



UNIVERSITY OF WARSAW

Faculty of Economic Sciences

WORKING PAPERS

No. 34/2016 (225)

PIERRE-ALEXANDRE MAHIEU
ROMAIN CRASTES
JORDAN LOUVIERE
EWA ZAWOJSKA

ON A WAY TO OVERCOME STRATEGIC
OVERBIDDING IN OPEN-ENDED STATED
PREFERENCE SURVEYS:
A RECODING APPROACH

WARSAW 2016



On a way to overcome strategic overbidding in open-ended stated preference surveys: A recoding approach

EWA ZAWOJSKA

Faculty of Economic Sciences
University of Warsaw
e-mail: ezawojka@wne.uw.edu.pl

PIERRE-ALEXANDRE MAHIEU

University of Nantes
LEMNA

ROMAIN CRASTES

University of Leeds
Centre of Choice Modelling

JORDAN LOUVIERE

University of South Australia
School of Marketing

Abstract

Stated preference (SP) surveys often use open-ended questions to elicit individuals' willingness-to-pay values for goods, services, or policy projects. However, an open-ended format may encourage strategic overbidding, and so lead to biased value estimates. We propose a new approach, based on economic theory, to limit strategic overbidding in open-ended SP surveys: prior to the valuation question, respondents are told that their insincere responses will be (unfavourably) recoded as zeros. We develop a theoretical model and verify its predictions in a field SP study. We find that the approach works: respondents aware of subsequent unfavourable recoding of their insincere answers state significantly lower willingness-to-pay values.

Keywords:

stated preferences, contingent valuation, open-ended survey, strategic overbidding, recoding approach

JEL:

C80, D01, D11, D12, D61

Working Papers contain preliminary research results.

Please consider this when citing the paper.

Please contact the authors to give comments or to obtain revised version.

Any mistakes and the views expressed herein are solely those of the authors.

On a way to overcome strategic overbidding in open-ended stated preference surveys: A recoding approach

1. Introduction

Stated preference (SP) surveys often employ an open-ended format to elicit individuals' willingness-to-pay values for goods, services, or policy projects. Because such stated values are thought to express individuals' preferences, they are used to assess whether it is worth introducing the goods, services, or policy projects. That is, the values provide estimates of benefits for cost-benefit analyses. However, a reliable cost-benefit assessment requires these values to truthfully represent the individuals' preferences.

Despite the open-ended format being widely used in SP preference surveys, it is also known that under some circumstances this format may lead to strategic overbidding and, thus, to biased value estimates. Respondents who are interested in seeing a good introduced, and who believe that the higher the values they state in the survey, the higher the probability of introducing the good, may want to overstate their actual values of the good. Incentives to overstate the true value of the good to the individual can be enhanced if a survey uses a non-coercive payment mechanism (for example, voluntary contributions)¹ because then, respondents may think that when the good is introduced, they can free-ride on payments made by others.

The fact that open-ended surveys encourage strategic overbidding was pointed in the recommendations of the National Oceanic and Atmospheric Administration (Arrow et al., 1993, p. 20), which set guidelines for SP surveys: "... an open-ended request for willingness to pay ... invites strategic overstatement. The more seriously the respondent takes the question, the more likely it is that he or she will see that reporting a large response is a costless way to make a point." This statement from the NOAA report suggests a way to avoid strategic overstatement could be rendering a survey inconsequential in order for respondents not to treat it seriously. In line with the definition by Carson and Groves (2007), a survey is inconsequential when a respondent does not care about the survey's outcome or does not believe in the actual consequences following from how they answer in the survey. In inconsequential surveys incentives to answer strategically (instead of revealing true preferences) are substantially weakened, if not eliminated: when respondents are indifferent towards any possible survey's outcome or do not think that the survey's results will affect the final decision, they may not take the survey seriously enough, and so would be unlikely to exert effort to figure out the optimal strategic response. However, this clearly is not a satisfactory solution because the data collected in inconsequential surveys are of little value and usefulness due to their lack of reliability. On the basis of economic theory, there are no expectations about respondents' behaviour in inconsequential surveys; i.e., a rational individual participating in an inconsequential survey should view any answer as equally good because the answer is not tied to actual consequences or the individual does not care about the consequences.

¹ Payment is coercive when it is required from every individual in the case of the good introduction. For instance, a tax is a coercive payment vehicle.

Our research focuses on consequential surveys. We develop a novel way to discourage respondents who participate in a consequential, open-ended SP survey from strategic overbidding, and, thus, to make their elicited values more reliable. The existing literature provides several ways to help solve the problem of strategic value overstatement (more on that in Section 3), but what distinguishes our approach is that it is based on economic theory: our approach incentivises a rational respondent to answer truthfully. The approach we propose is as follows: prior to the valuation question(s) and in addition to the standard notion that the survey responses will be made available to policy makers, we inform respondents that insincere (or suspected of being insincere) responses will be recoded as zeros. Then, responses suspected of being insincere are actually recoded as zeros before communicating the results of the survey to policy makers. We call the new approach we propose in this paper “a recoding approach”.

We develop a theoretical model to examine the incentives faced by respondents who answer a SP survey that uses our proposed recoding approach. Next, we empirically test the theoretical predictions and the effectiveness of the approach in reducing stated willingness-to-pay values in a consequential survey where respondents may be willing to strategically overbid. We find that our approach significantly reduces stated willingness-to-pay values in comparison with a consequential survey that does not employ the approach. We also compare the effectiveness of the recoding approach with another technique, recently proposed and often used, namely with an approach that relies on asking respondents to make an oath to tell the truth and sign it before stating preferences. The recoding approach appears similarly effective in reducing strategic overbidding compared with the oath approach.

The remainder of the paper is structured as follows. Section 2 shows the scale of the use of open-ended questions as preference elicitation formats in surveys. Section 3 discusses existing approaches used to enhance the reliability of SP value estimates. Section 4 describes the recoding approach. Section 5 develops a theoretical model and examines the behaviour of a rational respondent answering a SP survey with a recoding approach. Section 6 provides details about the empirical test of our proposed approach, while the results of the test are presented in Section 7. Section 8 concludes with proposing directions for future research.

