Ecological Economics 81 (2012) 44–47

Contents lists available at SciVerse ScienceDirect

Ecological Economics

journal homepage: www.elsevier.com/locate/ecolecon

Commentary

Task familiarity and contextual cues predict hypothetical bias in a meta-analysis of stated preference studies

Felix Schläpfer [a,b,*], Baruch Fischhoff [c,d,1]

[a] Institute for Environmental Decisions, Federal Institute of Technology ETH, Zurich, Switzerland
[b] Department of Economics and Management, Kalaidos University of Applied Sciences, Zurich, Switzerland
[c] Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA
[d] Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA

ARTICLE INFO

Article history:
Received 30 September 2011
Received in revised form 22 February 2012
Accepted 13 June 2012
Available online 5 July 2012

Keywords:
Consistency
Constructed preferences
Contingent valuation
Hypothetical bias
Meta-analysis
Stated preferences
Validity

ABSTRACT

Traditional stated preference theory does not provide useful guidance on hypothetical bias because the bias lies outside the theory. Not surprisingly, meta-analyses using variables from the theory have not revealed consistent patterns of bias. We find that these models are substantially improved by including two additional variables drawn from positive research on choice behavior. These results lead us to recommend survey procedures that address these concerns.

1. Introduction

In stated preference research the difference of hypothetical and actual statements of value is known as *hypothetical bias* (e.g. Ehmke et al., 2008). Several recent meta-analyses have used regression analysis to predict the extent of hypothetical bias in stated preference studies based on characteristics such as whether a study involved private or public goods and whether it posed open-ended vs. closed-ended questions (List and Gallet, 2001; Little and Berrens, 2004; Murphy et al., 2005). These analyses have, however, found few consistent patterns in the data. Indeed, a recent review in the *Journal of Economic Surveys* concludes (Loomis, 2011, p. 363), "We still do not know why people may give a different WTP on a survey than the same people would be actually willing to pay in an experiment that involves real money," then suggests (p. 368):

> "[…] an entirely new theory of respondent behavior might be found in behavioral economics and psychology. For example, psychologists suggest that preferences are first constructed by the respondent for unfamiliar goods during the interview process itself (Schkade and Payne, 1994). Understanding how preferences are constructed might shed light on the factors influencing the magnitude of hypothetical bias and how to mitigate it."

We follow this suggestion, considering two central variables from psychological studies of preference construction (Fischhoff, 1991): how familiar respondents are with (a) the task and (b) the context. Both variables have no obvious representation in stated preference theory, which assumes articulated preferences for all possible tasks and contexts.

The specific question we examine is whether additional explanatory patterns emerge if measures of task familiarity and context are included in meta-analysis regressions for explaining hypothetical bias. We use an existing dataset, compiled by Murphy et al. (2005), to avoid potential bias due to a self-serving selection of the study sample.

2. *Homo oeconomicus*, Task Familiarity and Context

Although a central topic in the literature on stated preferences, hypothetical bias lies outside the theory underlying that research — which applies standard consumer theory with its assumption of complete, transitive, and fully articulated preferences for all

* Corresponding author at: Department of Management and Economics, Kalaidos, University of Applied Sciences, Hohlstrasse 535, 8048 Zurich, Switzerland. Tel.: +41 44 200 19 02; fax: +41 44 200 19 15.
E-mail addresses: felix.schlaepfer@kalaidos-fh.ch (F. Schläpfer), baruch@cmu.edu (B. Fischhoff).
[1] Tel.: +1 412 268 3246.

0921-8009/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolecon.2012.06.016

Table 1
Explanatory variables.

