College of Business and Public Policy
UNIVERSITY of ALASKA ANCHORAGE

Department of Economics Working Paper
WP 2010-06
March 2010

A Comparison of Induced Value and Home-Grown Value Experiments to Test for Hypothetical Bias in Contingent Valuation

James J. Murphy
University of Alaska Anchorage

Thomas H. Stevens
University of Massachusetts Amherst

Lava Yadav
University of Massachusetts Amherst

UAA DEPARTMENT OF ECONOMICS
3211 Providence Drive
Rasmuson Hall 302
Anchorage, AK 99508

http://econpapers.uaa.alaska.edu

Revised March 2010

A Comparison of Induced Value and Home-Grown Value Experiments to Test for Hypothetical Bias in Contingent Valuation

JAMES J. MURPHY [a,b,*]

THOMAS H. STEVENS [b]

LAVA YADAV [b]

[a] Department of Economics, Rasmuson Hall, 3211 Providence Drive, University of Alaska Anchorage, Anchorage, AK 99508

[b] Department of Resource Economics, Stockbridge Hall, 80 Campus Center Way, University of Massachusetts-Amherst, Amherst, MA, 01003, USA

[*] Corresponding author.
Phone: 907-786-1936
Fax: 907-786-4115
Email: murphy@uaa.alaska.edu

This paper has not been submitted elsewhere in identical or similar form, nor will it be during the first three months after its submission to the Publisher.

A Comparison of Induced Value and Home-Grown Value Experiments to Test for Hypothetical Bias in Contingent Valuation

Abstract: This study tests the hypothesis that hypothetical bias may not be related to value elicitation; rather it may be a value formation problem. When participants are asked to indicate their willingness to pay for an induced value good, we find no evidence of hypothetical bias for three different commodity types (public good, private good, and publicly provided private good). However, when these same subjects are asked to value homegrown goods with no pre-assigned induced value using the same elicitation mechanism, hypothetical values are roughly double actual payments in all three cases. These results support the hypothesis that the process of forming values in a homegrown setting may be a key contributor to hypothetical bias.

Keywords: contingent valuation, hypothetical bias, experiments, induced values, home-grown values.

JEL Classification: C91, H41, Q26, Q28.

1. Introduction

Contingent valuation (CV) is one of the primary tools for estimating the value of non-market goods. The methodology can play a key role in policy decisions that rely on benefit-cost analysis, and as the *Exxon Valdez* case highlighted, CV can be instrumental in determining penalties for environmental damages. Despite CV's prominent role, its use as a decision making tool is often quite controversial. Because CV surveys are hypothetical in both the payment for and provision of the good, it is impossible to determine whether the choices made in a hypothetical survey are consistent with how the individual would behave if actually given an opportunity to do so. This inconsequential nature of CV surveys, i.e., the lack of a salient economic commitment, can lead to hypothetical bias in which CV overestimates the true economic value of the good.

Although the existence of hypothetical bias is well-documented (List and Gallet 2001, Murphy, *et al.* 2005a, Harrison and Rutström 2008), its underlying causes are not well understood (Murphy and Stevens 2004). Without a theory to help explain why hypothetical bias persists, the general applicability of calibration techniques designed to offset these effects may be limited. The cheap talk approach (Cummings and Taylor 1999), for example, is a popular calibration technique in which the hypothetical bias problem is described to subjects. Yet, the effectiveness of this approach may be sensitive to key variables such as script length (Poe, *et al.* 2002; Aadland and Caplan 2003), subject experience (List 2001; Lusk 2003; Aadland and Caplan 2003) and payment amounts (Brown, *et al.* 2003; Murphy, *et al*. 2005b). In fact, Aadland and Caplan 2006 find that a neutral cheap talk script can actually exacerbate hypothetical bias.[1]

[1] See Murphy and Stevens 2004, and Horowitz *et al*. forthcoming, for a more comprehensive discussion of this literature.

The vast majority of contingent valuation hypothetical bias experiments elicit subjective, homegrown values, rather than experimenter-controlled, pre-assigned induced values. Unlike induced values, the researcher cannot know these homegrown values with certainty. Because of this, when researchers observe that values in a hypothetical payment scenario are higher than the corresponding treatment with actual, consequential payments, it is impossible to know for sure whether hypothetical values are overstated or actual values are understated (or possibly a combination of the two). Researchers typically make the reasonable assumption that the responses in the real settings accurately represent the true economic value, yet it is entirely possible that the reverse is true. For example, when payments for public good provision are consequential, responses could be biased downward due to factors like free-riding or the desire to only pay one's fair share. In the case of private goods, both Harrison, *et al.* 2004 and Murphy and Stevens 2004 hypothesize that responses to actual payment questions for private goods may be censored by the market price. The linkage between value elicitation for public and private goods is particularly important considering that many experimental studies of non-market valuation techniques use private goods, and as with public goods, hypothetical bias is often observed.

