Journal of Choice Modelling, 3(2), pp 113-133
www.jocm.org.uk

Do you do what you say or do you do what you say others do?

Fredrik Carlsson [1,*] Dinky Daruvala [2,†] Henrik Jaldell [3,Ŧ]

[1]Department of Economics University of Gothenburg, Box 640, SE-40530 Göteborg, Sweden
[2]Department of Economics, Karlstad University, SE-651 88 Karlstad, Sweden
[3]Department of Economics, Karlstad University, SE-651 88 Karlstad, Sweden

Received 29 May 2009, received version revised 28 February 2010, accepted 19 July 2010

Abstract

We design a donations vs. own money choice experiment and compare the results from three different treatments. In two of the treatments the pay-offs are hypothetical. In the first of these, a short cheap talk script was used and subjects were required to state their own preferences in this scenario. In the second treatment, subjects were asked to state how they believed the average student would respond to the choices. In the third treatment the pay-offs were real, allowing us to use the results to compare the validity of the two hypothetical treatments. Our hypothesis is that when subjects are asked to state how they believe an average person would respond, they will use their own preferences in their responses without using the survey situation for self-enhancement. However, we find a large difference in the results from both hypothetical treatments compared to the real money treatment. We find that the marginal willingness to pay for donations is higher when subjects state their own preferences but lower when subjects state what they believe are other people's preferences. We also find that it is mainly women who are prone to these differences in the study.

Keywords: Stated preferences, cheap talk, third person approach, choice experiment

* Corresponding author, T: + 46 31 7864174, Fredrik.Carlsson@economics.gu.se
† T: + 46 54 7001511, dinky.daruvala@kau.se
Ŧ T: + 46 54 7001369, henrik.jaldell@kau.se

Licensed under a Creative Commons Attribution-Non-Commercial 2.0 UK: England & Wales License
http://creativecommons.org/licenses/by-nc/2.0/uk/

ISSN 1755-5345

1 Introduction

Survey questions are frequently used to elicit information on a variety of personal preferences. While this is relatively straightforward in most cases, there is some concern regarding the validity of responses in hypothetical survey settings associated with self-image situations such as risk, donations, and the provision of public goods. Kahneman and Knetsch (1992) describe donations as a ”purchase of moral satisfaction” and while Andreoni (1990) shows that donations to public goods may be due to either altruism or warm-glow (egoism) both reasons are related to a positive self-image motive. With this in mind, it appears likely that the hypothetical survey situation may provide a cheap opportunity for the respondents to enhance their self-image.

One criticism against using survey questions concerns incentives for the truthful revelation of preferences (Carson et al. 1996). The problem is that respondents may not place enough emphasis on the contingent part of the survey. In the hypothetical set-up where no actual payment is required, subjects may tend to focus mainly on the benefits of the project while largely ignoring the costs. The survey results would then reveal the attitudes rather than the preferences of the respondents (Kahneman and Sugden 2005).

A number of studies have tested the possible disparities between hypothetical survey responses and subjects’ responses in actual situations, in particular within the stated preference literature; see e.g. Bishop and Heberlein (1979), Carson et al. (1996) and Harrison and Rutström (2002) for studies testing hypothetical responses the contingent valuation method, and e.g. Carlsson and Martinsson (2001), Ding et al. (2005) and Lusk and Schroeder (2004) for studies testing hypothetical responses in choice experiments. Although findings from meta analysis studies support the belief that hypothetical situations lead to a higher stated WTP compared with non-hypothetical situations, the evidence is mixed (List and Gallet 2001; Murphy et al. 2005).

Another line of research in the stated preferences literature focuses on ways to overcome or at least reduce the differences between responses given in a hypothetical survey situation and behavior in a real situation. One method is the use of cheap talk scripts where subjects are informed that a propensity to exaggerate stated WTP has been found in previous similar studies. Thus, by introducing the notion into the consciousness of the subject, the effect of a self-image bias is thought to be reduced. The success of cheap talk scripts has shown varied results. Using private goods, classroom experiments, or closely controlled field settings, the use of cheap talk has proven to be potentially successful (Cummings and Taylor 1999; List 2001; Murphy et al. 2005). Similarly, short cheap talk scripts have been effective in reducing the marginal WTP in choice experiments (Carlsson et al. 2005). However, mixed results have been found when incorporating a public good with private good attributes (Aadland and Caplan 2003, 2006), and one explanation put forward for this difference is the effect of the length and structure of the cheap talk script.

In this paper, we suggest another method that could potentially be effective in reducing the difference between hypothetical and real situations: the third-person perception approach, where we ask subjects what they believe an average person would do. This is analogous to the false consensus notion in social psychology, and implies that in many situations, people tend to believe that others think and behave like themselves (Fields and Schuman 1976; Ross et al. 1977). The notion behind this is that people will use their own preferences to predict that of others and

consequently state their own preferences in their responses. The assumption we make here is that when respondents use their own preferences as a proxy for others, they will not use the survey situation for self-enhancement or social desirability. In our experiment, the respondents made their choices anonymously, so any direct effect of social desirability should be small. At the same time, because subjects know that they are participating in an economic experiment they might feel that they are under scrutiny by the researcher and thus social desirability could be important (cf. Levitt and List 2007). Our hypothesis is that social desirability will also be less influential when we use the indirect question approach (cf. Fisher 1993).

