An investigation of individual preferences: consistency across incentives and stability over time

Emmanouil Mentzakis and Jingjing Zhang

April 2012

An investigation of individual preferences:
consistency across incentives and stability over time

Emmanouil Mentzakis[a] and Jingjing Zhang[b*]

[a] Health Economics Group, Faculty of Medicine and Health Sciences, University of East Anglia, Norwich, NR4 7TJ, UK

[b] Chair for Organizational Design, Department of Economics, University of Zurich, Blümlisalpstrasse 10, Zurich, CH-8006, Switzerland

Abstract

This study compares individual preferences across incentives (i.e., hypothetical vs. real incentives) and over time (i.e. elicitation at two different points in time) in a choice experiment involving charitable donating decisions. We provide evidence of hypothetical bias but little evidence of instability of individual giving. There is significant heterogeneity in individual preferences, with real incentives either dampening or pronouncing the observed donating behaviour. Neither hypothetical bias nor instability is observed when we examine the propensity of individuals to make internally consistent decisions over identical choices.

An investigation of individual preferences:
consistency across incentives and stability over time

1. Introduction

Economists have used discrete choice experiments (DCEs) to elicit individual preferences for a number of non-market goods and services, including public transportation systems, environmental policy, health care services and marketing (Ben-Akiva and Lerman 1985; Truong and Hensher, 1985; Hanley et al., 1998; Swait and Adamowicz, 2001; Ryan et al., 2008). DCEs are founded upon consumer (Lancaster, 1966) and random utility theory (McFadden, 1974) and postulate that utility is not derived from the consumption of a good per se, but from the characteristics of this good. Individuals are presented with alternative scenarios and asked to indicate their preferred one. Each alternative is described by a defined set of attributes. The attributes are common across the alternatives but the values they take on can differ. Using principles of experimental design to systematically vary the levels of the attributes across alternatives, the trade-offs between attributes and their relative importance in people's decision making can be elicited.

Although DCEs have firm roots theoretically, some skepticism has been generated in relation to some methodological aspects of the design and implementation of choice experiments. Commonly, DCEs evaluate individuals' preferences based on responses to hypothetical questions with no direct consequences for respondents or anyone else. However, such lack of saliency in the experiment might be a cause for concern as responses to hypothetical scenarios may overestimate real preferences and willingness to pay values

(WTPs). [1] To-date the results from the literature are mixed, with one study (Carlsson and Martinsson, 2001) reporting no bias, while others find either partial (i.e. difference in WTPs but not underlying preferences) (Cameron et al., 2002; , List et al., 2006) or full bias (Taylor et al., 2010; Broadbent et al., 2008; Ready et al., 2010).

A second aspect of DCEs that has come under criticism relates to the assumption of stability of preferences in repeated elicitation. The underlying assumption is that individuals can consistently reveal their well-defined preferences when asked repeatedly without significant variations (Stigler and Becker, 1977). This is not to question individuals' time preference or preference change as a result of following external cues but to test whether individuals facing the same choice at different points in time will express the same preference. Despite its importance, this issue has received only minimum attention, with a couple of studies (in hypothetical environments) reporting no significant biases (San Miguel et al., 2002; Ryan and San Miguel, 2003).

This paper examines both the existence of hypothetical bias and the stability of preferences and two main features distinguish it from previous literature. First, our novel experimental design with both within- and between-subject treatments allows us to test for hypothetical bias both across incentives and over time, as well as, to test the stability of individual preferences. Specifically, the same experiment was conducted in two different environments using a *between-subject* design. The first was a hypothetical environment in which subjects' decisions had

---

[1] Hypothetical bias problems have been raised in related non-market valuation instruments, namely contingent valuation, for a while now (Cummings *et al.*, 1995; List, 2001; Blumenschein *et al.*, 2008). Recent studies reviewing the literature report that hypothetical preferences overstate real preferences by 2.6 (Murphy et al. 2005) to 3 times (List and Gallet, 2001; Little and Berrens, 2004).

no financial consequences and the second was a real environment in which subjects were given monetary incentives (in Canadian dollars) and asked to participate in the same exercise. The difference in choices between the two environments reflects hypothetical bias. The same treatments (both hypothetical and real) were also repeated, with the same subjects, a week later to test for stability of preferences and potential changes in the direction and magnitude of hypothetical bias over time (i.e. a *within-subject* design). Second, we measure the degree of internal consistency of preferences by having a choice set repeated three times in more or less equally spaced intervals within each treatment. We, subsequently, test for differences in the degree of internal consistency across treatments.

Briefly, our analysis provides several important results. We find individual preferences and WTPs to be relatively stable over time but there are significant differences between hypothetical and real incentive treatments. Neither hypothetical bias nor instability is observed when measuring the degree of internal consistency of preferences. Section 2, discusses the experimental settings, tests and econometric analysis. Section 3 presents the results, while Section 4 discusses the findings and concludes.

