Demand revelation and hypothetical bias in discrete choice experiments.*

Verity Watson

December 13, 2011

Work in progress: Please do not cite or quote without the author's permission.

*Health Economics Research Unit, University of Aberdeen, Polwarth Building, Aberdeen AB25 2ZD, United Kingdom.

Abstract

A discrete choice experiment (DCE) is multi-attribute, stated preference method used to value the benefits of health care interventions. However, some researchers remain sceptical about the validity of DCE results because they are based on responses to hypothetical questions. To date studies have investigated the presence of hypothetical bias in DCE responses using field experiments. This study takes a novel approach to this question and investigates hypothetical bias (and demand revelation) in DCE responses using an experimental economics method known as an induced value experiment. Our induced value experiment mirrors the salient features of a DCE task. In the experiment, preferences are induced for a laboratory good with four attributes. The value of each attribute is presented on a value card, which explains how to calculate the total value of the good based on its attributes. Participants can buy the laboratory good during the experiment and at the end it is exchanged for money at the specified rates. At the end of the experiment, participants exchange the goods that they buy for money. To mimic the repeated choice feature of DCEs, participants answer nine DCE questions. In each, participants choose between buying one of two goods or no good (the opt-out). There are 201 participants across 20 sessions. In order to investigate hypothetical bias, sessions are randomly assigned as being real or hypothetical. In hypothetical sessions, participants' earnings from the experiment do not depend on their choices. In real sessions, participants' earnings are determined by the choices they make; otherwise the experiments are identical. If DCEs are demand revealing, in each choice participants should buy the good with the highest net value (value of attributes minus cost of good). We find no evidence that respondents made different choices in the hypothetical and real sessions. However, overall only 65.4% were demand revealing. The proportion of demand revealing choices were affected by choice difficulty, learning, fatigue, and possibly the use of fast and frugal decision rules. Choice difficulty, as measured by the value difference across tokens in the choice sets (from 0.5 to 5). This has a positive and significant effect on making a demand revealing choice. Respondent are significantly more likely to make a demand revealing choice as they progress through the choices (learning) but this probability decreases (fatigue). We use time taken to complete the choice as a proxy for effort. We find that faster respondents are significantly less likely to make a demand revealing choice. While, we find no evidence of hypothetical bias, our results question the incentive compatibility of DCEs. Choice difficulty affects response: In the most difficult choices respondents do no better than choosing at random. Future work will analyse the data for evidence of decision heuristics.

Introduction

UK government decision making uses cost benefit analysis (CBA) to compare the costs, benefits, and risks of policies and proposals. CBA requires that costs and benefits are measured in monetary terms. For many costs and benefits, such as the environmental benefits of reducing greenhouse gas emissions or the health benefits of new drug treatments for cancer, no market price exists. If a value for these benefits is not included in government decisions, then public spending will not reflect all that society believes to be important. Stated preference techniques are used to infer values for these non-market costs and benefits (Bateman et al , 2002).

Stated preference techniques use questionnaires to infer monetary valuations for non-market goods by asking participants a variant of the following question: What is the maximum you are willing to pay for good y? Two standard methods are Contingent Valuation (CV) and Discrete Choice Experiments (DCEs). Opponents of stated preference techniques challenge whether hypothetical responses are accurate representations of 'real-world' behaviour (Diamond and Hausman , 1994). This has lead to a wealth of methodological work investigating issues in the development and application of CV and DCEs, including studies investigating the correspondence between hypothetical and real valuations. With respect to CV these studies typically report higher valuations when questions asked are hypothetical rather than real: this finding is known as hypothetical bias. (See Liljas and Blumenschein (2000) and List and Gallet (2001) for a review.)