2. The use of open-ended valuation questions in SP surveys

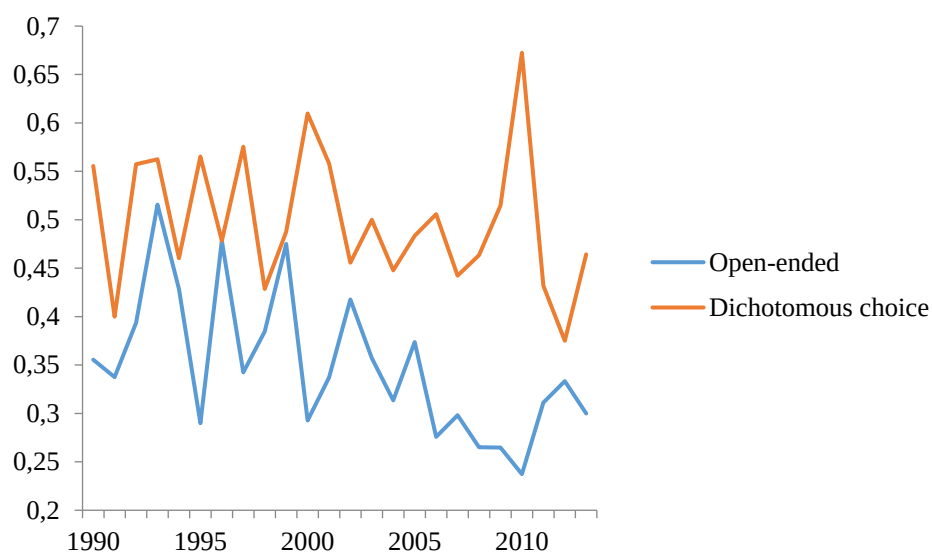
The SP literature has long recommended using a single dichotomous valuation question in surveys because of the simplicity of the format and its well-known incentive properties (Arrow et al., 1993). Recent research has placed even more emphasis on the impact of a preference elicitation format on preference disclosure. In the wake of this academic discussion, Carson and Groves (2007) postulated incentive compatibility² conditions for SP research, one of which suggests using a single dichotomous question to elicit respondents’ preferences. In spite of that, researchers commonly use alternative preference elicitation formats, which include an open-ended question. This state of affairs does not appear to be due to the researchers’ ignorance of the recommendations in the SP literature. For example, Özdemir and Johnson (2013) show that to environmental and health researchers, the question format is, respectively, the most and the second most important attribute while designing a SP survey. The use of a non-incentive-compatible, open-ended format often seems related to the goal of obtaining more precise value estimates: an answer to

² Incentive compatibility implies that revealing preferences truthfully in a valuation survey is the respondent’s optimal strategy.

an open-ended question reveals more information about the value of a good to a respondent than a single dichotomous choice question. Consequently, this format can be viewed by some as more statistically efficient, because a smaller sample size is needed to reach a given level of statistical precision of the value estimates.

Figure 1 illustrates the use of the open-ended format and the dichotomous choice formats³ in SP surveys conducted in environmental economics and health economics from 1990 to 2013. For each year in this period, we calculated the share of SP surveys that used an open-ended format and a dichotomous choice format based on the database of the Environmental Valuation Reference Inventory (EVRI). EVRI provides information on a large number of economic valuation studies that investigate environmental benefits and human health effects. From the EVRI database, 1,681 contingent valuation surveys were conducted within the considered period.⁴ As shown in Figure 1, the share of studies that use an open-ended format oscillates around 35% with a significant drop in the years 2005-2010. The share of dichotomous choice valuation surveys oscillates around 50% and is fairly stable over the period. The largest changes in fluctuations of the two shares appeared around 2009, with a large increase in use of dichotomous choice questions and a substantial decrease in use of open-ended questions. This result may mirror academic discussions about the most appropriate preference elicitation format, and may be related to the influential paper of Carson and Groves (2007) who strongly recommended using a single dichotomous choice question.

Figure 1. The share of SP surveys employing the open-ended format and the dichotomous choice format, for every year from 1999 to 2013.



Source: The authors' calculation based on the database of the Environmental Valuation Reference Inventory

³ The Environmental Valuation Reference Inventory, on the basis of which the shares in Figure 1 are calculated, does not provide a distinction between single and double-bounded dichotomous choice formats in the search tool option. However, responses to the first question in the double-bounded dichotomous choice format are equivalent to responses to the single dichotomous question, and they are often investigated separately in studies.

⁴ Only primary studies on a given data set are taken into account to avoid double counting. Benefit transfer and meta-analysis are excluded.

The non-wavering popularity of the open-ended format in SP surveys emphasises the need for developing techniques such as our proposed recoding approach that will produce more reliable value estimates from surveys that employ this format.

3. Existing approaches to obtain truthful value estimates in SP surveys

The literature proposes several approaches to limit and control non truthful preference disclosures in SP surveys. Depending on whether they involve researchers' actions before or after conducting a survey, we distinguish ex-ante and ex-post techniques, respectively.⁵ In this section we discuss the two groups one by one.

Among the ex-ante approaches, “cheap talk” (Cummings & Taylor, 1999) is probably one of the oldest techniques, and has received broad attention. This approach relies on the use of specific scripts placed before the valuation question(s) to inform individuals in the survey about the tendency of respondents to overstate their true valuations of the good considered. More recently, researchers have used evidence from psychology studies to design valuation surveys: De-Magistris, Gracia, and Nayga (2013) test the impact of honesty priming tasks on SPs, and Jacquemet, Joule, Luchini, and Shogren (2013) propose an “oath” approach. Surveys that use the concept of honesty priming ask participants to complete a task which content is related to honesty, fairness and truthfulness, before answering the valuation question(s). Surveys that use the oath approach begin by asking respondents to commit to telling the truth (typically, respondents sign a short written oath that they will answer sincerely). Ladenburg and Olsen (2014) suggest and test an approach that involves repetitively reminding respondents about an available opt-out option (in the context of our open-ended study, it could be a reminder that it is possible to state zero willingness to pay for the good considered).