2. Methods

2.1 Experimental treatments and procedures

In order to test for the consistency of preferences across incentives and over time we employ a 2x2 design with hypothetical versus real environment defining one dimension and first versus second time play defining the other (see Graph 1). Four treatments H1 (hypothetical first time play), R1 (real first time play), H2

(hypothetical second time play), and R2 (real second time play) were conducted. A total of 71 subjects from a Canadian university participated in three H1 and three R1 sessions. A week later 61 out of 71 subjects returned and participated in another three H2 and three R2 sessions. There were 9 to 12 subjects in each of the twelve sessions. Subjects who participated in H1 (R1) sessions also participated in H2 (R2) sessions. It was clearly stated in the recruiting email that subjects were required to return exactly one week later and participate again at the same time if they chose to participate in the experiment [2], which was instrumental for the within subjects treatment aspect of the design. However, to avoid any potential strategic answering (i.e. striving to be consistent to what they answered last time), participants were not told they would be playing the exact same experiment in the repeated session. Despite the fact that subjects would be able to recognize the similarities in the format and context of the two instances, strategic answering should not be a problem for the experiment as it would be virtually impossible to remember prior choices.

The computerized sessions were run using z-Tree (Fischbacher, 2007). Each session lasted about 15 minutes. At the beginning of each session subjects were given audio instructions, completed a practice round, and then proceeded to complete 24 charitable donation decisions. In each of these 24 decisions subjects were given a $12 endowment and asked whether they were willing to donate pre-specified amounts to a health charity, a social service charity or to not donate at all (for more details see Section 2.2). Subjects were told that one of their choice decisions would be randomly selected at the end of the experiment and the donation specified for their chosen alternative would be deducted from their endowments and donated to the charity, while they would receive the remaining

---

[2] For the second-time session each subject was seated in the same computer as the first time session.

4

amount of their endowment. In the hypothetical treatments, all decisions involving money were hypothetical and no donations were made. In the real treatments, subjects earned the portion of the endowments that were not donated. To ensure the credibility of donations, subjects were informed that after each session one of the experimenters would carry out the donation online in front of them. All subjects were paid a flat fee of $12 ($6 for each of the two sessions) for participation.

2.2 Choice experimental context, attributes and levels

The choice experiment involved charitable donation decisions. Such environment was deemed realistic enough for a choice experiment and complex enough to approximate common stated preferences experiments. Most importantly, it was an environment that could facilitate both hypothetical and real treatments with no change in the design or presentation apart from the saliency of payments. Each alternative was described by three characteristics: the type of the charity, the matching policy and the donation amount.

Type of Charity: Charities were identified as either a health or a social service charity according to their official mission and purpose. Health charities served causes directly linked to a health issue, while Social Service charities targeted a social. [3] Additionally, subjects were informed that all charities provided nationwide services and were comparable in terms of the percentage of funding

---

[3] Health charities: Heart and Stroke Foundation of Canada, Muscular Dystrophy Canada. Social service charities: Opportunity International Canada and Boys and Girls Clubs of Canada.

received from the government (45%-55%) and the percentage of funds spent on administrative/managerial purposes (40%-60%). [4]

Matching Policy: The matching policy indicated the percentage by which subjects' donations would be matched by the experimenters and took three levels: 0%, 25% and 50%.

Amount Donated: The donation amount was the dollar amount that subjects were asked to donate within a choice scenario. This could be $4, $8 or $12. These values were deemed to be large enough to create saliency and to elicit truthful behavior for the individuals.

An example of the exact language and instructions of the experiment is given in Appendix 1.

2.3 DCE experimental design

Absorbing the type of charity as a label in the choice experiment allows for the estimation of type-of-charity specific effects (i.e. separate parameters estimated for health and social charities). A D-efficient fractional factorial labelled design with two three-level attributes was generated ($L^{MA} = 3^{2*2}$), which consisted of 21 pair-wise choices. To ensure elicitation of realistic behaviors an opt-out alternative (i.e., an individual does not donate and keeps the full endowment) was included in the choice sets (Hensher et al., 2005). An extra choice set was also added and repeated three times throughout the experiment, in equally spaced intervals, to allow measurement of the degree of internal validity with which

---

[4] To ensure well-defined preferences (or at least well-informed choices), participants in the current study were given information on the charities' background characteristics, purpose and official mission, which information was accessible at all times during the experiment.

subjects responded[5], bringing the total number of choice sets subjects saw to 24. All aspects of the experimental design were performed using SAS 9.1.3 built-in capabilities (Kuhfeld, 2005).

2.4 Testing consistency and stability

The three main research questions are:

1. Does hypothetical bias exist in individual decisions in choice experiments? If yes, what is the direction and magnitude of the bias?
2. Can individual preferences be consistently revealed in repeated elicitation settings (i.e. stability of the preference) and if no, do they also result in changes in the direction and magnitude of hypothetical bias?
3. Do individuals make the same decision over identical choice sets in a DCE within the same treatment? Does the degree of internal validity differ across incentives and over time?

To address these questions five tests are conducted.

Test 1 - The same model specification is run for all four treatments and equality of underlying preferences is tested with LR-tests (the econometric model estimated are discussed in Section 2.5; results reported in Tables 1 and 2).

