REWARDING TRUTHFUL-TELLING IN STATED PREFERENCE STUDIES
Rewarding truthful-telling in stated preference studies

PIERRE-ALEXANDRE MAHIEU
University of Nantes
LEMNA
e-mail: pamahieu@gmail.com

ROMAIN CRASTES
University of Leeds
Centre of Choice Modelling

JORDAN LOUVERIE
University of South Australia
School of Marketing

EWA ZAWOJSKA
Faculty of Economic Sciences
University of Warsaw

Abstract
Stated preference surveys rarely provide conditions where a respondent’s optimal strategy is to answer truthfully. As a result, the reliability of stated preference data is often questioned. We consider a new approach economic theory-based approach to incentivize respondents to answer truthfully. Our approach is based on a lie detector coupled with a reward. We discuss theoretical predictions of the approach and test them empirically in a split sample choice experiment dealing with a tree planting program. We find lie detection (i) increases time spent to complete the valuation tasks and (ii) decreases the variance of the error term by using a hybrid choice model that accounts for possible endogeneity. Our results are encouraging but more research is needed to assess the validity of this new approach.

Keywords:
stated preferences; truth-telling; choice experiment

JEL:
Q51, Q30
Rewarding truthful-telling in stated preference studies

1. Introduction

Although widely used since late 80’s (after the well-known the Exxon Valdez oil spill), doubts remain about whether stated preference (SP) methods provide true value estimates. The main critique is that respondents usually have no incentive to answer truthfully in SP surveys. To try to make SP data as reliable as possible researchers have proposed various ex-ante techniques and formulated conditions under which SP surveys are incentive compatible. That is, seek to create conditions under which a respondent’s optimal strategy is to reveal preferences truthfully. In particular, few prior approaches were based on economic theory and limitations imposed on SP studies by incentive-compatibility requirements are key shortcomings; so, we propose a new approach that has economic theoretical properties and can improve SP data reliability.

Based on mechanism design theory, researchers recently defined a set of conditions that assures SP survey incentive compatibility. Specifically, Carson and Groves (2007) identified these conditions: (a) the authority can enforce the payment by voters on the program implementation (namely a coercive payment mechanism), (b) the valuation question is viewed as a take-it-or-leave-it offer, (c) the valuation question has a single, binary format (involves a yes-or-no answer), (d) participants care about the survey outcome(s), and (e) the probability that the proposed program is implemented increases monotonically with the proportion of votes in favor of it. The last two conditions jointly produce an SP survey’s consequentiality, and Carson and Groves (2007) note that respondents who care about a survey’s outcome view their responses as potentially influencing the final action undertaken. Vossler, Doyon, and Rondeau (2012) further develop conditions for cases when a survey involves a sequence of binary valuation questions, and set two more requirements to maintain survey incentive compatibility. Thus, prior work lays out conditions for SP survey incentive compatibility that impose strict limitations on survey design. In turn, these conditions decrease statistical efficiency (less information is revealed by a single respondent than in surveys involving multiple questions with several choice alternatives) and increase research costs. Furthermore, these conditions cannot be satisfied in all SP research; for example, when a survey is not viewed as consequential or the coercive payment mechanism is not credible this can trigger a large proportion of protest responses in some surveys. Thus, a way to obtain truthful responses in SP surveys when incentive-compatibility conditions are not fulfilled would make an important contribution to the existing literature.

Ex-ante approaches are thought (and often empirically observed) to enhance incentivenss of SP respondents to truthfully reveal their preferences. For example, the (i) “oath” approach (Jacquemet, Joule, Luchini, & Shogren, 2013) asks survey respondents to sign a form at the beginning of the survey stating that they swear to tell the truth, (ii) the “ten commandments” approach (Lim, Grebitus, Hu, & Nagya, 2015; Mazar, Amir, & Ariely, 2008) asks respondents to adhere, to the best of their ability, to ethical guidelines about lying and (iii) the “honesty priming” approach (De-Magistris, Gracia, & Nayga, 2013) encourages honest respondent behavior by exposing them to a task involving honesty concepts prior to the valuation question. Although approaches (i) and (iii) may seem to be effective in making respondents’ hypothetical answers closer to their true valuations, the incentives they provide to encourage truthful preference revelation are not based on economic theory. That is, according to economic theory, behaviors of respondents who answer surveys with and without
any of the proposed approaches should not differ as these approaches do not change the incentive structure.