Despite the cited research literature, the question remains as to how effective existing ex-ante approaches are in reducing strategic overstatement in open-ended SP surveys. As the last three approaches discussed above were proposed in recent years, there is little evidence about their efficiency; in particular, to the best of our knowledge, these approaches have not been tested in the context of an open-ended SP survey. Conversely, a long history of cheap talk in SP research has led to numerous studies that test the efficiency of this approach, and the evidence from these tests is mixed. For instance, Carlsson, Martinsson, and Akay (2011) and Silva, Nayga Jr, Campbell, and Park (2011) report that the cheap talk approach worked in their open-ended valuation studies,⁶ while Ami, Aprahamian, Chanel, and Luchini (2011) found the efficiency of cheap talk depended on the scenario implemented.

Ex-post approaches include mainly data screening and calibration. Data screening helps identify questionable observations of respondents, such as those whose willingness to pay for a good constitutes an unbelievably high share of their income. In order to control the effect of such observations, researchers may report an “alpha trimmed mean” (Mitchell & Carson, 1989, p.227), identify and remove outliers from a regression of willingness to pay

⁵ Relatedly, Loomis (2014) reviews the existing ways to mitigate hypothetical bias in SP research. Hypothetical bias typically results in overestimation of true values of a good to respondents, and in this respect it is related to overbidding incentives investigated in our paper. Nevertheless, approaches to mitigate hypothetical bias are designed for hypothetical (that is, inconsequential) surveys. We, however, examine a consequential survey.

⁶ In contrast to our study, in which the open-ended survey invites strategic overbidding, the open-ended survey of Carlsson et al. (2011) encourages strategic underbidding. The cheap talk effectively reduced the proportion of respondents stating a zero value.

used as a dependent variable, or employ a so-called Orbit model (Davies & Loomis, 2010). The Orbit model combines an ordered probit model and a tobit model, and typically uses two thresholds: zero and the mean or median of the willingness-to-pay data; stated willingness-to-pay values above the mean or median are treated as ordinal indicators (rather than cardinal amounts). Champ, Bishop, Brown, and McCollum (1997) suggest calibrating stated willingness-to-pay values based on the degree of respondents' uncertainty, and recoding "yes" answers to a dichotomous choice question as "no" when "yes" respondents indicate that they are uncertain about their response. In order to transfer this approach to an open-ended valuation study, as the one in our paper, one could consider recoding stated highly uncertain positive willingness-to-pay values as zeros. Recent advances ignore simple data recoding in favour of incorporating information on respondents' uncertainty directly into an econometric model (Moore, Bishop, Provencher, & Champ, 2010). Fox, Shogren, Hayes, and Kliebenstein (1998) propose calibration of stated willingness-to-pay values on the basis of actual behaviour observed in a market or laboratory experiment: first, a SP survey is conducted; then, a subset of respondents is chosen to participate in a validity experiment in a market or laboratory with the same good as used in the survey; responses from the experiment are used for calibration. The main problem with such ex-post approaches is they require arbitrary assumptions, such as what the threshold level is beyond which responses should be recoded. As SP survey results are sensitive to the study context and the study characteristics related to their design and implementation (e.g., the mode of the survey or the interview time), recommendations about assumptions coming from other research on data screening and calibration may not be easily transferrable between studies.

Given the large array of available approaches for making value estimates from SP data more reliable, and the fact that the effectiveness of some approaches has yet to be fully verified, one could well ask why to propose a new approach (the recoding approach). Two characteristics of the recoding approach which differentiate it from the other approaches we reviewed are important motives for developing it. First, to the best of our knowledge, the recoding approach is the first technique that integrates ex-ante and ex-post approaches: before answering a survey, respondents are informed that insincere responses will be recoded as zeros (an ex-ante approach); after the survey is conducted, responses suspected of being insincere are recoded as zeros (an ex-post approach). Second, in contrast to existing ex-ante techniques, the recoding approach is based on economic theory, which implies that when our approach is used, a rational respondent is incentivised to respond truthfully. Approaches like cheap talk, honesty priming tasks, or oath rely on respondents' moral values and on reminding them about what constitutes good behavioural patterns. Perfectly rational respondents participating in a SP survey will not change their behaviour depending on whether the survey employs any existing ex-ante approach (or not). However, our approach provides economic incentives to encourage respondents to truthfully reveal their preferences.

4. The recoding approach

We propose a new two-step approach to limit and control overbidding in consequential SP surveys that use an open-ended format. We consider only such surveys that by their construction, invite overbidding; for example, because it is believed that the value stated in a survey affects mainly the probability of having a good introduced, and has a negligible (if any) impact on the probability of imposing a payment upon the good introduction. The first stage of our approach places a script before the open-ended valuation question in a survey to explain that an attempt will be made to detect insincere responses that

subsequently will be recoded as zero responses. In the second step, a researcher attempts to detect deception and recodes responses suspected of being insincere before providing them to policy makers. In what follows we consider only consequential surveys; that is, surveys in which respondents care about the survey's outcome and believe in the actual consequences following from how they answer the survey.

The first step is the most important for encouraging truthful responding. If detection of insincere behaviour is perceived as 100% effective, a rational respondent will answer truthfully and no subsequent recoding will be needed. In real life, however, it is difficult to assure 100% perceived effectiveness of a mechanism to detect insincere behaviour, so the first step is typically insufficient. When the effectiveness of the detection mechanism is perceived to be very low and the benefits from lying are substantial (for instance, a respondent likes the good considered and strongly believes that high stated valuations of the good will largely increase the probability of the good being introduced), a rational respondent may not be sufficiently incentivised by the recoding approach to answer truthfully. As well, researchers should be aware that deception of research participants is not allowed for ethical reasons (Jones & Sigall, 1971); thus, researchers need to exert real effort to detect insincere responses.

