

Valuing Non-Marketed Goods and Services Using a List Field Experiment

John A. List
University of Chicago and NBER

First draft: Spring 1993
This draft: Winter 2025

Abstract

Contingent valuation is a widely used method for estimating the value of nonmarket commodities. Yet, a persistent issue is whether responses to Contingent Valuation Method (CVM) questions accurately reflect true values. Recent studies indicate that hypothetical bias is a significant factor that creates a gap between intentions and actions. I use a novel approach within non-market valuation—a List Experiment in the field—to test whether it can attenuate the hypothetical bias observed within CVM surveys. Using data from 400 subjects in a field experiment, I find initial promising results.

Keywords: CVM, benefit cost, hypothetical bias, List Experiment

JEL: C93.

Acknowledgements

This study was originally written for a class project in Ralph d'Arge's Environmental Economics course in the Spring of 1993 at the University of Wyoming. The paper received a strong A, so I figured I should revise it for a research publication. I tried to maintain as much of the original data analysis and writing as possible. Yet, the results remain unpublished to date. For correspondence: jlist@uchicago.edu

The core aim of environmental policy is straightforward: safeguard human well-being and protect the ecosystems crucial to our survival. With benefit-cost analyses now a staple at the federal level and increasingly at the state level, there's been a significant push toward developing robust, credible

methods for estimating the benefits and costs of environmental programs. While policymakers have become adept at quantifying the explicit costs of enhanced environmental protection, they currently rely on a variety of methods to measure the economic value of environmental goods and services. Among these, the Contingent Valuation Method (CVM) stands out as particularly contentious. This method, which allows researchers to gauge the total value of the commodity in question, has faced scrutiny, most notably in the critical review by Diamond and Hausman (1994).

Chief among these concerns is the question of whether hypothetical bias is inherently present in Contingent Valuation Method (CVM) responses. Emerging evidence from recent studies highlights notable differences between responses to real and hypothetical valuation questions. To tackle this issue, researchers have employed both *ex ante* and *ex post* methods to mitigate hypothetical bias (see, for example, Cummings and Taylor, 1999; List, 2001; Hofler and List, 2004; Carson et al., 2014 and for meta-analyses, see List and Gallet, 2001 and Murphy et al., 2005).

This study diverges from existing literature by employing a List Experiment. This technique was pioneered by political scientists in the 1980s to measure racial prejudice and attitudes towards affirmative action. I draw my inspiration from Sniderman et al. (1992), which I believe was the first published List Experiment in political science. Since then, there have been hundreds of studies in the social sciences using the List Experiment. Imai (2011) offers a comprehensive review of List Experiments in political science, discussing their statistical analysis and methodological considerations. Blair and Imai (2012) present a review of List Experiments in social science research,

focusing on their application in measuring sensitive attitudes and behaviors. Treibich and Lépine (2020) provide an overview of List Experiments in health research, specifically examining their use in estimating the prevalence of sensitive health behaviors. Coutts and Jann (2011) review the use of List Experiments in sociology, discussing their effectiveness in reducing social desirability bias. Glynn (2013) offers a critical review of List Experiments across various social science disciplines, examining their strengths and limitations.

These studies collectively demonstrate the widespread use of List Experiments across the social sciences, including political science, sociology, and public health research. They highlight the method's effectiveness in measuring sensitive topics while also discussing its methodological challenges and potential improvements.

My study represents a novel application in the realm of economics experiments on non-market valuation. To the best of my knowledge, this is the first application of a List Experiment in the field to value non-marketed goods and services. The List Experiment holds significant promise for benefit estimation within CVM, as it potentially mitigates social desirability bias by embedding the sensitive question within a list of non-sensitive items.

My design is straightforward and follows the general approach of Sniderman et al. (1991). In the Spring of 1993, I conducted an experiment at a sports card show in Denver, CO that had two treatments: the sports card treatment and the Yellowstone Park treatment. For each treatment, I developed two sub-treatments: a control group list of four items and a treatment group list of five items. The treatment group list

included all the items from the control list plus one additional item.

For the Yellowstone Park treatment, the added item was "I would pay a \$100 one-time tax to reintroduce wolves into Yellowstone Park."¹ For the sports card treatment, the added item was "I would pay \$50 for a Ken Griffey Jr. 1989 Upper Deck PSA 10." To explore external validity, I also conducted a Vickrey second-price auction for the Griffey Jr. sports card treatment subjects by holding a real auction for the card. I am unaware of such an incentive-compatible external validity test being done in the literature on List Experiments, either within economics or the broader social sciences.

The pilot results, drawn from 400 subjects, are encouraging. For the Yellowstone Park List treatment, the data suggest that 31% of respondents were willing to pay the \$100 one-time tax to reintroduce wolves into Yellowstone Park. In the sports card treatment, 26% of subjects revealed that they would pay \$50 for the Griffey card. Notably, in the auction for the Griffey Jr. sports card, approximately 22% of bidders bid at least \$50. A test of proportions shows that these percentages are not significantly different at conventional levels. While these findings mark a step forward, they also present a challenge for future researchers to replicate these pilot results and to design List Experiments that can accurately recover true values in environments that have extraneous factors that can influence stated preferences.