Much is known about the presence of hypothetical bias with homegrown goods, even though the underlying causes are not fully understood. There is a recent increase in the use of induced value experiments to shed light on the hypothetical bias observed when eliciting homegrown values. In induced value experiments, the true underlying value is known and therefore deviations from this value are more readily identified. In an induced value experiment that uses a referendum for provision of a public good, Taylor, *et al*. 2001 find no evidence of hypothetical bias in aggregate, even though they observe errors at the individual level. They

conclude that hypothetical bias is not a value elicitation problem, but rather that value formation may be at the heart of hypothetical bias. Their conclusion is certainly reasonable, and dovetails nicely with the literature about value uncertainty.[2] However, respondents misvoted—that is, voted in a manner inconsistent with their induced values—over 16% of the time in both the real and hypothetical payment treatments. It turns out that even though roughly one in six responses was a "decision error," there was no hypothetical bias in aggregate.

Even in the absence of any hypothetical bias, it is still possible that the WTP estimates could be inaccurate. This is precisely what Mitani and Flores (2009) observe. They find no evidence of hypothetical bias in aggregate for a threshold public good with an open-ended elicitation procedure. However, only 8% of respondents truthfully revealed their true value, 75% underreported their values and 17% overreported, leading to WTP estimates that are significantly below induced values. Vossler and McKee 2006 investigate value uncertainty as a possible cause of hypothetical bias by testing four elicitation mechanisms: dichotomous choice, dichotomous choice with a follow-up certainty question, payment card, and multiple-bounded discrete choice. They find no evidence of hypothetical bias in aggregate with any of these mechanisms, although similar to Taylor *et al*., they do find evidence of decision errors at the individual level. There are also two studies that use induced values to test the properties of contingent valuation when individual costs are uncertain (Burton, *et al*. 2003; Polomé 2003), but neither has an inconsequential payment condition to draw inferences about hypothetical bias.

---

[2] The relevant literature regarding uncertain preferences includes Opaluch and Segerson 1989, Gregory, *et al.* 1995, Li and Mattsson 1995, Gregory and Slovic 1997, Wang 1997. Subsequent studies which use a respondent's self-reported level of uncertainty as an *ex post* calibration mechanism include Champ, *et al.* 1997, Blumenschein, *et al.* 1998, Johannesson, *et al.* 1998, Ethier, *et al.* 2000, Champ and Bishop 2001, Poe, *et al.* 2002. For a review of this literature, see Murphy and Stevens 2004.

Interestingly, the three aforementioned induced value studies report no evidence of hypothetical bias in aggregate for a public good provided via either a referendum or a voluntary contribution with a provision point. However, Cherry *et al*. 2004 do report the presence of hypothetical bias for a private good in an induced value experiment that uses a theoretically incentive-compatible second price auction to elicit values (we revisit this result later). Moreover, there was substantial individual decision error in their no outside option treatment (which is the closest parallel to our study).

This small but growing literature suggesting the absence of hypothetical bias in aggregate in induced value settings, but a nontrivial level of biased responses persists at the individual level, even with theoretically incentive compatible elicitation mechanisms. In most induced value studies, the individual decision errors have been consistent between real and hypothetical payments, leading to no hypothetical bias in aggregate. Nevertheless, these individual errors could still lead to inaccurate WTP estimates and also raise concerns about the conditions under which these individual errors arise.

Both Taylor *et al*. and Vossler and McKee focused solely on eliciting induced values for public goods. Our experiments extend this line of research by linking not only the public and private good literatures, but also the induced value and homegrown value literatures, so we can better understand how inferences from one context can provide insights into another. We do not observe hypothetical bias in aggregate in induced value experiments for any good type, but when subjects are asked to value homegrown goods, hypothetical bias arises. This provides some support for Taylor *et al*.'s 2001 conclusion that hypothetical bias is a value formation problem, not one of value elicitation. Like Taylor *et al*., we also observe the presence of individual decision errors, or misvotes, in induced value referenda. However, when subjects are given a

purely private decision with no strategic interaction or potential for other regarding behavior, these individual decision errors disappear. This would suggest that value elicitation, particularly in a referendum, may affect responses at the individual level.

3. Experimental Design

A total of 169 students were recruited from the general student population at the University of Massachusetts, Amherst. Subjects were informed about the experiments through flyers that were placed around campus and a recruiting table that was set up in the campus center. Subjects were paid $5 for showing up on time and were told they would be given an opportunity to earn additional money during an experiment that would last no more than 90-minutes (actual experiment time was less than one hour).

