



Handelshögskolan
VID GÖTEBORGS UNIVERSITET

WORKING PAPERS IN ECONOMICS

No 252

**Hypothetical bias in choice experiments:
Within versus between subject tests**

by

Olof Johansson-Stenman and Henrik Svedsäter

April 2007

ISSN 1403-2473 (print)
ISSN 1403-2465 (online)

SCHOOL OF BUSINESS, ECONOMICS AND LAW, GÖTEBORG UNIVERSITY

Department of Economics

Visiting address Vasagatan 1 Postal address P.O. Box 640, SE 405 30

Göteborg, Sweden

Phone + 46 (0) 31 786 0000

Hypothetical bias in choice experiments: Within versus between subject tests¹

Olof Johansson-Stenman
Department of Economics, Göteborg University

Henrik Svedsäter
Department of Psychology, Göteborg University and
Organisational Behaviour, London Business School

Abstract

A choice experiment eliciting environmental values is set up in order to test for hypothetical bias based on both within and between sample designs. A larger hypothetical bias was found in the latter case, which explains parts of the previous diverging results in the literature. People seem to prefer to do what they say they would do.

Key words: Stated-preference methods, choice experiment, hypothetical bias, internal consistency, non-market valuation

JEL-classification: C91, Q28

¹ We are grateful for useful comments on earlier versions from Fredrik Carlsson, George Loewenstein, Peter Martinsson and from seminar participants at Göteborg University, London Business School, London School of Economics, Toulouse University, the European Economic Association conference in Lausanne and an experimental economics workshop at Copenhagen University.

1. Introduction

The validity of Stated Preference (SP) methods for valuing non-market goods is of vital importance for policy purposes, and in the US the law prescribes that all major new regulations should be preceded by a cost-benefit analysis. An obvious validity test of an SP method is to compare hypothetical statements with people's real maximum willingness to pay (WTP). Empirical evidence of the Contingent Valuation (CV) method suggests a non-negligible hypothetical bias in that people's *stated* WTPs most often exceed their real-money WTPs; see meta-analyses by List and Gallet (2001) and Murphy et al. (2005). These studies also found a lower hypothetical bias in studies that relied on within-subject tests of hypothetical and actual WTP than in studies that made comparisons between subjects.²

It has been argued that Choice Experiments (CEs), where people make repeated choices between hypothetical bundles of goods, may be a more promising SP method (e.g. List et al. 2006). Previous results are mixed, however. Carlsson and Martinsson (2001) used a within-subject CE design without finding any significant hypothetical bias. Using a between-subject design, Cameron et al. (2002) report that the mean WTPs from CEs were between 30 and 330 percent larger than the corresponding real-money treatments, but a common underlying preference structure could not be rejected due to large error terms. Lusk and Schroeder (2003) and List et al. (2006) also found a significant hypothetical bias in a between subject design, although the bias disappeared once a cheap-talk script was introduced in the latter study.

The present study is the first that employs both within and between subject designs in the same CE. Drawing on an extensive body of research in psychology, following Festinger, (1957), we hypothesise that people strive for consistency in their answers across various domains, hence resulting in a higher real WTP when being preceded by hypothetical

² Murphy et al. found a large and significant effect between within and between subject studies, whereas List and Gallet only found a directional but not significant difference.

statements. The results are consistent with this prediction, implying that within subject tests may not be appropriate to assess hypothetical bias in CEs.

2. Experimental Design

Seventy (mainly graduate) students from various courses at the London School of Economics volunteered, and were randomly divided into two groups. The subjects in the first group made hypothetical choices followed by choices that involved real payments, whereas the subjects in the other group made real-money choices directly. The subjects faced 15 choice sets that were identical in all contexts (an extra choice set was added in order to test for inconsistency, but was not used in the analysis). The design largely follows that of Carlsson and Martinsson (2001), although we used both within and between-sample designs. The experiment was conducted in six sessions with 10-15 subjects in each session within the same week. Sessions with subjects who made both hypothetical and real choices were run first. The sub-samples were very similar in terms of identifiable characteristics.

Each session started with some questions about socio-economic characteristics. The subjects then received verbal and written instructions about the CE and the nature and purpose of the environmental projects, which were two campaigns run by the World Wildlife Fund (WWF). In the hypothetical setting the instructions additionally read:

The choices are hypothetical but it is still very important that you answer them truthfully and as if they involved real money. There are altogether 16 choices for you to make. Try to consider each of them in isolation as if that was the only choice you have to make.

The subjects did not know that they would face real-money choices after these hypothetical ones. In the sub-sequent real conditions, the subjects were given the following information:

In the following you will be presented similar choices as before, although now your choices will in fact determine how much money you earn in this experiment, as well as how much money is contributed to the campaigns. It thus involves real money. The procedure is the following:

- you will again make 16 pair-wise choices*
 - afterwards one of these will be drawn randomly as the actual choice set*
 - you will be paid the amount of money according to the alternative chosen in this particular choice set, whereas the corresponding contribution is paid anonymously by us to the WWF*
- Thus, your choices will determine how much money you earn in this experiment, and how much money that is contributed to the campaigns.*

Essentially the same information was presented to the subjects being faced with real choices directly. An example of a choice set is presented below.