Various studies suggest that the third person approach may have potential, especially in situations where there is little social distance between the predictor and the target. For example, Epley and Dunning (2002) found in a series of experiments that student participants consistently tend to overstate their own generosity, but were relatively accurate when predicting the generosity of other students. Other studies have found that when predicting the risk-behavior of others, subjects tend to believe others have the same risk preferences as themselves (Chakravarty et al. 2005; Hsee and Weber 1997). Similarly, Henriksen and Flora (1999) studied the perceived influence of cigarette advertising on children. They found that the discrepancy between perceived influence on themselves and others was smaller when children compared themselves with their best friends than with other peers. Fisher (1993) found that there are differences between direct and indirect questioning when the issues are subject to social influence, and that indirect questioning is insensitive to the degree of anonymity. Johansson-Stenman and Martinsson (2006) find that people state that the environmental performance of cars is important for them, but not for their neighbors. The authors argue that this stated difference is due to a positive self-image attained by responding to the survey. The studies that are closest to our own are Lusk and Norwood (2009a,b), in whose model people may derive utility from stating that they are willing to pay for a good due to the effects of social desirability or warm-glow. They also argue that a third person approach, or what they called inferred valuation, can mitigate this bias. In Lusk and Norwood (2009a), lab experiments are used where subjects vote for the provision of unique Swiss chard plants to everybody in the group. If more than half the group vote yes, all subjects receive a plant at a certain cost. If more than half vote no, the plants are destroyed. They find that predictions of others' voting behavior are similar to actual voting behavior. Lusk and Norwood (2009b) compare three goods that vary in terms of people's familiarity and social normative motivation. One of the main findings is that for goods with high normative consequences, own stated willingness to pay is higher than the predicted willingness to pay of others.

In order to test the performance of the cheap talk and third-person perception approaches described above, we designed a donations vs. own money choice experiment using three different treatments and a between-subject comparison. The main reason for using a between-subject design is to avoid the risk of subjects trying to be consistent between the treatments that are very similar in design. In each treatment, the subjects were required to make 12 pair-wise choices where the characteristics of each choice were personal money, donation to a charity, and type of charity. In two of the treatments the pay-offs were hypothetical. In one of these, a short cheap talk script was used, and subjects were required to state their own preferences in this hypothetical scenario. In the second, subjects were asked to state how they believed an average student would respond to the choices. In the third treatment, the pay-offs were instead real, thereby allowing us to use the results to

compare the validity of the two hypothetical treatments. All three treatments were conducted using student respondents at Karlstad University, Sweden.

Measuring the difference between real and hypothetical situations even in a specific experimental context is not without problems. Studies have shown that the experimental situation itself can lead to bias even if real-payoffs are involved; see for example Alpizar et al. (2008) and List et al. (2004). Thus, the real money treatment in this setting may have induced responses that are biased due to for example, self-enhancement or social desirability effects.[1] The treatment with real money in this experiment is not designed to exactly replicate an actual donation situation. Instead, it is designed to be a reference case to which the two hypothetical scenarios may be compared. Thus, although we cannot be sure that the real money treatment used in this study is the objective standard, we assume that this treatment is the one most closely aligned to true preferences, and for simplicity, we will denote the difference between the real and hypothetical treatments as hypothetical bias. Thus, this should not be interpreted as a belief that behavior in the real money experiment is not affected by contextual factors or that there are no self-enhancement effects.

Our results find a large difference between the hypothetical and real treatments, where the marginal willingness to pay for donations is higher when subjects state their own preferences but lower when subjects state what they believe are other students preferences, compared with the treatment with real money. Our explanation is that self-image effects are at play in both cases. In the cheap talk script treatment, the results reveal that it is mainly women who drive the difference between the hypothetical and real treatment.

The remainder of the paper is organized as follows: section 2 provides a description of the experimental design and procedure. The results from the study are presented in section 3 followed by a discussion in section 4.

2 The Experiment

2.1 Experimental Design

In order to test the performance of the *hypothetical-cheap-talk* and *hypothetical-other* approaches discussed above we designed a donations vs. own money choice experiment using three different treatments within which subjects were required to make the same 12 pair-wise choices where the characteristics of each choice were personal money, donation to a charity and type of charity. The experiment is essentially a repeated pair-wise choice version of a dictator game. In each alternative a certain amount of money is donated to a specific charity, and a certain amount of money is given to the subject for his/her personal use. The charities included were World Wildlife Fund, UNICEF Children's Fund and The Red Cross disaster relief. All these charities are well known to the subjects. The attributes, donations and

[1] There is empirical evidence that behavior in the lab involving real money does not mimic behavior outside the lab. For example, Shogren et al. (1999) conducted a hypothetical mail survey and a lab experiment concerning irradiated food, and compared the results with actual store purchases. They found that both the survey and the lab experiment resulted in a larger market share prediction of irradiated chicken than the grocery store prediction. Lusk et al. (2006) compare a framed field experiment with actual retail sales. They find that the results of the framed field experiment predict consumer behavior in the store, although there is some evidence of more pro-social behavior in the framed field experiment.

personal money had four different levels respectively; the levels of the donation were 200, 300, 400, and 500 SEK, and the levels of personal money were 0, 20, 50, and 100 SEK.[2] A simple cyclical design or a so-called fold-over design was used. First, an orthogonal main effects design was generated, consisting of 12 attribute level combinations. These combinations represent one alternative in each set. The attribute levels in the second alternative were obtained by adding two levels to the attribute levels in the first alternative, and when the highest level was reached, we started over from the lowest level. Thus, a subject made in total 12 pair-wise choices. We use a split sample design, with three treatments. The main reason for using a split is to avoid the problem of subjects trying to be internally consistent in their responses (Ariely et al. 2003; Johansson-Stenman and Svedsäter 2008). The three treatments used were:

> *i) Real-Money*: In this treatment the subjects made choices with real pay-offs to both the charity as well as themselves. They were informed that one of these would be randomly drawn as the actual choice set. An example of one of the choices in this treatment is given in Figure 1.

We use the results from this treatment in order to assess the performance of the hypothetical approaches.

> ii) *Hypothetical-cheap-talk*: In this treatment all pay-offs are hypothetical, both to the subjects and to the charities. The subjects were given a short cheap-talk script in which they were informed, verbally, using the overhead as well as in the text that

"Experiences from similar studies have found that peoples' responses in a survey situation often differ to how they actually act in real life. It is especially common for people to state that they are willing to donate money to a worthy charity, but later do not do so."