Our 3×2 (type of good × payment type) experimental design is shown in Table 1. Each of the six cells represents a unique treatment. Each subject participated in only one of the six treatments. The data were collected on six separate days, one treatment each day. The three rows indicate the three good types (private, publicly-provided private, or public). As shown in Table 1, the switch from a private to public good context entails changes to both the good used to elicit homegrown values and the elicitation mechanism. To control for potential confounding effects, we also included a publicly-provided private good as an intermediate step between the two good types. The payment type columns reflect the two payment conditions (real or hypothetical). The real and hypothetical treatments differed only in that outcomes in the former were consequential, whereas in the latter subjects received a pre-announced fixed payment regardless of the experiment results. The instructions in the hypothetical payment treatments were modified slightly to reflect the inconsequential nature of the payments. Subjects in these hypothetical

treatments were told: "Each of you will receive exactly $13.50 at the end of the experiment today (in addition to the $5 you already received for participating), regardless of your decisions and the outcomes... Although your earnings in today's experiment are fixed at $13.50, we ask you to suppose that your earnings were based on your decisions and the auction outcomes..." The $13.50 fixed earnings in the hypothetical payment treatments equal the expected earnings in the real payment treatments.[3]

In each of the six treatments in Table 1, subjects participated in a total of five rounds. In each round of a particular treatment, the type of good, payment type and elicitation mechanism remained constant; the only factor that varied across the five rounds for a particular subject was whether induced or homegrown values were elicited. As shown in Table 2, the first two rounds were practice rounds using induced values to familiarize subjects with the instructions and the elicitation mechanism.[4] After both practice rounds were completed, the transaction prices for each practice round were determined and results announced. The results from the practice rounds did not count towards an individual's earnings and are not included in the data. After completion of the two practice rounds (P1 and P2), stage 1 (rounds 1 and 2) was introduced. Like the practice rounds, stage 1 also used induced values. Stage 2 (round 3) followed the same procedures as the previous rounds with the exception that induced values were no longer used. Instead, homegrown values for a specific good (coffee mug or a contribution to Heifer International) were elicited.

---

[3] It turns out that the additional earnings in the real payment treatments were higher, averaging $17.43 ($\sigma$=2.78). This difference between actual and expected earnings is due to the transaction prices that were randomly selected using a BDM-style elicitation mechanism described shortly.

[4] Induced values were described as the purchase of a token in the private and publicly provided private good experiments, and a benefit from a project in the public good experiments.

Hence, for each subject, we have three WTP observations (two induced values from stage one, and one homegrown value from stage two), all of which were elicited in same manner under the same payment condition (i.e., real or hypothetical). This design allows us to make within-subject comparisons of outcomes in induced vs. homegrown values settings while holding the payment condition and good type constant, and between-subject tests for hypothetical bias holding constant the type of good and value type (i.e., induced or homegrown).

Private good. In the two private good treatments (one each for real and hypothetical payment), values were elicited using a slightly modified version of the Becker-DeGroot-Marschak (BDM) mechanism (Becker *et al*. 1964).[5] For both the induced value (stage one) and homegrown value (stage two) elicitation, subjects received a purchase offer slip that contained each of the 15 possible prices, ranging between $1 and $15 in whole dollar increments. Table 3 shows an example of the purchase offer slip for a subject with an induced value of $4.50 in the private good real payment treatment. For each of these possible prices, subjects indicated whether they were willing to make a purchase at that price. Bids had to be consistent in that if they indicated "yes" to a particular price, then they had to indicate yes to all lower prices. With this payment card approach, an individual's maximum WTP lies between the highest value to which she indicated "yes" and the next highest amount. The transaction price was determined by randomly selecting a marble from a basket containing 15 marbles numbered 1 through 15; the number on the marble was the price. If the participant indicated "yes" to making a purchase at

[5] Horowitz (2006) shows that the BDM is not theoretically incentive compatible for some behavioral models which are outside an expected utility framework. The empirical results are mixed. Both Irwin *et al*. (1998) and Vossler and McKee (2006) provide experimental evidence supporting the demand revealing properties of the BDM, and Plott and Zeiler (2005) cite the BDM's incentive compatibility, particular with subjects who are trained in a mechanism which might otherwise be unfamiliar. On the other hand, Noussair *et al*. (2004) find that the second-price auction is a more effective at eliciting willingness-to-pay than the BDM. Lusk and Rousu (2006) also found that the second-price and random *n*th price auctions are more accurate than the BDM on average.

the randomly selected price, then a transaction occurred at that price. Earnings were the difference between the induced value of the token and the transaction price. If the participant indicated "no" for the selected price, then no transaction occurred and earnings were zero. This process was repeated for each participant.

The first stage used induced values and consisted of two separate individual decision tasks using tokens with pre-assigned values. Before the stage one tasks, all subjects were told they had $10 in an initial cash balance and that any earnings would be added to this. All subjects had an induced value of $4.50 in the first round and $11.50 in the second. These values were private information and subjects were told to make no assumptions about the others' token values. Assuming truthful demand revelation, expected earnings in the first round are $0.53 and $4.03 in the second.