Many tests of hypothetical bias compare 'homegrown values' for real goods i.e. for goods or services that are available in reality (such as organic beef steaks, household electrical goods, contributions to charitable organisations). These studies may be conducted in realistic field settings but are also conducted in more artificial laboratory settings. Most studies do not test the relationship between stated payments, hypothetical and real, and true values, because in field studies true values are unobservable. Studies of hypothetical values assume that real values are an unbiased measure of true values. However this is only true if incentive compatible revelation mechanisms are used to elicit the real payments (Mitani and Flores , 2009). More recently, research has split the study of hypothetical bias into two parts [1]: that concerned with the elicitation mechanism and that concerned with valuation formation. Polomé (2003) and List et al (2006) argue that respondents social networks may influence their value formation, and lead to an overstatement of value for the goods

[1] This dichotomy is also emerging in studies of other stated preference response biases see for example

provided in CV studies because in general the goods provided are socially desirable, and individuals take their cues from the behaviour of others in their social network. This occurs in the hypothetical questions is because there is no penalty to tell the truth. An experimental economics method, induced value experiments, allow the research to focus on the elicitation mechanism by removing the valuation formation phase of the CV study (Smith , 1976). This provides the advantage that because the researcher induces the individual's value for the good the exact deviation between the actual value and state value is known in all settings.

This study uses an induced value experiment to investigate: first, if responses to DCEs are demand revealing, and second, if responses differ when choices are hypothetical or real. To implement this, we induce in participants preferences for a private multi-attribute laboratory good: a token. We find no statistical evidence of hypothetical bias. However, we find that fewer than 70% of respondents make demand revealing choices. In a probit analysis of response deviations we find that task difficultly, task completion speed influence the probability of errors, and that respondents learn as they complete the task.

Background

Incentive compatibility is a theoretical construct that implies that a the mechanism provides respondents with the incentive to reveal their preferences truthfully. In other words, telling the truth is the dominant strategy. Incentive compatibility assumes that respondents have perfect information, know their preferences with certainty and omniscient optimisers. The Gibbard-Sattherwaite theorem provides a theoretical framework for the analysis of voting mechanisms and is transferable to alternative stated preference elicitation methods. The theorem indicates that a single binary choice format is the only non-dictatorial mechanism that is incentive compatible (**??**). Empirically, one can test if responses to a mechanism are demand revealing behaviour.

In the context of a stated preference study Carson and Groves (2007)provide a list of requirements for stated preference studies to be incentive compatible: the survey must be consequential; the task must be transparent; the policy must be credible; the scenario must be plausible; the issue must be relevant to the respondent; information about how the responses will be used must be included; and respondents must believe that their responses will influence the outcome. A close inspection of these casts doubt on the incentive compatibility of many existing fields studies. All surveys are inconsequential in that respondents will not be held to their answers. However, Carson and Groves (2007) allow a less strict interpretation which is that respondents must perceive that their responses will influence the findings of the study and these finding will be used to inform policy. However, DCEs ask respondents to make a series of choices, typically between 8 and 32, but never explain how these responses will be used, i.e. how they are aggregated across choices and across respondents to arrive at a set of results and conclusions. This a far removed from a simple majority voting rule on a binary choice (for example: dichotomous choice contingent valuation). required for incentive capability clearly indicates that the omission or vagueness of this information casts serious doubts on the validity of the results of previous stated preference studies. **?** in a study valuing a public good find no evidence to suggest that in the absence of a provision rule that respondents will assume the provision rule is a plurality voting rule. It is more reasonable for respondents to assume that the choices over all choices situations would be aggregated in some way. Which leads to more complicated strategic behaviour. For instance finding the 'best' alternative across all choices voting for this and opting out of all other markets.

To test if a mechanism has demand revealing properties experimental economists use induced value theory. Induced value theory is based on the idea that by using (financial) rewards the experimenter can create preferences in subjects over goods. These goods are usually tokens, or other objects, that have negligible value outside the experiment. By doing this, the researcher knows subjects' preferences over these 'laboratory' goods, and can observe if subjects make utility maximising decisions. Several studies have used induced value experiments to explore demand revelation and hypothetical bias in contingent valuation type questions. Taylor et al (2001) use an induced value experiment to test if the referendum voting mechanism is demand revealing and to test for hypothetical bias. The referendum was a secret ballot with a majority rule. The language in the hypothetical and real referenda were almost identical, the hypothetical referendum asked subjects to "...put yourself in the following situation. Suppose you were all going to vote in a referendum..." and then subjunctive language was used throughout the instructions such as "each of you would pay" — and each person "would" receive. The authors hypothesize that there should be no "mis-votes" in the real referendum as it is theoretically incentive compatible. 21 of 130 (16.1%) of votes were mis-votes in the real referendum and 24 out of 143 (16.8%) of votes were mis-votes in the hypothetical referendum. There is no significant difference between referenda. A probit model of mis-voting indicates that a $1 increase in earnings reduces the probability of mis-voting by 5%.