Recently, Prelec (2004) proposed rewarding truthful answers in multiple type questions with a general approach called (iv) the Bayesian Truth Serum (BTS). In this approach scores based on personal answers and predictions are combined into an aggregate score used to reward respondents. The BTS can be used when the “objective truth is unknowable” (Prelec, 2004). This includes question such as “is Picasso your favorite twentieth-century painter” (yes/no) or “will you vote in the next presidential election (definitely / probably / probably not / definitely not)?” This approach has been used in very few SP studies (Barrage & Lee, 2010; Menapace & Raffaelli, 2015; Weaver & Prelec, 2013) with mixed results. Barrage and Lee (2010) found the BTS approach was not fully effective at mitigating hypothetical bias and some people may have doubted that BTS could detect deception according to the authors.

We propose a novel approach to reward participants for providing truthful answers in SP surveys. The idea behind our approach is to use a lie detection device and reward participants who answer truthfully. The main advantages of our approach over approaches (i) to (iii) are that economic theory predicts how our approach revises the strategies of respondents as to how to optimally answer a SP survey. Our approach is more flexible than BTS as it can be used for any type of elicitation questions, including multiple choice question, open-ended question and ranking tasks; BTS only applies to multiple choice questions. Also, our approach does not increase survey length: in BTS, the number of valuation questions is twice as high, which can significantly increase survey length and respondent burden (e.g., the number of choice sets is twice as high using BTS in choice experiments, see Menapace and Raffaelli (2015). Finally, thanks to cinema and television, lie detector devices are well known, even in countries where they are seldom used.

Inconsequential surveys are commonly conducted¹. A possible strategy for rational individuals in this type of surveys is to minimize the level of effort invested in the valuation task (e.g., random responses). Obviously, it takes some effort and time to respond truthfully to complex questions and it is unclear why rational people would invest such levels of effort without incentives. Similarly, the literature on searching preferences (Guzman & Kolstad, 2007) proposes people are uncertain about their willingness-to-pay but can provide more accurate and truthful responses by searching their preferences and investing more time and effort in the valuation task. An especially complex approach is discrete choice experiments (DCEs) where people can be asked to make complex trade-offs.

We will discuss theoretically how a lie detection approach contributes to enhance reliability of stated preference data in inconsequential surveys, and we will empirically test the approach with data from a lab study in France in February 2015 where an oximeter is employed to record cardiac pulse. More precisely, we test if people take more time to respond to the valuation question and provide less random responses when their cardiac pulse is recorded. Our results indicate that people spend more time completing the valuation task in the lie detection condition, and the variance of the error term is lower in a hybrid choice model that accounts for possible heterogeneity. Overall, the lie detection approach seems to increase the validity of the SP data.

¹ Carson and Groves (2007) identify at least four situations: “Inconsequential preference questions can most often be identified by having one or more of the following characteristics: (a) being asked of a population or at a location that is irrelevant from the perspective of an agency seeking input on a decision, (b) providing few, if any, details about the goods and how they would actually be provided, (c) asking about goods that are implausible to provide, or (d) about an implausible prices for them”.

2
The remainder of the article is structured as follows. Section 2 describes the reward approach. Section 3 presents the model. Section 4 reviews available approaches to lie detection. Section 5 provides details of the study we conducted. Section 6 presents the results, which are subsequently discussed in Section 7. Section 8 concludes the paper.