The second stage (round 3) elicited homegrown values. Each subject was presented with an opportunity to purchase one 16-ounce stainless steel travel coffee mug with the logo of a popular on-campus student-run coffee shop. The mug was described as follows:

> We will sell these 16-ounce stainless steel travel mugs from People's Market, which I will now pass around for you to see. As many of you probably know, People's Market is a non-profit student run co-op located in the Student Union. This mug entitles you to a 20-cent discount on each of your future coffee purchases at People's Market. This discount only applies to coffee purchases; it does not apply to any other beverage, including tea. People's Market does not have a refund policy, so if you buy a mug you will not be able to return it. (If it is defective, you will be able to exchange it for a replacement).[6]

Other than the change from induced values to homegrown values, the second stage proceeded in exactly the same manner as the first. After the second stage was completed, all three prices (two for stage one, and another for the second stage) were determined using the BDM mechanism.

[6] The no-refund policy removes the possibility of purchasing the mug with the sole intent of returning it for the $10 retail price.

Outcomes were announced at the end of the experiment to avoid the possibility that later decisions could be conditioned on prior results.

Public good. In the two public good treatments, values were elicited through a BDM-style referendum that was modified to parallel the mechanism used for private goods as closely as possible (Vossler and McKee, 2006). Subjects received a voting slip similar to the purchase offer slip in Table 3. For each of the 15 possible "project costs," subjects had to indicate how they would vote in a referendum that required all participants to pay this amount, and as with the BDM, votes had to be consistent. The project cost was determined by randomly selecting a marble from a basket containing 15 marbles numbered 1 through 15; the number on the marble was the cost. After determining the cost, we tallied the votes for that cost. If more than half the subjects voted yes to the randomly selected cost, then the project was implemented and each subject paid that amount.

The first stage used induced values and was presented as a pair of group projects from which each participant received some pre-assigned benefit. The initial cash and induced values were the same as the private good BDM. In the second stage (with homegrown values), the induced values were replaced with a referendum on whether the group would make a contribution to Heifer International to purchase a flock of chickens for needy families. Subjects were given promotional materials from Heifer International, and the referendum was described as follows:

> Heifer International is a nonprofit, humanitarian organization dedicated to ending world hunger and saving the earth by providing livestock, trees, training and other resources to help poor families around the globe become self-reliant. Since 1944, Heifer International has worked directly with millions of families in more than 125 countries to alleviate hunger and increase self-reliance. The livestock given to poor families (such as cows, llamas, goats, chickens, etc.), provide milk, eggs, plowing power and other benefits that can mean improved nutrition, higher income, education for children, health care, improved housing and a new way of life. Heifer requires all livestock recipients to "Pass on the Gift" by sharing offspring of their livestock with others in need. This common sense approach leads to sustainable development of one family at a time.

> You will vote to decide whether everyone in the group will contribute part of their earnings to provide needy families with a small flock of chickens. Each dollar per person will provide an additional needy family with a flock of chickens.

Publicly-provided private good. Although careful attention was paid to make the public and private good treatments as similar as possible, they differ in two important dimensions: the elicitation mechanism (BDM vs. referendum) and the stage two good (coffee mug vs. Heifer International). In the second stage, a vote in support of a group contribution to Heifer International imposes costs on all group members and also provides a public good. The decision to purchase a mug, however, is a private transaction that does not impose a cost on other group members and the benefits also accrue only to the individual. To determine whether any differences in behavior between the public and private good contexts are attributable to the elicitation mechanism or the good itself, we also conducted a pair of publicly-provided private good treatments.

The first stage was the same as that for the public good with a slightly different frame. Subjects were asked to vote in a referendum (public provision) to determine whether everyone in the group would buy a token (private good) at a randomly selected price. The pair of induced values was the same as the other two treatments. In the second stage, subjects were asked to vote in a referendum to determine whether everyone would buy a coffee mug. If a majority of subjects voted yes to a randomly selected price, then each person would have to buy a mug regardless of how she voted.

4. Results

Induced values. We begin using the stage one results to test whether hypothetical bias exists with induced values. Because bids were constrained to whole dollar increments, a demand-