Test 2 - Following estimation, WTP values are calculated and comparisons across treatments are drawn using t-tests (results reported in Tables 1 and 2).

---

[5] Choice sets 4, 12 and 21 were identical. These three choice sets were added to the 21 of the DCE experimental design

Test 3 – Having a thrice repeated choice set, internal validity of individual preferences in completing a DCE within each treatment, was measured by the number of times the individual did, in fact, choose the same alternative in all three instances (an individual with well-defined, consistent preferences would be expected to prefer the same alternative in all three repetitions). The internal validity indicator takes the value of zero if an individual's choices match in none of the three choice sets, the value of one if two out of the three choices match and the value of two if all three choices are identical. The distribution of this variable is compared across treatments using a Pearson χ^2 statistic, testing the hypothesis that the rows and columns in a frequency table are independent. Failing to reject the hypothesis implies homogeneous distributions across the treatments, i.e., internal validity is, on average, similar when playing in real or hypothetical treatments and for the first or second time played (results reported in Table 4).

Test 4 and Test 5 are facilitated by the fact that we are able to match the individuals' responses from the first and the second time they completed the experiment (i.e. similar to a panel data structure).

Test 4 - We compare the degree of internal validity (measured in Test 3) an individual displays when playing for the first time with the corresponding internal validity when they play for the second time. Hence, a new stability variable is created that takes the value of one when the individual displays the same degree of internal validity on both times they play and zero otherwise. A Pearson χ^2 statistic is used to test whether the stability variable values are the same between real treatments and hypothetical treatments (the results are reported in Table 5).

Test 5 - We compare each individual's decision in each choice set in the first time play with the corresponding choice in the second time played. Again, we would expect an individual who exhibits well-defined preferences to give identical answers in the two time periods. Hence, this allows for the creation of another stability score taking values from zero (if no choice from period one matches the corresponding choice from period two) to 24 (if all choices match). In addition to using a Pearson χ^2 statistic, a Kolmogorov-Smirnov test formally tests for differences in the distributions of the stability scores between real and hypothetical treatments (results reported in Table 6).

2.5 Estimation models for the choice experimental data

For the analysis a semi-parametric latent-class model (LCM) (Greene and Hensher, 2003) is used, where individuals are probabilistically sorted into classes following a multinomial distribution (Greene and Hensher, 2003). Within each class a conditional logit is fitted (McFadden, 1974). The LCM avoids problems often associated with the standard conditional logit model (i.e. restrictive IIA, ignoring panel nature of data, ignoring preference heterogeneity) and presents an intuitive and easily interpretable alternative. The number of latent classes is determined a priori, based on the performance of different models with respect to information criteria, i.e. BIC (Swait and Adamowicz, 2001). Following estimation and using the cost attribute (i.e. monetary donation) as a numéraire, average WTP values for the attributes can be obtained.

3. Results

A total of 71 university students participated in the experiment: 35 for H1 and 36 for R1. Of those 30 returned for H2 and 31 for R2. In the real sessions subjects

earned about $16 on average (out of possible $24 for playing twice and in addition to the $12 flat fees) and a total amount of $353 was donated to charities. The LCM outperformed the conditional logit, while convergence was achieved for all estimations only when two latent classes. [6] Interestingly, none of the demographics significantly affected the class probabilities and hence they were dropped from the analysis.[7]

Before proceeding with the hypothetical bias and preference stability results we will briefly discuss the interpretation of the H1 model to familiarise the reader with inference from choice models. Estimated coefficients depict part-worth utilities, with a positive coefficient implying an increase in the alternative's attractiveness (i.e. bringing higher utility) as the attribute's level increases. Hence, in Class 1, from the constants, we find a preference for health charities relative to a social charity. All other attributes possess a positive sign implying that as the percentage of matching increases or as the amount to donate required increases, the probability of choosing such charity increases. [8] The counter-intuitive nature of the contribution attributes (which is in contrast with the marginal disutility of increasing price that would be expected a priori) will be briefly discussed later. Turning to Class 2, contribution attributes display the expected negative signs confirming the disutility associated with increasing cost, whereas matching is as before increasing in utility.

---

[6] For models that converged with three classes, the results of the second and third classes were largely comparable.

[7] Insignificant demographic effects are common in such models, e.g. Hanley et al. (2005).

[8] Only the sign and the relative importance of the raw coefficients within each model can be interpreted. The size of the raw coefficients cannot be interpreted directly and cannot be compared across models. WTPs can be used for this.

Although the latent nature of the model does not make known which participant falls in which class we can estimate the size of such class probabilities, with Class 2 being more likely with 63% relative to 37% of Class 1.

Moving away from the raw coefficients, the WTP column of the H1 model presents the willingness-to-donate values for each class. WTP values themselves indicate the amount of money by which a subject increases/decreases their donation for a percentage increase in the matching policy. For Class 1, on average, individuals imply a reduction in their willingness-to-donate by $0.08 and $0.07 for the health and social charities, respectively, for every unitary increase in the matching percentage (i.e. going from 1% to 2% matching policy), whereas matching has the opposite effect for Class 2, with the corresponding WTPs indicating a willingness to increase donations by $0.29 and $0.41 for health and social charities, respectively.