Vossler and McKee (2006) use an induced value experiments to compare demand revelation and hypothetical bias in dichotomous choice (DC) and payment card contingent valuation formats. The provision rule for the DC format is a majority voting rule and in the real payment setting the PC is a version of the Random Price Voting Mechanism of Messer et al (2006). Decisions are converted into yes and no votes for a randomly determined cost amount and then a majority voting rule is used determine whether the proposal passes or not. The authors offer three reasons why an individual might deviate from the optimal responses: First, the respondent was not able to decipher what the optimal responses was, because of an unfamiliarity with the type of question or the incorrect belief that they are better off to act strategically. Second, respondents have other regarding preferences. Third, respondents are not risk neutral utility maximisers. The authors find no difference in deviations between hypothetical and real treatments and that the PC has more deviations than DC. The authors use a probit to predict the probability of deviations and find that the probability of a deviation decreases by 2.6-7.1% for every increase of $1 in value-cost spread, and suggest that deviations are related to the difficulty of the decision task.

Mitani and Flores (2009) state that most induced value tests of demand revelation and hypothetical bias in the environmental economics literature use incentive compatible mechanisms (binary referenda) and one-shot decision tasks. Two exceptions are Mitani and Flores (2009) and Collins and Vossler (2009) who consider a threshold public goods game and a discrete choice experiment, respectively. Mitani and Flores (2009) test a threshold public goods game with continuous contributions, both real and hypothetical, along with a money back guarantee. Subjects were members of the general public who live in Tokyo. The authors create and index of demand revelation that identifies understatement, truthful statement and overstatement and find significant difference between real and hypo treatment. Real treatment induces understatement and hypothetical induced truthful statement. However, mechanism is not incentive compatible, and the real behaviour is inline with the theoretical predictions of behaviour in this type of mechanism.

Collins and Vossler (2009) use an induced value experiment to investigate demand revelation in DCEs when valuing public goods. The authors compare dichotomous choice (DC) and trichotomous choice (TC) DCEs and three difference provision rules: plurality vote[2]; random selection whereby the probability than an option is implemented is proportional to the number of votes it receives; random selection with regulator whereby a regulator has a preferred option [3]. The laboratory good has

[2]The DC question in this setting is incentive compatible

[3]The authors claim that this mimics the advisory role that preference studies play in decision making. The regulator has a number of votes equal to number of respondents and the regulators choice is predetermined (before the experiment)

three attributes, one of which is cost. Collins and Vossler (2009) recruited 196 undergraduate students from University of Tennessee. The authors find fewer deviations from optimal choice with TC compared to DC and more deviations with the random voting rules. Collins and Vossler (2009) find the probability of deviation decreases with the difference in value between the alternatives offered in the DCE.

Methods

In this study, we test demand revelation and hypothetical bias in DCE responses. Our study differs from that of Collins and Vossler (2009) in that we use a private good this implies that mechanism is theoretically incentive compatible and we rule out other regarding preferences. To implement the induced value experiments, preferences are induced for a multiattribute laboratory good (token) with respondents offered the opportunity to purchase tokens at an advertised cost. The good in the experiment was a token with 4 attributes each with three levels: its colour (red, yellow, blue); its shape (circle, triangle, square); its size (small, medium, large); and its cost (Table 1). The reward for each token was dependent on its attributes. Respondents were informed of the rule used to calculate their total reward based on the token's attributes (a linear additive function of the attributes' rewards). All participants had the same induced values. These were certain and did not change during the experiment and these were presented to respondents in the instruction sheet.

The choice questions were designed to resemble typical discrete choice experiment questions used to value health and health care. Accordingly, in each choice set respondents were asked to choose between two alternative tokens and no purchase (an opt-out). The tokens to include in the choice sets were chosen using a fractional factorial design. To further replicate the DCE procedure, respondents were asked to complete a series of choices. Each participant answered the same nine choice questions, however the question order was randomised across respondents.