revealing individual with an induced value of $4.50 would indicate a maximum WTP of $4.00. Similarly, when induced values are $11.50, the maximum WTP should be $11.00. Indeed, the median WTP in both the real and hypothetical treatments conform to predictions for all three good types and both payment amounts. Table 4 shows that with induced values of $4.50, mean hypothetical values are slightly higher than real payments, but median values were identical. Mann-Whitney-Wilcoxon rank-sum (Wilcoxon 1945; Mann and Whitney 1947) tests indicate that this between-subject difference is not significant for either the private or public goods.[7] For the publicly-provided private good, the difference is weakly significant (p=0.08); this result is largely driven by four individuals in the hypothetical treatment who indicated a $5 WTP which would result in a $0.50 loss. It is possible that earnings less than $1 are not salient which could help explain this outcome (Taylor, *et al*. 2001). Vossler and McKee 2006 use a similar threshold to distinguish between small and large errors. When induced values are $11.50, mean hypothetical values are slightly lower than their real counterparts in two of the three treatments, but again these differences are not significant. Hence, we cannot reject the hypothesis of no hypothetical bias in five of the six comparisons of induced values (two induced values × three good types), and there is one case in which the null hypothesis is rejected, largely due to four subjects willing to incur a 50-cent loss. This result is consistent with those of both Taylor, *et al*. 2001 and Vossler and McKee 2006 in a public goods context, and with Cherry *et al*. 2004 for private goods.

Another measure for comparing bidding behavior between real and hypothetical payment conditions is the mean absolute deviation (MAD) of the observed bid from the induced value (Table 5). The MAD for subjects who fully reveals demand would be $0.50 (resulting from a

[7] Unless otherwise specified, all hypothesis tests conducted using Mann-Whitney-Wilcoxon rank-sum tests.

$4.00 bid with a $4.50 induced value or an $11.00 bid with an $11.50 induced value). Note that with this approach, a bid that incurs a $0.50 loss is treated identical to a demand revealing bid. Using this metric, there is no statistically significant difference between real and hypothetical bids for any of the comparisons in Table 5.

An advantage of the induced value experiments is that we can observe whether behavior is consistent with these values, whereas with homegrown values we must *assume* that behavior in the real payment treatments reflects the individual's true WTP. With the stage one induced value data, we can identify misvotes or decision errors, *i.e.*, those instances in which individuals are not behaving in a manner consistent with truthful demand revelation. In the private good treatments, Table 6 shows that were no misvotes in the real payment condition—which leaves little doubt that subjects fully understood the instructions, the elicitation mechanism and how to accurately state their values. Similarly, with hypothetical payments, only 3 of 56 subjects (5%) misvoted and a Fisher exact test indicates that there is no statistically significant difference (p=0.25) between misvotes in these hypothetical and real treatments.

This result of near-perfect bidding behavior using a BDM to elicit values for a private good provides strong support for the demand revealing properties of the mechanism. With real payments, Irwin *et al*. (1998) found that 62% of their BDM bids were optimal, and there was no significant difference between bids and induced values; they did not have a hypothetical payment condition for comparison. The no outside option treatments in Cherry *et al*. (2004) do allow for a comparison of hypothetical and real payments in an induced value experiment for a private good, but they use a standard second-price auction not a BDM.[8] When averaged over the course of 10 rounds, they do observe hypothetical bias: hypothetical bids are 38.5% higher than real

[8] Shogren *et al*. (2001) note that second price auctions are accurate in aggregate, but are susceptible to insincere bidding by off-margin traders.

bids on average. Much of this is attributable to inflated hypothetical bids in later rounds, possibly due to boredom. Their first two rounds represent a closer parallel with our experimental design. Somewhat surprisingly, they report substantial *underbidding* in both real and payment conditions for the first two rounds. Although real bids were 12.8% *higher* than hypothetical bids, as in our study this difference was not statistically significant.[9]

When we introduced the referendum, however, the rate of misvotes increased. In our public good referendum, we observe 19% and 12% misvotes in the real and hypothetical treatments, respectively, and a Fisher exact test indicates that there is no statistically significant difference between these (p=0.44). This result is consistent with those of Taylor *et al*. who observe about a 16% misvote rate in both treatments. In the publicly-provided private good treatment, hypothetical misvotes (34%) are significantly higher than in the real payment scenario (12%). Much of this difference in the rate of misvotes between the real and hypothetical payment scenarios is due to subjects voting yes to hypothetical outcomes that would yield a $0.50 loss. Both Taylor *et al*. and Vossler and McKee suggest that such small losses may not be salient; using a "loose" definition of misvotes (which does not classify a vote leading to a 50 cent loss as misvote) we find that the difference between real and hypothetical payments for the publicly provided private good is much smaller and only weakly significant (p=0.10). As in Taylor *et al*., we find that although there are some misvotes at the individual level, those who overbid and underbid tend to cancel each other out so aggregate results are approximately demand revealing with both the real and hypothetical payments. However, this also raises questions about the conditions under which these voting errors in a referendum might be asymmetric, possibly leading to a bias in some referenda.

---

[9] Thanks to Todd Cherry for sharing their data.