3.1 Results for Test 1

We first estimate parameters using the latent class model (Table 1). We regress the probability of choosing an alternative on the level of matching percentage (Matching Health, Matching Social), the amount to donate (Contribution Health, Contribution Social), the dummy variable which equals to 1 if health charity is chosen (i.e., Donate to Health Charity) and the dummy variable which equals to 1 if the opt out option is chosen (i.e., Do Not Donate). We then test for equality of parameters across treatments using LR tests and allowing for scaling differences (Swait and Louviere, 1993). Equality of parameters is rejected when comparing treatments H1 and R1 and treatments H2 and R2, suggesting evidence of hypothetical bias. Comparing treatments H1 and H2 we are unable to reject the null hypothesis that the parameters are not different, suggesting stability of

preferences in repeated elicitation. However, such stability is rejected when comparing treatments R1 and R2 (failing to reject only at higher than conventional significance levels, i.e. 1‰).

3.2 Results for Test 2

We compare WTPs across models and classes. WTPs are given in the second column of each treatment in Table 1, while Table 3 presents *t*-test values for all possible pair-wise comparisons. Those playing with real incentives are less responsive to changes in the matching policy (4 out of 8 *t*-tests indicate significant differences), while such difference are minimised when comparing subjects between first and second time played (none out of 8 t-tests is significant). Specifically, for Class 1, comparing hypothetical and real treatments we find a crowding-out effect which is more pronounced for those playing for real stakes (e.g. WTP_{R1}=-0.57 vs WTP_{H1}=-0.07), implying that subjects playing for real tend to reduce their donation for every increase in the matching by more than 8 times the reduction observed in those playing hypothetically. For Class 2, comparing hypothetical and real treatments, we find that matching induces donation but this donation is much smaller in the real treatment. Subjects in the real treatments tend to increase their donation by less (2 to 7 times less) for every percentage increase in matching compared to the subjects in the hypothetical treatments.

3.3 Results for Test 3, Test 3 and Test 5

Tests 3 to 5 are feasible due to the fact that the same choice set is repeated three times for each individual. These tests are not testing differences in the choice experiment estimations (parameters or preferences) but rather the internal validity in completing the DCE across treatments and the intertemporal stability of such

internal validity across incentives. Tables 4, 5 and 6 present the results. In general, we find that subjects do not display varying degrees of internal validity according to the treatment they were assigned to or between the two times they played, i.e. the levels of observed internal validity in the choice experiments were the same for all four of our treatments. None of the χ^2 tests detect any significant differences, while a Kolmogorov-Smirnov statistic confirms the lack of systematic differences between the distributions of the stability scores (D = 0.2903, p-value = 0.153).

Acknowledging the potential biases in testing for the stability of preferences over time when including in the analysis subjects who participated in the "1st time" but did not participate in the "2nd time" of each treatment we repeat the analyses excluding such participants (5 from treatment H1 and 5 from R1). [9] Results indicate little difference from the full sample with the constants in Class 1 now becoming insignificant as does contribution for social charity across four treatments, whereas the findings from the LR-tests and other tests remain unchanged.

4. Conclusions

We examine consistency of individual preferences across incentives (i.e. hypothetical bias) and their stability in repeated elicitation in a choice experiment involving decision to make charitable contributions within a controlled laboratory setting. We provide some evidence of hypothetical bias based on the underlying preferences and the derived WTP (or more correctly willingness-to-donate) values. On the other hand, we find partial support for instability, with preferences

---

[9]The potential biases are only related to Tests 1, 2 and 3, as for Tests 4 and 5 which required the panel aspect of our design such subjects were excluded anyway.

being unstable for the incentivised treatments but with WTP values stable across both hypothetical and real environments.

In particular, looking at the equality of parameters across models we find significant differences between first time and second time of play in the real treatments but no difference for the hypothetical treatments or any of the WTP values (real or hypothetical). Two distinct preference patterns emerge from our sample. For some subjects, matching policy increases donations with the effect being smaller for real than hypothetical scenarios, while for others a crowding-out effect is observed with the effect being more pronounced for those playing with real incentives.

Finally, neither hypothetical bias nor time instability is observed when we focus on the degree of internal validity in completing the choice experiment elicited from a thrice repeated choice set.

Hence, contrary to past literature, we observe a more persistent presence of hypothetical bias irrespective of the testing method (List et al., 2006; Lusk and Schroeder, 2004), while we partially confirm past findings in relation to the stability of preference over time (San Miguel *et al.*, 2002).

Regarding the theoretical validity of our substantive results, our findings are consistent with past experimental findings on the observed heterogeneity in people's pro-social preferences (Anderoni and Vesterlund, 2001). The positive signs for contribution (and subsequent negative WTPs given increases in the matching rate) can be attributed to pure altruism (a preference for other's well-being), warm glow (a pleasurable feeling from giving) or impure altruism (a combination of pure altruism and warm glow) (Anderoni 1989, 1990), while the

negative WTPs to a motivational crowding-out effect, where the amount experimenters matched may be perceived as reducing in self-determination or as controlling and undermining the intrinsic motivation of giving (Deci and Ryan, 1985; Frey, 1997). However, thorough analysis and discussion of the issue is beyond the scope of this paper.

