Journal of Economic Behavior & Organization 84 (2012) 879–891

Contents lists available at SciVerse ScienceDirect

Journal of Economic Behavior & Organization

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

Self-image and valuation of moral goods: Stated versus actual willingness to pay

Olof Johansson-Stenman[a,*], Henrik Svedsäter[b]

[a] Department of Economics, School of Business, Economics and Law, University of Gothenburg, Box 640, SE 40530 Gothenburg, Sweden
[b] Organisational Behaviour, London Business School, Regent's Park, London NW1 4SA, United Kingdom

ARTICLE INFO

Article history:
Received 9 January 2011
Received in revised form 14 March 2012
Accepted 9 October 2012
Available online 22 October 2012

JEL classification:
C91
D63
Q5

Keywords:
Stated-preference methods
Choice experiment
Hypothetical bias
Self-image
Non-market valuation
Warm glow

ABSTRACT

Hypothetical bias in stated-preference methods appears sometimes to be very large, and other times non-existent. This is here largely explained by a model where people derive utility from a positive self-image associated with morally commendable behavior. The results of a choice experiment are consistent with the predictions of this model; the hypothetical marginal willingness to pay (*MWTP*) for a moral good (contributions to a WWF project) is significantly higher than the corresponding real-money *MWTP*, whereas no hypothetical bias is seen for an amoral good (a restaurant voucher). Moreover, the evidence suggests that also the real-money *MWTP* for the moral good is biased upwards, in the sense that it appears to be higher within than outside the experimental context.

1. Introduction

What determines people's responses in stated-preference (SP) surveys that target issues with a perceived ethical dimension, such as valuation of environmental and various other types of public goods? And to what extent can we interpret those responses as being representative of underlying preferences? These questions are crucial from a policy perspective, in particular in the US and an increasing number of European countries, where cost-benefit analysis, often making use of SP methods, is compulsory for all major proposed regulations. Although most researchers probably agree that there is potential scope for overstatement in various kinds of SP studies, no consensus exists on whether this is a major problem, or on how hypothetical estimates could or should be calibrated to better represent underlying preferences. Perhaps more importantly, few studies have investigated for which types of goods and under what circumstances hypothetical bias is likely to occur, and why this is the case.

In this paper, we develop and test a theoretical model aimed at explaining variations of hypothetical bias in the literature. Drawing on papers by Andreoni (1989, 1990), Kahneman and Knetsch (1992), Akerlof and Kranton (2000), Brekke et al. (2003), Santos-Pinto and Sobel (2005), and Nyborg and Brekke (2010), the model proposes that people, in addition to the instrumental benefits associated with a good, derive utility from a positive self-image. This, in turn, is influenced by (*i*)

* Corresponding author. Tel.: +46 31 786 25 38; fax: +46 31 786 10 43.
E-mail addresses: Olof.Johansson@economics.gu.se (O. Johansson-Stenman), hsvedsater@london.edu (H. Svedsäter).

0167-2681/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jebo.2012.10.006

the degree to which stated or real behavior coincides with the respondents' ethical views, and (*ii*) the extent to which respondents are honest with themselves. The model predicts that in SP studies people overstate their marginal willingness to pay (*MWTP*) for goods with a perceived ethical dimension, denoted moral goods, but not for morally neutral goods. The model furthermore suggests that also the elicited real-money *MWTP* exaggerates people's valuation of a moral good, although to a lesser extent.

In order to test these predictions, we conduct a choice experiment (CE) assessing people's valuation of what we refer to as a moral and an amoral good, respectively. A CE is an SP method where the respondents make repeated choices between bundles of goods. The method has been increasingly used to value non-market goods (see, e.g., Louviere et al., 2000; List et al., 2006). The moral good is here represented by a donation to a campaign administered by the World Wildlife Fund (WWF) to help save the Asian Elephant, and the amoral good is a voucher valid at a local Italian restaurant in Gothenburg, Sweden. The CE is then compared with the outcome of a similar exercise, based on another but similar sample drawn from the same underlying student population, only this time using real instead of hypothetical monetary trade-offs. The empirical results are consistent with the predictions of our model; the stated *MWTP* for the moral good (the WWF campaign) is significantly higher than the corresponding real-money *MWTP*, whereas no difference is found between stated and real-money WTP for the amoral good (the restaurant voucher). In following up on these findings, we illustrate how also the real-money CE exaggerates people's valuation of the moral good, in the sense that the experimental situation per se seems to induce a positive bias.

Section 2 presents a brief review of hypothetical bias in SP studies and of relevant psychological and behavioral economics literature that helps to explain past empirical results. Section 3 presents a formalized model and derives testable hypotheses, whereas Section 4 outlines the CE design for assessing the value of our moral and amoral good. The empirical results are presented in Section 5, while Section 6 discusses the findings in a broader context.

2. Literature review

2.1. The existence of hypothetical bias

The extent to which *WTP* statements correspond with real-money payments is often seen as the ultimate validity test of SP methods. List and Gallet (2001) and Murphy et al. (2005) conducted meta-studies on observed disparities between hypothetical and real-money *WTP* in contingent valuation (CV) studies, and reported that hypothetical *WTP* generally exceeds real-money *WTP*, and that the difference tends to be larger for public than for private goods. Murphy et al. (2005) also found a much lower hypothetical bias in studies that relied on a within-subject test of hypothetical and real-money *WTP* than in studies making split-sample comparisons between subjects.[1]

However, some other studies report no statistically significant differences between hypothetical and real-money *WTP*. Of particular relevance to our work are Carlsson and Martinsson (2001) and Cameron et al. (2002), who used CEs to value what we here denote moral goods. In Carlsson and Martinsson (2001), the respondents first made 16 hypothetical pair-wise choices and then 16 similar (but not identical) pair-wise choices with real-money implications. No significant difference was found between hypothetical and real-money marginal *WTP* for donations to a variety of environmental projects, although the former was 10–15 percent higher than the latter. Cameron et al. (2002) tested several elicitation formats in a comprehensive study and found that the mean *WTP* was between 30 and 330 percent larger in hypothetical CEs. However, due to large error terms, a common underlying preference structure could not be rejected.[2]

For some environmental goods, such as access to recreation sites or hunting rights, it is possible to compare SP methods with revealed preference (RP) methods, for instance by using travel-cost or hedonic-pricing methods. In a meta-analysis by Carson (1996), values obtained from RP studies were found to be of the same order of magnitude as those derived using dichotomous-choice CV studies. Risk and time valuations are other examples where both SP and RP methods are routinely used. In another large meta-analysis, Kochi et al. (2003) found that CV studies on average result in significantly lower values of statistical lives than studies based on the hedonic-pricing method. Finally, Wardman (2001) performed a meta-study of British value-of-time studies and found relatively small differences, although SP studies on average yielded somewhat lower values.

2.2. Possible explanations behind hypothetical biases

The most frequently assumed reason for a positive hypothetical bias is that respondents simply do not take hypothetical questions seriously. However, if this assumption were correct, we would expect to see a greater variance of bids and not a systematic bias upwards. Given the above empirical patterns and the high policy relevance, it appears worthwhile to investigate more systematically for which goods and under what circumstances hypothetical statements are and are not likely to be biased. Moreover, we would like to have an intuitively plausible theory for *why* overstatements frequently occur in some contexts but not in others.

[1] For a direct test of hypothetical bias in within- and between-subject designs, see Johansson-Stenman and Svedsäter (2008).

[2] Presumably, this was partly due to the fact that there was much less variation in the real-money bids; cf. Carlsson and Johansson-Stenman (2010).