Variable	Description and coding	Mean
Variables used in Murphy et al. (2005)		
Student	Subject pool; 1 if college students; 0 if adults	0.421
Group[a]	Setting of the decisions; 1 if there was a group interaction (e.g. through a Vickrey auction); 0 if values were elicited individually	0.554
Private	Type of good; 1 if private good; 0 if public good	0.494
Choice	Elicitation mechanism (question format); 1 if choice based (dichotomous, polychotomous, referendum, payment card, conjoint); 0 otherwise	0.373
Within	Type of comparison; 1 if within-subject; 0 if between-subject	0.337
Calibrate	Use of any type of ex ante or ex post calibration; 1 if the observation is based on any such technique; 0 otherwise	0.205
Additional variables		
Familiar	Familiarity of the good; 1 if familiar; 0 if unfamiliar	0.230
Cues	Presence/absence of contextual cues; 1 if present; 0 if absent	0.349

[a] Group and Student are highly correlated. Murphy et al., therefore, used only one of the two in any model, finding no difference in results. Our models use Student (omitting Group).

Table 2
Correlations between the explanatory variables.[a]

	Group	Private	Choice	Within	Calibrate	Familiar	Cues
Student	0.766	−0.258	−0.054	0.474	−0.131	−0.001	−0.319
Group		−0.132	−0.159	0.486	−0.266	0.316	−0.512
Private			−0.115	−0.195	0.115	0.551	0.489
Choice				−0.340	0.102	0.172	0.061
Within					−0.299	−0.024	−0.202
Calibrate						−0.134	0.254
Familiar							0.142

[a] Pearson correlation coefficients, using all 83 observations.

possible goods and contexts (e.g. Fisher, 1996, p. 19ff.).[2] Following the challenge posed by Loomis (2011) and other recent studies (e.g. Carlsson, 2010; Robinson and Hammitt, 2011), we look to behavioral research for guidance lacking in standard economic accounts. That research typically finds (a) that choices are more consistent with familiar tasks than with unfamiliar ones (e.g. Coupey et al., 1998) and (b) that people rely on contextual cues to formulate consistent choices (e.g., Druckman, 2001; Morgan et al., 2001; Huffman et al., 2007; Spash, 2007). Contextual cues can be provided by interacting directly with others (Clemen and Reilly, 2002; Macmillan et al., 2002; Druckman and Nelson, 2003; Spash, 2007), observing their choices (Slembeck and Tyran, 2004), or learning the positions of organizations with known positions (Lupia, 1994; Druckman, 2001; Huffman et al., 2007; Schläpfer et al., 2008). If familiarity and context prove to have explanatory power, then stated preference research requires a significant reorientation, from patching the problems associated with individual biases (e.g. Hanley and Shogren, 2005) to adopting tasks better suited to the cognitive processes of respondents. After examining the evidence, we consider such implications.

3. Materials and Methods

3.1. Data and Variables

Murphy et al. (2005) included 28 studies, with 83 observations, selected because of their accessibility in the peer-reviewed literature (see Appendix Table A).[3] They examined six explanatory variables: (i) whether the good is public or private, (ii) whether the preferences are from students or other (perhaps more strongly motivated) individuals, (iii) whether preferences are elicited from individuals or (possibly more thoughtful) groups, (iv) whether preferences were expressed on an open-ended scale or by choosing among options that might suggest answers (e.g., possible prices), (v) whether the design was between-subjects or within-subjects (which might have induced sensitivity to the factors involved in consistency tests), and (vi) whether the researchers used any 'calibration' techniques to control for hypothetical bias.

We coded two additional variables for each observation: (a) whether the good was familiar or unfamiliar and (b) whether there were contextual cues available to the respondents (Table 1), such as some knowledge of market prices (for private goods) or political debate (for public goods). We coded the tasks based on their methods sections and without knowledge of their outcomes. Table A in the Appendix shows the resulting codes, for readers' examination, along with brief descriptions of the goods and choice contexts. Among the 28 studies, 13 had unfamiliar goods without informative context, 7 had unfamiliar goods with informative context, 5 familiar goods without informative context, and 3 familiar goods with informative context. We used the estimates of hypothetical bias provided by Murphy et al. (2005). Based on the dataset of Murphy et al. (2005; see footnote 3) and the information in the Appendix, readers may repeat the analysis with their own coding.

Table 2 shows correlations among the variables. Because the explanatory (dummy) variables "Group" and "Student" are highly correlated, Murphy et al. used only one at a time (in alternative models with similar results). As a result, we used just "Student" (omitting "Group").