As a general remark, despite acknowledging potential limitations in the representativeness and generalisation of our small samples, this study has shown the presence of strong hypothetical bias in choice experiments but less problems of time instability of preferences. Caution is suggested in drawing inferences from hypothetical DCEs, while their implementation should particularly focus on enhancing the realism of the question and its context. In addition to contributing to the emerging literature of the external and internal validity of choice experiments, this paper further demonstrates that choice experiments can be a useful tool in examining charitable donations decisions, especially in disentangling income and substitution effects, although it leaves more in depth examination of the issue to future research.

References

Andreoni, J. (1989). Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence, *Journal of Political Economy*, 97(6), 1447-158

Andreoni, J. (1990). Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving? *Economic Journal*, 100 (401), 464-477.

Andreoni, James and Lise Vesterlund (2001). Which Is the Fair Sex? Gender Differences in Altruism. *Quarterly Journal of Economics* 116. 293-312.

Ben-Akiva, M., and S.R. Lerman. (1985), *Discrete choice analysis: Theory and application to travel demand*. Cambridge, MA: The MIT Press.

Benz, M. and Meier, S. (2008), Do People Behave in Experiments as in the Field?-Evidence from Donations, *Experimental Economics*, 11 (3), 268-281.

Blumenschein, K., Blomquist, G.C., Johannesson, M., Horn, N., and Freeman, P. (2008). Eliciting willingness to pay without bias: Evidence from a field experiment. *Economic Journal*, 118(525), 114-137.

Broadbent, C.D., Grandy, J.B. and R.P. Berrens (2010). Testing for hypothetical bias in a choice experiment using a local public good: riparian forest restoration. *International Journal of Ecological Economics*, 19

Cameron, T.A., Poe, G.L., Ethier, R.G., and Schulze, W.D. (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 Martinsson, P. (2001). Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments? *Journal of Environmental Economics and Management*, 41, 179–92.

Cummings, R.G., Harrison, G.W., and Rutstrom, E.E. (1995). Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive-compatible? *American Economic Review*, 85(1), 260-266.

Deci, Edward L. and Richard M. Ryan (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Press.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics*, 10, 171-178.

Frey, Bruno S. (1997). *Not Just for The Money. An Economic Theory of Personal Motivation*. Cheltenham, UK and Brookfield, USA: Edward Elgar.

Greene, W.H. (2003), *Econometric analysis*. 5th Edition ed. Upper Saddle River, NJ: Prentice Hall.

Greene, W.H., and D.A. Hensher (2003), 'A latent class model for discrete choice analysis: Contrasts with mixed logit', *Transportation Research Part B: Methodological,* 37(8): 681-98.

Hanley, N., Adamowicz, W.L., Wright, R.E. (2005). Price vector effects in choice experiments: an empirical test. *Resource and Energy Economics*, 27, 227–234

Hensher, D.A., J.M. Rose, and W.H. Greene. (2005), *Applied choice analysis: A primer*. Cambridge, UK: Cambridge University Press.

Kuhfeld, W.F. (2005), 'Marketing research methods in SAS', TS-722, SAS Institute Inc., Cary, NC, USA.

Lancaster, K.J. (1966), A New Approach to Consumer Theory. *Journal of Political Economy*, 74, 132-157.

List, J. A. (2001). Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auctions for sportscards. *American Economic Review*, 91(5), 1498-1507.

List, J.A. and Gallet, C.A. (2001) What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values, *Environmental Resource Economics,* 20(30), 241-254.

List, J.A., Sinha, P., and Taylor, M.H. (2006). Using choice experiments to value non-market goods and services. *Advances in Economic Analysis and Policy*, 6, 1-37.

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

Lusk, J.L., and Schroeder, T.C. (2004). Are choice experiments incentive compatible? A test with quality differentiated beef steaks. *American Journal of Agricultural Economics*, 85, 840-56.

McFadden, D. (1974). 'Conditional logit analysis of qualitative choice behavior', in P. Zarembka (ed.), *Frontiers in econometrics*, New York: Academic Press.

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

Ready R.C., Champ, P.A., Lawton, J.L. 2010. Using respondent uncertainty to mitigate hypothetical bias in stated choice experiments. Land Economics 86(2), 363-381.

Ryan, M., K. Gerard, and M. Amaya-Amaya, eds. 2008. *Using discrete choice experiments to value health and health care*. The economics of non-market goods and resources., ed. I. J. Bateman. The Netherlands: Springer.

Ryan, M., and F.S. Miguel (2003), 'Revisiting the axiom of completeness in health care', *Health Economics,* 12(4): 295-307.

San Miguel, F., M. Ryan, and A. Scott (2002), 'Are preferences stable? the case of health care', *Journal of Economic Behavior & Organization,* 48(1): 1-14.

Stigler, G. J. and Becker, G. S. (1977), De Gustibus Non Est Disputandum. *American Economic Review*, 67, 76-90.

