hypothetical bias is likely to be individual-specific and a composite of a variety of factors. Experiments can play a key role in developing a better understanding of what causes hypothetical bias and can assist researchers in designing CV instruments which incorporate these considerations.

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.

Anderson, C. M. (2004, October). "How Institutions Affect Outcomes in Laboratory Tradable Fishing Allowance Systems." *Agricultural and Resource Economics Review* 33(2), 193–208 [this journal issue].

Arrow, K., R. Solow, P. Portney, E. E. Leamer, R. Radner, and H. Schuman. (1993). "Report of the NOAA Panel on Contingent Valuation." *Federal Register* 58(10), 4602–4614.

Bishop, R. C., and T. A. Heberlein. (1979). "Measuring Values of Extramarket Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics* 61(5), 926–930.

———. (1986). "Does Contingent Valuation Work?" In R. Cummings, D. Brookshire, and W. Schulze (eds.), *Valuing Environmental Goods: A State of the Art Assessment of the Contingent Valuation Method* (pp. 123–147). Totowa, NJ: Rowman and Allenheld.

Blackburn, M., G. W. Harrison, and E. E. Rutström. (1994). "Statistical Bias Functions and Informative Hypothetical Surveys." *American Journal of Agricultural Economics* 76, 1084–1088.

Blumenschein, K., M. Johannesson, G. C. Blomquist, B. Liljas, and R. M. O'Connor. (1998). "Experimental Results on Expressed Certainty and Hypothetical Bias in Contingent Valuation." *Southern Economic Journal* 65(1), 169–177.

Bohm, P. (1972). "Estimating the Demand for Public Goods: An Experiment." *European Economic Review* 3, 111–130.

———. (1984). "Revealing Demand for an Actual Public Good." *Journal of Public Economics* 24(2), 135–151.

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, 353–361.

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

Champ, P. A., 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., T. C. Brown, and K. J. Boyle. (2004). *A Primer on Nonmarket Valuation.* Dordrecht: Kluwer Academic Publishers.

Coursey, D. L., J. L. Hovis, and W. D. Schulze. (1987). "The Disparity Between Willingness-to-Accept and Willingness-to-Pay Measures of Value." *Quarterly Journal of Economics* 102(3), 679–690.

Crocker, T. D., and J. F. Shogren. (1991). "Preference Learning and Contingent Valuation Methods." In F. Dietz, F. V. d. Ploeg, and J. V. d. Straaten (eds.), *Environmental Policy and the Economy* (pp. 85–107). Amsterdam: North-Holland.

Cummings, R. G., G. W. Harrison, and L. L. Osborne. (1995). "Can the Bias of Contingent Valuation Surveys Be Reduced?" Economics Working Paper No. B-95-03, Division of Research, College of Business Administration, University of South Carolina.

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

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.

Diamond, P. A., and J. A. Hausman. (1994). "Contingent Valuation: Is Some Number Better Than No Number?" *Journal of Economic Perspectives* 8(4), 45–64.

Dickie, M., A. Fisher, and S. Gerking. (1987). "Market Transactions and Hypothetical Demand Data: A Comparative Study." *Journal of the American Statistical Association* 82, 69–75.

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.

Fox, J. A., J. F. Shogren, D. J. Hayes, and J. B. Kliebenstein. (1998). "CVM-X: Calibrating Contingent Values with Experimental Auction Markets." *American Journal of Agricultural Economics* 80(3), 455–465.

Gregory, R., S. Lichenstein, T. C. Brown, G. L. Peterson, and P. Slovic. (1995). "How Precise Are Monetary Representations of Environmental Improvements?" *Land Economics* 71(4), 462–473.

Gregory, R., and P. Slovic. (1997). "A Constructive Approach to Environmental Valuation." *Ecological Economics* 21(3), 175–181.

Haab, T. C., and K. E. McConnell. (2003). *Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation.* Northampton, MA: Edward Elgar.

Hanemann, W. M. (1984). "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." *American Journal of Agricultural Economics* 66(3), 332–341.

———. (1994). "Valuing the Environment Through Contingent Valuation." *Journal of Economic Perspectives* 8(4), 19–43.

Harrison, G. W. (2002, June). "Experimental Economics and Contingent Valuation." Working paper, Department of Economics, University of Central Florida. Online. Available at http://www.bus.ucf.edu/gharrison/wp/Experimental%20Economics%20and%20Contingent%20Valuation.pdf. [Accessed August 18, 2004.]

Harrison, G. W., R. M. Harstad, and E. E. Rutström. (2004). "Experimental Methods and Elicitation of Values." *Experimental Economics* 7(2), 123–140.

Harrison, G. W., and E. E. Rutström. (Forthcoming). "Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods." In C. Plott and V. L. Smith (eds.), *Handbook of Results in Experimental Economics*. New York: Elsevier Science.

Heberlein, T. A., and R. Bishop. (1986). "Assessing the Validity of Contingent Valuations: Three Field Experiments." *Science of the Total Environment* 56, 434–479.