2 Our approach

In order to incentivize respondents in SP surveys to answer truthfully, we propose a new ex-ante approach involving a lie detection mechanism coupled with a reward approach. The lie detection mechanism may not be sufficient, which is why we propose combining it with a reward. Lie detection on its own would reveal that a given answer is probably not truthful, but, this would not necessarily encourage respondents to tell the truth. The reward approach may help to solve this problem.

The reward approach monetarily rewards participants who tell the truth. It differs from a typical participation fee used to increase participation rates, but does not provide any economic incentives to encourage respondents to answer truthfully in a SP survey. In contrast, the reward approach provides the latter. Both payoff mechanisms, namely the standard participation fee and a reward for being truthful are compatible in the sense that respondents can be rewarded separately for participation and honesty. The standard participation fee incentivizes people to complete the survey but does not incentivize them to respond to it seriously (Campbell, Mørbak, & Olsen, 2012).

We distinguish two types of monetary rewards: a “weak” and a “strong” reward. Weak rewards can be used when respondents are indifferent between two situations / behaviors (here, lying and not lying), and then, the reward makes one of the two optional situations / behaviors more attractive. In contrast, strong rewards can be used when respondents prefer one situation / behavior to another (here, prefer lying to being truthful); then the reward can be used to motivate them to choose their less preferred situation (but more preferred from the perspective of the researcher; here, being truthful). Thus, depending on the expected preference of respondents towards possible situations / behaviors (that is, whether the respondent is indifferent, or has a strict preference), rewards can be relatively “high” or “low”.

Standard economic theory does not make predictions about how rational people should behave in inconsequential surveys (Carson & Groves, 2007). We consider two possible situations / behaviors: (i) Responding truthfully and not truthfully requires no effort. Hence, people are indifferent between telling the truth and not telling the truth (i.e., both answers imply no effort and no benefit). (ii) Responding truthfully requires an effort unlike responding not truthfully. Hence, people prefer not telling the truth (i.e., both answers imply no benefit but responding truthfully implies a cost). As an illustration for (i), consider a rational individual with truthful willingness to pay (WTP) of A EUR. He/she will be indifferent between stating A EUR or any other amount in an open-ended SP question because the decision affects neither benefit nor cost. A weak reward will make the truthful response more attractive. As an illustration for (ii) and the searching preference approach, consider the same individual willing to pay A EUR when investing an effort in the valuation task and who bears in mind a range between B and C EUR otherwise. He/she will prefer not investing an effort in

2 Various available techniques of lie detection are discussed in Section 4.
3 The survey is inconsequential when a respondent does not care about the survey’s outcome and / or she does not believe that her vote in a survey will have any effect on the final action undertaken by an agency.
the valuation task and pick a random value in the range [B;C] or any random value outside the range. A strong reward will make the truthful response more attractive. The same reasoning applies with other formats, like DCEs. In (i), the individual will indicate the alternative she prefers the most or any other alternative. In (ii), she will pick a random alternative or will use another strategy requiring no effort (e.g., always choose the same alternative).

Note that a truthful response is not necessarily a truthful valuation. A truthful valuation corresponds to what an individual would respond if the survey was incentive compatible. For instance, in a DCE an individual provides a “true valuation” if she chose her most preferred alternative given her budget constraint and own preferences, regardless of others’ votes (i.e., as if the decision was only based on her vote). However, DCE surveys generally are not incentive compatible because votes are expected to be conditional on other voters. The voting literature (see Carson & Groves, 2007) indicates that people should identify the two alternatives that receive the highest number of votes and then vote for the alternative they prefer among these two alternatives under certain conditions. From the perspective of a researcher it would be valuable to get to know the true valuation. We propose that lie detection coupled with appropriate question formulation may help. That is, instead of asking participants to vote in a DCE, they can be asked to indicate the alternative they would prefer to be implemented.