Choice 1. Which of the two following alternatives do you choose?

	Alternative 1	**Alternative 2**
Type of project	World Wildlife Fund	Unicef Childrens Fund
Money to the project	500 SEK	300 SEK
Money to yourself	0 SEK	200 SEK

I choose

☐ Alternative 1

☐ Alternative 2

Figure 1. Example choice task

[2] At the time of the survey 1 SEK = 0.16 USD. Due to a typing error, one choice sets with the level 200 SEK for own money, was instead 20 SEK. We will still include this choice set in the analysis.

We used a cheap talk script that is shorter than in many other studies, because a similar script has been used with good results for choice experiments (Carlsson et al. 2005). The choice sets were exactly the same as in the real-money treatment but they were instead asked "*Which of the two following alternatives would you choose?*"

> iii) *Hypothetical-others:* In this version subjects were required to state how they believed an average student would respond to the choices. The choice of reference group is not straightforward. Our hypothesis is that subject will not use the experiment situation for self-enhancement and will base their answer on their true preferences. The social distance between the subject and the group should not be too great. Thus, to use for example an average citizen, would make it hard for the subjects to transmit her own preferences; see for example Hsee and Weber (1997). On the other hand, one should not choose a group that is too close, since there is then a risk of group-image bias; see for example Henriksen and Flora (1999). We believe the choice of using the "average student" as the reference group was a reasonable compromise.

The subjects were required to answer the same choice sets as the other treatments but were asked the question "*Which of the two alternatives would the average student choose?*"

The full scripts are presented in the appendix. Each respondent made a total of 12 pair-wise choices, and these were the same in all three treatments. By comparing the responses in the three treatments, we can assess the performance of the two hypothetical treatments. Since we only observe the choices and not the preferences of the respondents, we apply a standard random utility model in the analysis. The utility of alternative *i* for individual *k* is

$$V_{ik} = \beta_{jk}' Donation_{ij} + \lambda_k' \mathrm{Money}_{\mathrm{i}} + \varepsilon_{ik} \qquad (1)$$

where, $Donation_{ij}$ is the amount of money donated to project *j*, $\mathrm{Money}_{\mathrm{i}}$ is the amount of money the individual receives, ε_{ik} is an error term, and β_{jk} and λ_{k} are parameters. In order to allow for unobserved heterogeneity we assume that the donation parameters β_{jk} are randomly distributed with a normal distribution. Since we have repeated observations, we assume that the parameters are constant across choice sets for a given individual. We assume that the errors terms are normally distributed. The model is estimated as a random parameter logit model with simulated maximum likelihood using Nlogit 4.0; see Train (2003) for details on simulated maximum likelihood. From the utility specification in (1), we can estimate the marginal willingness to pay for a donation to a project; this is simply the ratio of the donation and money coefficients.

2.2 Experimental procedure

A total of 268 undergraduate students from Karlstad University took part in the experiments that were conducted at the beginning of a lecture. The participants studied courses in business administration and economics. 103 men and 165 women participated in seven separate experimental sessions, each of which lasted around 20 minutes. Due to budget limitations, we were required to limit the number of subjects

in the treatment with real money. There were 64 subjects in the real-money treatment, 108 subjects in the hypothetical-cheap-talk treatment and 96 subjects in the hypothetical-others treatment.

Verbal instructions with supporting overheads were used in addition to the written instructions in the questionnaire. The questionnaire consisted of two parts, the choice experiment and questions regarding the respondents' socioeconomic status. The responses were anonymous in all three treatments. In the *Real-money* version, the participants were given an identity number which was also printed on the back of the questionnaire. The session began with the experimenter explaining how the payment procedure guaranteed a large degree of anonymity. After the session, each respondent took their identity card to a room where another person, not involved in the study, randomly picked a number in order to establish for which of the twelve questions the pay-off would occur. The respondent was immediately paid the sum corresponding to the choice made and the corresponding donation registered and later paid anonymously. Although the payments were made privately, there is nevertheless a small risk that the knowledge that a third party would be present at the payment procedure might have affected subject responses.

3 Results

The raw data from the three different treatments for the whole sample and for the male and female respondents are presented in Table 1. Swedish students are a fairly homogenous group, therefore not many questions were asked regarding their socio-economic characteristics. The two main characteristics that are of interest in the analysis are gender and whether the subjects currently contribute to charities. Using a proportion test for these two characteristics we find no significant differences between the three treatments.[3] To begin with, we will focus on the results for the whole sample.

The share of subjects who chose the alternative which gave them the most money is consistently lower in the hypothetical-cheap-talk treatments compared with the two other treatments. This suggests that there is a hypothetical bias in the treatment with a cheap-talk script. There is no consistent difference in behavior between the hypothetical-other and the real-money treatment, and consequently we need the econometric analysis to determine whether there is a hypothetical bias in the hypothetical-other treatment as well.

We begin by estimating three separate models for the three different treatments. Table 2 reports the results of the random parameter models for the three treatments. All models are estimated with simulated maximum likelihood using 500 Halton draws; see Train (2003) for details on simulated maximum likelihood. We also report the estimated marginal WTPs for a donation, the standard errors for marginal WTP are calculated with the Krinsky and Robb (1986) method with 2,000 draws.

[3] The shares of females are 0.59, 0.58 and 0.68 in the hypothetical-other, hypothetical cheap talk and the real money treatment respectively. The shares of subjects currently donating to charities are 0.29, 0.24, and 0.31 respectively.