Hoehn, J. P., and A. Randall. (1987). "A Satisfactory Benefit-Cost Indicator from Contingent Valuation." *Journal of Environmental Economics and Management* 14(3), 226–247.

Hofler, R. A., and J. List. (2004). "Valuation on the Frontier: Calibrating Actual and Hypothetical Statements of Value." *American Journal of Agricultural Economics* 86(1), 213–221.

Johannesson, M. (1997). "Some Further Experimental Results on Hypothetical versus Real Willingness to Pay." *Applied Economics Letters* 4, 535–536.

Johannesson, M., G. C. Blomquist, K. Blumenschein, P. Johansson, and R. M. O'Connor. (1999). "Calibrating Hypothetical Willingness-to-Pay Responses." *Journal of Risk and Uncertainty* 8, 21–32.

Johannesson, M., B. Liljas, and P. Johansson. (1998). "An Experimental Comparison of Dichotomous Choice Contingent Valuation Questions and Real Purchase Decisions." *Applied Economics* 30, 643–647.

Kurz, M. (1974). "Experimental Approach to the Determination of the Demand for Public Goods." *Journal of Public Economics* 3, 329–348.

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." *American Economic Review* 91(5), 1498–1507.

List, J. A., and C. Gallet. (2001). "What Experimental Protocols Influence Disparities Between Actual and Hypothetical Stated Values?" *Environmental and Resource Economics* 20, 241–254.

List, J. A., and J. F. Shogren. (1998). "Calibration of the Difference Between Actual and Hypothetical Valuations in a Field Experiment." *Journal of Economic Behavior and Organization* 37(2), 193–205.

———. (2002). "Calibration of Willingness to Accept." *Journal of Environmental Economics and Management* 43(2), 219–233.

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

Lusk, J. L. (2003). "Willingness to Pay for Golden Rice." *American Journal of Agricultural Economics* 85(4), 840–856.

Mansfield, C. (1998). "A Consistent Method for Calibrating Contingent Valuation Survey Data." *Southern Economic Journal* 64(3), 665–681.

Mitchell, R. C., and R. T. Carson. (1989). *Using Surveys to Value Public Goods: The Contingent Valuation Method.* Washington, DC: Resources for the Future.

Murphy, J. J., T. H. Stevens, P. G. Allen, and D. Weatherhead. (2003). "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation." Working Paper No. 2003-8, Department of Resource Economics, University of Massachusetts, Amherst.

Neill, H. R., R. G. Cummings, P. T. Ganderton, G. W. Harrison, and T. McGuckin. (1994). "Hypothetical Surveys and Real Economic Commitments." *Land Economics* 70(2), 145–154.

Opaluch, J. J., and K. Segerson. (1989). "Rational Roots of Irrational Behavior, New Theories of Economic Decision-Making." *Northeastern Journal of Agricultural and Resource Economics* 18, 81–95.

Poe, G. L., J. E. Clark, D. Rondeau, and W. D. Schulze. (2002). "Provision Point Mechanisms and Field Validity Tests of Contingent Valuation." *Environmental and Resource Economics* 23, 105–131.

Polomme, P. (2003). "Experimental Evidence on Deliberate Misrepresentation in Referendum Contingent Valuation." *Journal of Economic Behavior and Organization* 52(3), 387–401.

Portney, P. (1994). "The Contingent Valuation Debate: Why Should Economists Care?" *Journal of Economic Perspectives* 8(4), 3–17.

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.

Roth, A. E. (1988). "Laboratory Experimentation in Economics: A Methodological Overview." *Economic Journal* 98, 974–1031.

Samuelson, W., and R. Zeckhauser. (1988). "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty* 1(1), 7–59.

Sinden, J. A. (1988). "Empirical Tests of Hypothetical Biases in Consumers' Surplus Surveys." *Australian Journal of Agricultural Economics* 32(2/3), 98–112.

Smith, V. K., and C. Mansfield. (1998). "Buying Time: Real and Hypothetical Offers." *Journal of Environmental Economics and Management* 36, 209–224.

Smith, V. L. (1976). "Experimental Economics: Induced Value Theory." *American Economic Review* 66(2), 274–279.

———. (1994). "Economics in the Laboratory." *Journal of Economic Perspectives* 8(1), 113–131.

Taylor, L. (1998). "Incentive Compatible Referenda and the Valuation of Environmental Goods." *Agricultural and Resource Economics Review*, 27(2), 132–139.

Taylor, L. O., M. McKee, S. K. Laury, and R. G. Cummings. (2001). "Induced-Value Tests of the Referendum Voting Mechanism." *Economics Letters* 71(1), 61–65.

Wang, H. H. (1997). "Treatment of 'Don't-Know' Responses in Contingent Valuation Surveys: A Random Valuation Model." *Journal of Environmental Economics and Management* 32(2), 219–232.