3.2. Statistical Analyses

Following Murphy et al. (2005, p. 314), we sought to "evaluate the effect of several SP formats and other factors on the degree of hypothetical bias." We used (natural log-transformed) hypothetical bias – the ratio of hypothetical and actual values – as the dependent variable, as it improves model fit given a right-skewed distribution of calibration factors (Fig. 1 in Murphy et al.).

We estimated two models. Model 1 uses the explanatory variables considered by Murphy et al. Model 2 adds terms for "task familiarity," "contextual cues" and their interaction.

Many of the 83 observations in the original dataset are problematic due to within-subject designs, calibration, or pseudo-replication. Within-subject designs can signal the features that interest the researchers, thereby inducing consistency (Fischhoff and Bar-Hillel, 1984). Calibration techniques attempt to eliminate the bias, sometimes ex post, thereby complicating further analysis. Therefore, we run the regressions both with and without the 44 observations having within-subject designs or calibration. The reduced set also decreases the problem of pseudo-replication by dropping the multiple observations in several studies (Sinden, 1988; Blumenschein et al., 2001; Johannesson et al., 1998; List, 2001, 2003).

4. Results

The left-hand side of Table 3 shows regression estimates for the full 83-observation dataset. In Model 1 (with predictors from Murphy et al.), the only significant effects are for *Within* (indicating less bias with within-subject designs, which may signal the desired behavior) and for *Calibrate* (indicating less bias when stated values are calibrated, so

[2] One exception is that standard survey approaches assume that respondents do *not* answer strategically (e.g. Arrow et al. 1993, p. 4606), even though they should, as rational actors (see Flores and Strong, 2007).

[3] The original dataset is available online at: http://faculty.cbpp.uaa.alaska.edu/jmurphy/meta/meta.html#data (accessed 22 February, 2012).

Table 3
Ordinary least squares estimates of models explaining the ratio of hypothetical to actual willingness to pay.[a]

Variable	Full dataset				Trimmed dataset[b]			
	Model 1		Model 2		Model 3		Model 4	
	Coeff.	t-Ratio	Coeff.	t-Ratio	Coeff.	t-Ratio	Coeff.	t-Ratio
Constant	0.887***	4.910	0.889	4.678***	1.390	5.031***	1.426***	5.210
Student	0.143	0.784	0.050	0.259	0.450	1.523	0.247	0.840
Private	0.076	0.460	0.753	3.340***	−0.573	−2.051**	0.307	0.773
Choice	−0.186	−1.064	0.031	0.173	−0.630	−2.256**	−0.426	−1.559
Within	−0.719***	−3.469	−0.626	−3.251***				
Calibrate	−0.533**	−2.633	−0.662	−3.369***				
Cues (C)			−0.552	−2.302**			−0.830**	−2.233
Familiar (F)			−1.071	−3.503***			−1.187**	−2.610
C×F			0.503	1.184			0.788	1.363
n	83		83		39		39	
R^2	0.177		0.328		0.185		0.359	
Adj. R^2	0.124		0.256		0.115		0.239	

Notes: Significance levels: *: p<0.1; **: p<0.05; ***: p<0.01.
[a] The dependent variable is the natural logarithm of hypothetical bias (hypothetical value/actual value).
[b] Within-subject comparisons and observations with "calibrated" values excluded.

as to reduce the bias). The adjusted R^2 is only 0.12. Adding *Familiar* and *Cues* (Model 2) increases the adjusted R^2 to 0.26. Both variables are highly significant, with their signs indicating that the bias is less when the tasks are familiar and the context is meaningful. The interaction of *Familiar* and *Cues* is non-significant. There is now also a significant coefficient for *Private*, indicating smaller bias with private goods than with public ones — once familiarity and contextual richness are controlled.