Table 1. Share of subjects who chose the alternative which gave them the most money

							Share of subjects who chose the alternative which gave them the most money		
	Alternative 1			**Alternative 2**			Hypo-	Hypo-	Real money
Set	Project	Donation	Money	Project	Donation	Money	cheap	other	
1	WWF	500	0	Unicef	300	200	0.57	0.93	0.73
2	WWF	400	100	Unicef	200	50	0.54	0.61	0.73
3	WWF	300	20	Unicef	500	100	0.79	0.84	0.95
4	WWF	200	50	Unicef	400	0	0.31	0.51	0.36
5	Unicef	500	100	Red Cross	300	50	0.74	0.77	0.94
6	Unicef	400	0	Red Cross	200	200	0.42	0.79	0.59
7	Unicef	300	50	Red Cross	500	0	0.40	0.50	0.47
8	Unicef	200	200	Red Cross	400	100	0.38	0.59	0.55
9	Red Cross	500	50	WWF	300	0	0.70	0.82	0.92
10	Red Cross	400	200	WWF	200	100	0.78	0.84	0.92
11	Red Cross	300	0	WWF	500	200	0.61	0.78	0.88
12	Red Cross	200	100	WWF	400	50	0.37	0.57	0.45

Table 2. Results of random parameter models for the three treatments, *p*-values in parentheses.

	Hypothetical-cheap-talk		Hypothetical-other		Real-money	
	Coefficient (p-value)	**MWTP (s.e.)**	**Coefficient (p-value)**	**MWTP (s.e.)**	**Coefficient (p-value)**	**MWTP (s.e.)**
			Random parameters			
Donation to WWF	0.0020	1.231	0.0007	0.105	0.0039	0.474
	(0.000)	(0.216)	(0.000)	(0.031)	(0.000)	(0.046)
Std dev. donation to WWF	0.0022		0.0001		0.0039	
	(0.000)		(0.998)		(0.000)	
Donation to Unicef	0.003	1.847	0.0017	0.260	0.0054	0.645
	(0.000)	(0.328)	(0.000)	(0.035)	(0.000)	(0.040)
Std dev. donation to WWF	0.0012		0.0001		0.0001	
	(0.000)		(0.999)		(0.703)	
Donation to Red Cross	0.0027	1.661	0.0017	0.261	0.0045	0.546
	(0.000)	(0.287)	(0.000)	(0.027)	(0.000)	(0.029)
Std dev. donation to Red Cross	0.0016		0.0001		0.0008	
	(0.000)		(0.991)		(0.015)	
			Fixed parameters			
Personal money	0.0016	0.0014	0.0066			
	(0.000)	(0.000)	(0.000)			
Number of observations	1296		1152		768	
Number of individuals	108		96		64	
Pseudo R^2	0.02		0.01		0.04	

In terms of the sign and statistical significance of the mean parameters, the three treatments yield similar results. However, the standard deviation parameters are not statistically significant in all models, indicating that we are only to a limited extent, capturing unobserved heterogeneity. Most importantly, the magnitude of the WTP is quite different across the three treatments. In the hypothetical-cheap-talk treatment the marginal WTP for a donation varies from 1.23 SEK for WWF to 1.85 SEK for Unicef, which means for example that a subject states that they would be willing to pay 1.82 SEK in order to increase a donation to Unicef by 1 SEK. Thus, the model results suggest that subjects are willing to sacrifice more money than what is actually donated; despite the inclusion of the short cheap-talk script. The marginal WTP in the treatment with real money varies between 0.47 to 0.65 SEK. The differences in WTP between the hypothetical and real-money treatments are all significant, and there is thus a significant hypothetical bias. This result is in contrast with the earlier findings of Carlsson and Martinsson (2001), and Lusk and Schroeder (2004) who both find that choice experiments tend not to suffer from hypothetical bias for marginal trade-offs. Johansson-Stenman and Svedsäter (2008), on the other hand, found a hypothetical bias for marginal WTP in a similar experiment.[4] One should of course be careful when comparing different studies since the nature of the good, the sample and designs vary considerably across studies. However, the experiments in both Carlsson and Martinsson (2001), and Johansson-

[4] One reason why Johansson-Stenman and Svedsäter (2008) find a hypothetical bias, while Carlsson and Martinsson (2001) do not, could be that the latter use a within-sample design, where the same respondents answer both a hypothetical and a real-money experiment.

Stenman and Svedsäter (2008) are similar to our experiment. In a meta-analysis of hypothetical bias the hypothetical value was about 2.5-3 times the real-money value for public goods, in choice experiments with cheap-talk scripts, which is in line with our results (Murphy et al. 2005).[5] Our results are therefore also in contrast with for example Carlsson et al. (2005), who find that a short cheap-talk script reduced hypothetical bias, while we find no evidence of the same.

The hypothetical-other treatment also results in a hypothetical bias in the sense of a difference in WTP compared with the real-money treatment. The marginal WTP in this treatment is around 0.21 SEK, which is statistically significantly lower than the marginal WTP in the real-money treatment. In this case, the bias is in the other direction, since the marginal WTP for this treatment is lower than the marginal WTP using real money. One explanation is that subjects use themselves as a reference point when evaluating others (Dunning and Hayes 1996), and use the survey situation to bolster their self-image and validate a positive sense of self by predicting the generosity of others to be less than their own (Dunning 1996). Thus, a respondent with an own marginal WTP of say 0.3 SEK, is able to enhance her own self-image by stating a lower value for her peers[6]. Our results contrast the findings of Lusk and Norwood (2009a) where people's predictions about the voting behavior of others for a public good are similar to their own actual behavior. One explanation for the difference in the results is that the good used in our study, donations to charities, has stronger self-image effects and stronger relative self-image effects than the good used in Lusk and Norwood (2009a), a relatively unique plant. That is to say, it is more important to believe that you would donate more to charities than others, than it is to believe that you are more willing to protect a unique pot plant. This is consistent with the findings of Lusk and Norwood (2009b), where the difference between stated own WTP and predicted WTP of others is significant when the good is associated with normative motivations. However, in our experiment the predicted WTP of others is even significantly lower than the WTP from the hypothetical experiment with a cheap talk script.