3 Model

We successively study the two possible behaviors in a more formal way. (i) Responding truthfully requires no effort. Hence, people are indifferent between telling the truth and not telling the truth. (ii) Responding truthfully requires an effort. Hence, people prefer not telling the truth. We assume that the monetary payoff linearly increases a respondent’s utility although our main conclusions are robust to other specifications (see Appendix A for a different assumption involving an additional parameter related to risk aversion).

3.1 Responding truthfully requires no effort.

If an agent is indifferent between telling the truth and lying, it is easy to show that their optimal strategy is to provide a true answer when a reward approach is employed as long as the perceived probability of being caught lying \((P_c \in [0,1])\) is strictly higher than the perceived probability of being wrongly caught lying when telling the truth \((P_{wc} \in [0,1])\). Let there be a participation fee equal to \(x \geq 0\) paid regardless of whether a respondent is not perceived as a liar or not, and let there be a monetary reward for being truthful (paid to a participant when they are observed not to lie) equal to \(\alpha > 0\). Then, given that the monetary payoff linearly increases an agent’s utility (see Appendix A for a different assumption), the agent (respondent) will choose not to lie as long as the probability of lie detection being right is larger than the probability of lie detection being wrong, regardless of the value of a which can be very low (weak reward).

The choice between lying and not lying is equivalent to choosing between two lotteries, as shown in Figure 1. A respondent’s expected utility from lying is given by \(EU_L = P_c x + (1 - P_c)(x + \alpha)\) and their expected utility from not lying is given by \(EU_{NL} = P_{wc} x + (1 - P_{wc})(x + \alpha)\). If \(P_c > P_{wc}\), then \(EU_{NL} > EU_L\). If the lie detector is perceived as not effective at all, \(P_c = P_{wc} = P\), where \(P\) can take any value between 0 and 1.
It follows that $EU_{NL} = EU_L$ since $Px + (1 - P)(x + a) = Px + (1 - P)(x + a)$ and a rational individual is indifferent between lying and not lying. The reward approach is not useful. If the lie detector is perceived as 100% effective, it follows that $P_{wc} = 0$ and $P_c = 1$ and also that $EU_{NL} > EU_L$ since $(x + a) > x$. The reward approach is useless. If the lie detector is not perceived as 100% nor 0% effective, the reward approach is effective as long as $P_c > P_{wc}$.

**Figure 1. An inconsequential survey perceived as a choice between the two lotteries**

The weak reward approach can be applied as long as $P_c > P_{wc}$. Let us illustrate (i) with a single dichotomous choice question applied in different inconsequential settings. Consider a new private good that is not available in the market (e.g., a new pair of shoes). If the respondent does not like it, the survey is inconsequential and the reward approach will make the “no” more attractive to the question “would you buy the good at $A$ if it was available?”. If he/she likes it, the survey is inconsequential as long as the survey does not influence the probability that the new good will be available in the market. The reward approach will make the “no” or “yes” more attractive, depending on whether the person’s WTP is inferior or superior to $A$. The same logic applies for a public good (e.g., air quality). If a person does not care about the good and the payment is voluntary, the survey is inconsequential and the reward will make it “no” more attractive. If they care about the good but the outcome of the survey has no effect on the probability to implement the project, the reward approach will make the “no” or “yes” more attractive, depending on whether their WTP is inferior or superior to $A$. If the payment is coercive (e.g., tax), the survey is inconsequential if the outcome of the survey has no effect on the probability to implement the project. The person will care about the action undertaken by the agency (even if they dislike the good as its budget will be impacted). If they dislike the good, then they would respond “no”, while if they like it, they will compare their WTP and the bid amount (the survey is incentive compatible).