Models 3 and 4 provide analogous analyses for the trimmed dataset (without within-subject and calibrated observations). Model 3 reveals significantly less bias with *Private* goods, rather than public ones, and when the *Choice* involves fixed options (which might suggest expected preferences), rather than open-ended responses (where subjects must generate the options). These two coefficients are no longer significant when *Familiar* and *Cues* are added (in Model 4), suggesting that they are weak surrogates for these more powerful predictors, both significant in the expanded model, whose adjusted R^2 increases from 0.12 to 0.24.

5. Discussion

Task familiarity and contextual cues explain much of the variation in hypothetical bias across studies, beyond what was possible in previous analyses without these variables. Moreover, they absorbed variance previously attributed to variables potentially associated with experimental artifacts (within-subject designs, use of calibration techniques). In brief, stated and revealed choices converge more when tasks are familiar or contextual cues are available. This general result is nicely illustrated in two specific studies included in the meta-analysis. Both Vossler and Kerkvliet (2003) and Vossler et al. (2003) found little hypothetical bias with public goods with a task that produced the familiarity and context of a referendum.

If our findings are robust, they set clear boundary conditions for stated preference studies: conduct them only when the good and the context can be made familiar and meaningful to respondents. Meeting those conditions should be relatively easy with private goods having natural contextual cues, such as prices in competitive markets or popularity in relevant populations, or with public goods deliberated in political campaigns and policy debates. Novel public and private goods require creating the familiarity needed for stable preferences. Developing the needed communications is relatively straightforward (e.g., Fischhoff, 2011). Indeed it would be straightforward for the field to establish evidentiary standards for demonstrating that research participants have understood tasks as intended. It may be more challenging to identify the context that participants need in order to create meaningful preferences — unless researchers already have some specific context in mind (rather than being interested in the good per se).

In creating the conditions needed for stable preferences, researchers can draw on the extensive behavioral research, perhaps best understood in collaboration with behavioral researchers (Fischhoff et al., 1980, 1999; Keeney and Raiffa, 1993; Lichtenstein and Slovic, 2005; Payne et al., 1999).

6. Conclusion

Economists have traditionally made a strong distinction between revealed and stated preferences. The former are usually held to meet the axiomatic assumptions for preferences, while the latter are often dismissed for their perceived inconsistency. The meta-analyses conducted here find that stated preferences can be consistent if researchers ensure that the goods and the context are familiar — conditions that are met naturally with revealed preference studies of widely marketed goods. Our analyses used a sample of studies collected by Murphy et al. (2005) for other purposes, so as to reduce the chances of a selection bias toward confirming our hypotheses.

Our hypotheses were based on interdisciplinary research into how people articulate (or "construct") preferences when confronted by novel tasks, showing the roles of "task familiarity" and "contextual cues," two variables absent from standard consumer theory which assumes universally applicable preference functions. These variables absorbed much of the variance accounted for by two variables from the standard account: the private character of the good and use of choice-based tasks. These "traditional" predictors seem to have mattered because they were more common in studies involving better understood tasks and more informative contextual cues. In this light, previous analyses of choice consistency confounded the effects of the stated vs. revealed choice tasks and private vs. public goods with the effects of task familiarity and informative contextual cues. Considering these variables can guide us in deciding which existing results are trustworthy and how to create better ones in the future.

Appendix Table. A Goods and choice contexts in the study sample of the meta-analysis[1].