As discussed in the introduction, the critical assumption of the third-person perception approach is that because respondents were not asked to explicitly state their own preferences, they would not use the survey situation as an opportunity for self-enhancement. However, the results of our study suggest that this assumption is erroneous. Research has shown that individuals who report attitudes and behavior for themselves and others are motivated to believe they possess various desirable attributes not only on an absolute level, but also on a relative level when compared to others (Alicke and Klotz 1995; Messick et al. 1985; Moore and Kim 2003). We believed that the respondents would not engage in such comparison as they were not explicitly asked to state their own preferences relative to others, but the results suggest that this was probably not the case. However, it is important to remember that we are comparing three situations where the subjects make their decisions in a laboratory. Hence they

[5] However, the total number of choice experiment studies included in Murphy et al. (2005) is small, and they were only able to measure the effect of hypothetical bias of calibration techniques in general, not only cheap talk scripts.

[6] This is of course under the premise that the responses in the real money treatment reflect true preferences.

are aware that they are being observed by a researcher. This has an effect on behavior, not only in the two hypothetical treatments, but also in the treatment involving real money. As we discussed in the introduction, there are problems associated with using results from laboratory experiments involving real money in order to predict actual behavior outside the lab; see e.g. Shogren et al. (1999), List et al. (2004) and Lusk et al. (2006). One explanation for this is that although there are self-image effects from donations in the real world, these may be augmented due to the laboratory situation.

Using two-sided *t*-tests we can in all cases reject the hypothesis of equal WTP and for all models we can reject the hypothesis of equal parameters.[7] Thus, looking at the aggregate data, we see a strong indication of hypothetical bias in both the hypothetical-cheap-talk and the hypothetical-other treatments, but that they go in opposite directions. This is most likely due to strong self-image effects.

Considering the rank of the projects we find that the UNICEF children's fund is the most preferred project for the real-money treatment, while there is no statistical difference between UNICEF and the Red Cross disaster relief for the hypothetical treatments. The WWF is least preferred in all treatments.

3.1 Differences among subject groups

Although students tend to be fairly homogenous with respect to observable characteristics, two of these characteristics may affect responses in the experiment: the gender of the subject and current contributions to charitable organizations. There is some empirical evidence that women are less egoistic than men (List 2004), offer more in dictator games (Eckel and Grossman 1998), and express more concern with the environment than men (Zelezny et al. 2000) although the results are mixed. There is also a recent paper that found that women are more prone to starting point bias in a choice experiment than men are (Ladenburg and Olsen 2008). In order to check for whether the subjects that are currently donating money to charitable organizations are more likely to have a higher WTP than other subjects, we estimate the random parameter logit models for the three treatments, where all the random parameters interact with two dummy variables. The first one, female, is equal to one if the subject is a female. The second one, donated today, is equal to one if the subject is currently contributing money to a charitable organization. The results are reported in Table 3. All models are estimated with simulated maximum likelihood using 500 Halton draws.

From table 3 we see that there are differences between male and female subjects and between those subjects who currently donate to charitable organizations and those that do not. Moreover, the differences between subject groups vary between the treatments. In order to get a clearer picture of the

[7] These are tested with likelihood ratio tests. When performing this test we need to account for the fact that the estimated parameters are confounded with the respective scale parameters. One way of dealing with this problem is to first test for a difference in scale between the data sets. We do this using the grid search procedure by Swait and Louivere (1993). Given the estimated scale parameter, one can then test the hypothesis of equal parameters. When estimating the random parameter model with the grid search procedure, 25 draws were used instead of 500.

differences we estimate the marginal WTP for the different subject groups; these are presented in Table 4. The standard errors for marginal WTP are calculated with the Krinsky and Robb (1986) method with 2000 draws.

There are substantial differences between male and female responses in the hypothetical-cheap-talk treatment, and more importantly, differences in the relation between hypothetical-cheap-talk and the real-money treatment. For males, the marginal WTP is between 0.46 and 1.27 SEK in the hypothetical treatment and the difference between hypothetical-cheap-talk and real-money is not statistically significant using a two-sided t-test (p-values are 0.92, 0.09 and 0.33 respectively). For females, the estimated marginal WTP is between 1.76 and 2.3 in the -cheap-talk treatment, which means that for all charities, females have a substantially higher WTP. However, this is not the case for the real-money treatment, which in turn means that there are large and significant differences between marginal WTPs for females.

Table 3. Results of random parameter models for the three treatments, *p*-values in parentheses.

	Hypothetical-cheap-talk	**Hypothetical-other**	**Real-money**
	Coeff. (*p*-value)	**Coeff. (*p*-value)**	**Coeff. (*p*-value)**
Random parameters			
Donation to WWF	0.0006	0.0008	0.0047
	(0.067)	(0.015)	(0.000)
Std dev donation to WWF	0.0022	0.0001	0.0039
	(0.000)	(0.998)	(0.000)
Donation to Unicef	0.0018	0.0020	0.0073
	(0.000)	(0.000)	(0.000)
Std dev donation to WWF	0.0012	0.0001	0.0001
	(0.000)	(0.999)	(0.992)
Donation to Red Cross	0.0011	0.0020	0.0051
	(0.000)	(0.000)	(0.000)
Std dev donation to Red Cross	0.0014	0.0001	0.0007
	(0.000)	(0.993)	(0.032)
Fixed parameters			
Donation to WWF × Female	0.0021	0.0067	-0.0001
	(0.000)	(0.000)	(0.891)
Donation to Unicef × Female	0.0016	0.0002	-0.0018
	(0.000)	(0.590)	(0.018)
Donation to Red Cross × Female	0.0025	0.00001	-0.0002
	(0.000)	(0.995)	(0.700)
Donation to WWF × Give today	0.0005	-0.00001	-0.002
	(0.181)	(0.978)	(0.005)
Donation to Unicef × Give today	0.0008	-0.0008	-0.0019
	(0.039)	(0.032)	(0.037)
Donation to Red Cross × Give today	0.0005	-0.0010	-0.0013
	(0.190)	(0.032)	(0.033)
Personal money	0.0016	0.0010	0.0085
	(0.000)	(0.0210)	(0.000)
Number of observations	1296	1152	768
Number of individuals	108	96	64
Pseudo R^2	0.03	0.01	0.06

Both groups, subjects that currently give to charitable organizations and those who do not, show a strong hypothetical bias as well. The difference in marginal WTP is significant in all six cases.