We find it particularly noteworthy that bidding behavior with private goods is near-perfect, but when the BDM-style referendum is introduced some misvoting occurs. This raises the question as to why misvoting arises more frequently with a group decision via referenda than with a purely private transaction. Our design does not permit a testing of possible causes for this outcome and we can only provide some conjectures about this. We note the rate of misvotes observed is consistent with that reported by both Taylor *et al*. and Vossler and McKee. While we cannot rule out the possibility that misvotes in the referendum are due to confusion or misunderstanding, we feel that this is unlikely, particularly in light of the near-perfect outcomes with the private good BDM mechanism which was structured to be as close as possible to the BDM-style referendum we used. More likely, these misvotes are not random errors, but rather indicate that a referendum is highly demand revealing, but not perfectly so. Vossler and McKee suggest that some misvoting could be due to other-regarding behavior, in particular a willingness to incur a small loss if it is in the best interest of the group. They conclude that this effect is likely to be small and we tend to agree. Without information about the distribution of others' induced values, subjects have no way of knowing what is in the group interest. Another possible explanation is a misguided attempt at strategic behavior or coordination. The private good BDM is a purely individual decision making exercise with no potential for strategic interactions or other-regarding behavior. However, the referendum does entail a group decision and subjects could conceivably make errors in recognizing that this mechanism is also incentive compatible.

Homegrown values. Whereas we observe no hypothetical bias with induced values, when we elicit homegrown values this bias does emerge. The bottom of Table 4 shows that hypothetical bias is present for all three types of good. Median WTP is \$3 to \$4 higher in the three hypothetical payment treatments than their real payment counterparts, and these differences

are statistically significant. Thus, the same subjects who exhibited no hypothetical bias in stage one with induced values, do exhibit hypothetical bias when asked to value a specific good in stage two. These results clearly support Taylor *et al*.'s conjecture that hypothetical bias is not a value elicitation problem in aggregate, but rather a value formation problem.

An unexpected result with the homegrown values is the higher willingness-to-pay for the exact same coffee mug in the private good treatment when compared to the publicly-provided context. The main difference between the two is the elicitation procedure: individual purchase via an individual decision task versus a referendum requiring all group members to buy a mug. The difference is significant for real payments (p=0.08), but not with hypothetical payments (p=0.22). If anything, we anticipated a lower willingness-to-pay in the referendum because participants might be reluctant to require others to involuntarily purchase a private good.

5. Conclusions

Induced valuation respondents are confronted with known values for the commodity being examined. Hypothetical bias is therefore unlikely with induced values, but it could arise; in a post-experiment survey some individuals in our real payment treatments with referenda expressed concern about forcing others to pay. On the other hand, homegrown values formulated during the CV exercise are likely to be quite uncertain and the potential for hypothetical bias is much greater.

Although few studies have examined the process by which individuals make decisions in the CV framework, Schkade and Payne's (1994) verbal protocol analysis suggests that respondents in a hypothetical homegrown valuation focus on a variety of considerations. Among these are the notion of paying ones' fair share, consideration of previous charitable contributions,

and a concern for "how the world should be", all largely detached from the context of money. But, when asked to make actual payment, respondents appear to focus on monetary considerations, and this can give rise to hypothetical bias.

Of course, in homegrown situations many other factors, like free riding, may contribute to this problem. And, in some cases, hypothetical bias may even be seen as a general cognitive response; people like to think of themselves as being generous and this is costless when payment is not required. Even our everyday speech hints at hypothetical bias—"actions speak louder than words," "practice what you preach", and the "road to hell is paved with good intentions."

To better understand and to correct for hypothetical bias, a necessary first step is to be able to turn it off and on in the laboratory which is the major contribution of this paper. We observe that hypothetical bias at the aggregate level emerges during the process of homegrown value formation for private, publicly-provided private, and public goods. The presence of individual decision errors in our referenda using induced values, but not in the BDM, suggests that although value formation is a major factor in hypothetical bias, the elicitation mechanism may also play a secondary role, at least at the individual level.

References

Aadland, D. and Caplan, A. J. 2003. Willingness to pay for curbside recycling with detection and mitigation of hypothetical bias. *American Journal of Agricultural Economics*, 85(2), 492-502.

Aadland, D. and Caplan, A. J. 2006. Cheap talk reconsidered: Evidence from CVM. *Journal of Economics and Behavioral Organization*, 60(4), 562-578.

Becker, G., DeGroot, M., and Marschak, J. 1964. Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3), 226-232.

Blumenschein, K., Johannesson, M., Blomquist, G. C., Liljas, B. and O'Conor, R. M. 1998. al results on expressed certainty and hypothetical bias in contingent valuation. *Southern Economic Journal*, 65(1), 169-77.

Brown, T. C., Ajzen, I. and Hrubes, D. 2003. Further tests of entreaties to avoid hypothetical bias in referendum contingent valuation. *Journal of Environmental Economics and Management*, 46(2), 353-361.