Study[1]	Description of good and choice context	Coding	
		Familiar	Cues
Blumenschein et al. (1997)	Lab sunglasses; economics students	0	0
Blumenschein et al. (2001)	Asthma management program to patients in "scientific study"; general population	0	0
Bohm (1972)	Preview of TV program; general population at broadcast station	1	0
Botelho and Costa Pinto (2002)	Contribution to NGO information campaign on river otters; university students	0	0
Boyce et al. (1989)	Common house plant; experiment with university staff	1	0
Brown et al. (1996)	Contribution to remove roads along Grand Canyon rim; mail survey/collection by university	0	0
Brown and Taylor (2000)	Contribution to NGO environmental program; at university	0	0
Cameron et al. (2002)	Donation to actual green electricity and tree planting program; fixed $6 surcharge, general population	0	1
Carlsson and Martinsson (2001)	Contribution to 3 environmental projects ("rain forest," "Mediterranean," "Baltic Sea"); students prior to lecture	0	0
Champ et al. (1997)	Contribution to remove roads along Grand Canyon rim; mail survey/collection by university	0	0
Champ and Bishop (2001)	Green electricity donation; actual wind power program of electric company, fixed surcharge	0	1
Duffield and Patterson (1992)	Contribution to Montana river mgt.; contingent valuation survey	0	0
Frykblom (1997)	Swedish national atlas, volume "the environment"; prior to lecture	1	0
Frykblom (2000)	Swedish national atlas, volume "the environment"; prior to lecture	1	0
Heberlein and Bishop (1986)	Hunting permits issued by Department of Natural Resources; hunters	1	1
Johannesson et al. (1998)	Displayed box of chocolates; students	1	1
Kealy et al. (1988)	Familiar candy bars (Cadbury); students	1	1
List (2001)	Sports cards (list price $200–250); participants in existing market	0	1
List (2003)	Sports cards (book value $12); participants in existing market	0	1
List and Shogren (1998)	Sports cards (book value $350); participants in existing market	0	1
Loomis et al. (1997)	Wildlife art print displayed; auction to subjects, clerical and administrative staff on campus	1	0
MacMillan et al. (1999)	Contribution to nature reserve; mail survey mimicking regular collection	0	0
Murphy et al. (2002)	Contribution to environmental NGO; students	0	0
Neill et al. (1994)	Painting (list price $75), historical map replica (list price $20); students	0	0
Sinden (1988)	Contribution to soil conservation program; students	0	0
Spencer et al. (1998)	Contribution to pond water monitoring program; students	0	0
Vossler et al. (2003)	Pre-election poll on collectively provided public good, subject to public debate; general population	0	1
Vossler and Kerkvliet (2003)	Pre-election poll on collectively provided public good, subject to public debate; general population	0	1

[1]The complete dataset is available from the authors on request.

References

Arrow, K., Portney, P.R., Leamer, E.E., Radner, R., Schuman, H., Solow, R., 1993. Report of the NOAA Panel on contingent valuation. Federal Register 58, 4601–4644.

Blumenschein, K., Johannesson, M., Yohoyama, K.K., Freeman, P.R., 2001. Hypothetical versus real willingness to pay in the health care sector: results from a field experiment. Journal of Health Economics 20, 441–457.

Carlsson, F., 2010. Design of stated preference surveys: is there more to learn from behavioural economics? Environmental and Resource Economics 46, 167–177.

Clemen, R.T., Reilly, T., 2002. Making Hard Decisions. Duxbury, Boston.

Coupey, E., Irwin, J.R., Payne, J.W., 1998. Product category familiarity and preference construction. Journal of Consumer Research 24, 459–468.

Druckman, J.N., 2001. Using credible advice to overcome framing effects. Journal of Law, Economics & Organization 17, 62–82.

Druckman, J.N., Nelson, K.R., 2003. Framing and deliberation: how citizens' conversations limit elite influence. American Journal of Political Science 47, 729–745.

Ehmke, M.D., Lusk, J.L., List, J.A., 2008. Is hypothetical bias a universal phenomenon? A multinational investigation. Land Economics 84, 489–500.

Fischhoff, B., Slovic, P., Lichtenstein, S., 1980. Knowing what you want: measuring labile values. In: Wallsten, T. (Ed.), Cognitive Processes in Choice and Decision Behavior. Erlbaum, Hillsdale, N.J., pp. 117–141.

Fischhoff, B., 1991. Value elicitation: is there anything in there? American Psychologist 46, 835–847.

Fischhoff, B., Bar-Hillel, M., 1984. Focusing techniques: a shortcut to improving probability judgments? Organizational Behavior and Human Performance 34, 175–191.

Fischhoff, B., Welch, N., Frederick, S., 1999. Construal processes in preference assessment. Journal of Risk and Uncertainty 19, 139–164.

Fischhoff, B., 2011. Applying the science of communication to the communication of science. Climatic Change 108, 701–705.