Overall our results therefore suggest that it is the female respondents that drive the results of a strong hypothetical bias in the aggregate model. It is difficult to come up with any one single explanation to this. It cannot be the case that women, in this setting, care more about the good as men actually have a higher WTP in the real-money experiment. Additionally, we should not expect any large income differences between men and women in a student-sample. It could be that women are more socially oriented than men are, and that generosity and altruism is an important part of their self-image. Thus, when responding in the hypothetical-own treatment they may tend to overstate their generosity to a greater degree than men in order to conform to their own self-image and perhaps also the image they believe that society has of women. For example in a field experiment on blood donations and monetary compensation Mellström and Johanneson (2008) find a significant crowding out effect for women but not for men. They argue that this is because women are more concerned with social esteem than men, and that the behavior in the experiment is a way to signal generosity. It could also be due to differences in effects of the cheap talk script between men and women. The script could have been effective in reducing hypothetical bias for men, but not for women.

Table 4. Marginal WTP estimates for subject groups, in SEK.

		Hypothetical-cheap-talk	**Hypothetical-other**	**Real-money**
Males	Donation to WWF	0.458	0.086	0.479
		(0.189)	(0.044)	(0.068)
	Donation to Unicef	1.267	0.261	0.792
		(0.268)	(0.049)	(0.077)
	Donation to Red Cross	0.759	0.261	0.561
		(0.197)	(0.049)	(0.050)
Females	Donation to WWF	1.760	0.261	0.468
		(0.334)	(0.039)	(0.046)
	Donation to Unicef	2.223	0.261	0.581
		(0.414)	(0.055)	(0.053)
	Donation to Red Cross	2.298	0.262	0.537
		(0.418)	(0.045)	(0.039)
Give today	Donation to WWF	1.428	0.018	0.296
		(0.315)	(0.054)	(0.077)
	Donation to Unicef	2.144	0.151	0.489
		(0.426)	(0.067)	(0.099)
	Donation to Red Cross	1.860	0.156	0.437
		(0.259)	(0.064)	(0.065)
Do not give today	Donation to WWF	1.107	0.141	0.546
		(0.218)	(0.039)	(0.044)
	Donation to Unicef	1.659	0.306	0.713
		(0.302)	(0.044)	(0.046)
	Donation to Red Cross	1.551	0.305	0.590
		(0.266)	(0.037)	(0.036)

Our findings are completely contrary to two previous studies on hypothetical bias and gender. Brown and Taylor (2000) find, using an open-ended contingent valuation survey on donations to the Nature Conservancy, hypothetical bias for both males and females. However, the hypothetical bias for males was three times larger than the one for females. Mitani and Flores (2007) find in an induced value public good game that females were more likely to reveal their true value than males when hypothetical payments are used. Since we get the opposite results, neither of the studies can be generalized.

The difference between WTP for the hypothetical-other and the real-money treatments is significant for both males and females, as well as for subjects that currently donate and those who do not donate to charitable organizations. Thus, our suggestion to use a third-person approach in order to reduce hypothetical bias has not proven to be successful in this particular experiment. Interestingly, those subjects who donate to charitable organizations, actually state an even lower marginal WTP in the hypothetical-other treatment, than other subjects.

4. Discussion

If we accept the premise that the real money treatment is the experiment that is most closely aligned to true preferences then the results from the third-person perception approach imply that the expected false consensus effect where the respondents should project their own preferences on others was unsuccessful. Instead we may have observed a self-enhancement effect where the respondents derive satisfaction from favorable social comparison.

The implication that the success of the third person perception approach is dependent on the normative motivations associated with the good is consistent with findings of Lusk and Norwood (2009b). Further, in our experiment the predicted WTP of others is even significantly lower than the WTP from the hypothetical experiment with a cheap talk script. Although we do not explicitly ask the respondents to state their own preferences as well, they may well have used their own preferences as an anchor and when predicting others to be less generous than themselves tip the generosity scale in their own favor. This better-than average-effect (Alicke and Klotz 1995), can be viewed as a type of self-serving bias in which people evaluate their own characteristics more favorably than those of others. This self-enhancement motive is central in the psychological downward comparison theory where people validate a positive sense of self by engaging in social comparison thereby obtaining feelings of well-being and self-esteem (Wills 1981). This effect may also have been augmented by the ambiguous nature of the comparison target, the “average student” permits a high level of subjectivity in the comparison process thereby allowing respondents more latitude to select downward comparison targets (Alicke and Klotz 1995; Perloff and Fetzer 1986; Weinstein 1980).

Kahneman and Sugden (2005) discuss the risk that survey questions may elicit responses that reflect attitudes rather than preferences. While this is a problem usually associated with open-ended contingent valuation surveys, the results from the hypothetical own treatment confirm this risk even for choice experiments. The higher marginal WTP for the hypothetical own treatment compared to the other treatments reflects that own money has less influence on the observed choices and that donated money and the choice of project play a

much more important role in the decision. Since the participants apparently contemplate less over own money, their responses appear to reflect attitudes rather than preferences between donated and own money. One reason for this attitude effect is probably the self-enhancement obtained from donations. This problem is especially apparent for female respondents in our survey, which is actually contrary to two previous studies (Brown and Taylor 2000; Mitani and Flores 2007). Coupled with the fact that we have quite a small student sample one should be cautious with generalizing the results and further, it indicates the need for further studies that consider differences between men and women with respect to hypothetical bias.