Burton, A. C., Carson, K. S., Chilton, S. M., Hutchinson, W. G. 2003. An experimental investigation of explanations for inconsistencies in responses to second offers in double referenda. *Journal of Environmental Economics and Management*, 46(3), 472-489.

Champ, P. A., R. C. Bishop, T. C. Brown, and D. W. McCollum. 1997. Using Donation Mechanisms to Value Nonuse Benefits from Public Goods. *Journal of Environmental Economics and Management*, 33(2): 151-162.

Champ, P. A. and Bishop, R. C. 2001. Donation payment mechanisms and contingent valuation: an empirical study of hypothetical bias. *Environmental and Resource Economics*, 19(4), 383-402.

Cherry, T. L., Frykblom, P., Shogren, J., List, J. and Williams, M. 2004. Laboratory testbeds and nonmarket valuation: the case of bidding behavior in a second price auction with an outside option. *Environmental and Resource Economics*, 29(3), 285-294.

Cummings, R. G., and Taylor, L. O. 1999. Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. *American Economic Review*, 89(3), 649 - 665.

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

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): 462-73.

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

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. 2008. Experimental Evidence on the Existence of Hypothetical Bias in Value Elicitation Methods. In: Plott, C.R. and Smith, V.L. (eds). *Handbook of Experimental Economics Results*, volume 1. New York: North-Holland.

Horowitz, J.K. 2006. The Becker-DeGroot-Marschak mechanism is not necessarily incentive compatible, even for non-random goods. *Economics Letters*. 93:6-11.

Horowitz, J.K., McConnell, K.E. and J.J. Murphy. Forthcoming. Behavioral Foundations of Environmental Economics and Valuation. In: List, J.A. and Price, M.F. (eds). *Handbook on Experimental Economics and the Environment*. Northampton, MA: Edward Elgar.

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

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.

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 sports cards. *American Economic Review*, 91(5), 1498-1507.

List, J. A. and Gallet, C. 2001. What experimental protocol influence disparities between actual and hypothetical stated values? *Environmental and Resource Economics*, 20(3), 241-254.

Lusk, J.L. 2003. Willingness-to-pay for Golden Rice. *American Journal of Agricultural Economics*. 85(4): 840-856.

Lusk, J.L. and M. Rousu. 2006. Market price endogeneity and accuracy of value elicitation mechanisms. In: List, J.A. (ed). *Using Experimental Methods in Environmental and Resource Economics*. Northampton, MA: Edward Elgar.

Mann, H. B., and D. R. Whitney. 1947. On a test whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*. 18: 50-60.

Mitani, Y. and Flores, N.E. 2009. Demand Revelation, Hypothetical Bias, and Threshold Public Goods Provision. *Environmental and Resource Economics*. 44(2):231-243.

Murphy, J. J. and Stevens, T. H. 2004. Contingent valuation, hypothetical bias and experimental economics. *Agricultural and Resource Economics Review*, 33(2), 182-192.

Murphy, J. J., Stevens, T. H., Allen, P. G. and Weatherhead, D. 2005a. A Meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics*, 30(3), 313-325.

Murphy, J. J., Stevens, T. H., and Weatherhead, D. 2005b. Is cheap talk effective at eliminating hypothetical bias in a provision point mechanism? *Environmental and Resource Economics*, 30(3), 327-343.

Noussair, C., Robin, S. and B. Ruffieux. 2004. Revealing consumers' willingness-to-pay: A comparison of the BDM mechanism and the Vickrey auction. *Journal of Economic Psychology*. 25:725-741.

Opaluch, J.J., Segerson, K., 1989. Rational roots of 'irrational' behavior: new theories of economic decision-making. *Northeastern Journal of Agricultural and Resource Economics* 18(2):81-95.

Plott, C.R. and K. Zeiler. 2005. The willingness to pay–willingness to accept gap, the "endowment effect," subject misconceptions, and the experimental procedures for eliciting evaluations. *American Economic Review*, 95(3):530-545.

Poe, G. L., Clark, J. E., Rondeau, D. and Schulze, W. D. 2002. Provision point mechanisms and field validity tests of contingent valuation. *Environmental and Resource Economics*, 23(1):105-131.

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

Schkade, D., and J.W. Payne. 1994. How People Respond to Contingent Valuation Questions: A Verbal Protocol Analysis of Willingness to Pay for an Environmental Regulation. *Journal of Environmental Economics and Management*. 26: 88-109.

Shogren, J.F., Margolis, M., Koo, C. and J.A. List. 2001. A Random *n*th-Price Auction. *Journal of Economic Behavior and Organization*. 46(4):409-421.

Taylor, L. O., McKee, M., Laury, S. K. and Cummings, R. G. 2001. Induced-value tests of the referendum voting mechanism. *Economics Letters*, 71(1):61-65.

Vossler, C. A. and McKee, M. 2006. Induced-Value Tests of Contingent Valuation Elicitation Mechanisms. *Environmental and Resource Economics*, 35(2):137-168.