**3.2 Responding truthfully requires an effort.**

Consider now that truth-telling requires an effort unlike deception (e.g., random response). A new lottery is depicted in Figure 2 where the effort is perceived as a cost, called $c$, where $c > 0$\(^4\) if the individual invests an effort in the valuation task and zero otherwise. The person’s expected utility from lying remains the same ($EU_L = P_c x + (1 - P_c)(x + a)$) and the expected utility from not lying is now $EU_{NL} = P_{wc}(x - c) + (1 - P_{wc})(x - c + a)$. This new lottery is less interesting than the previous in term of expected gain when the individual decides to tell the truth.

\(^4\) This modelling is consistent with Guzman and Kolstad (2007).
If the lie detector is perceived as fully ineffective ($P_c = P_w = P$ where $1 > P > 0$), the reward approach is again useless. Indeed, $EU_{NL} < EU_L$ since $Px + (1 - P)(x + a) = P(x - c) + (1 - P)(x - c + a)$. Conversely, if it is fully effective ($P_w = 0$ and $P_c = 1$), the reward approach is useful as long as $a > c$ (i.e., $EU_{NL} < EU_L$). If the lie detector is neither perceived as fully effective nor ineffective, then the best strategy depends on the level of $P_w$, $P_c$, $a$ and $c$. The reward approach will be effective if $EU_{NL} > EU_L$, that is if $a > c e^{-c e^{-pc}}$ after re-arrangement. The condition $P_c > P_w$ is insufficient. If truth-telling requires an important effort and the reward is low, the best strategy is not to respond truthfully. Therefore, the higher the cost, the higher the reward should be.

Figure 2. An inconsequential survey perceived as a choice between the two lotteries

![Diagram of two lotteries](image)

In situation (ii), a “strong” reward and the same reasoning as (i) apply. As an illustration, again consider a new private good that does not exist (e.g., a new pair of shoes). If a person does not like it, a low reward (“weak reward”) will not necessarily make the “no” more attractive to the question “would you buy the good at $A$ if it was available?”. The cost implied by the effort in the valuation task will not be compensated by the benefit.

The cost $c$ and therefore the level of reward $a$ needed to incentivize responses will depend on the questionnaire design. A complex DCE will lead to a higher $c$ than a straightforward CV. Likewise, when dealing with DCEs, choosing the most preferred alternative is probably easier than predicting the two alternatives that will receive the highest number of votes and then choosing the most prefer alternatives among the two predicted highest vote-getting alternatives.

4 Lie detection: existing approaches

Various devices can be used to detect deception in SP based on bodily reactions, but none are 100% effective, so can be combined with more traditional approaches. Ideally, the actual and perceived effectiveness of a device should be high and the lie detector should be in use throughout the survey as people have no incentive to respond truthfully to the non-WTP questions. If people misreport their income or their age, sampling (e.g., national sample) can be biased.

---

5 Deceiving participants is banned in experimental economics and can be criticized on ethical grounds, so one must make a real attempt to detect lying behavior in respondents. So, we really tried to monitor peoples’ body reactions when responding to the survey.

6 People may voluntary lie in socio-demographic questions. Say a tax will be implemented for an air pollution program; those with high WTP may be tempted to overstate their income so it applies only to high income people. If so, both the sampling and the estimated income elasticity between income and WTP can be biased.
4.1 Device

In face-to-face surveys, polygraphs can be used because it is well-known (e.g., in crime detection, see Kleiner (2002), and provides relatively high perceived effectiveness. Typically, polygraphs rely on several measures including heart rate, respiration rate and blood pressure. In internet surveys, web-cams can be used to try to capture eye-tracking because eye movement, pupil dilation and response time have been shown to indicate deception (Lim et al., 2015). Furthermore, facial expressions also can be examined, and voice stress analysis can provide evidence of lying in phone surveys. In phone interviews, pitch, frequency and intensity are sometimes used to detect lies. For example, Streeter, Krauss, Geller, Olson, and Apple (1977) found the average voice frequency was higher when a person lied than when they were truthful.