Swait, J., and W. Adamowicz (2001), 'The influence of task complexity on consumer choice: A latent class model of decision strategy switching', *Journal of Consumer Research,* 28(1): 135-48.

Swait, J., and J. Louviere (1993), 'The role of the scale parameter in the estimation and comparison of multinomial logit models', *Journal of Marketing Research,* 30(3): 305-14.

Taylor L.O., Morrison, M.D., Boyle, K.J. (2010). Exchange rules and the incentive compatibility of choice experiments, *Environmental and Resource Economics,* 47, 197–220.

Truong, T.P., and D.A. Hensher (1985), 'Measurement of travel time values and opportunity cost from a discrete-choice model', *The Economic Journal,* 95(378): 438-13.

Graph 1. Experimental Design and Treatment Labels

2x2 Design

Between-Subject Design			
		Hypothetical Incentive	Real Incentive
Within-Subject Design	First-time play	H1	R1
	Second-time play	H2	R2

Note: Internal validity is measured in every treatment

Table 1. Latent Class model estimations and willingness-to-pay calculation by treatment.

	(1) Hypothetical, 1st time (H1)			(2) Real, 1st time (R1)			(3) Hypothetical, 2nd time (H2)			(4) Real, 2nd time (R2)	
	Coeff.	WTP	.	Coeff.	WTP	.	Coeff.	WTP	.	Coeff.	WTP
Class 1											
Donate to Health Charity	1.421**			1.003**			0.435			0.963	
	(0.703)			(0.498)			(0.978)			(0.727)	
Do not donate	0.221			0.587			1.607*			0.134	
	(0.686)			(0.363)			(0.906)			(0.521)	
Matching Health	0.014*	-0.0821*		0.046***	21.3610		0.055***	-0.1553***		0.038***	-1.4888
	(0.008)	(0.045)		(0.006)	(334.0)		(0.013)	(0.032)		(0.006)	(2.22)
Contribution Health	0.17***			-0.002			0.353***			0.025	
	(0.054)			(0.034)			(0.087)			(0.038)	
Matching Social	0.022***	-0.0671**		0.058***	-0.5737***		0.047***	-0.0940***		0.04***	-0.2544***
	(0.008)	(0.026)		(0.006)	(0.209)		(0.011)	(0.0247)		(0.008)	(0.095)
Contribution Social	0.326***			0.100***			0.504***			0.155***	
	(0.054)			(0.035)			(0.081)			(0.048)	
Class 2											
Donate to Health Charity	0.299			-0.273			0.452			-0.305	
	(0.359)			(0.756)			(0.437)			(0.626)	
Do not donate	-0.17			0.393			-0.112			-1.008**	
	(0.198)			(0.417)			(0.265)			(0.444)	
Matching Health	0.042***	0.2905***		0.052***	0.1532***		0.062***	0.3133***		0.044***	0.0844***
	(0.003)	(0.0603)		(0.009)	(0.035)		(0.004)	(0.0467)		(0.008)	(0.017)
Contribution Health	-0.146***			-0.338***			-0.198***			-0.521***	
	(0.03)			(0.053)			(0.028)			(0.067)	
Matching Social	0.035***	0.4068***		0.031***	0.1199***		0.051***	0.4737***		0.041***	0.0682***
	(0.004)	(0.103)		(0.008)	(0.035)		(0.005)	(0.127)		(0.009)	(0.0147)
Contribution Social	-0.086***			-0.255***			-0.107***			-0.594***	
	(0.023)			(0.048)			(0.028)			(0.079)	
Prob. of Class 1	0.369			0.668			0.341			0.259	
Prob. of Class 2	0.631			0.332			0.659			0.741	
# of subjects	35			36			30			31	
BIC	1.79			1.62			1.55			1.42	

Standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$
Donate to Health Charity and *Not donate* are alternative specific constants, where the vase category is *Donate to Social Charity*. Such constants capture differences in utility based solely on the type of charity.

Table 2. Likelihood ratio tests for equality of model parameters across models

		.	Sum LL Model 1 + LL Model 2	.	Sum LL Model 3 + LL Model 4	.	Sum LL Model 1 + LL Model 3	.	Sum LL Model 2 + LL Model 4	.	LR test: 2*(LL Sum – LL Pooled)
LL Model 1: Hyp, 1st	-615.741										
LL Model 2: Real 1st	-570.325										
LL Model 3: Hyp, 2nd	-446.510										
LL Model 4: Real, 2nd	-419.135										
LL Pooled Model 1 & 2 [a]	-1209.4		-1186.066								46.66
LL Pooled Model 3 & 4 [a]	-922.078				-865.645						112.86
LL Pooled Model 1 & 3 [b]	-1065.510						-1062.251				6.51
LL Pooled Model 2 & 4 [b]	-1005.760								-989.460		32.59

Critical value for χ^2 statistic at 5% =23.69, 1%=29.14, 1‰=36.12

[a] Indicates test relevant to hypothetical bias (i.e. across incentives comparisons)

[b] Indicates test relevant to time stability (i.e. across time comparisons)

Table 3. Values of the t-statistics for *t*-test of equality of WTPs across models/samples and classes