The literature contains various methods to detect deception. Detection of insincere behaviour may rely on monitoring body reactions with the use of special devices such as polygraphs (Kleiner, 2002), eye trackers (Lim, Friedrich, Radun, & Jokinen, 2013), voice stress evaluators (Streeter, Krauss, Geller, Olson, & Apple, 1977), or simply on an assessment of how reliable respondents' choices seem to be (e.g., based on debriefing questions or a comparison of the stated willingness-to-pay amount and a respondent's income (Loomis, 2014)). The choice of method should consider actual and perceived effectiveness of the mechanism to detect insincere responding, as well as the costs of different detection mechanisms. In particular, although the subjective, for instance, income-based assessment of the reliability of respondents' answers can be tied to weak perceived effectiveness, using special devices to detect insincere responding may considerably increase research costs. Furthermore, use of such devices should accommodate the attitudes and opinions of respondents; for example, some respondents may not agree to be examined with a special device, or some devices can cause respondents to be more stressed.

When using a recoding approach, one should take into account the fact that respondents may be tempted to answer in a way consistent with their expectations formed a priori. For instance, the recoding approach may induce respondents with high actual willingness-to-pay values to lower their stated values if they are afraid that the stated values will be perceived by a researcher as suspiciously high and they risk their answers being recoded as zeros. Thus, it is important to assure respondents that the assessment of whether the answer is sincere or not is independent of their exact stated value of willingness to pay.

5. A theoretical model

Let there be a population of individuals (respondents) who participate in a SP survey. Let the survey elicit the individuals' preferences towards a specific good via a single open-ended contingent valuation question. Let the individuals believe that the value stated in the valuation question positively affects the probability of providing the good (the higher the stated value, the higher this probability), and that this value does not impact the cost needed to be paid by every individual upon the good provision, because, for example, the cost is

determined solely by the cost of the good provision. Consequently, rational individuals participating in the survey who like the good being valued and who want to have the good provided will have an incentive to overstate their value of the good.

In order to avoid strategic overbidding by respondents in such surveys, researchers need to take certain actions. Specifically, we propose a novel approach, namely unfavourable recoding of the responses of individuals who answer (or are suspected of answering) insincerely to the valuation question. In this section we develop a theoretical model to examine individual's incentives while answering a SP survey that uses a single open-ended valuation question and employs the proposed recoding approach. On the basis of economic theory, we show that the approach alters the incentives of respondents and, thus, reduces the bias of value estimates caused by strategic overbidding.

Consider an individual from the surveyed population who likes the good valued and who wants to have the good provided. Let WTP^T denote the individual's true valuation of the good (his true willingness to pay for the good), and let WTP denote the amount that the individual states in the survey. Both WTP^T and WTP are non-negative. Given the survey context and the individual's potential incentive to overstate his true value, we assume that $WTP^T \leq WTP$. We measure the degree of the value overstatement by $D = WTP / WTP^T - 1$, and clearly $D \geq 0$, with $D = 0$ when the individual answers sincerely. Let $B(D)$ be a twice differentiable function that represents the benefits in the perception of the individual from how he answers the survey; in other words, $B(D)$ measures the utility that the individual derives from stating in the survey a value related to overbidding at the level D . We assume that $\partial B(D) / \partial D > 0$ because the higher the overbidding, the higher the probability of having the wanted good provided, so benefits increase in the degree of value overstatement.

Let a mechanism that allows for detection of insincere responding be employed in the survey. The individual perceives that the stated value related to overbidding D will pass the test for sincere responding (will not be suspected of being an insincere response) with probability $S(D)$. $S(D)$ is a twice differentiable function that takes values from the interval $(0, 1]$, where 0 is excluded because if $S(D) = 0$, the individual would have no incentive to answer truthfully (no value he states would pass the test, that is, would be treated as a sincere answer). We assume that the probability of passing the test for sincere responding is equal to 1 when the answer is truthful, namely $S(0) = 1$, and that this probability decreases with the degree of the value overstatement, namely $\partial S(D) / \partial D < 0$.

When the recoding approach is used, an answer that does not pass the test for sincere responding is recoded as 0. Then, the individual is assumed to experience a utility of 0 from his (recoded) response. Consequently, the expected utility, EU , of the individual from the degree of the value overstatement D in the survey can be expressed by

$$EU(D) = S(D) \cdot B(D) + [1 - S(D)] \cdot 0 = S(D) \cdot B(D). \quad (1)$$

For ease of derivation, we assume that

$$B(D) = D + 1, \quad (2)$$

and that

$$S(D) = 1 - AD, \quad (3)$$

where $A > 0$. To ensure that $S(D) \in (0, 1]$, we have that

$$1 - AD \leq 1, \quad (4)$$

and that

$$1 - AD > 0. \quad (5)$$

Rearrangement of (4) yields $0 \leq AD$, which is fulfilled because by assumption $A > 0$ and $D \geq 0$. When $D = 0$, (5) is satisfied for any $A \in \mathbf{R}$; when $D > 0$, we obtain that $A < 1/D$. In combination, when $D > 0$, the parameter A needs to be such that $A \in (0, 1/D)$, and when $D = 0$, it is enough to have $A > 0$. We note that $\lim_{D \rightarrow 0} (1/D) = +\infty$ and that $\lim_{D \rightarrow +\infty} (1/D) = 0$, implying that the range of possible values of A substantially decreases with the degree of overstatement.

The parameter A in the function $S(D)$ can be interpreted as a measure of the effectiveness of the mechanism to detect insincere responding. Namely when A is high, $S(D)$ decreases quickly in D , that is, the probability of passing the test for sincere responding decreases quickly with the degree of value overstatement. Then, even a small divergence from the true willingness-to-pay value is strongly punished in terms of a substantial reduction of the probability of passing the test for sincere responding. Thus, a high A can be interpreted as high perceived effectiveness of the insincere responding detection mechanism. This does not contradict the point made in the preceding paragraph that the range of possible values of A decreases with a degree of overstatement. The correlation between the perceived effectiveness of insincere responding detection A and the degree of overstatement D is intuitively expected: an individual is likely to overstate by far his true good valuation when he does not believe in the effectiveness of the insincere responding detection.