Fisher, A., 1996. The conceptual underpinnings of the contingent valuation method. In: Bjornstad, D.J., Kahn, J.R. (Eds.), The Contingent Valuation of Environmental Resources. : Methodological Issues and Research Needs. Edward Elgar, Cheltenham, U. K, pp. 19–37.

Flores, N.E., Strong, A., 2007. Cost credibility and the stated preference analysis of public goods. Resource and Energy Economics 29, 195–205.

Hanley, N., Shogren, J.F., 2005. Is cost–benefit analysis anomaly-proof? Environmental and Resource Economics 32, 13–34.

Huffman, W.E., Rousu, M., Shogren, J.F., Tegene, A., 2007. The effects of prior beliefs and learning on consumers' acceptance of genetically modified foods. Journal of Economic Behavior & Organization 63, 93–2006.

Johannesson, M., Liljas, B., Johansson, P., 1998. An experimental comparison of dichotomous choice contingent valuation questions and real purchase decisions. Applied Economics 30, 643–647.

Keeney, R.L., Raiffa, H., 1993. Decisions with Multiple Objectives: Preferences and Tradeoffs. Cambridge University Press, New York.

Lichtenstein, S., Slovic, P. (Eds.), 2005. The Construction of Preference. Cambridge University Press, New York.

Little, J.A., Berrens, R., 2004. Explaining disparities between actual and hypothetical stated values: further investigation using meta-analysis. Economics Bulletin 3 (6), 1–13.

List, J.A., Gallet, C.A., 2001. What experimental protocol influence disparities of hypothetical and actual willingness to pay. Environmental and Resource Economics 20, 241–254.

List, J.A., 2001. Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards, American Economic Review 91, 1498–1507.

List, J.A., 2003. Using random nth price auctions to value non-market goods and services. Journal of Regulatory Economics 23, 193–205.

Loomis, J., 2011. What's to know about hypothetical bias in stated preference valuation studies? Journal of Economic Surveys 25, 262–370.

Lupia, A., 1994. Shortcuts versus encyclopedias — information and voting behavior in California insurance reform elections. American Political Science Review 88, 63–76.

Morgan, M.G., Fischhoff, B., Bostrom, A., Atman, C., 2001. Risk Communication: A Mental Models Approach. Cambridge University Press, New York.

Murphy, J.J., Allen, P.G., Stevens, T.H., Weatherhead, D., 2005. A meta-analysis of hypothetical bias in stated preference valuation. Environmental and Resource Economics 30, 313–325.

Macmillan, D.C., Philip, L., Hanley, N., Alvarez-Farizo, B., 2002. Valuing the non-market benefits of wild goose conservation: a comparison of interview and group-based approaches. Ecological Economics 43, 49–59.

Payne, J.W., Bettman, J.R., Schkade, D.A., 1999. Measuring constructed preferences: towards a building code. Journal of Risk and Uncertainty 19, 243–270.

Robinson, L.A., Hammitt, J.K., 2011. Behavioral economics and the conduct of benefit-cost analysis: towards principles and standards. Journal of Benefit-Cost Analysis 2, Article 5.

Schläpfer, F., Schmitt, M., Roschewitz, A., 2008. Competitive politics, simplified heuristics, and preferences for public goods. Ecological Economics 65, 574–589.

Slembeck, T., Tyran, J.R., 2004. Do institutions promote rationality? An experimental study of the three-door anomaly. Journal of Economic Behavior & Organization 54, 337–350.

Spash, C.L., 2007. Deliberative monetary valuation (DMV): issues in combining economic and political processes to value environmental change. Ecological Economics 63, 690–699.

Vossler, C.A., Kerkvliet, J., 2003. A criterion validity test of the contingent valuation method: comparing hypothetical and actual voting behavior for a public referendum. Journal of Environmental Economics and Management 45, 631–649.

Vossler, C.A., Kerkvliet, J., Polasky, S., Gainutdinova, O., 2003. Externally validating contingent valuation: an open-space survey and referendum in Corvallis, Oregon. Journal of Economic Behavior & Organization 51, 261–277.