	(H1)WTP HealthC1	(H1)WTP SocialC1	(H1)WTP HealthC2	(H1)WTP SocialC2	(R1)WTP HealthC1	(R1)WTP SocialC1	(R1)WTP HealthC2	(R1)WTP SocialC2	(H2)WTP HealthC1	(H2)WTP SocialC1	(H2)WTP HealthC2	(H2)WTP SocialC2
(R1) WTP Health C1	**0.064 [a]**	0.064	0.063	0.063	0.000							
(R1) WTP Social C1	-2.300	***-2.406 [a]***	-3.973	-4.209	-0.066	0.000						
(R1) WTP Health C2	4.127	5.053	**-1.969 [a]**	-2.331	-0.063	3.431	0.000					
(R1) WTP Social C2	3.543	4.289	-2.447	***-2.637 [a]***	-0.064	3.274	-0.673	0.000				
(H2) WTP Health C1	**-1.326 [b]**	-2.139	-6.530	-5.212	-0.064	1.979	-6.505	-5.803	0.000			
(H2) WTP Social C1	-0.232	**-0.750 [b]**	-5.901	-4.728	-0.064	2.280	-5.771	-4.993	1.516	0.000		
(H2) WTP Health C2	6.097	7.117	**0.299 [b]**	-0.827	-0.063	4.142	2.743	3.314	8.277	7.710	0.000	
(H2) WTP Social C2	4.125	4.172	1.303	**0.409 [b]**	-0.063	4.283	2.433	2.686	4.803	4.388	1.185	0.000
(R2) WTP Health C1	-0.634	-0.640	-0.801	-0.853	**-0.068 [b]**	-0.410	-0.740	-0.725	**-0.601 [a]**	-0.628	-0.812	-0.883
(R2) WTP Social C1	-1.639	-1.902	-4.843	-4.719	-0.065	**1.391 [b]**	-4.026	-3.697	-0.989	**-1.634 [a]**	-5.363	-4.591
(R2) WTP Health C2	3.461	4.877	-3.290	-3.088	-0.064	3.139	**-1.768 [b]**	-0.912	6.615	5.950	***-4.606 [a]***	-3.038
(R2) WTP Social C2	3.175	4.530	-3.582	-3.254	-0.064	3.064	-2.239	**-1.362 [b]**	6.347	5.643	-5.006	***-3.172 [a]***

The letters in parentheses at the beginning of the names indicate treatments of the WTP value tested, while the "C1" or "C2" at the end indicates whether it corresponds to the first or second class of the LCM.

[a] Indicates t-test values relevant to hypothetical bias (i.e. across incentives comparisons)
[b] Indicates t-test values relevant to time stability (i.e. across time comparisons)
  Numbers in italic indicate significance at 5%

Table 4. χ^2 tests of degree of internal validity across treatments[a]

Internal validity measure values	(1) Hypothetical, 1st time	(2) Real, 1st time	(3) Hypothetical, 2nd time	(4) Real, 2nd time
0	0	1	1	0
1	11	12	9	6
2	24	23	20	25
# of subjects	35	36	30	31

Hypothetical bias tests

(H1) vs. (R1) $\chi^2 = 1.05$

(H2) vs. (R2) $\chi^2 = 2.14$

Time stability tests

(H1) vs. (H2) $\chi^2 = 1.19$

(R1) vs. (R2) $\chi^2 = 2.72$

[a] Internal validity is based on comparison of choices across three repetitions of a choice set (i.e. consistency in completing the DCE)

Critical value for $\chi^2 = 5.99$

Table 5. χ^2 **tests for the stability measure across incentives[a]**

Stability measure values	Real	.	Hypothetical	.	χ^2 test for stability of preference between Real treatments vs. Hypothetical treatments
0	13		11		
1	18		19		
# of subjects	31		30		
					$\chi^2 = 0.177$

[a] Stability is based on the individuals' degree of internal validity displayed in the two times they played.

Critical value for $\chi^2 = 3.841$

Table 6. χ^2 tests for the stability of answers in each choice set between the two times of play across incentives

Stability score	.	Real	.	Hypothetical	.	χ^2 test for stability of preference between Real treatments vs. Hypothetical treatments
2		1		0		
5		1		0		
6		1		0		
7		2		0		
9		4		0		
10		1		1		
11		0		1		
12		3		2		
13		1		2		
14		1		2		
15		2		0		
16		1		2		
17		1		3		
19		2		4		
20		1		4		
21		2		3		
22		4		2		
23		0		2		
24		3		2		
# of subjects		31		30		
						$\chi^2 = 19.72$

Critical value for $\chi^2 = 28.87$

An investigation of individual preferences: consistency across incentives and stability over time

Appendix

Instructions (Individual Treatment—REAL SESSIONS)

Introduction

This survey is concerned with charitable donating decisions, i.e. how much to donate and to which type of charity. We are interested in gaining an understanding of the relative importance individuals place on the different characteristics of a charity in the donation decision.