An experimental economics norm is to pay subjects a meaningful amount determined by their performance in the task. DCEs, however, are a hypothetical task. The experiment had four treatments: the choices were hypothetical or real and the payoff differences were narrow and wide. In the hypothetical treatment, participants were paid \$10 no matter what they chose in the experiment. In the real treatment, respondents earning from the experiment would depend on the choices that they made. Specifically, at the beginning of the experiment participants in the real treatment were told that they had an account with \$4 in it and that they could use this money to buy tokens that were offered for sale in the experiment. To ensure that participants' behaviour was not constrained by their ability to pay, all tokens cost less than \$4. To ensure that all choices were treated separately and there were no threshold wealth or strategic effects, respondents were told that their earnings depended on only one of their choices that was chosen at random at the end of the experiment. To ensure that participants in the real treatment would not earn \$0, participants were given \$2 for showing up on time and participating.

In the *narrow* and *wide* treatments the potential difference in value between two tokens in each choice set varied. To obtain this the value of each attribute level differed across each treatment (Table 1). In the *narrow* treatment, the difference between the lowest and highest value level with in each attribute was smaller than in the *wide* treatment. Consequently, the payoff difference in the *narrow* is likely to be less than in the wide treatment. The payoff differences across alternative within each choice set are presented in table 2 column 4 [4].

Participants were recruited from students at the University of Aberdeen. Participants were recruited using the Exlab software [5]. Experiments were conducted ar the Scottish Experimental Economics Laboratory.The Laboratory is a large room with 20 personal computers. Each computer is numbered and privacy screens prevent participants from seeing the choices of other participants.

Before the experiment began each respondent receive a consent form, a set of instructions for the experiment, and a payment form [6]. Respondents read and signed the consent form which was collected before the experiment started. Following this the researcher read the instruction sheet aloud to the group and answered any questions that participants had. The instruction sheets for each experiment were identical, except subject language was used in the hypothetical treatment: Respondents were asked to "Put yourself in a situation where your account balance at the end of the experiment would depend on the choice you made...", whereas in the real treatment

[4] The choice set number refers to a unique choice from the experimental design. These choices were randomised across respondents. This choice number 1 would be presented first to some respondents, second to others etc

[5] When students log on to University computers they see a virtual notice board. Occasionally, this notice board include adverts encouraging students to register to participate in economics experiments. Students who are registered with Exlab receive email notifying them of new experiments that they can participate in. While many participants had participated in other experiments, no participant had participated in any similar experiments at University of Aberdeen

[6] To re-enforce that hypothetical treatment earnings were unaffected by the choices, the payment form was pre-completed

respondents were told "Your account balance at the end of the experiment will depend on the choice you made..." At the end of the experiment respondents were called by seat number and paid their earnings in private.

Random utility theory [22] assumes that when faced with a choice between a set of alternatives respondents will choose the alternative which maximises utility. In this case respondents will choose the token providing the highest profit, and if both tokens provide a loss the respondents will choose the 'do nothing' alternative. For each choice, it is possible to identify the payoff/utility maximising alternative; responses are compared to this prediction and the proportion of 'correct' choices calculated. Test if this is significantly different from random using a test of proportions.

Respondents make a series of nine choices (within the *narrow* and *wide* treatments all respondents answered the same nine choices, but the order of the choices was randomised. We calculate for each respondent the number of correct choices that they make and test if this is significantly better than at random using a binomial test. In a series of 9 choices having 6 or more 'correct' choices is significantly better than random at the 5% significance level.

Respondents might fail to make the correct choice for a number of reasons. Respondents might make optimisation errors and pick the wrong alternative. We expect that this is more likely when choices are difficult. In other words, we expect respondents are less likely to make 'correct' choices when payoff differences are small. Respondents may not understand the question/task at the beginning of the experiment but learn about the institutional setting as they complete the experiment. In this case, we would expect that respondents would improve as they proceed. Respondents may become bored or fatigued with the repetitive task and expend less effort to select the correct alternative. In this case, we would expect respondents would be less likely to select the 'correct' alternative in later choices. Respondents in the hypothetical treatment may be unwilling to expend the effort to select the 'correct' alternative when this has no impact on their pay-off. In the case, we would expect hypothetical respondents to be less likely to select the 'correct' alternative. Respondents may be generally unwilling to make the effort the correct choice and just pick any alternative to complete the task as quickly as possible. In this case, we would expect respondents who complete the choice quickly to be less likely to select the 'correct' alternative. We test these different reasons using a random effect probit model where the dependent variable is whether or not the individual make the 'correct' choice and the dependent variables are pay-off difference between the two tokens, the position of the choice within the experiment (we include a squared term to allow for learning diminishing and possible also fatigue within the question sequence), hypothetical treatment, time taken to answer the question (Time is measured in seconds, we plot the distribution of this across all respondents and all choices and construct a dummy variable with three levels: quicker than the 25th percentile, between the 25th and 75th percentile, and slower than the 75th percentile.