There is much evidence from psychology, and more recently from behavioral economics, that people like to have a positive self-image, and that they try to maintain this image in various ways (Gilovich, 1991; Baumeister, 1998). Consistent with this, it has been found that most people believe that they perform a variety of tasks better than the average person (e.g., perceiving themselves as better drivers, or that they are smarter). Central to our argument here is that moral identity is part of an individual's self-image (e.g., Aquino and Reed II, 2002), which is often associated with certain beliefs, attitudes, and types of behavior (Shih et al., 1999; Forehand et al., 2002). The fact that people prefer to see themselves as more socially responsible than others (e.g., Gilovich, 1991; Taylor and Brown, 1994) only serves to illustrate the importance people attach to moral identity. In this context, Johansson-Stenman and Martinsson (2006) asked people about what characteristics they considered to be important when buying a car. Whereas most people claimed environmental characteristics to be very important, very few emphasized the status associated with a specific brand or model. Interestingly, when asked about what characteristics they believed were important *for others*, the reverse pattern emerged insofar as status became much more important and environmental aspects less important. This underlines the desirability of this trait, and the tendency to view oneself as "better" than others in this respect. Similar findings are reported by Brekke et al. (2003). When investigating people's motivation behind recycling in a Norwegian survey, as many as 73 percent of the respondents answered that one of their main reasons was that they would like to see themselves as responsible citizens.

Provided that a high *WTP* is seen as something honorable, and hence improving a person's self-image, it follows that people have an incentive based on self-deception to overstate their *WTP*. This obviously applies to both stated and real-money *WTP*, but since nothing needs to be paid in a hypothetical context, a positive hypothetical bias is logical. Our reasoning also corresponds with psychological theories arguing that people derive value from merely expressing certain opinions or attitudes (e.g., Katz, 1960; Herek, 1986), particularly under circumstances when these are not binding or directly tied to outcomes (Kahneman and Knetsch, 1992; Bodner, 1995). Hence, when a verbal statement is free of charge, more emphasis will be placed on maintaining a positive self-image than when economic costs are involved. The self-image motive proposed here also helps explain the observed pattern of higher hypothetical bias for public goods, since these goods often have moral implications. Arguably, the preference for saving wild animals from extinction is built on different premises and is more strongly associated with a moral code of conduct than, say, private access to fishing or hunting rights or consumption of chocolate bars.[3]

However, since we do not observe infinite *WTP*s in hypothetical SP studies even for clearly ethical issues, some moderating factor must play a role. Our presumption is that people simultaneously want to be honest with themselves, knowing that there is a limit as to how much they are able or willing to commit. Assuming that you won 100,000 USD in a lottery, how much of this would you donate to charity? Even if we agree that the most honorable thing to do would be to donate all of it, most of us would not and, hence, claiming to be willing to do so in a hypothetical survey would probably make us feel dishonest. People's desire to be honest to themselves can also contribute to our understanding of why so-called cheap-talk scripts tend to result in lower hypothetical bias, since they aim to make respondents more honest and realistic in their answers; see, e.g., Cummings and Taylor (1999) and List et al. (2006), respectively, for CV and CE applications of such scripts.

It is worth emphasizing that the self-image motive proposed here should not be confused with preference falsification or with the willingness to impress or provide informative signals to other people (Bernheim, 1994; Kuran, 1995; Neilson, 2009). List et al. (2004), for example, showed that CV respondents are much more willing to vote in favor of a costly environmental project if others are informed about their choice. In the present paper the assumed driving force can instead be seen as self-signaling (Bodner and Prelec, 2003), implying that opinions and actions provide signals to ourselves as to what kind of person we are, including our intentions toward the matter at stake. Like Adam Smith (1759) and Benabou and Tirole (2006), we may think of an individual who makes moral decisions by assessing his/her own conduct from the perspective of how an ideal person would act in a certain situation, regardless of whether or not his/her actions are being observed. The assumed mechanism is therefore effective also in highly anonymous contexts, as in the experiment conducted here, and does not rely on whether actions are publicly known.[4]

Finally, we are not arguing that there can be no role for pure altruism in motivating people to contribute to charity. Indeed, Harbaugh et al. (2007) present brain scanning evidence suggesting that also mandatory transfers to a charity elicit neural activity in areas linked to reward processing. However, they also found that neural activity further increases when people make transfers voluntarily, suggesting that warm-glow motives seem to matter beyond pure altruism.

3. The theoretical model

Our model can be seen as an extension of Andreoni's (1989, 1990) model, where people derive a "warm glow" from contributing to a "good cause" (which public goods are often seen as), and of the idea developed by Kahneman and Knetsch

[3] Hypothetical bias for purely private goods such as chocolate bars or sunglasses can obviously not be explained by the self-image effects discussed here (e.g., Cummings et al., 1995; Lusk and Schroeder, 2003). One plausible interpretation of such findings is that some respondents are actually answering a slightly different question than the one being asked. In order to make sense of the inquiry raised, they may for example ask themselves "How much would I be willing to pay if I were to buy a pair of sunglasses today?"

[4] For recent applications and theorizing of similar motivational sources in economics, see e.g., Murnighan et al. (2001) and Benabou and Tirole (2002, 2004, 2006). Likewise, for measures of the relative strength of extrinsic versus intrinsic motives of voluntary acts, see Alpizar et al. (2008) and Lacetera and Macis (2010).

(1992) that people's value statements in SP surveys represent the "purchase of moral satisfaction." A key feature of these models is that people gain utility intrinsically from their own contributions but not from those of others. The extensions suggested here consist of specifying why and when people receive such a warm glow, and of considering the limits posed by people's contention to be honest.

3.1. The conventional baseline model

Consider first a conventional strictly increasing and strictly quasi-concave utility function as follows:

$$U = u(Money, Rest, WWF), \tag{1}$$

where *Money* is private income and *Rest* and *WWF* represent money for a restaurant voucher and a WWF campaign, respectively. The true marginal willingness to pay for *Rest* in terms of *Money* is then given by:

$$MWTP_{Rest}^{true} \equiv -\frac{dMoney}{dRest}\bigg|_u = \frac{\partial u/\partial Rest}{\partial u/\partial Money}, \tag{2}$$

i.e., the marginal rate of substitution between *Rest* and *Money*. Similarly, the true marginal willingness to pay for *WWF* in terms of *Money* is given by:

$$MWTP_{WWF}^{true} \equiv -\frac{dMoney}{dWWF}\bigg|_u = \frac{\partial u/\partial WWF}{\partial u/\partial Money}. \tag{3}$$

$MWTP_{WWF}^{true}$ can then be interpreted as the amount of money that can be withdrawn from the individual per unit of money added to WWF *by someone else* in order to keep utility constant for the individual. As such, it can be seen as an individual utility measure that would result from a change outside the experimental context, e.g., through a governmental policy change toward WWF.

In most SP experiments aimed to measure the value of a change in a public good, respondents are assumed to maximize a function u as above, so that for a good i, the subjects' stated *MWTP*, $MWTP_i^{stated}$, equals the true value outside the experimental context, i.e., such that $MWTP_i^{stated} = MWTP_i^{true}$.