There are two parts to this survey. In Part I you are asked to make 24 donating decisions and for each one you are given an initial monetary endowment of $12. You are then presented with the option to donate a fraction of your endowment to one of two charities or to not donate to either charity. Part II asks you for some basic demographic information to help us analyze the data from the survey. This information will not be used in any way to identify you. For Part II you will be compensated a flat fee of $6.

Completing the survey should take somewhere between 15 and 30 minutes.

Charity and Scenario description

Each charity is described by three characteristics: 1) the type of the charity (social service or health); 2) the matching policy (0%, 25%, 50%); and 3) the amount to donate ($4, $8, $12).

(1) Type of Charity:

Each charity is identified as either a Health or a Social Service charity according to the mission and the purpose of the charity.

- Health charities imply serving causes that are directly linked to some health issue (e.g. Cardiovascular diseases, Muscular Dystrophy).

- Social Service charities target causes that deal with some social issue (e.g. Poverty, Support for youth and families at risk)

The charities that have been chosen are comparable in terms of the percentage of funding received from the government (45%-55%) and the percentage of funds spent on administrative/managerial purposes (40%-60%). Two Health charities and two Social Service charities are identified. All

charities provide nationwide services. More details about the exact causes and missions of each charity are given on separate sheets on your desk. This information comes directly from the web pages of the charities.

(2) Matching Policy:

For each charity, you will also be presented with our matching policy. This policy indicates the percentage with which we will match your donations. If you donate $8 (out of your initial endowment of $12) to a charity with a matching policy of 25%, we will automatically add $8*25%=$2 to your donation. Hence, the total amount that will be donated to the charity will be $2+$8=$10. If you donate $8 with a matching policy of 50%, we will automatically add $8*50%=$4 to your donation for a total of $4+$8=$12.

There are three levels of matching policies:

- 0%
- 25%
- 50%

(3) Amount Donated

Donations can be made in three different amounts:

- $4
- $8
- $12

Payments and Online Donations

When you start the survey, you will be, sequentially, asked to make 24 donating decisions/scenarios. In each of these 24 scenarios you will be given a $12 endowment and asked how much you are willing to donate to a Health charity, or a Social Service charity or not to donate at all. At the end of your 24 decisions, one of these decisions will be randomly chosen for calculating your payment.

The first two options (i.e. donate to Health or Social Service charity) will each have a donation of $4, $8 or $12 to be deducted from your endowment of $12, while the third option (i.e. No Donation) requires no dollar donation and you will keep all $12.

Hence:

- If you chose the third option (i.e. No Donation), you will be given $12 and no money will be donated to the charities.

- If you have chosen either of the other two options, we will deduct the donation amount from your endowed $12 and you will keep the rest. We will then match your donation according to the matching policy, add it up to your donation and the sum will be donated to a charity in the specified category (social or health).

If you have chosen a Health charity, the donations will be sent to one of the two Health charities that are described on the separate sheets. The specific charity is chosen at random and will be the same for all people in your session.

If you have chosen a Social Service charity, the donations will be sent to one of the two Social Service charities that are described on the separate sheets. The specific charity is chosen at random and will be the same for all people in your session.

Once the donated amounts are determined and calculated and the charity is randomly drawn, one of the experimenters will carry out the donation online in front of you. No one will be claiming a charitable donation tax credit.

PART I

In this part you are presented with 24 donating decisions. You start with a new endowment of $12 in each scenario. Your decision in one scenario should **not** affect your decision in another scenario.
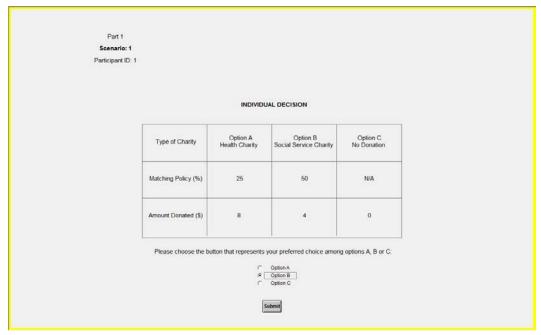
In each scenario you will be asked to choose between donating to a Health charity, donating to a Social Service charity or no donation.

All aspects of the charities apart from those presented should be assumed identical.

THERE ARE NO RIGHT OR WRONG ANSWERS.

We are interested in your donation decisions.

Example

Part 1
Scenario: 1
Participant ID: 1

INDIVIDUAL DECISION

Type of Charity	Option A Health Charity	Option B Social Service Charity	Option C No Donation
Matching Policy (%)	25	50	N/A
Amount Donated ($)	8	4	0

Please choose the button that represents your preferred choice among options A, B or C:

- ○ Option A
- ◉ Option B
- ○ Option C

[Submit]

In the example above, option B has been selected. Therefore, the individual has chosen to donate $4 of his $12 endowment to a Social Service charity with a matching policy of 50%. Hence if this scenario is randomly drawn for payment, the individual will receive $8 and the charity will receive $6 (i.e. $4 + (50%*$4)).

Are there any questions?

In the section below, we will start with a practice round, followed by 24 donating decisions. In each case please indicate whether you prefer Option A, B or C by placing a tick in the appropriate radio button and then click "Submit".

4