<< Insert Table 1 about here >>

Each of the two alternatives were thus characterised by three attributes. There were four payment levels (£0, £5, £10, £15), four donation levels (£0, £7, £14, £21), and two campaigns: the African Elephant and the Green Sea Turtle. The chosen attribute levels were based on pre-tests. The real choices that followed the hypothetical ones were presented in a different order. Moreover, to counterbalance and enable tests for order effects, the order of the choice sets was reversed for half the subjects in each context.

Once the experiment was completed, all respondents left the classroom. They were then called back individually, whereby a draw was made to decide which of the 15 choice sets that would determine how much money the subject would earn, how much would be donated, and to which campaign.

3. The Model and Results

Consider a standard random-utility framework consisting of a systematic part V and a random unobservable part ε . The utility for individual i from choosing alternative 1 then becomes:

$$u_{i1} = V_{i1} + \varepsilon_{i1} \quad . \quad (1)$$

The probability that individual i chooses alternative 1 is then given by:

$$\Pr(A_i = 1) = \Pr(V_{i1} + \varepsilon_{i1} > V_{i2} + \varepsilon_{i2}) = \Pr(\varepsilon_{i1} - \varepsilon_{i2} > V_{i2} - V_{i1}) \quad , \quad (2)$$

where the differences between the error terms are assumed to be logistically distributed, and V to be linear in the interval considered:

$$V_{i1} = \alpha + \beta^x x + \beta^E D^E + \beta^T D^T + \beta^z z_i (D^E + D^T) \quad , \quad (3)$$

where D^E and D^T are the donations given to the Elephant and Turtle projects, respectively, and where z is a vector of dummy variables reflecting treatment and gender. The parameters associated with this model (except for the intercept that cancels out) can be estimated with a conditional logit model (e.g. Louviere et al. 2000). The *MWTP* for an additional pound donated to the Elephant Campaign is then given by:

$$MWTP_i^E = \frac{\partial u_i / \partial D^E}{\partial u_i / \partial x_i} = \frac{\partial V_i / \partial D^E}{\partial V_i / \partial x_i} = \frac{\beta^E + \beta^z z_i}{\beta^x} \quad . \quad (4)$$

Assuming men to constitute the base case, their *MWTP* for the elephant project is given by β^E / β^x . The *MWTP* difference between the two campaigns is given by $(\beta^E - \beta^T) / \beta^x$, and the *MWTP* difference between females and males by $(\beta^{Female} - \beta^{Male}) / \beta^x$. Similarly, in the pooled estimation, $(\beta^{Hypothetical} - \beta^{Real}) / \beta^x$ is the *MWTP* difference between the hypothetical and the real-money treatment, and $(\beta^{RealAfterHyp} - \beta^{Real}) / \beta^x$ is the *MWTP* difference between the real after hypothetical and the real-money directly treatment.

Table 2 below presents the parameter estimates of a pooled sample and each of the three samples, whereas Table 3 presents the corresponding *MWTPs*;³ the *t*-values are calculated based on the delta method. Men's mean *MWTP* for the Elephant Campaign is always statistically different from zero at the 1% level, but women still turned out to have a much higher *MWTP* across all contexts, consistent with some experimental evidence (e.g. Eckel and Grossman 1998).

MWTP (for men) is clearly highest in the hypothetical context (*MWTP*=0.64), lower in the real-money experiment that was preceded by the hypothetical experiment (0.56), and lowest in the real-money directly context (0.23). From the pooled model, where real-money-directly is the reference case, we find again that the *MWTP* differs between the samples in the expected direction. We also ran a model in order to test whether the *MWTPs* differ significantly between the hypothetical and the real-after-hypothetical contexts (not presented in Table 3), and it turned out that the difference is significant at the 5% level. The results from likelihood-ratio tests are similar. We can reject the hypotheses that the parameter estimates of the different contexts come from the same underlying distribution as follows: Real-directly versus Hypothetical (1%, $\chi_4^2 = 23.36$), Real-after-hypothetical versus Hypothetical (1%, $\chi_4^2 = 13.84$), and Real-after-hypothetical versus Real-directly (10%, $\chi_4^2 = 8.08$).

In order to control for unequal variance among the three samples, we also estimate a heteroscedastic full information maximum likelihood model, where the relative scale factors and the utility parameters are estimated simultaneously (see e.g. Hensher et al. 1999; Louviere

³ Ten subjects' responses were lexicographic in at least one context. Following Carlsson and Martinsson (2001), these observations are included in the analysis; however, excluding them implies no major alterations of our results. Five subjects had inconsistent preferences within one of the three contexts, of which one was inconsistent in both the hypothetical and the subsequent real-money experiment. This individual was excluded from the analysis.

et al. 2000). From this we could not reject equal variances among the sub-samples, and the parameter estimates are very similar to those of the conventional model.