3.2. An extended model with concerns for self-image

We will here conjecture that people, when responding to the choice experiments, maximize a utility function that in addition to the direct material effects also depends on the subjects' self-image, s. In our case we have two treatments, one with hypothetical money and one with real money. We will denote the corresponding stated marginal willingness to pay measures in these treatments as $MWTP_i^{hyp}$ and $MWTP_i^{real}$, respectively. By including self-image effects, s, into the model (following Akerlof and Kranton, 2000, 2002; Brekke et al., 2003; Santos-Pinto and Sobel, 2005; Johansson-Stenman and Martinsson, 2006; Alpizar et al., 2008; Nyborg and Brekke, 2010), we have instead the following utility function:

$$V = v(U, s) = v(u(Money, Rest, WWF), s), \tag{4}$$

where $\partial v/\partial s > 0$. It is thus assumed that people's utility, in addition to changes in *Money* and *WWF*, depends on how their self-image is affected by their intentions and actions. An individual's self-image is of course influenced by many factors. In the model proposed here, s is assumed to depend on (i) the degree to which individuals act in accordance with their ethical beliefs (*ethics* for short), and (ii) the extent to which they are honest with themselves (*honesty* for short).[5] These entities are then modeled as differences between the stated *MWTP* and the morally superior and true *MWTP*, respectively, as follows:

$$s = f(d_{Rest}^{ethics}, d_{Rest}^{honesty}, d_{WWF}^{ethics}, d_{WWF}^{honesty}), \tag{5}$$

where $d_i^{ethics} \equiv \left|MWTP_i^{stated} - MWTP_i^{moral}\right|$ is the absolute value of the difference between the stated *MWTP* for good i and its corresponding ethically superior value, $MWTP_i^{moral}$, defined as the value that would maximize the respondent's self-image should there be no conflicts with other determinants of self-image. Similarly, $d_i^{honesty} \equiv \left|MWTP_i^{stated} - MWTP_i^{true}\right|$ is the absolute value of the difference between stated and true *MWTP*, where the latter is defined as the resulting *MWTP* when holding s constant (or, equivalently, the amount of money that could be taken away from the individual per dollar given by *someone else* to the good). Let us also assume that $\partial f/\partial d_i^{ethics} < 0$ for $d_i^{ethics} > 0$, $\partial f/\partial d_i^{ethics} = 0$ for $d_i^{ethics} = 0$, $\partial f/\partial d_i^{honesty} < 0$ for $d_i^{honesty} > 0$, $\partial f/\partial d_i^{honesty} = 0$ for $d_i^{honesty} = 0$, $\partial^2 f/\partial (d_i^{ethics})^2 < 0$, $\partial^2 f/\partial (d_i^{honesty})^2 < 0$, where the shape of the second derivatives ensures a unique optimum. Thus, in a survey of a particular good, a statement deviating from both the morally superior value and the true value causes a disutility for the individual. In our case we have two goods to be valued, so that

[5] While self-image is modeled as a direct argument in the utility function, there may of course also exist indirect benefits of a positive self-image, for example through status and signaling effects (cf. Ball et al., 2001).

$$s = f\left(\left|MWTP_{Rest}^{stated} - MWTP_{Rest}^{moral}\right|, \left|MWTP_{Rest}^{stated} - MWTP_{Rest}^{true}\right|, \left|MWTP_{WWF}^{stated} - MWTP_{WWF}^{moral}\right|, \left|MWTP_{WWF}^{stated} - MWTP_{WWF}^{true}\right|\right). \quad (6)$$

Let us furthermore define a good i to be a *moral good* if and only if $MWTP_i^{moral} > MWTP_i^{true}$. Correspondingly, a good i is defined to be an *amoral good* if and only if $MWTP_i^{moral} = MWTP_i^{true}$.[6] Note that the model for analytical simplicity is static in nature, and hence we do not model the *process* in which self-image is created, but focus instead on the equilibrium conditions. Moreover, in the model it seems that people implicitly know the morally superior value as well as the true value, and more generally who they are and what they believe in. This is of course not fully realistic, and should not be taken literally. Instead, the model should either be seen as reflecting self-deception, where the utility maximization in part reflects unconscious processes, or as a reduced form of a model with self-signaling, as in Benabou and Tirole (2006). In the latter case, people have only imperfect knowledge of themselves and use statements as well as real actions to signal to themselves who they are. Naturally, signaling to themselves that they are good and responsible people is more costly in the real-money treatment than in the hypothetical treatment.

3.3. Derived hypotheses

Assuming money to the WWF campaign to be a moral good and the restaurant voucher to be an amoral good, we are able to derive the following hypotheses (see Appendix for proofs):

Hypothesis 1. Given that money to the WWF campaign constitutes a moral good, we have: $MWTP_{WWF}^{true} < MWTP_{WWF}^{real} < MWTP_{WWF}^{hyp} < MWTP_{WWF}^{moral}$.

Hypothesis 2. Given that the restaurant voucher constitutes an amoral good, we have: $MWTP_{Rest}^{true} = MWTP_{Rest}^{real} = MWTP_{Rest}^{hyp} = MWTP_{Rest}^{moral}$.

Thus, the model predicts that hypothetical *MWTP* exceeds real-money *MWTP* for goods with moral implications, but not otherwise. Since the hypotheses are tested in a between-subject design, it excludes the influence of peoples' desire to appear consistent as demonstrated in previous within-subject designs of such hypotheses (e.g., Johansson-Stenman and Svedsäter, 2008). Also, the experimental set-up provides a clear distinction between public and private goods that can help to clarify some conflicting results observed previously (e.g., List and Gallet, 2001; Carlsson and Martinsson, 2001). The model further adds to similar theorizing in the literature (e.g., Andreoni, 1989, 1990; Kahneman and Knetsch, 1992), most importantly by setting hypothetical against real-money WTP and by invoking the concept of self-image as a key driver of ethically commendable behavior. Another interesting feature of the model is that it calls into question the validity of real monetary trade-offs assessed in an experimental context and the potential problems that arise for cost-benefit analysis, since the model predicts that $MWTP_{WWF}^{true} < MWTP_{WWF}^{real}$. In other words, for a small increase in WWF, the individual would be willing to pay more in the real-money experimental context than he/she would implicitly be willing to pay outside this context where the changes in WWF are independent of the individual's own action. The intuition is as follows: Consider first an individual who obtains a positive self-image from making a monetary contribution toward an environmental improvement in an experimental context. Consider next the same individual, but this time facing a compulsory tax increase that is combined with an identical environmental improvement. Would he/she then also obtain the same self-image benefit? Presumably not since this is a forced outcome, implying that experimentally elicited values may lead to an overestimation of the benefits most relevant for cost-benefit analysis.

In the next section we outline an experimental design in order to test key parts of Hypotheses 1 and 2, namely whether $MWTP_{WWF}^{real} < MWTP_{WWF}^{hyp}$ and whether $MWTP_{Rest}^{real} = MWTP_{Rest}^{hyp}$. In contrast, $MWTP_{WWF}^{true} < MWTP_{WWF}^{real}$ will not be tested directly, since we cannot observe $MWTP_{WWF}^{true}$. Instead we will, starting from (1), use conventional economic theory to derive some implications of our $MWTP_{WWF}^{real}$ as measured experimentally. It is then possible to perform a thought experiment and judge whether or not these implications are reasonable. Suppose we obtain that $MWTP_{WWF}^{real} = a$, i.e., that a representative subject is willing to forsake at most a USD for an additional dollar to WWF. Given that the subject's preferences are given by (1), which is strictly quasi-concave, and that these preferences reflect utility changes inside as well as outside of the experiment, we can conclude that the individual should then be better off in status quo compared to in a situation where the individual receives a 1000 dollar bills falling from the sky and where WWF receives a budget cut of 100,000 USD due to changed priorities of a large external donor. To see this, consider the status quo situation A in Fig. 1 given by the allocation $\{Money^0, WWF^0\}$, where the slope of the indifference curves in this point equals $-a$.