Wang, 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.

Wilcoxon, F. 1945. Individual comparisons by ranking methods. *Biometrics*. 1: 80-83.

Table 1. Experimental Design

	Payment Type (number of subjects)			
Type of Good	**Real**	**Hypothetical**	**Homegrown Value Good**	**Elicitation Mechanism**
Private good	25	28	Coffee mug	BDM
Publicly-provided private good	29	29	Coffee mug	Referendum
Public Good	29	29	Heifer International	Referendum

Table 2. Sequence of events within a treatment

	Round	Value Type	Good offered
Practice	P1	Induced	Token value $6.25
	P2	Induced	Token value $11.60
	Practice results announced		
Stage 1	1	Induced	Token value $4.50
	2	Induced	Token value $11.50
Stage 2	3	Homegrown	Coffee mug or Heifer International
	Stage 1 and 2 results announced		

Table 3. Example Purchase Offer Slip in the Induced Value Private Good Real Payment Treatment

Your Token Value is **$4.50**.

Possible Prices	**Indicate whether you are willing to buy a token at each possible price**			
		If you <u>do</u> buy a token at this price, you will earn…		If you <u>do not</u> buy a token at this price, you will earn…
$1	□ **Yes**	**$3.50**	□ **No**	**$0**
$2	□ **Yes**	**$2.50**	□ **No**	**$0**
⋮	⋮	⋮	⋮	⋮
$14	□ **Yes**	**– $9.50**	□ **No**	**$0**
$15	□ **Yes**	**– $10.50**	□ **No**	**$0**

Please remember that **your Purchase Offers must be consistent**.

Table 4. Mean and Median Willingness-to-Pay

	Mean		Median		
Type of Good	**Real**	**Hypothetical**	**Real**	**Hypothetical**	**Wilcoxon (p-value) [a]**
$4.50 Induced Value					
Private	4.00 (0)	4.29 (0.25)	4.00	4.00	0.18
Public Private	3.93 (0.05)	4.34 (0.21)	4.00	4.00	0.08 *
Public	3.83 (0.13)	4.24 (0.18)	4.00	4.00	0.15
$11.50 Induced Value					
Private	11.00 (0)	10.86 (0.14)	11.00	11.00	0.34
Public Private	10.76 (0.11)	10.21 (0.40)	11.00	11.00	0.57
Public	10.59 (0.31)	10.93 (0.23)	11.00	11.00	0.92
Homegrown Value					
Private	2.08 (0.45)	4.43 (0.68)	1.00	5.00	0.02 ***
Public Private	3.41 (0.56)	5.55 (0.62)	3.00	6.00	0.01 ***
Public	5.72 (0.86)	8.17 (0.94)	5.00	8.00	0.07 *

[a] Results of a Mann-Whitney-Wilcoxon test of the null hypothesis that the real and hypothetical values were drawn from the same distribution. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Standard errors in parentheses.

Table 5. Mean Absolute Deviations of Bids from Induced Values

Type of Good	Mean Absolute Deviation[a] Real	Hypothetical	Wilcoxon (p-value) [b]
$4.50 Induced Value			
Private	0.50 (0.00)	0.71 (0.21)	0.34
Public Private	0.57 (0.05)	0.81 (0.15)	0.13
Public	0.71 (0.13)	0.71 (0.13)	1.00
$11.50 Induced Value			
Private	0.50 (0.00)	0.64 (0.14)	0.34
Public Private	0.74 (0.11)	1.50 (0.37)	0.16
Public	1.02 (0.29)	0.81 (0.20)	0.66

[a] Mean absolute deviation equals the induced value minus the observed bid. Standard errors in parentheses. If a subject submits a perfectly demand revealing bid, then this value is 0.50.

[b] Results of a Mann-Whitney-Wilcoxon test of the null hypothesis that the real and hypothetical values were drawn from the same distribution. * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$.

Table 6. Percent of Misvotes with Induced Values (Stage 1)

Type of Good	Percent Misvotes (both induced values combined) Real	Hypothetical	Fisher exact [a]	
Strict Definition [b]				
Private	0%	5%	0.25	
Public Private	12%	34%	0.01	***
Public	19%	12%	0.44	
Loose Definition [c]				
Private	0%	4%	0.50	
Public Private	12%	26%	0.10	*
Public	12%	10%	1.00	

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$

[a] p-values from a Fisher exact test comparing the rate of misvotes in the Real and Hypothetical treatments.

[b] A misvote occurs if a respondent indicates a WTP other than $4 with a $4.50 induced value, or a WTP other than $11 with an $11.50 induced value .

[c] A misvote occurs if a respondent indicates a WTP other than $4 or $5 with a $4.50 induced value, or a WTP other than $11 or $12 with an $11.50 induced value .