Substituting into (1) the specific functional forms of $B(D)$ and $S(D)$ given in (2) and (3), respectively, yields the following expected utility function of the individual

$$EU(D) = (1 - AD)(D + 1). \quad (6)$$

We have that

$$\frac{\partial [EU(D)]}{\partial D} = -A(D + 1) + (1 - AD) = 1 - A(1 + 2D), \quad (7)$$

and that

$$\frac{\partial^2 [EU(D)]}{\partial D^2} = -2A. \quad (8)$$

Because $A > 0$ by assumption, and, thus, the right-hand side of (8) is negative, $\partial [EU(D)] / \partial D = 0$ will yield the expected utility maximising degree of value overstatement, D^* . Based on (7), we have that $\partial [EU(D)] / \partial D = 0$ if and only if

$$1 - A(1 + 2D^*) = 0, \quad (9)$$

which, upon rearrangement, gives

$$D^* = \frac{1-A}{2A}. \quad (10)$$

Because $D^* \geq 0$ by assumption, and given that $A > 0$, it follows that $1 - A \geq 0$, which is equivalent to $A \leq 1$. We now check whether this last condition on A is consistent with the range of possible values of parameter A for which $S(D)$ is defined. From (5) we know that

$$A < 1/D. \quad (11)$$

Substitution of (10) into (11) yields $A < 2A/(1-A)$, which, after simplification, becomes $0 < A+1$, and this is surely satisfied because $A > 0$. Thus, there exist such values of parameter A that fulfil the model assumptions.

(10) implies that the individual states his value of the good truthfully ($D^* = 0$) only when $A=1$; that is, when the perceived effectiveness of the mechanism of insincere responding detection is possibly highest. Any divergence from this level of the perceived effectiveness of insincere responding detection results in the individual being incentivised to overstate.

When no recoding approach is applied, the expected utility of the individual from the value overstatement in the survey is simply

$$EU(D) = B(D), \quad (12)$$

because there is no risk that his answer will be recoded. Given that $\partial B(D)/\partial D > 0$, the individual may be incentivised to state an infinitely high willingness-to-pay value for the good. This shows the effectiveness of the recoding approach, which, especially when the mechanism of insincere responding detection is perceived as credible, substantially reduces the degree of value overstatement.

6. Survey design and implementation

We designed a survey to assess the value of an entertainment event in the Plant Garden in Nantes, France, to the population of citizens of Nantes and the town neighbourhood. The entertainment event considered was similar to the event “Enchanting Nights” organised in 2014 that had included colourful lighting and music in the park every Friday night in July and August. Our survey asked how much people would be willing to pay to participate in a similar event in the coming summer if it was repeated. The willingness-to-pay question was open-ended. We made it explicit that the survey would not influence the entrance fee paid by the visitors (which would be determined by the cost of organising the event), but only the decision whether to organise it: the script informed respondents that low stated values, being a sign of low interest in the event, would make the organisation event less likely. Therefore, for those interested in the event, the format invited overstating their own valuation of the project to increase the chances of having it implemented. We designed the survey to encourage a single dominant strategy to overbid to be able to test if the recoding approach can efficiently decrease the stated willingness-to-pay amounts. The survey script assured respondents about actual consequences following from the survey, as the results were said to be made available to the town council; interviews revealed that respondents believed in the survey consequentiality.

Face-to-face interviews were conducted in May 2015 in the Plant Garden in Nantes. The choice of the survey mode was guided primarily by two aspects. First, a significant share of the survey respondents were elderly, and pre-testing indicated that they did not feel comfortable using electronic devices like computers or tablets. Second, conducting the survey on the spot enhanced its consequentiality, which requires that respondents care about the good being valued (Carson & Groves, 2007). Those being entertained in the park could be assumed to care about the events organised there. We verified the perceived consequentiality, and the pre-test showed that people believed that the outcome of the study would be sent to policy makers and that they could influence the final result by answering the survey.

The survey was structured as follows. First we asked a screening question about whether respondents live in Nantes or nearby, as the aim was to estimate the value of the event for the population in the neighbourhood of the park. For those living in this area, we continued the survey by showing them pictures of the event “Enchanting Nights” from 2014 with a brief description of the event. Then, we explained to respondents that the town council was not sure whether to repeat the event in the coming summer because of the cost of organising the event. A possible solution to this problem could be introducing a fee to enter the park for “Enchanting Nights”. We emphasised that the amount of the fee would depend only on the cost of the event organisation.⁷ Respondents were told that the results of the survey would be made available to the policy makers (the town’s authorities), so, by answering the survey, they could influence the decision to repeat the event or not. After this full explanation, respondents were asked how much they would pay maximally to enter the park for the event “Enchanting Nights” if it was repeated the following summer. Finally, we collected socio-demographic data about the respondents.

We introduced three split-sample conditions: baseline, recoding, and oath. We compared the effectiveness of our new approach with the oath approach for the following reasons. First, we wanted to compare the effectiveness of the recoding approach with an existing ex-ante technique. The recently proposed oath approach has been empirically shown to produce promising results regarding its effectiveness, while evidence for the effectiveness of cheap talk is mixed. Thus, we chose the oath approach for comparison. Second, and as already mentioned, to the best of our knowledge, the oath approach has not been tested in an open-ended SP survey, so we wanted to provide preliminary evidence.

In comparison with the baseline survey, surveys used in the recoding condition and the oath condition included extended scripts before the valuation question. In the recoding condition, we used the following script: “If we doubt about the sincerity of the amount you have declared in the survey, we will change your response to 0 EUR. This will decrease the chance of the event taking place”.⁸ No details were provided about how lying would be detected. When respondents asked about the recoding procedure, interviewers explained that it would be done by an associate professor from the University of Nantes. No respondents objected to answering the survey that would be subjected to a subsequent examination of insincere behaviour detection and recoding. Pre-tests suggested that insincere behaviour detection appeared reliable as people believed that it was possible to detect deception. In the oath condition, prior to the valuation question, respondents were asked to pledge that they would answer sincerely in the survey. Respondents needed to sign a special form that stated: “I swear that I will tell the truth throughout the survey”. In the final survey, three participants refused to sign the form, but, despite that completed the survey.

⁷ The pre-test showed that people believed that the entry fees would not depend on the outcome of the survey.

⁸ The script was originally in French. Here we present translations.