If the indifference curves would be straight lines, an individual would be indifferent between A and B, i.e., between status quo and a combined change such that the WWF would experience a budget cut of X USD and the individual would obtain

[6] It is then natural to define a good as an *immoral good* if and only if $MWTP_i^{moral} < MWTP_i^{true}$, although we will not make use of this case in the current paper.

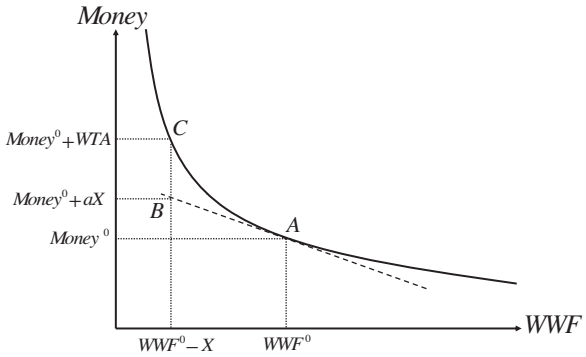

Fig. 1. Indifference curve in *Money*–*WWF*-space.

an income increase of aX USD (for all positive a and X).[7] Given strict quasi concavity, we know that the indifference curves are instead convex toward the origin, implying that the individual would be indifferent between A and C, such that the individual would strictly prefer status quo to the combined change in money to the WWF and the individual. At the end of Section 5, we will reflect on whether or not this seems plausible for the estimated parameter a.

4. Experimental design

The group of subjects consisted of students at the University of Gothenburg enrolled in a wide range of different undergraduate and graduate courses. They were recruited from a pool of students volunteering at the beginning of each semester to participate in experiments run by the university. A total of 160 students chose to participate in the experiment reported here, which was conducted in individual sessions with one subject at a time. The average age of the subjects was 26.84 years (SD = 7.59 years), and 56 (35.0 percent) were men and 104 (65 percent) women.

The design is largely based on those used by Carlsson and Martinsson (2001) and Johansson-Stenman and Svedsäter (2008), with some important modifications. Carlsson and Martinsson (2001) utilized a within-subject design CE with a moral good, such that each respondent first made a number of hypothetical pair-wise choices followed by the same number of real-money pair-wise choices. Johansson-Stenman and Svedsäter (2008) hypothesized that the fact that Carlsson and Martinsson (2001) could not reject the null hypothesis of no hypothetical bias may in part be due to the fact that they used a within-subject design, and that real-money WTP may be influenced by previously expressed hypothetical WTP. In other words, subjects may have a preference for acting consistent over time, i.e., they want to behave in accordance with previously expressed statements of how they would act. Johansson-Stenman and Svedsäter (2008) tested this hypothesis by using both a within- and a between-sample design. They found, as hypothesized, a much larger hypothetical bias in the between-sample test compared to the within-sample test. Overall, this means that there seems to be a substantial hypothetical bias in CEs with moral goods.

Yet, as discussed previously in this paper, many studies have found no or small hypothetical biases for amoral goods. In order to experimentally test the hypothesis that there is a hypothetical bias for moral but not for amoral goods in CEs, we clearly need a CE design with two different goods, one moral and one amoral, and for each good one hypothetical and one real-money treatment. Such a design is therefore used here, but has, as far as we know, not been utilized before. Moreover, in order to avoid confounding effects of individual consistency across experiments, we use a between-sample design.

The subjects were randomly divided into two sub-samples. Subjects in one sub-sample made hypothetical choices, whereas subjects in the other made choices involving real monetary trade-offs. The sessions started with two questions about gender and age. The subjects then received verbal and written instructions about the choice experiment, and were informed that its main purpose was to assess the value people place on various goods and services, in this case a voucher valid at a local restaurant and a donation to the WWF to help save the Asian Elephant.[8] Accordingly, they were later presented a number of trade-offs associated with these goods. The characteristics of the restaurant and the features and purposes of the WWF campaign were explained. In the hypothetical setting, the subjects were instructed to answer the questions as truthfully as possible, carefully taking into account how much each good was actually worth to them and how much they could afford to contribute. They were also informed that hypothetical valuation of environmental goods and services

[7] Algebraically, we can express the individual's minimum willingness to accept a WWF budget of X USD as follows: $WTA = -\int_{WWF^0-X}^{WWF^0} \frac{dMoney}{dWWF}\Big|_{u=u^0} dWWF \geq -X \frac{dx(Money^0, WWF^0)}{dWWF}\Big|_{u=u^0} = aX$, where we have used the fact that the slope of the indifference curve is most flat in the interval at $(Money^0, WWF^0)$, which in turn follows directly from the quasi-concavity assumption.

[8] The specific restaurant used in this study was an Italian mid-priced restaurant that the vast majority of the students at the university were familiar with.

	Alternative A	*Alternative B*
Money that you will receive	40	120
Voucher at local restaurant	120	0

	Alternative A	*Alternative B*
Money that you will receive	160	40
WWF donation	0	240

Fig. 2. Examples of choice sets used.

is commonly used as a means to inform public policy making, thereby emphasizing the importance of providing realistic answers. The instructions to the subjects are presented in Appendix B.

There were five cash payment levels offered to the subjects (SEK 0, 40, 80, 120, 160) and five levels of the donation or voucher value (SEK 0, 60, 120, 180, 240). The specific amounts were chosen on the basis of a pre-test carried out prior to the main study, involving 20 respondents. Altogether 32 unique choice sets were constructed from these amounts, 16 that valued a restaurant visit and 16 a WWF donation. The 32 choice sets were divided into two blocks of choice sets, Block A and Block B. Each subject was faced with 16 choice sets (i.e., either with Block A or Block B). Eight choice sets in each block concerned a trade-off between cash payment and a restaurant voucher, and the remaining eight a trade-off between a cash payment and a donation to the WWF campaign.

Finally, the order of the choice sets within each block was varied by either presenting eight choice sets valuing the restaurant visit first and then eight valuing the WWF campaign, or vice versa. The two blocks and the different orders of choice sets were balanced across the two sub-samples, hence occurring with the same frequency in the hypothetical and real-money treatment. In order to minimize any potential differences in behavior between men and women, subjects were divided across samples in order to achieve identical gender ratios in the hypothetical and real-money scenarios (28 men and 52 women in each sample).

In each choice-set, the subjects were asked to choose between two alternatives, A and B. In eight choice sets, each alternative specified the amounts of money that the subject and the WWF would receive. In the other eight choice sets, each alternative instead specified how much money the subject would receive, and the value of the restaurant voucher. The subjects were specifically told to make each choice independently of the others, and that there were no direct associations between them. Fig. 2 presents two examples of the choice sets used. The complete list of choice sets can be found in Appendix C.

Subjects in the hypothetical treatment were compensated with a fixed show-up fee of SEK 50. The participants in the real-money treatment, on the other hand, were informed prior to the task informed that the actual monetary payoffs and donation or restaurant voucher would be based on one of the sixteen choice sets, and that after completing the task, a number from 1 to 16 would be randomly drawn under the supervision of the participant. They were further told that this number would specify the choice set that would decide the actual payoffs. For example, if number 2 were drawn, then the choice made by the subject in the second choice set counting from the beginning of the questionnaire would determine that person's cash payment, and similarly for the value of the restaurant voucher and the size of the WWF donation. In this way it was in the subjects' interest to treat each choice in isolation, i.e., as if it were the only choice that needed to be made.