¹ This was a timely discussion of the day, and wolves were reintroduced to Yellowstone National Park on January 12, 1995.

I view the overarching set of results as important for policymakers, as they potentially provide a reliable method for valuing non-market goods and services, ensuring accurate and informed decisions about resource allocation and policy priorities. By understanding public support for environmental initiatives, such as the reintroduction of wolves to Yellowstone, policymakers can better gauge public preferences and justify expenditures. This approach therefore aids in designing policies that align with constituents' values and enhances the credibility and transparency of policy decisions, fostering greater public trust. Additionally, reliable valuation methods improve benefit-cost analyses, leading to better societal outcomes.

The remainder of the study proceeds as follows. The next Section describes the List field experimental design. Section III presents the experimental results. Section IV concludes.

II. Experimental Design

I conducted the experiment at a sportscard show in Denver, CO in March 1993. For similar experiments conducted in this setting please see List and Shogren (1998) and List (2001, 2002, 2003), though this List field experiment pre-dated all others. The experiment followed a three- or four-step process. In Step 1, either I or my assistant approached a person entering the show and asked if they would like to participate in a study that would last no more than 10 minutes. If the individual agreed, in Step 2, I placed them into one of two treatments.

If they were placed in the Yellowstone Park treatment, then I further randomized if they were in treatment or control. For those in the control group, I asked one simple

question: Please indicate how many (not which) of the following four statements apply to you:

- o I recycle every week.
- o I support environmental initiatives around Denver.
- o In the past, I have donated to wildlife conservation causes, such as the NWF or the Sierra Club.
- o I have tried to educate people, either through chatting or writing, on environmental issues in the past.

For those in the treatment group, I asked the same question, but I included a fifth statement:

- o I would pay a \$100 one-time tax to re-introduce wolves into Yellowstone Park

Since reintroduction of wolves into Yellowstone Park was a current local topic of interest for Coloradans, this fifth option needed little further description.

If the subject was placed into the sports card treatment, then I also further randomized them into treatment or control. For the control group, I included the following list:

- o I own a baseball, or a baseball card, autographed by a current star player.
- o I have traded sports cards with dealers at a card show in the last year.
- o I have friends who are collectors.
- o I have attended more than ten baseball card shows in the past year.

And, for the treatment group, I augmented that list by adding one more statement:

- o I would pay \$50 for a Ken Griffey 1989 Upper Deck PSA 10.

For subjects who were placed in the sports card treatment group, they proceeded to Step 3. In Step 3, I told the participant that he or she now had the chance to bid on the Griffey card from the survey. I then retrieved the card from under my dealer's table and the participant could pick up and

visually examine the Griffey card, which was sealed with the PSA grade clearly marked on the cardholder. I further told the subjects that they would be bidding in a sealed bid *second-price auction* and that for this type of auction it was best for them to bid their *true value* for this item because overbidding may cause them to pay too much, and underbidding decreases their odds of winning the item.

Finally, in Step 4 we debriefed the participant (Step 3 for non-auction subjects). The monitor explained that the actual bids would be ranked, and the participant would be contacted within 3 days after the show if he or she was the highest bidder. For those not in this treatment arm, they were thanked for their participation.

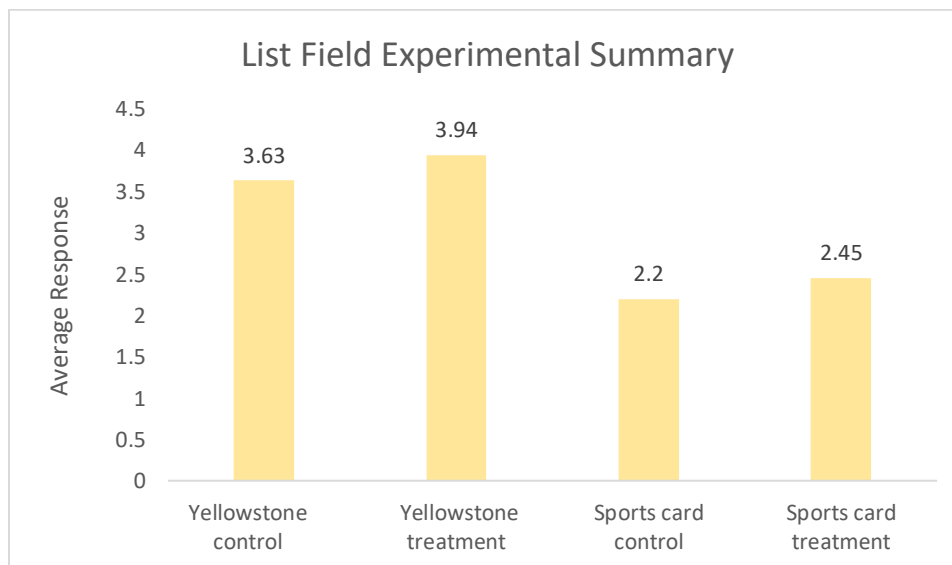
The treatments took approximately 18 hours to complete over two days. On the top of each hour the treatment was switched. I conducted each treatment until 100 subjects participated. Since no participant took part in more than one treatment, I had 400 subjects in total.