In total, 185 respondents were interviewed. Seven observations were excluded due to protest answers (two observations in the baseline condition, two observations in the recoding condition, and three observations in the oath condition)⁹, so the final sample consists of 178 observations, equally split between the three conditions, as the first (head) line of Table 1 shows. The table also presents socio-demographic characteristics of each condition sample, and the test results for whether samples differ in terms of characteristics. We conducted t-tests of equality of means. In every comparison the null hypothesis for lack of significant differences could not be rejected, meaning that the samples do not differ with respect to gender, income, age, and the number of visits to the park per month.

Table 1. Means of socio-demographic characteristics in condition samples and results of t-tests of equality of the means between the conditions

		Means			T-statistics (t) and p-values (p)		
		Baseline N = 61	Recoding N = 57	Oath N = 60	Baseline versus Recoding	Baseline versus Oath	Recoding versus Oath
<i>Female</i>	1 if the respondent is a female, 0 otherwise	0.541 (0.502) N = 61	0.579 (0.498) N = 57	0.667 (0.475) N = 60	t = -0.412 p = 0.681	t = 1.413 p = 0.160	t = -0.975 p = 0.332
<i>Income</i>	Net personal monthly income	1509.434 (1037.705) N = 53	1576.923 (873.828) N = 52	1401.786 (814.190) N = 56	t = -0.360 p = 0.720	t = -0.604 p = 0.547	t = 1.078 p = 0.283
<i>Age</i>	Age in years	43.250 (18.407) N = 60	43.607 (15.224) N = 56	40.75 (17.296) N = 60	t = -0.113 p = 0.910	t = -0.767 p = 0.445	t = 0.942 p = 0.348
<i>Visit</i>	Average number of visits to the park per month	9.475 (1.090) N = 61	8.356 (1.075) N = 57	8.717 (1.025) N = 60	t = 0.728 p = 0.468	t = -0.507 p = 0.613	t = -0.241 p = 0.810

Note: N indicates the number of observations. Standard errors are given in brackets.

We took a conservative approach and recoded stated willingness-to-pay values as zeros only if we had a strong suspicion of deception. This decision was driven by two main considerations: first, to avoid the risk of recoding answers of respondents who actually told the truth, and, second, to avoid possible overcorrection (Loomis, 2014). As a result, we recoded only one observation prior to communicating the results to the town authorities. The recoded observation, suspected of being insincere, was of a participant with a very low reported income (100 EUR) and a very high willingness to pay (20 EUR).

7. Results

In this section we compare the willingness-to-pay (WTP) values across the three conditions: baseline, recoding, and oath, in order to empirically verify the effectiveness of the proposed new approach for reducing bias related to strategic overbidding. For the recoding

⁹ The protest answers were identified on the basis of a follow-up question for the reason of stating a zero value. Most of these participants argued that the event should be free.

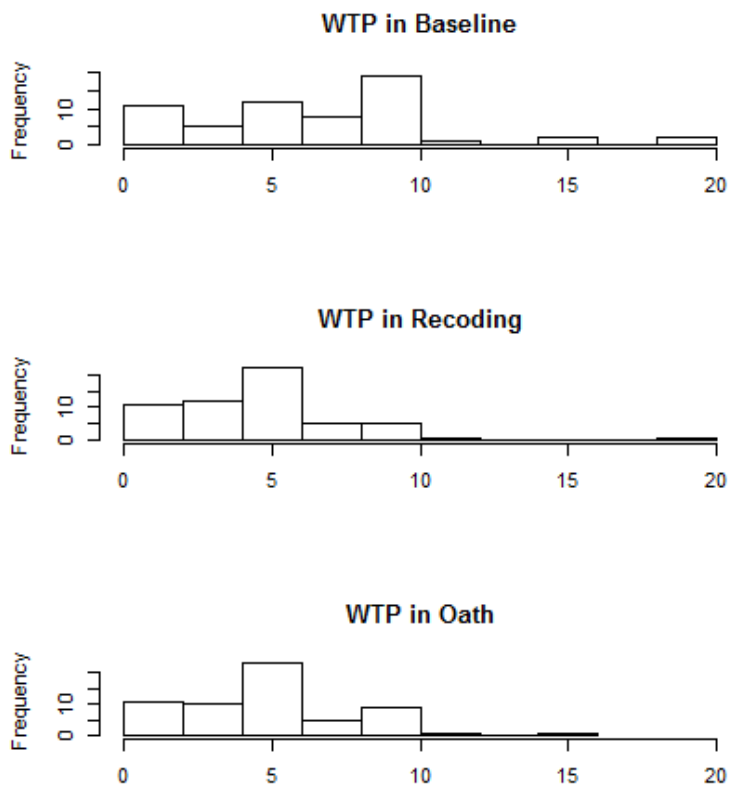
condition, we also examine the WTP values prior and post recoding of the response suspected of being insincere.

Result 1. On average, respondents in the recoding and oath conditions state lower WTP than respondents in the baseline condition.

Result 2. Stated WTPs in the recoding and oath conditions have smaller variances than in the baseline condition.

Support: Figure 2 and Figure 3 illustrate the distributions of stated WTP for each condition, and Table 2 summarises the basic statistics. Clearly, stated WTPs are lower and less dispersed in the recoding and oath conditions compared with the baseline condition.

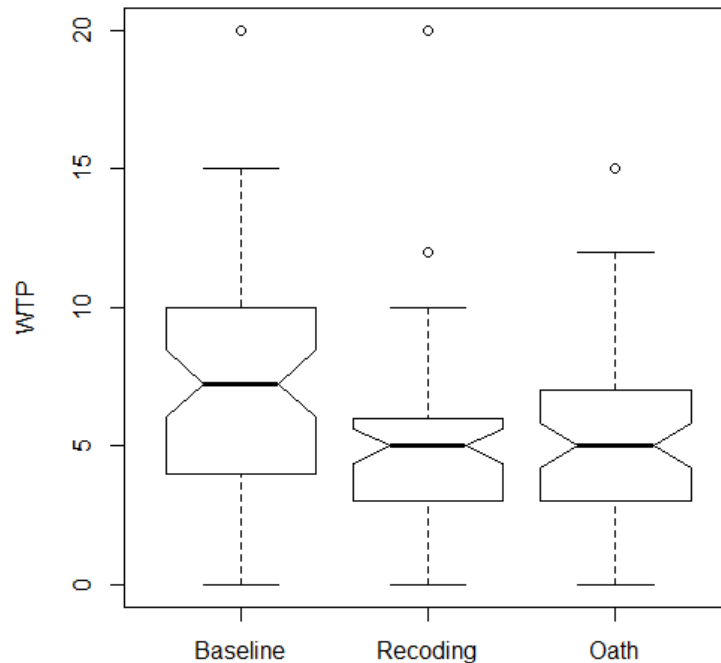
Figure 2. Histograms of WTP by condition