Before placing the questionnaire in an envelope, the subjects had the opportunity to check how much money they were to receive and whether a restaurant voucher was to be issued or a donation was to be made (and the exact amount of this). They were also informed that the actual donation would be administered and sent to the WWF by the research team once the data collection had been completed. A double-blind procedure was used in order to ensure respondent anonymity. Each subject was given a ticket with the same number as printed on the envelope containing his/her questionnaire. The sealed envelopes were then transported to a department secretary, who did not know anything about the purpose of the experiment or about who each subject was. This person opened the envelope, checked how much money was owed, and put cash and potentially a restaurant voucher in another envelope with the same number written on it. The subjects were then given a date when they could exchange their tickets for the envelopes containing their compensation.

Note that there are no direct free-riding problems associated with our design with respect to measuring preferences for public goods. In contrast, it is widely claimed in the environmental economics literature that people tend to overstate their true WTP for a public good if they believe that they would not have to pay (their share) for the good in reality. Yet, in this study the subjects knew that in order to increase the amount of money given to the WWF projects, they had to pay for it.

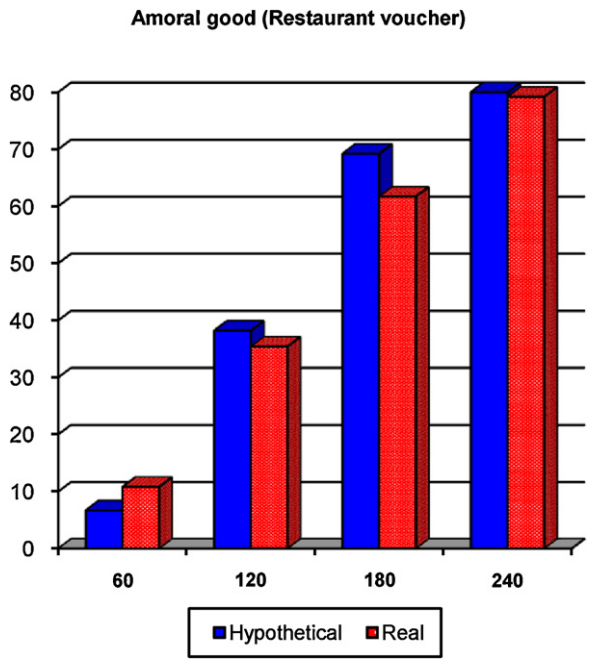

Fig. 3. Relative frequency of choices of the amoral good (restaurant voucher) versus cash payment. The horizontal axis displays the value of the voucher.

5. Results of the choice experiment

5.1. Descriptive results

Figs. 3 and 4 display the relative frequency of choices favoring a restaurant voucher and a donation to the WWF campaign over cash payment, summarized across all subjects. The horizontal axes indicate the value of the restaurant voucher and the size of the donation to the WWF (in SEK), respectively. The frequencies represent the average over all levels of trade-offs in cash payments.

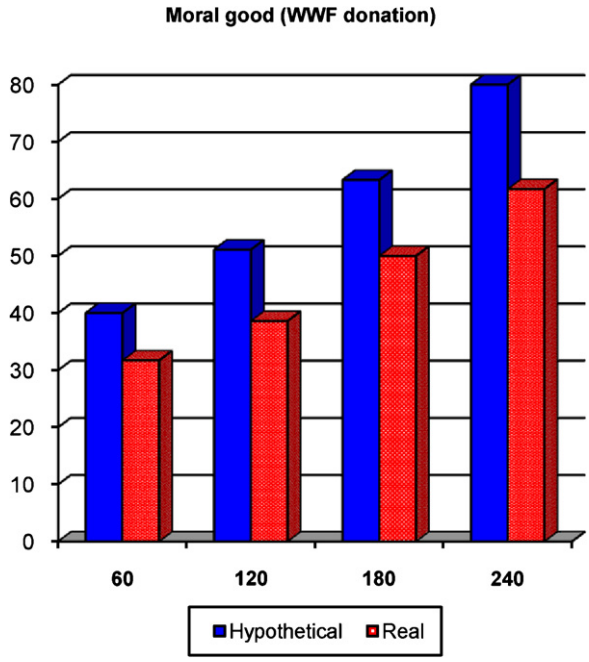

Fig. 4. Relative frequency of choices of the moral good (WWF donation) versus cash payment. The horizontal axis displays the value of the donation.

As expected, the greater the value of the restaurant voucher or the WWF donation for all levels of cash payments taken together, the more often this alternative is chosen. More importantly for our purposes, apart from the SEK 180 level where the restaurant voucher is slightly (but not significantly) more favored in the hypothetical than in the real-money scenario, the frequency of choosing the amoral good is roughly the same in the hypothetical and in the real-money setting across all levels of voucher value. For the moral good, on the other hand, the frequency is always greater in the hypothetical than in the real-money context.

Chi-square tests indicate that the relative frequency of hypothetical choices favoring the WWF campaign over cash payment is significantly higher than equivalent choices made in the real-money context for all but one donation level ($p < 0.01$ for WWF donations of SEK 120 and 240, $p < 0.05$ for a donation of SEK 180, and $p = 0.11$ for a donation of SEK 60), thus broadly confirming our hypotheses. Yet, these results are just indications, and in the next sub-section we will across treatments simultaneously account for the *difference* between money to the individual and to the goods (i.e., the value of the restaurant voucher and the WWF donation, respectively) based on a conventional random utility framework and pooled logit regressions.

5.2. Random utility model and logit regressions

In the econometric analysis we rely on a standard random-utility framework, assuming that each individual has a utility function consisting of a systematic part, V, and a random unobservable term, ε. The utility derived for individual i from choosing a given alternative, say Alternative 1, therefore becomes:

$$u_{i1} = V_{i1} + \varepsilon_{i1}. \tag{7}$$

The probability that i chooses Alternative 1 then equals the probability that the utility from this alternative is greater than the utility of Alternative 2, i.e.,

$$\Pr(A_i = 1) = \Pr(V_{i1} + \varepsilon_{i1} > V_{i2} + \varepsilon_{i2}) = \Pr(\varepsilon_{i1} - \varepsilon_{i2} > V_{i2} - V_{i1}), \tag{8}$$

where the differences between the error terms in (8) are assumed to be logistically distributed. The systematic part of the utility function, associated with either alternative, is assumed to be linear in the attributes in the interval considered:

$$V_i = \alpha + \beta Money_i + \rho Rest_i + \gamma WWF_i + \lambda HYP\,Rest_i + \mu HYP\,WWF_i, \tag{9}$$

where $Money_i$, $Rest_i$, and WWF_i represent money to the respondent, money to the respondent in the form of a restaurant voucher, and money to the WWF project resulting from choosing alternative 1. *HYP* is a dummy variable that takes the value 1 in the hypothetical treatment and 0 in the real-money treatment. Given the assumed error distribution, the parameters associated with this model (except for the intercept, which cancels out) can be estimated with a logit model (see, e.g., Louviere et al., 2000, for a good state-of-the-art overview of the analysis of stated choice models). Note that (8) and (9) together imply that the probability that a certain alternative is chosen will depend on (i) the difference in money to the respondent, (ii) the difference in the value of the restaurant vouchers offered, and (iii) the difference in the WWF donation between the alternatives.