In mail surveys handwriting can provide information about deception. For example, Luria and Rosenblum (2010) find differences in mean and spatial measures in an experiment in which 34 participants had to write true and false sentences on a digitizer. Other approaches include functional magnetic resonance imaging (FMRI), electroencephalography, and thermal imaging Rusconi and Mitchener-Nissen (2013), but they may be too complex or costly to use on larger sample sizes.

4.2 No Device

More traditional approaches also can be used to try to detect deception. For example, one can use the time to respond as a way to try to detect deception (Sandorf, 2016); and one can check for suspicious answers. In DCEs, when faced with several alternatives, always picking the same alternatives (e.g., alternative A) is very suspicious. In open-ended or payment card surveys, people for whom the stated willingness to pay represents a very large portion of their income could be suspected of deception (Loomis, 2014). For example, a student with no income who states very high willingness to pay can be suspected of lying.

5 What we Did: Survey

To verify how the proposed reward approach works, we conducted an inconsequential laboratory DCE in February 2015 in Nantes, France, that asked respondents to state their preferences for tree planting. We created three conditions: “baseline”, “oath” and “lie detection”. The “oath” condition differs from “baseline” only in that it involves asking respondents to sign a form. “Lie detection” differs from “baseline” in employing a special mechanism for detecting insincere behavior in respondents. We did not use a real payment condition due to the purpose of the survey (inconsequentiality) and a research budget constraint.

A computer-based DCE on reforestation was conducted among undergraduate economics students, and 424 students participated. All subjects were recruited from classes and surveys were conducted during regularly scheduled class-meetings in a computer room. Prior to the

---

7 We used the same oath as in Jacquemet et al. (2013)
8 People could discard the invitation to participate in the experiment.
DCE, we pre-tested the perceived effectiveness of the lie detection devices in November 2014 at the University of Nantes. Specifically, we asked 149 undergraduate economics students in computer-based courses which type of lie detection devices they saw as the most effective: voice, cardiac pulse or facial expressions. In total, 65.10%, 25.50% and 9.40% indicated the cardiac pulse, facial expressions and voice, respectively, were more effective. Therefore, we used a special device to record cardiac pulse in the final survey. We did not use a polygraph due to the high cost, an unclear gain in effectiveness and a potential strong effect on stress.

In the final survey, the level of reward was set using the literature and the pre-test results that showed participants did not find it very difficult/burdensome to complete the survey. Empirical evidence shows that a single prize lottery is preferred to a multiple prize lottery. For instance, Gajic, Cameron, and Hurley (2012) studied four types of payments in SP studies: i) “no incentive”, ii) prepaid cash incentive ($2), (iii) a low lottery (10 prizes of $25) and (iv) a high lottery (2 prizes of $250). They found the high lottery was the most cost-effective; the prize level was worth the effort.

The baseline survey was constructed as follows. In the first part, participants were told that a lottery would be organized at the end of the whole experiment and one participant would win a gift voucher of 50 EUR, with no information provided on the total number of survey participants. To highlight the inconsequentiality of the experiment, it was explained that only students would participate in the experiment and that it did not involve real transactions. In the second part, the role of forests in the ecosystem was described. Among other things, it was explained that forests slow climate change, limit desertification by hydrating soil, and preserve biodiversity by providing many ecological services. Next, the activities of an existing NGO called “reforest action” were described using information available on the official website of the NGO that conducts several tree planting programs in Peru and Senegal. Then, four characteristics of the proposed program were described, which were the attributes used in the choice tasks. They include the place of tree planting (Senegal or Peru), whether online information about the program implementation was available, whether the program provides restoration or conservation of lands, and the cost attribute, which was phrased as the price of planting a tree. Table 1 provides details of the attributes and their levels.