Both, the mean and median WTP in the baseline condition are considerably higher than the means and medians in the other conditions. The statistical significance of the differences in the means is confirmed by t-tests, the results of which are in Table 3. The null hypothesis of equality of WTP means is rejected at the 5% significance level for every comparison with the baseline condition. The result that the respondents in the baseline condition state considerably higher WTP than the respondents in the recoding condition coincides with the expectation derived from the theoretical model outlined in Section 5. Actually, in line with the model, perfectly rational respondents in the baseline condition

should state infinitely high WTP values, but we need to take into account that behavioural-economy-based considerations could induce them to lower their stated values.

Figure 3. Boxplots of WTP by condition



At the same time, the difference between the means in the oath condition and the recoding (both prior and post recoding) condition is not large, and there is no difference in the medians between the two conditions. The lack of a significant difference is statistically confirmed: the null hypothesis of mean WTP equality of the t-test cannot be rejected for the comparison of the oath and recoding conditions (both prior and post recoding). The WTP means in the recoding condition also are not significantly different for the samples prior and post recoding.

Table 2. Basic statistics of WTP by condition

	Baseline	Recoding (prior recoding)	Recoding (post recoding)	Oath
Mean	7.033	4.982	4.632	5.225
Standard deviation	4.395	3.441	2.851	3.215
Minimum	0	0	0	0
1st quartile	4	3	3	3
Median	7.25	5	5	5
3rd quartile	10	6	5.5	7
Maximum	20	20	12	15

Figure 2 and Figure 3 show higher dispersion of the stated WTP in the baseline condition compared with the other two conditions, which is confirmed by differences in the standard deviations reported in Table 2. This finding may suggest that the precision of the stated WTP improves once respondents are faced with a direct request to give a thoughtful answer (either in a form of signing an oath, or in a form of being informed about the risk of having the answer unfavourably recoded).

Table 3. Results of t-tests of equality of WTP means across conditions

	Baseline	Recoding (prior recoding)	Recoding (post recoding)
Recoding (prior recoding)	-2.801***		
Recoding (post recoding)	3.487***	0.593	
Oath	2.573**	0.394	1.054

Note: *** and ** indicate 1% and 5%, respectively, significance levels.

Result 3. The recoding condition reduces stated WTP.

Result 4. The effect of the recoding condition on stated WTP is not statistically significantly different from the effect of the oath condition.

Support: We verify the effect of the recoding condition on stated WTP with ordinary least squares (OLS) estimation. We run two regressions: one with the stated WTP as the dependent variable and another with the stated and recoded WTP as the dependent variable. In both regressions, the set of explanatory variables includes condition dummies and socio-demographic characteristics. The results of the two regressions are in Table 4. Among the socio-demographic characteristics, only income and the average number of visits per month to the park appear to be significant drivers of stated WTP; both variables are positively correlated with stated WTP, as expected. Most importantly in the context of our research, the dummy for the recoding condition has a negative and statistically significant coefficient, implying that it decreases stated WTP. Similarly, the coefficient of the oath condition dummy is negative and statistically significant, although the effect appears weaker compared with the effect of the recoding condition. However, when the coefficients of the condition dummies are compared, the null hypothesis of equality of the two coefficients cannot be rejected for either regression. The insignificant statistical difference between the effects of the recoding and the oath conditions on stated WTP is illustrated in Figure 4, where we portray 95% confidence intervals for the two coefficients in each regression.

Table 4. Results of the OLS regressions of WTP as the explained variable

	WTP	WTP recoded
<i>Recoding</i>	-2.013*** (0.731)	-2.435*** (0.690)
<i>Oath</i>	-1.598** (0.719)	-1.593** (0.679)
<i>Female</i>	0.230 (0.616)	0.051 (0.582)
<i>Income</i>	0.813** (0.368)	1.051*** (0.347)
<i>Age</i>	-0.018 (0.020)	-0.023 (0.018)
<i>Visit</i>	0.074** (0.034)	0.051 (0.034)
<i>Constant</i>	5.680*** (1.071)	5.838*** (1.011)
Number of observations	160	160
R ²	0.112	0.145
Adjusted R ²	0.077	0.111
F-statistics	3.203***	4.321***

Note: *** and ** indicate 1% and 5% significance levels, respectively. Standard errors are given in brackets.

Figure 4. 95% confidence intervals of the OLS estimates of the condition effects



8. Discussion and conclusions

SP surveys using an open-ended question to elicit individuals' preferences do not provide incentives for the individuals to reveal their preferences truthfully (unless a specific, complex mechanism of translating the individuals' responses into a final decision of policy makers is employed, as shown by (Holladay & Vossler, 2016)). If, in the perception of respondents, the surveys are linked to actual consequences, respondents may be willing to strategically overbid their true valuation of the considered good to send a signal to policy makers so that the probability of the good introduction will be increased. This behaviour of respondents biases the mean willingness-to-pay value upwards.

We propose a new approach to avoid strategic overbidding in consequential SP surveys that use an open-ended preference elicitation format. What is particular about the proposed approach is that it relies on economic theory: it provides incentives for a rational individual to answer truthfully. Prior to the valuation question, respondents are informed that their responses will be communicated to policy makers and that responses which will be considered as insincerely overbid will be recoded as zeros. We developed a theoretical model in which we showed that the proposed approach could efficiently reduce the willingness-to-pay overstatement. We then tested the approach empirically and found that the willingness-to-pay values were substantially lowered upon the implementation of the approach when compared with a survey which did not use the approach.