Since each individual made 16 choices in total, the statistical observations are not independent, implying that the standard errors of a basic logit regression, as outlined above, are biased downwards. We deal with this in two different ways: by clustering the error terms at the individual level (using the cluster-command in Stata) and by utilizing Random Effects and Fixed Effects Models.[9]

Table 1 presents the parameter estimates of pooled regression models, with and without gender interaction variables.[10] The results are in the expected direction according to our theoretical predictions. The parameters associated with money given to subjects, a restaurant voucher, and a donation to WWF, respectively, are all positive and significant at the 0.01 level across all models. Moreover, the interaction effect between the hypothetical treatment and a WWF donation is significant at the 0.01 level. This implies that, across all models, the likelihood of trading off a cash payment in favor of a donation to the WWF is significantly greater in the hypothetical than in the real-money context. Conversely, no significant difference between the hypothetical and the real-money treatment is found for the restaurant voucher.[11]

Furthermore, we find no significant gender differences in hypothetical bias in any of our three model specifications. Thus, according to the results here, the hypothetical bias demonstrated above for the public (but not the private) good seems gender independent. This can be compared to the mixed pattern in the existing literature: Brown and Taylor (2000) found in a CV study that men on average had a much higher hypothetical bias, whereas Carlsson et al. (2010) found the opposite

[9] Where a choice favoring the good over money is always coded "1" and a choice favoring money over the good is always coded "0."

[10] Since the analysis is based on a student sample, we do not include age as an explanatory variable in the analysis. Moreover, although age is relatively high on average, with a non-negligible standard deviation, this is largely driven by a few outliers in the sample.

[11] As suggested by a referee and in order to check the robustness of our findings, we have included in the specifications (not shown) a dummy variable on the hypothetical question itself (except for the fixed effect specifications where such a dummy variable would clearly eliminate the within-group variation). Yet, the results regarding both parameter sizes and standard errors with such a dummy variable are almost identical to the ones reported in Table 1, both with and without gender interaction variables.

Table 1
Estimated parameters from pooled regression models (standard errors in parentheses).

	Logit, clustered, robust standard errors		Random effects logit		Fixed effects logit	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Main variables						
Money to subjects, β	0.027***	0.027***	0.035***	0.035***	0.035***	0.035***
	(0.0016)	(0.0016)	(0.0016)	(0.0016)	(0.0018)	(0.0018)
Restaurant voucher money, ρ	0.014***	0.014***	0.019***	0.018***	0.019***	0.018***
	(0.0013)	(0.0015)	(0.0012)	(0.0014)	(0.0015)	(0.0018)
WWF donation money, γ	0.013***	0.012***	0.017***	0.015***	0.017***	0.016***
	(0.0012)	(0.0014)	(0.0012)	(0.0014)	(0.0015)	(0.0026)
Hypothetical treatment times restaurant voucher, λ	0.0014	0.0018	0.0021	0.0022	0.0023	0.0016
	(0.0014)	(0.0018)	(0.0015)	(0.0015)	(0.0021)	(0.0021)
Hypothetical treatment times WWF donation, μ	0.0049***	0.0063***	0.0066***	0.0078***	0.0068***	0.0072***
	(.0017)	(.0020)	(0.0015)	(0.0019)	(0.0022)	(0.0026)
Gender interaction variables						
Male times restaurant voucher		0.0003		0.0021		0.0052
		(0.0021)		(0.0022)		(0.0032)
Male times WWF donation		0.0021		0.0041*		0.0075**
		(0.0027)		(0.0022)		(0.0032)
Hypothetical treatment times male times restaurant voucher		−0.0012		−0.0004		−0.0012
		(0.0028)		(0.0031)		(0.0048)
Hypothetical treatment times male times WWF donation		−0.0039		−0.0037		−0.0019
		(0.0037)		(0.0032)		(0.0048)
Log likelihood			−1203.705	−1201.57	−781.594	−776.73
Log pseudolikelihood	−1367.686	−1365.126				
Statistical observations	2560	2560	2560	2560	2560	2560
Number of subjects	160	160	160	160	160	160

* Significance at the 0.1 level.
** Significance at the 0.05 level.
*** Significance at the 0.01 level.

pattern in a CE study; Mitani and Flores (2007) found in an induced value public good game that females were more likely to reveal their true value than males when hypothetical payments are used. Yet, due to the limited sample size, this particular finding should of course be interpreted with caution.[12]

5.3. Estimating marginal willingness to pay and hypothetical bias

Although the results from the logit regressions are interesting, we are fundamentally interested in the corresponding *MWTP* for each good across treatments. An individual's *MWTP* for an additional dollar given to the WWF campaign in the real-money treatment is given by:

$$MWTP^{real}_{WWF} = \frac{\partial u_i/\partial WWF}{\partial u_i/\partial Money} = \frac{\partial V_i/\partial WWF}{\partial V_i/\partial Money} = \frac{\gamma}{\beta}. \tag{10}$$

The corresponding *MWTP* in the hypothetical treatment is given by $MWTP^{Hyp}_{WWF} = (\gamma + \mu)/\beta$. Therefore, our measure of hypothetical bias, i.e., the *MWTP* difference between the two treatments, is given by:

$$MWTP^{hyp}_{WWF} - MWTP^{real}_{WWF} = \frac{\mu}{\beta}. \tag{11}$$

Similarly, the real-money *MWTP* for an additional dollar in the form of a restaurant voucher is given by:

$$MWTP^{real}_{Rest} = \frac{\rho}{\beta}, \tag{12}$$

whereas the corresponding *MWTP* in the hypothetical treatments is $MWTP^{hyp}_{Rest} = (\rho + \lambda)/\beta$. Hence, the *MWTP* difference for the restaurant voucher between the two treatments is given by:

$$MWTP^{hyp}_{Rest} - MWTP^{real}_{Rest} = \frac{\lambda}{\beta}. \tag{13}$$

Table 2 presents the *MWTP*s corresponding to Eqs. (10)–(13) associated with the basic models without gender interaction variables, where the standard errors are calculated using the delta method.

[12] The same applies of course to the gender differences in MWTP itself, where, as can be seen in Table 1, men are actually willing to pay more for an increase in the WWF donation, although this effect is not statistically significant based on the clustered logit regression.

Table 2
Estimated *MWTPs* and *MWTP*-based measures of hypothetical bias (standard errors calculated using the delta method in parentheses).

	Logit, clustered, robust standard errors	Random effects logit	Fixed effects logit
Baseline real-money marginal willingness to pay estimates			
$MWTP^{real}_{Rest}$	0.540*** (0.038)	0.539*** (0.030)	0.545*** (0.045)
$MWTP^{real}_{WWF}$	0.474*** (0.043)	0.471*** (0.029)	0.474*** (0.043)
Measures of hypothetical bias			
$MWTP^{hyp}_{Rest} - MWTP^{real}_{Rest}$	0.053 (0.052)	0.058 (0.042)	0.064 (0.061)
$MWTP^{hyp}_{WWF} - MWTP^{real}_{WWF}$	0.184*** (0.063)	0.188*** (0.043)	0.193*** (0.061)

* Significance at the 0.1 level.
** Significance at the 0.05 level.
*** Significance at the 0.01 level.

The (sample mean) real-money *MWTP* for an additional dollar in the form of a restaurant voucher is hence equal to 0.54 dollars, whereas it is 0.47 dollars for an additional dollar donated to the WWF campaign. Yet, our main concern here is the difference between the two treatments. For the restaurant voucher, we only have a small (about 0.06) and statistically non-significant hypothetical bias.[13] For the donation to the WWF campaign, in contrast, there is a sizable and statistically significant hypothetical bias of about 0.19 ($p < 0.01$ in all models). This means that the hypothetical *MWTP* for a donation to the WWF is approximately 40 percent larger than the corresponding real-money *MWTP*.