III. Experimental Results

Experimental results for List experiments are straightforward to present. Figure 1 shows the summary results for the average number of statements the subject agreed with in each treatment type. As Figure 1 reveals, the average number of statements the control group subjects agreed with in the Yellowstone Park treatment was 3.63 whereas the comparable figure for the treatment group was 3.94. The difference ($3.94 - 3.63 = 0.31$) suggests that approximately 31% of respondents would be willing to pay a \$100 one-time tax for wolves to be re-introduced to Yellowstone Park.

For the sports card treatment, Figure 1 shows that the average number of statements the control group subjects

agreed with was 2.20 whereas the comparable figure for the treatment group was 2.45. The difference ($2.45 - 2.19 = 0.26$) suggests that approximately 26% of respondents are willing to pay \$50 for the Ken Griffey 1989 Upper Deck PSA 10.² Interestingly, this figure compares well to the 22% of subjects who bid \$50 or more for the Ken Griffey 1989 Upper Deck PSA 10 card in the second-price auction. When doing a test of proportions, I fail to reject the homogenous null of proportional equivalence at conventional levels.



Note: Figures summarize results from the List Field Experiment.

IV. Epilogue

The application of a List Field Experiment to address hypothetical bias in CVM holds significant promise for improving the accuracy of economic valuations for nonmarket commodities. Accurate valuations are critical for informed

² The card ended up selling in the auction for \$100. The same card today regularly auctions for roughly \$2500.

policymaking and resource allocation, particularly in environmental economics, where the true value of public goods like clean air, water, and natural landscapes must be estimated. By reducing the gap between stated intentions and actual behaviors, the List Experiment enhances the credibility and reliability of survey results, thereby leading to more effective and justified policy decisions that better reflect the public's preferences and willingness to pay for non-marketed goods and services. This, in turn, supports sustainable development and the efficient management of environmental resources.

References

Blair, Graeme, and Kosuke Imai. "Statistical Analysis of List Experiments." *Political Analysis* 20, no. 1 (2012): 47-77.

Carson, Richard T., Theodore Groves, and John A. List. 2014. "Consequentiality: A Theoretical and Experimental Exploration of a Single Binary Choice." *Journal of the Association of Environmental and Resource Economists* 1, no. 1/2: pp. 171-207.

Coutts, Elisabeth, and Ben Jann. "Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (UCT)." *Sociological Methods & Research* 40, no. 1 (2011): 169-193.

Cummings, Ronald, G., and Laura O. Taylor. 1999. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review*, 89 (3): pp. 649-665.

Diamond, Peter and Jerry Hausman, 1994. Contingent valuation: Is some number better than no number? *Journal of Economic Perspectives* 8, pp. 45-64.

Glynn, Adam N. "What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment." *Public Opinion Quarterly* 77, no. S1 (2013): 159-172.

Hofler, Richard A., and John A. List. 2004. "Valuation on the Frontier: Calibrating Actual and Hypothetical Statements of Value." *American Journal of Agricultural Economics* 86 (1): 213-223.

Imai, Kosuke. "Multivariate Regression Analysis for the Item Count Technique." *Journal of the American Statistical Association* 106, no. 494 (2011): 407-416.

List, John, A. 2001. "Do Explicit Warnings Eliminate the Hypothetical Bias in Elicitation Procedures? Evidence from Field Auctions for Sportscards." *American Economic Review*, 91 (5): pp. 1498-1507.

List, John, A. 2002. "Preference Reversals of a Different Kind: The "More Is Less" Phenomenon." *American Economic Review*, 92 (5): pp. 1636-1643.

List, John A., 2003. Does Market Experience Eliminate Market Anomalies?, *The Quarterly Journal of Economics*, 118(1), pp. 41-71.

List, John A., and Craig A. Gallet. 2001. "What Experimental Protocol Influence Disparities Between Actual and Hypothetical Stated Values? Evidence from a Meta-Analysis." *Environmental and Resource Economics* 20 (3): 241-254

List, John A., and Jason F. Shogren, 1998. "Calibration of the difference between actual and hypothetical valuations in a field experiment," *Journal of Economic Behavior & Organization*, 37(2), 1998, pp. 193-205.

Murphy, James J., P. Geoffrey Allen, Thomas H. Stevens, and Darryl Weatherhead. 2005. "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation." *Environmental and Resource Economics* 30 (3): 313-325.

Sniderman, P. M., P. E. Tetlock, and T. Piazza. 1992. Codebook for the 1991 National Race and Politics Survey. Survey Research Center, Berkeley, California.

Treibich, Carole, and Aurélia Lépine. "Estimating Misreporting in Sensitive Health Behaviours: Evidence from Condom Use of Female Sex Workers in Senegal." *PLoS ONE* 15, no. 4 (2020): e0231670.