Collins and Vossler (2009) found that the status quo, no token alternative, was selected more than expected and that determinants of selecting the wrong token differed depending on whether respondents chose the wrong token or chose the no token alternative when at least one token in the choice set offered a positive payoff. We test for this by estimating a multinomial probit where the dependent variable indicates is respondents chose the 'correct' token, the 'wrong' token or 'no token' when a positive payoff token was included in the choices task [7]

Results

200 students (both undergraduate and taught postgraduate)from the University of Aberdeen participated in the study in May 2009 and June 2010. The number of participants taking part in a session ranged from 8-18. Respondents were approximately evenly split across treatments (Table 2).

In the *narrow* treatment the proportion of correct responses varied across choices in both the hypothetical and real treatments (Table 2 panel A, columns 5 and 6). In both cases, the choice C had the lowest proportion of 'correct' choices 30.2% and 34% in the real and hypothetical treatments respectively. In both cases, this is no better than choosing at random. Choice C has the smallest pay-off difference. Nevertheless, in choice F, which also has a pay-off difference of £0.5, 60.4% and 65.9% of respondent made 'correct' choice across the real and hypothetical treatments respectively. Apart from choice C all other choices had better than random proportions of 'correct' choices.

In the *wide* treatment the proportion of correct responses varied across choices in both the hypothetical and real treatments(Table 2 panel B, columns 5 and 6). In both cases, the choice C had the lowest proportion of 'correct' choices 27.7% and 14.9% in the real and hypothetical treatments respectively. In the real treatment this is no better than choosing at random and in the hypothetical treatment this is significantly different (and worse) than choosing at random. This is despite choice

[7] In both the *narrow* and *wide* treatments each choice set had at least one token with a positive payoff. Thus, selecting the no token alternative was never best response.

C having a payoff difference of £4.5. However, in the *wide* treatment, responses to choice A were significantly worse than choosing at random in both the real (9.3%) and hypothetical (14.9%) treatments, and responses to choice B were no better than choosing at random. The payoff difference in choice A is £1, however the payoff difference in choice B is £7. Overall, the proportion of correct responses are lower in the *wide* treatment compared to the *narrow* treatment.

In the *narrow* treatment, the number of correct choices made by a participants ranged from 0 to 9 (Table 3). In both the hypothetical and the real treatments the highest number of respondents made 8 correct choices. In the *wide* treatment, the number of correct choices made by a participants ranged from 0 to 9 (Table 3). In both the hypothetical and the real treatments the highest number of respondents made 6 correct choices.

Table 4 presents the results of the random effects probit model exploring what influences the probability that respondents will make a pay-off maximising choice. In the *narrow* treatment, respondents are less likely to make a payoff maximising choice in the real treatment (REAL), and when they answer the choice quickly. The position of the choice in sequence has a significant influence on the probability of respondents making a payoff maximising choice, as respondents progress through the experiment the probability of making a pay-off maximising choice increases (CHOICE NUMBER) but at a decreasing rate (Choice number^2). In the *wide* treatment, we do not find evidence of a significant difference across real and hypothetical treatments. Respondents are more likely to make a pay-off maximising choice when the larger the pay-off difference and in later choices. Similar to the narrow treatment, the probability of a payoff maximising choice increases at a decreasing rate as respondents progress through the experiment and respondents who make their choices quickly are less likely to make a payoff maximising choice.

Table 5 presents the results of the multinomial probit model that differentiates between choosing the wrong token and choosing the no token alternative when at least one in the choice set offered a positive pay-off. Table 5 panel A presents the results for a wrong token error. In the *narrow* treatment, respondents were more likely to make a narrow token error if they completed the choices quickly or slowly. Respondents were less likely to make a wrong token error as the progressed through the nine choices, but the effect of experience was diminishing. We find no evidence that the choices being hypothetical or real influence probability of making a wrong token error. We also find no evidence that the pay-off difference between tokens influenced the probability of a wrong token error. In the *wide* treatment, respondents were less likely to make a wrong token error the larger the pay-off difference between tokens and as they progressed through the experiment. Respondents who made their choices quickly were more likely to make a wrong token error. Again, we find no evidence that the choices being hypothetical or real influence probability of making a wrong token error.