<<Insert Table 2 and Table 3 about here>>

4. Conclusion

The results suggest the existence of a hypothetical bias also for CEs, and that the size of this bias is underestimated when quantified based on within subject tests. It appears that hypothetical statements in part work as internal commitment devices, and that people prefer to behave internally consistent and actually do what they say they would do. This is also in line with psychological theory and recent experimental evidence suggesting that people prefer not to lie (Gneezy 2005).

References

- Cameron, T. A., G. L. Poe, R. G. Ethier, and W. D. Schulze. (2002). Alternative Non-Market Value-Elicitation Methods: Are the Underlying Preferences the Same?, *Journal of Environmental Economics and Management*, 44, 391-425.
- Carlsson, F., and P. Martinsson. (2001). Do Hypothetical and Actual Willingness to Pay Differ in Choice Experiments? Application to the Valuation of the Environment, *Journal of Environmental Economics and Management*, 41, 179-92.
- Eckel, C. C., and P. J. Grossman. (1998). Are Women Less Selfish Than Men? Evidence from Dictator Experiments, *Economic Journal*, 108, 726-35.
- Festinger, L. (1957). *A Theory of Cognitive Dissonance*, Evanston, IL: Row, Peterson.
- Gneezy, U. (2005). Deception: The Role of Consequences, *American Economic Review*, 95, 384-94.

- Hensher, D, Louviere, J.J. and J. Swait. (1999). Combining sources of preference data, *Journal of Econometrics* 89:197-221.
- List, J. A., and C. A. Gallet. (2001). What Experimental Protocol Influence Disparities Between Actual and Hypothetical Values?, *Environmental and Resource Economics*, 20, 241-254.
- List, J. A., P. Sinha and M. H. Taylor. (2006). Using choice experiments to value non-market goods and services, *Advances in Economic Analysis & Policy*, 6(2), 1-37.
- Louviere, J.J, Hensher, D.A. and Swait, J.U.D. (2000). *Stated Choice Methods: Analysis and Applications*, Cambridge: Cambridge University Press.
- Lusk, J. L., and T. C. Schroeder. (2003). Are choice experiments incentive compatible? A test with quality differentiated beef steaks, *American Journal of Agricultural Economics*, 85, 840-56.
- Murphy, J. J., P. G. Allen, T. H. Stevens, and D. Weatherhead. (2005). A Meta Analysis of Hypothetical Bias in Stated Preference Valuation, *Environmental and Resource Economics*, 30, 313-25.

Table 1. Example of a choice set

Choice number 3	Alternative A	Alternative B
Money given to you	5	10
Contribution to campaign	14	7
Campaign	Elephant	Sea Turtle

Table 2. Estimated parameters from conditional logit models (*t*-values in parentheses).

Parameter	Hypothetical 1	Real after hypothetical	Real directly	Pooled sample	Pooled sample Heteroscedastic
β^{Money}	0.151*** (7.31)	0.155*** (7.68)	0.153*** (7.50)	0.152*** (12.97)	0.153*** (12.50)
$\beta^{Elephant}$	0.097*** (5.54)	0.087*** (5.20)	0.036** (2.35)	0.046*** (4.14)	0.046*** (4.07)
$\beta^{Elephant} - \beta^{Turtle}$	-0.010 (-1.09)	-0.003 (-0.32)	-0.014* (-1.66)	-0.009* (-1.77)	-0.008 (-1.57)
$\beta^{Female} - \beta^{Male}$	0.105*** (5.27)	0.056*** (3.23)	0.104*** (6.21)	0.088*** (8.56)	0.088*** (8.65)
$\beta^{Hypothetical} - \beta^{Real}$				0.060*** (4.76)	0.060*** (4.78)
$\beta^{RealAfterHyp} - \beta^{Real}$				0.020* (1.69)	0.020* (1.67)
Heteroscedasticity parameters:					
Hypothetical sample					0.0003 (0.04)
Real-after-Hypothetical sample					0.0043 (0.61)
Log-likelihood fcn	-276.09	-302.73	-307.88	-889.89	-889.73
Statistical observations	510	510	525	1545	1545

Note:***, ** and * denote significance at the 0.01 level, 0.05 level and 0.1 level, respectively.

Table 3. *MWTPs* based on the conditional logit models in Table 2 (*t*-values in parentheses).

Parameter	Hypothetical	Real after hypothetical	Real directly	Pooled sample	Pooled sample Heteroscedastic
$MWTP^{Elephant}$	0.64*** (7.05)	0.56*** (6.60)	0.23*** (2.78)	0.30*** (4.64)	0.30*** (4.65)
$MWTP^{Elephant} - MWTP^{Turtle}$	-0.06 (-1.08)	-0.02 (-0.32)	-0.09* (-1.64)	-0.06* (-1.76)	-0.05 (-1.55)
$MWTP^{Female} - MWTP^{Male}$	0.70*** (4.90)	0.36*** (3.12)	0.68*** (5.33)	0.58*** (7.86)	0.57*** (7.59)
$MWTP_{Hyp} - MWTP_{Real}$				0.39*** (4.63)	0.39*** (4.57)
$MWTP_{Real after Hyp} - MWTP_{Real}$				0.13* (1.68)	0.13* (1.65)

Note:***, ** and * denote significance at the 0.01 level, 0.05 level and 0.1 level, respectively.