Our study indicates several areas for future research that will help advance the recoding approach. First, we tested the approach in the context of an open-ended SP survey, while its effectiveness in surveys using other preference elicitation formats remains to be verified. We can easily think how the recoding approach could be introduced in SP surveys with other formats of valuation questions. Assuming, as in the case of our study, that respondents have incentives to overstate their true value of the good, in a payment card study, an insincere positive stated value could be recoded as zero or the other minimum amount on the card; in a binary-choice referendum, an insincere "yes" response could be recoded as a "no" response; in a multiple-alternative question, an insincere choice of a costly alternative could be recoded as a choice of a current state (status quo).

Second, effectiveness of unfavourable recoding of responses remains to be tested in SP surveys which are linked to other dominant strategies than overbidding. If the dominant strategy is to underbid, one could recode insincere responses to extremely high amounts (e.g., the highest amount of the payment card). When there is no clearly defined dominant strategy (e.g., some participants overbid because it increases the probability for the good to be provided while others underbid in an attempt to lower the price of the good when it becomes available in the market), a symmetrical recoding could be considered. For instance, given the open-ended survey analysed in this paper, insincere zero answers could be recoded as extremely high amounts, while insincere positive stated values could be recoded as zeros.

Third, although a field survey, which we conducted, provides more appropriate conditions for testing the proposed approach than a lab experiment because the actual application of SP surveys is in the field, the field character of our study does not allow us to identify if overbidding is eliminated or only reduced. This is because field studies elicit so-called home-grown values, that is, respondents' personal preferences which they bring with them to the study, thus, a researcher has no knowledge about true, underlying preferences of the respondents. In future research, the recoding approach could be tested in an induced-value experiment. In such a lab experiment, a researcher assigns values to specific experimental

actions or results about which the experiment's participants are informed. By this assignment of values, a researcher induces participants' preferences, so he perfectly knows their actual preferences. Consequently, an induced-value experiment could inform about whether the recoding approach leads to elimination or reduction of overbidding.

References

- Ami, D., Aprahamian, F., Chanel, O., & Luchini, S. (2011). A test of cheap talk in different hypothetical contexts: The case of air pollution. *Environmental and Resource Economics*, *Forthcoming*.
- Arrow, K., Solow, R., Portney, P., Leamer, E., Radner, R., & Schuman, H. (1993). Report of the NOAA panel on contingent valuation. *Federal Register*, *58*(10), 4602-4614.
- Carlsson, F., Martinsson, P., & Akay, A. (2011). The effect of power outages and cheap talk on willingness to pay to reduce outages. *Energy Economics*, *33*(5), 790-798. doi: <http://dx.doi.org/10.1016/j.eneco.2011.01.004>
- Carson, R. T., & Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and resource economics*, *37*(1), 181-210.
- Champ, P. A., Bishop, R. C., Brown, T. C., & McCollum, D. W. (1997). Using donation mechanisms to value nonuse benefits from public goods. *Journal of Environmental Economics and Management*, *33*(2), 151-162.
- Cummings, R. G., & Taylor, L. O. (1999). Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. *American Economic Review*, *89*(3), 649-665.
- Davies, S., & Loomis, J. (2010). An improved method for calibrating purchase intentions in stated preference demand models. *Journal of Agricultural & Applied Economics*, *42*(4), 679.
- De-Magistris, T., Gracia, A., & Nayga, R. M. (2013). On the use of honesty priming tasks to mitigate hypothetical bias in choice experiments. *American journal of agricultural economics*, *95*(5), 1136-1154.
- Fox, J. A., Shogren, J. F., Hayes, D. J., & Kliebenstein, J. B. (1998). CVM-X: calibrating contingent values with experimental auction markets. *American journal of agricultural economics*, *80*(3), 455-465.
- Holladay, S., & Vossler, C. A. (2016). *Alternative value elicitation formats in contingent valuation: A new hope*. Paper presented at the EAERE, Zurich.
- Jacquemet, N., Joule, R.-V., Luchini, S., & Shogren, J. F. (2013). Preference elicitation under oath. *Journal of Environmental Economics and Management*, *65*(1), 110-132.
- Jones, E. E., & Sigall, H. (1971). The bogus pipeline: a new paradigm for measuring affect and attitude. *Psychological Bulletin*, *76*(5), 349.
- Kleiner, M. (2002). *Handbook of polygraph testing*: Academic Press San Diego, CA.
- Ladenburg, J., & Olsen, S. B. (2014). Augmenting short cheap talk scripts with a repeated opt-out reminder in choice experiment surveys. *Resource and Energy Economics*, *37*, 39-63.
- Lim, K. K., Friedrich, M., Radun, J., & Jokinen, K. (2013). *Lying through the eyes: detecting lies through eye movements*. Paper presented at the Proceedings of the 6th workshop on Eye gaze in intelligent human machine interaction: gaze in multimodal interaction.
- Loomis, J. B. (2014). Strategies for overcoming hypothetical bias in stated preference surveys. *Journal of Agricultural and Resource Economics*, *39*(1), 34-46.
- Mitchell, R. C., & Carson, R. T. (1989). *Using surveys to value public goods: the contingent valuation method*. Washington: Resource for the Future.

- Moore, R., Bishop, R. C., Provencher, B., & Champ, P. A. (2010). Accounting for Respondent Uncertainty to Improve Willingness-to-Pay Estimates. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 58(3), 381-401.
- Özdemir, S., & Johnson, F. R. (2013). Estimating willingness to pay: do health and environmental researchers have different methodological standards? *Applied Economics*, 45(16), 2215-2229.
- Silva, A., Nayga Jr, R. M., Campbell, B. L., & Park, J. L. (2011). Revisiting cheap talk with new evidence from a field experiment. *Journal of Agricultural and Resource Economics*, 280-291.
- Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., & Apple, W. (1977). Pitch changes during attempted deception. *Journal of personality and social psychology*, 35(5), 345.



FACULTY OF ECONOMIC SCIENCES
UNIVERSITY OF WARSAW
44/50 DŁUGA ST.
00-241 WARSAW
WWW.WNE.UW.EDU.PL