Table 5 panel B presents the results for a no token error. In the *narrow* treatment, respondents were more likely to make a no token error in the real treatment and when they made their choices quickly. Respondents were less likely to make a no token error if they made their choices slowly. We find no evidence that the pay-off difference or the position within the choice sequence influences the probability that respondents make a no token error. In the *wide* treatment, we find that respondents were more likely to make a no token error when they made their choices quickly and less likely to make a no token error if they made their choices slowly. Respondents were also less likely to make a no token error the larger the pay-off difference across token and as they progressed through the experiment.

Discussion

In this study, we find little statistical evidence of hypothetical bias in responses to induced value experiments. However, worryingly we also find that less than 70% of respondents make demand revealing choices. The proportion of demand revealing choices differs across each choice set and in some cases is lower than would be expected if respondents were answering at random. Collins and Vossler (2009) find that more respondents than expected select the status quo alternative. We also find this, despite each choice set containing at least one positive pay-off token. This pattern of choice is present in both the narrow and wide treatments despite the greater potential payoff gains available from tokens in the wide payoff treatment. This may be evidence of a simplifying decision heuristic and warrants further investigation.

That we find no difference between hypothetical and real choices adds further evidence to the assertion that there is a value formation problem in field studies. Where respondents may be influenced by social peer pressure, norms or may spend relatively less time and draw on less information for a hypothetical payment decision.

We find robust evidence across treatments of learning. Vossler and McKee (2006) state that learning is ruled out in their study because results were not announced until after all rounds were completed. We find that learning is possible without feedback. This learning must be interpreted as learning about the mechanism not value formation. This has implications for field study DCEs, where both learning about the choice mechanism and value formation will be occurring simultaneously. Field experiments which find differences between choices early in the questionnaire and late cannot easily separate mechanism learning and value formation.

We find evidence that task difficultly influences the probability of deviations. Furthermore, the choice sets with the highest propotions of wrong answers are also those choice sets where the first token had a high positive payoff. These results indicate that respondents are likely to use decision heuristics to complete the task. Future work will investigate which heuristics best explain these results and test if these are present in field experiment data.

Acknowledgements

Financial support from the University of Aberdeen and Chief Scientist Office of the Scottish Government Health Directorate are acknowledged. The usual disclaimer applies.

References

Bateman, I.J., Carson, R.T., Day, B., Hanemann, M., Hanley, N., Hett, T., Jones-lee, M., Loomes, G., Mourato, S., Özdemiroglu, E., Pearce, D.W., Sugden, R., and Swanson, J. (2002). *Economic Evaluation with stated preference techniques: A manual.* Edward Elgar, Cheltenham, UK.

Carson, R.T. and T. Groves (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics*, **37**, 181–210.

Collins, X. and Vossler, C. (2009). Incentive compatibility tests of choice experiment value elicitation questions. *Journal of Environmental Economics and Management*, **58**,226–235.

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

Liljas, B., and Blumenschein, K., (2000). On hypothetical bias and calibration in cost-benefit studies. *Health Policy*, **52**, 53–70.

List, J.A., and Gallet, C.A., (2001). What experimental protocol influence disparities between actual and hypothetical stated values? *Environmental and Resource Economics*, **20**, 241–54.

List, J.A., Sinha,P. and Taylor, M.H. (2006) Using Choice Experiments to Value Non-Market Goods and Services: Evidence from Field Experiments. *The B.E. Journal of Economic Analysis and Policy*, **6**.

Mitani, and Flores, N., (2009). Demand revelation, hypothetical bias, and threshold pubic goods provision. *Environmental and Resource Economics*, , – .

Polomé, X. (2003) Experimental evidence on deliberate misrepresentation in referendum contingent valuation. *Journal of Economic Behaviour and Organisation*, **52**, 387–401.

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

Taylor, L.O., and (2001). Induced value tests of the referendum voting mechanism. *Economic Letters*, , – .

Vossler, C., and McKee, (2006). Induced value tests of contingent valuation elicitation mechanisms. *Environmental and Resource Economics*, **35**, 137–168.