The results are thus consistent with our theoretical model: There is a substantial and statistically significant hypothetical bias for the moral good, and no significant hypothetical bias for the amoral good.

5.4. The external validity of the real-money choice-experiment

Let us finally return to the other part of Hypothesis 1, namely that the MWTP estimates from our real-money choice experiment are biased upwards in the sense that an MWTP obtained in a real-money experiment tends to exceed the same MWTP assessed outside the experimental context, i.e., $MWTP^{true}_{WWF} < MWTP^{real}_{WWF}$. In our case, the mean MWTP in the real-money treatment is 0.47, implying that subjects at the margin are indifferent between receiving 0.47 dollars themselves and that 1 dollar is given to the WWF project. As shown in Section 3, this implies that, based on a utility function given by (1) and that these preferences reflect utility changes inside as well as outside the experimental context, we can conclude that the individual would then be better off in status quo, compared to a situation where he/she receives forty-seven 1000 dollar bills falling from the sky and where WWF simultaneously receives a budget cut of 100,000 USD due to changed priorities, say at a large external donor. We consider this welfare implication unlikely.

Consequently, we argue that people's *MWTP* in the real-money context of a CE need not be a good indicator of the true value people place on an improvement per se outside the experimental context. Thus, not only do hypothetical experiments fail to correctly estimate individual welfare effects when important ethical values are involved, the same appears to be true also for real-money experiments, although to a lesser extent. Note that we are not arguing that warm-glow effects from contributing to a good cause per se should be excluded from social welfare analyses. On the contrary, we believe that warm-glow feelings are as valid as other motives. However, it is not appropriate to generalize findings that arise solely in experimental or survey situations[14] to other valuation contexts.[15] That is, if the moral satisfaction occurs primarily when responding to survey questions, or when acting in an economic experiment, then those who are not included in the sample, who obviously constitute the vast majority of the population, do not enjoy this welfare improvement; see also Andreoni (2006, Section 4) for an insightful discussion of whether and when warm-glow effects should be included in social welfare analysis.

6. Discussion and conclusions

It is often argued in the environmental valuation literature that people will reveal their true preferences unless they have a strategic incentive not to do so. However, as argued by Cummings et al. (1997), such "epsilon truthfulness" is a very strong assumption for which there is not much empirical support. Indeed, a frequent criticism of SP methods is that, due to their hypothetical nature, such approaches are likely to result in overestimation of the true values people place on public goods. However, the empirical results of such tests differ, and from several reviewed meta-studies it appears far from correct to conclude that hypothetical survey methods always overstate the benefits of public goods.

[13] This is also true based on a logit model without clustering (not shown), i.e., without taking into consideration the fact that we use a repeated measures design.
[14] See Nunes and Schokkaert (2003) for a method to remove warm-glow from a CV survey in order to obtain a "cold" WTP measure.
[15] Still, it is of course possible to argue that there may be benefits that should be taken into account beyond human well-being, and that animal welfare, and perhaps also the environment, should be valued intrinsically (e.g., Singer, 1974).

As far as we know, the model developed here is the first aimed at explaining the observed variation of hypothetical bias across studies. It is also, to our knowledge, the first attempt to test differences between hypothetical and real-money WTP for public and private goods within the same experimental context. The central tenet of our model is that people derive utility from a positive self-image, which depends on the degree to which they act in accordance with their ethical beliefs and how honest they are to themselves. Thus, people have an incentive, through self-deception and/or self-signaling, to overstate their true MWTPs if a high value is in accordance with their ethical views, but not otherwise. In other words, there is a positive hypothetical bias for what is here denoted moral goods, but no such hypothetical bias for amoral goods. The empirical results presented in this paper are consistent with these hypotheses, and inconsistent with the conventional model typically assumed in the environmental valuation literature; the hypothetical MWTP is significantly higher than the real-money MWTP for a public good with moral implications, whereas no such difference is found for a private, morally neutral good.

Another hypothesis derived from our model is that even real-money CEs tend to exaggerate people's valuations of moral goods. This was shown to be consistent with our experimental results combined with a straightforward thought experiment, and it is also in line with doubts expressed by List et al. (2004), List (2007), and Levitt and List (2007) about whether real-money experiments really measure correctly the magnitudes of people's true preferences outside the experimental situation.[16]

The model also corresponds with important psychological insights as well as recent findings in the behavioral and experimental economics literature. It draws upon well-established arguments that attitudinal statements and actions not only reflect instrumental motives, but also rest on presentational concerns and underlying ideals toward which a person aspires. Yet, the fact that the model presented here is consistent with the experimental results and also is in line with insights from much psychological research does of course not mean that the model is necessarily the correct one, since there may be other models that are able to explain the experimental results here.

That attitudes and actions partly represent symbolic expressions raises some questions about what should and should not be accounted for in a benefit assessment. Some authors have argued that it is irrelevant whether people's preferences reflect selfish interests, instrumental considerations, moral judgments, or any other reasons for that matter, hence suggesting that all such value foundations are valid. Yet, a major problem arises for public policy analysis if these motivations are context specific and do not transcend from one situation to another.

It is arguably important to analyze various kinds of unselfish behavior experimentally. People are clearly not as selfish as the standard Homo economicus model suggests. Yet, people may sometimes not be as unselfish or altruistic in a day-to-day setting as some experimental findings seem to suggest either. Or they may indeed sometimes display remarkably unselfish behavior also in real life, but such behavior is often internally motivated and conditioned on the extent to which the individual herself receives credit for taking a certain action, and not on the actual consequences of the behavior per se.

Acknowledgements

We are grateful for very constructive comments from Fredrik Carlsson, Peter Martinsson, Co-Editor Catherine Eckel, an Associate Editor, and two referees. Financial support from the Swedish Research Council and FORMAS COMMONS is gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jebo.2012.10.006.

References

Akerlof, G.A., Kranton, R.E., 2000. Economics and identity. Quarterly Journal of Economics 115, 715–753.
Akerlof, G., Kranton, R.E., 2002. Identity and schooling: some lessons for the economics of education. Journal of Economic Literature 40, 1167–1201.
Alpizar, F., Carlsson, F., Johanson-Stenman, O., 2008. Anonymity, reciprocity and conformity: evidence from voluntary contributions to a Natural Park in Costa Rica. Journal of Public Economics 92, 1047–1060.
Andreoni, J., 1989. Giving with impure altruism: applications to charity and Ricardian equivalence. Journal of Political Economy 97, 1447–1458.
Andreoni, J., 1990. Impure altruism and donations to public goods: a theory of warm-glow giving. Economic Journal 100, 464–477.
Andreoni, J., 2006. Philanthropy. In: Kolm, S.C., Ythier, J.M. (Eds.), Handbook of the Economics of Giving, Reciprocity, and Altruism, vol. 2. North-Holland, Amsterdam.
Aquino, K., Reed I.I., A., 2002. The self-importance of moral identity. Journal of Personality and Social Psychology 83, 1423–1440.
Benabou, R., Tirole, J., 2002. Self-confidence and personal motivation. Quarterly Journal of Economics 117, 871–915.
Benabou, R., Tirole, J., 2004. Willpower and personal rules. Journal of Political Economy 112, 848–886.
Benabou, R., Tirole, T., 2006. Incentives and prosocial behaviour. American Economic Review 96, 1652–1678.
Ball, S., Eckel, C.C., Grossman, P.J., Zame, W., 2001. Status in markets. Quarterly Journal of Economics 116, 161–181.