Acknowledgements

Financial support from the Swedish Rescue Services Agency is gratefully acknowledged. The paper has benefited from comments by three anonymous referees, Olof Johansson-Stenman, and participants at the Second Nordic Workshop in Behavioral and Experimental Economics, University of Gothenburg and the International Choice Modelling Conference, Harrogate.

References

Aadland, D. and Caplan, A. J., 2003. Willingness to pay for curbside recycling with detection and mitigation of hypothetical bias. *American Journal of Agricultural Economics*, 85(2), 492-502.

Aadland, D. and Caplan, A. J., 2006. Cheap talk revisited: New evidence from CVM. *Journal of Economic Behavior and Organization*, 60(4), 562-578.

Alicke M. D. and Klotz, M.L., 1995. Personal Contact, Individuation, and the Better-Than Average Effect. *Journal of Personality and Social Psychology*, 68(5), 804-825.

Alpizar, F., Carlsson, F. and Johansson-Stenman, O. 2008. Does context matter more for hypothetical than for actual contributions? Evidence form a natural field experiment. *Experimental Economics*, 11(3), 299-314.

Andreoni, J., 1990. Impure Altruism and donations to public goods: A theory of warm glow giving. *The Economic Journal*, 100, 464-477.

Ariely, D., Loewenstein, G. and Prelec, D., 2003. Coherent arbitrariness: Stable demand curves without stable preferences, *The Quarterly Journal of Economics*, 118(1), 73-105.

Bishop, R. and Heberlein, T., 1979. Measuring values of extra-market goods: Are indirect measures of value biased?. *American Journal of Agricultural Economics*, 61(5), 926-930.

Brown, K. and Taylor, L., 2000. Do as you say, say as you do: evidence on gender differences in actual and stated contributions to public goods. *Journal of Economic Behavior and Organization*, 43(1), 127-139.

Carlsson, F., Frykblom, P., Lagerkvist, C. J., 2005. Using cheap-talk as a test of validity in choice experiments. *Economic Letters*, 89(2), 147-152.

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(2), 179-92.

Carson, R., Flores, N. and Martin, K. and Wright, J., 1996. Contingent valuation and revealed preference methodologies. Comparing estimates for quasi-public goods. *Land Economics*, 72(1), 80-99.

Chakravarty, S., Harrison, G., Haruvy, E. and Ruström, E., 2005. Are you risk averse over other people's money?. *Working paper*, Indian Institute of Management, Ahmedabad.

Cummings, R.G. and Taylor, L.O., 1999. Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. *American Economic Review*, 89(3), 649-65

Ding, M., Grewal., R. and Liechty, J., 2005. Incentive-aligned conjoint analysis. *Journal of Marketing Research*, 42(2), 67-82.

Dunning, D.A., 1999. A newer look: Motivated social cognition and the schematic representation of social concepts. *Psychological Inquiry*, 10(1), 1-11.

Dunning, D. A. and Hayes A.F., 1996. Evidence for egocentric comparison in social judgment. *Journal of Personality and Social Psychology*, 71(2), 213-229.

Eckel, C. and Grossman, G., 1998. Are women less selfish than men? Evidence from dictator experiments. *The Economic Journal*, 108(448), 726-735.

Epley, N. and Dunning, D., 2002. Feeling "Holier than thou": Are self-serving assessments produced by errors in self- or social prediction?. *Journal of Personality and Social Psychology*, 79(6), 861-75.

Fields, J. and Schuman, H., 1976. Public beliefs about the beliefs of the public. *The Public Opinion Quarterly*, 40, 427-448.

Fisher, R., 1993. Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, 20(2), 303-315.

Harrison, G. and Rutström, E., 2002. Experimental evidence on the existence of hypothetical bias in value elicitation methods, in Smith, V. L. (ed.) *Handbook of Results in Experimental Economics*, Elsevier Science, New York

Henriksen, L. and Flora, J. A., 1999. Third-person perception and children. *Communication Research*, 26(6), 643–665.

Hsee, C. K. and Weber, E. U., 1997. A fundamental prediction error: Self-others discrepancies in risk preference. *Journal of Experimental Psychology*, 126(1), 45-53

Johansson-Stenman, O. and Martinsson, P., 2006. Honestly why are you driving a BMW?. *Journal of Economic Behavior and Organization*, 60, 129-146

Johansson-Stenman, O. and Svedsäter, H., 2008. Measuring hypothetical bias in choice experiments: The importance of cognitive consistency. *The B.E. Journal of Economic Analysis & Policy*, 8(1), *Article 41*.

Kahneman, D. and Knetsch, J. L., 1992. Valuing public goods: The purchase of moral satisfaction. *Journal of Environmental Economics and Management*, 22(1), 57-70.

Kahneman, D. and Sugden, R., 2005. Experienced utility as a standard of policy evaluation. *Environmental and Resource Economics*, 32(1), 161-181.

Krinsky, I., Robb, R., 1986. On approximating the statistical properties of elasticities. *Review of Economics and Statistics*, 68(4), 715-719.

Ladenburg, J. and Olsen, S., 2008. Gender specific starting point bias in choice experiments: Evidence from an empirical study. *Journal of Environmental Economics and Management*, 56(3), 275-285.

Levitt, S. and List, J., 2007. What do laboratory experiments measuring social preferences reveal about the real world?. *Journal of Economic Perspectives*, 20(2), 153-174.

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

List J. A., 2004. Young, selfish and male: Field evidence of social preferences. *The Economic Journal*, 114(492), 121-149.

List, J., Berrens, R., Bohara, A. and Kerkvilet, J., 2004. Examining the role of social isolation on stated preferences. *American Economic Review*, 94(3), 741-752

List, J. A. and Gallet, C., 2001. What experimental protocol influence disparities between actual and hypothetical stated values? *Environmental and Resource Economics*, 20(3), 241-254.

Lusk, J., Pruitt, J. R. and Norwood, B., 2006. External validity of a framed field experiment, *Economics Letters*, 93(2), 285-290.