Table 1: Induced values

Token attributes	Level	Narrow (£)	Wide (£)
Size	Small	1.00	0.50
	Medium	2.00	2.50
	Large	3.00	4.00
Colour	Red	0.50	1.00
	Yellow	1.00	1.50
	Blue	1.50	2.00
Shape	Circle	1.00	1.50
	Triangle	1.50	3.00
	Square	2.00	6.00

Table 2: Propotion of correct responses across choices and treatments

Choice	Value A	Value B	Payoff Diff	Treatment - Narrow	
				Real N=53	Hypothetical N=47
				N (%)	N (%)
A	2.50	1.5	1	39 (73.6)*	31 (65.9)*
B	0	4	4	38 (71.6)*	37 (78.7)*
C	2	2.5	0.5	16 (30.2)	16 (34.0)
D	-1	4	5	40 (75.5)*	35 (74.5)*
E	2.5	1.5	1	32 (60.4)*	32 (68.1)*
F	0.5	1	0.5	32 (60.4)*	31 (65.9)*
G	3	1.5	1.5	42 (79.2)*	33 (70.2)*
H	1.5	-1	2.5	33 (62.3)*	34 (72.3)*
I	3	0.5	2.5	37 (69.8)*	39 (82.9)*
Overall (%)				64.78	66.19
				Treatment - Wide	
				Real N=54	Hypothetical N=47
				N (%)	N (%)
A	5.5	6.5	1	5 (9.3)*	7 (14.9)*
B	2.5	9.5	7	18 (33.3)	18 (38.3)
C	3.5	8	4.5	15 (27.7)	7 (14.9)*
D	-0.5	7	7.5	46 (85.2)*	36 (76.5)*
E	8	3	5	40 (74.1)*	34 (72.3)*
F	4.5	3	1.5	40 (74.1)*	34 (72.3)*
G	6	4	2	44 (81.5)*	35 (74.4)*
H	3	0.5	2.5	43 (79.6)*	32 (68.1)*
I	8	1	7	40 (74.1)*	35 (74.4)*
Overall (%)				59.87	56.26

Table 3: Number of correct choices made by respondents across treatments

	Narrow		Wide	
Number of correct choices	Real	Hypothetical	Real	Hypothetical
0	1	0	1	3
1	3	0	0	0
2	3	2	1	2
3	7	6	1	3
4	0	4	9	5
5	3	3	10	6
6	7	3	25	21
7	12	7	6	7
8	15	16	1	0
9	2	2	0	0
Total	53	43	54	47

Table 4: Determinants of payoff maximising choice (random effects probit with robust standard errors)

	Narrow		Wide	
	Coefficient	P	Coefficient	P
Payoff Difference	-0.019		0.042	**
Real treatment	-0.172	*	0.057	
Choice number	0.199	***	0.164	**
Choice number^2	-0.016	**	-0.013	*
Time taken less than 25 percentile	-0.778	***	-0.332	***
Time taken more than 75 percentile	-0.007		0.084	
Constant	0.250		-0.358	**
Number of obs.	900		909	
Log likelihood	-544.570		-604.878	

Table 5: Determinants of wrong token and no token errors (multinomial probit with robust standard errors)

	Narrow		Wide	
	Coefficient	P	Coefficient	P
	(Wrong token error)			
Real treatment	-0.102	0.479	-0.070	0.576
Pay off Difference	0.020	0.487	-0.099	0.000
Choice number	-0.353	0.004	-0.251	0.023
Choice number^2	0.028	0.021	0.019	0.075
Time taken less than 25 percentile	1.297	0.000	0.473	0.002
Time taken more than 75 percentile	0.349	0.055	-0.069	0.652
constant	-0.656	0.020	0.094	0.707
	(Neither token error)			
Real treatment	0.572	0.000	-0.107	0.543
Payoff Difference	0.029	0.310	-0.043	0.019
Choice number	-0.165	0.193	-0.339	0.027
Choice number^2	0.013	0.272	0.023	0.120
Time taken less than 25 percentile	0.752	0.000	0.529	0.009
Time taken more than 75 percentile	-0.342	0.071	-0.966	0.001
constant	-1.198	0.000	-0.515	0.118
Number of choice sets	900		909	
chi 2(12)	99.33		89.67	
Log likelihood	-742.579		-744.52943	