[16] Yet, there is evidence that people's behavior in laboratory experiments are positively correlated (sometimes strongly so) with behavior outside the lab, see e.g. De Oliveira et al. (2011, 2012) and references therein. Thus, according to this evidence people who act more pro-socially in the lab tend to do so also in real life.

Baumeister, R., 1998. The self. In: Gilbert, D., Fiske, S., Lindzey, G. (Eds.), Handbook of Social Psychology. McGraw Hill, Boston.
Bernheim, D.B., 1994. A theory of conformity. Journal of Political Economy 102, 841–877.
Bodner, R., 1995. Self knowledge and the diagnostic value of actions: the case of donating to a charitable cause. Unpublished Ph.D. Dissertation, MIT, Sloan School of Management.
Bodner, R., Prelec, D., 2003. Self-signaling and diagnostic utility in everyday decision making. In: Brocas, I., Carrillo, J.D. (Eds.), The Psychology of Economic Decisions. Volume 1: Rationality and Well-Being. Oxford University Press, Oxford.
Brekke, K.A., Kverndokk, S., Nyborg, K., 2003. An economic model of moral motivation. Journal of Public Economics 87, 1967–1983.
Brown, K.M., Taylor, L.O., 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, 127–139.
Cameron, T.A., Poe, G.L., Ethier, R.G., 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., Daruvala, D., Jaldell, H., 2010. Do you do what you say or do you do what you say others do? Journal of Choice Modeling 3, 113–133.
Carlsson, F., Johansson-Stenman, O., 2010. Scale factors and hypothetical referenda: a clarifying note. Journal of Environmental Economics and Management 59, 286–292.
Carlsson, F., Martinsson, P., 2001. Do hypothetical and actual willingness to pay differ in choice experiments? Application to the valuation of the environment. Journal of Environmental Economics and Management 41, 179–192.
Carson, R.T., 1996. Contingent valuation and revealed preference methodologies: comparing the estimates for quasi-public goods. Land Economics 72, 80–99.
Cummings, R.G., Elliot, S., Harrison, G.W., Rutstrom, E.E., 1995. Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive-compatible? American Economic Review 85, 260–266.
Cummings, R.G., Elliot, S., Harrison, G.W., Murphy, J., 1997. Are hypothetical referenda incentive compatible? Journal of Political Economy 105, 609–621.
Cummings, R., Taylor, L., 1999. Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. American Economic Review 89, 649–665.
De Oliveira, A.C.M., Croson, R.T.A., Eckel, C., 2011. The giving type: identifying donors. Journal of Public Economics 95, 428–435.
De Oliveira, A.C.M., Croson, R.T.A., Eckel, C., 2012. The stability of social preferences in a low-income neighborhood. Southern Economic Journal 79, 15–45.
Forehand, M., Deshpandé, R., Reed II, A., 2002. Identity salience and the influence of differential activation of the social self-schema on advertising response. Journal of Applied Psychology 87, 1086–1099.
Gilovich, T., 1991. Why We Know What Isn't So. The Free Press, New York.
Harbaugh, W.T., Mayr, U., Burghart, D.R., 2007. Neural responses to taxation and voluntary giving reveal motives for charitable donations. Science 316, 1622–1625.
Herek, G.M., 1986. The instrumentality of attitudes: toward a neofunctional theory. Journal of Social Issues 42, 99–114.
Johansson-Stenman, O., Martinsson, P., 2006. Honestly, why are you driving a BMW? Journal of Economic Behavior and Organization 60, 129–146.
Johansson-Stenman, O., Svedsäter, H., 2008. Measuring hypothetical bias in choice experiments: the role of cognitive consistency. B.E. Journal of Economic Analysis and Policy, 8, article 41.
Kahneman, D., Knetsch, J.L., 1992. Valuing public goods: the purchase of moral satisfaction. Journal of Environmental Economics and Management 22, 57–70.
Katz, D., 1960. The functional approach to the study of attitudes. Public Opinion Quarterly 24, 163–204.
Kochi, I., Hubbel, B., Kramer, R., 2003. An empirical Bayes approach to combining estimates of the value of a statistical life for environmental policy analysis. Unpublished report to the U.S. Environmental Protection Agency.
Kuran, T., 1995. Private Truths, Public Lies: The Social Consequences of Preference Falsification. Harvard University Press, Cambridge.
Lacetera, N., Macis, M., 2010. Social image concerns and prosocial behavior: field evidence from a nonlinear incentive scheme. Journal of Economic Behavior and Organization 76, 225–237.
Levitt, S., List, J., 2007. What do laboratory experiments measuring social preferences reveal about the real world? Journal of Economic Perspectives 21, 153–174.
List, J.A., Gallet, C.A., 2001. What experimental protocol influence disparities between actual and hypothetical values? Environmental and Resource Economics 20, 241–254.
List, J.A., Berrens, P., Bohara, A.K., Kerkvliet, J., 2004. Examining the role of social isolation on stated preferences. American Economic Review 94, 741–752.
List, J.A., Sinha, P., Taylor, M.H., 2006. Using choice experiments to value non-market goods and services. Advances in Economic Analysis & Policy 6, 1–37.
List, J.A., 2007. On the interpretation of giving in dictator games. Journal of Political Economy 115, 482–493.
Louviere, J.J., Hensher, D.A., Swait, J.U.D., 2000. Stated Choice Methods: Analysis and Applications. Cambridge University Press, Cambridge.
Lusk, J.L., Schroeder, T.C., 2003. Are choice experiments incentive compatible? A test with quality differentiated beef steaks. American Journal of Agricultural Economics 85, 840–856.
Mitani, Y., Flores, N., 2007. Does gender matter for demand revelation in threshold public good experiments? Economic Bulletin 3, 1–7.
Murnighan, J.K., Oesch, J.M., Pillutla, M., 2001. Player types and self-impression management in dictatorship games: two experiments. Games and Economic Behavior 37, 388–414.
Murphy, J.J., Allen, P.G., Stevens, T.H., Weatherhead, D., 2005. A meta analysis of hypothetical bias in stated preference valuation. Environmental and Resource Economics 30, 313–325.
Neilson, W.S., 2009. A theory of kindness reluctance, and shame for social preferences. Games and Economic Behavior 66, 394–403.
Nunes, P., Schokkaert, E., 2003. Identifying the warm glow effect in contingent valuation. Journal of Environmental Economics and Management 45, 231–245.
Nyborg, K., Brekke, K.A., 2010. Selfish bakers, caring nurses? A model of work motivation. Journal of Economic Behavior and Organization 75, 377–394.
Santos-Pinto, L., Sobel, J., 2005. A model of positive self-image in subjective assessments. American Economic Review 95, 1386–1402.
Shih, M., Pittinsky, T.L., Ambady, N., 1999. Stereotype susceptibility, identity salience and shifts in quantitative performance. Psychological Science 10, 80–83.
Singer, P., 1974. All animals are equal. Philosophical Exchange 1, 103–116.
Smith, A., 1759. The Theory of Moral Sentiments. Cambridge University Press, Cambridge.
Taylor, S.E., Brown, J.D., 1994. Positive illusions and well-being revisited: separating fact from fiction. Psychological Bulletin 116, 21–27.
Wardman, M., 2001. A review of British evidence on time and service quality valuations. Transportation Research Part E 37, 107–128.