Lusk, J. and Norwood, B., 2009a. An inferred valuation method. *Land Economics*, 85(3), 500-514.

Lusk, J. and Norwood, B., 2009b. Bridging the gap between laboratory experiments and naturally occurring markets: An inferred valuation method, Journal of Environmental Economics and Management, 58(2), 236-250.

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*, 86(2), 467-482.

Mellström, C. and Johannesson, M., 2008. Crowding out in Blood Donation: Was Titmuss Right?. *Journal of the European Economic Association*, 6(4), 845-863.

Messick, D. M., Bloom, S., Boldizar, J. P. and Samuelson, C. D., 1985. Why we are fairer than others. *Journal of Experimental Social Psychology*, 21(5), 469-485.

Mitani, Y. and Flores, N., 2007. Does gender matter for demand revelation in threshold public good experiments?. *Economics Bulletin*, 3(27), 1-7.

Moore, D. A. and Kim, T. G., 2003. Myopic social prediction and the solo comparison effect. *Journal of Personality and Social Psychology*, 85(6), 1121-1135.

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

Perloff, L. S. and Fetzer, B. K., 1986. Self-other judgements and perceived vulnerability to victimization. *Journal of Personality and Social Psychology*, 50(3), 502-510.

Ross, L., Greene, D. and House, P., 1977. The False Consensus Effect: An Egocentric Bias in Social Perception and Attribution Processes. *Journal of Personality and Social Psychology*, 13(3), 279-301.

Shogren, J. F., Fox, J. A. Hayes, D. J. and Roosen, J., 1999. Observed Choices for Food Safety in Retail, Survey, and Auction Markets. *American Journal of Agricultural Economics*, 81(5), 1192–1199.

Swait, J. and Louviere, J., 1993. The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of Marketing Research*, 30, 305-314.

Train, K., 2003. Discrete Choice Methods with Simulation. Cambridge University Press, New York

Weinstein, N., 1980. Unrealistic optimism about future life events. *of Personality and Social Psychology*, 39(5), 806-820.

Wills, T. A., 1981. Downward comparison principles in social psychology. *Psychological Bulletin*, 90(2), 245-271.

Zelezny, L., Chua, P. and Aldrich, C., 2000. Elaborating on gender differences in environmentalism. *Journal of Social Issues*, 56, 443-457.

Appendix 1 Experimental instructions

Hypothetical-cheap-talk

In this part of the survey we are interested in finding out how people regard different types of charities and donations to the same. We will present you with twelve different choice situations, each of which has two alternatives. Please use a cross to mark the alternative you would prefer if you found yourself in this situation. Each alternative has three characteristics.

- Type of project:
  There are three possible projects
  - The Red Cross Disaster Relief (Postal giro account 90 08 00 – 4).
    Donations are sent as relief aid to disaster areas.
  - UNICEF Children's Fund (Postal giro account 90 20 01-7).
    Donations are sent to children in need around the world.
  - World Wildlife Fund (Postal giro account 90 19 74-6).
    Donations are used for the protection of endangered species of animals, plants and the conservation of natural environments.
- Money given to the project:
  The sum of money we will donate to the given project.
- Money given to you:
  The sum of money you receive from us, with which you are free to do as you wish.

Your choice will influence how we would distribute the money between yourself and the charity as well as the type of charity that will receive the money.

For each of the choice situations below, we wish you to state which alternative **<u>you</u> would choose** if you were faced with the choice**.** Regard each question as a new situation.

Experiences from similar studies have found that peoples' responses in a survey situation often differ to how they actually act in real life. It is especially common for people to state that they are willing to donate money to a worthy charity, but later do not do so.

Hypothetical-other

In this part of the survey we are interested in finding out how people regard different types of charities and donations to the same. We will present you with twelve different choice situations where each situation has two alternatives. Please use a cross to mark the alternative you believe the average student would choose if faced with the situation. Each alternative has three characteristics.

- Type of project:
  There are three possible projects
  - The Red Cross Disaster Relief (Postal giro account 90 08 00 – 4).
    Donations are sent as relief aid to disaster areas.
  - UNICEF Children's Fund (Postal giro account 90 20 01-7).
    Donations are sent to children in need around the world.
  - World Wildlife Fund (Postal giro account 90 19 74-6).
    Donations are used for the protection of endangered species of animals, plants and the conservation of natural environments.
- Money given to the project:
  The sum of money we will donate to the given project.
- Money given to the person:
  The sum of money the person making the choice will receive from us, with which he/she is free to do as they wish.

A person can through his/her choices influence how the money is distributed between himself and the charity as well as which charity will receive the money.

For each choice situation below, we wish you to state which alternative you believe **the average student would choose** if faced with the choice**.** Regard each question as a new situation.

Real-money

In this part of the survey we are interested in finding out how people regard different types of charities and donations to the same. We will present you with twelve different choice situations where each situation has two alternatives. Please use a cross to mark the alternative you choose. Each alternative has three characteristics.

- Type of project:
  There are three possible projects
    - The Red Cross Disaster Relief (Postal giro account 90 08 00 – 4).
      Donations are sent as relief aid to disaster areas.
    - UNICEF Children's Fund (Postal giro account 90 20 01-7).
      Donations are sent to children in need around the world.
    - World Wildlife Fund (Postal giro account 90 19 74-6).
      Donations are used for the protection of endangered species of animals, plants and the conservation of natural environments.
- Money given to the project:
  The sum of money donated by us to the given project.
- Money given to you:
  The sum of money you receive from us, with which you are free to do as you wish.

You can influence through your choices how the money is distributed between yourself and the charity as well as which charity will receive the money.

For each choice situation below, we wish you to state which alternative **<u>you</u> choose.** Regard each question as a new situation as payment will be made only on one of the twelve questions.

After the session, one of the choice situations will be chosen randomly by drawing a number between 1 and 12. You will receive payment directly in accordance with your choice of alternative. The charity in the alternative you have chosen will also receive the money stated in the alternative.