In the third part of the survey, people were shown an example choice set that contained two unlabeled programs and a no-program option. After this example, 16 choice sets followed in which respondents were asked to choose their preferred alternative. Before making a choice, participants had to imagine that one of the 16 choice sets would be randomly drawn at the end of the exercise and the choice made in that set would be binding; they also had to imagine that the tree would not be planted if they chose the status quo (For a similar procedure see Bosworth & Taylor, 2012). This procedure limits free riding since the given tree is not planted if a participant refuses to pay. In the last part of the survey, participants were asked questions about their gender, age, income, and the perceived importance of the attributes. They also were asked about the stress they encountered and the perceived effectiveness of the lie detection; these two questions were not compulsory (unlike the other questions), which lead to missing observations. The distribution of scores related to stress and credibility is reported in Appendix B and Appendix C respectively.

9 Pre-testing did not suggest fatigue effects. Evidence, notably in transport, suggests that such concerns have been overstated.
Table 1. Attribute and attribute levels

<table>
<thead>
<tr>
<th>Description</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>Senegal, Peru</td>
</tr>
<tr>
<td>Online information</td>
<td>Yes, No</td>
</tr>
<tr>
<td>Ecosystem services</td>
<td>Conservation, Restoration</td>
</tr>
<tr>
<td>Cost</td>
<td>2, 5, 10, 15 EUR</td>
</tr>
</tbody>
</table>

In the “lie detection” condition, we put an oximeter on each participant’s finger at the very beginning of the survey. Then, the lottery was explained, as in the “baseline” condition, and the following message appeared on the screen: “People suspected of lying will be excluded from the lottery. In an attempt to identify them, we will use the device you have on your finger. This device records heart pulse. When people do not tell the truth, the body tends to react.”\(^{10}\). We used the term “attempt” because lie detection is not fully reliable. Also, the cardiac pulse of the respondents was briefly shown to the respondent on a screen. The screen was then hidden from the respondents.

Table 2. T-test mean comparison of socio-demographic characteristics

<table>
<thead>
<tr>
<th></th>
<th>Baseline N=146</th>
<th>Lies N=141</th>
<th>Oath N=137</th>
<th>Baseline versus Oath</th>
<th>Oath versus Lies</th>
<th>Lies versus Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>0.534 (0.501)</td>
<td>0.539 (0.500)</td>
<td>0.599 (0.492)</td>
<td>0.277</td>
<td>0.318</td>
<td>0.936</td>
</tr>
<tr>
<td>Income</td>
<td>0.291 (0.221)</td>
<td>0.264 (0.236)</td>
<td>0.267 (0.218)</td>
<td>0.358</td>
<td>0.903</td>
<td>0.312</td>
</tr>
<tr>
<td>Income_sup</td>
<td>0.418 (0.041)</td>
<td>0.326 (0.040)</td>
<td>0.358 (0.041)</td>
<td>0.301</td>
<td>0.5824</td>
<td>0.110</td>
</tr>
<tr>
<td>Age</td>
<td>19.884 (1.896)</td>
<td>19.773 (1.555)</td>
<td>19.737 (1.521)</td>
<td>0.476</td>
<td>0.846</td>
<td>0.590</td>
</tr>
</tbody>
</table>

*Note: Standard deviations are given in brackets.*

We used a main effects fractional factorial design (Louviere & Woodworth, 1983) to construct 16 choice sets. The choice sets were identical in the three conditions to facilitate comparisons between them. The time was recorded whenever a respondent moved to the next choice task. We did not allow participants to go back or change their answers. Each participant was randomly assigned to the conditions. We found no difference between conditions in socio-demographic characteristics, as shown in Table 2.\(^{11}\) No one refused to use

\(^{10}\) The message originally was in French. We provide a translated version.

\(^{11}\) This seems to suggest that neither oath nor lie detection affected responses to socio-demographic questions.
the oximeter or the oath sheet. Among the 141 and 137 students in the oath and lie detection conditions, respectively, none refused to use the oximeter or sign the oath sheet. Also, no student stopped the experiment. Of the 11 participants that always chose the status quo, none were identified as protestors.

We adopted a two-step procedure to identify deception in the lie detection condition. First, we identified the respondents who responded to the questionnaire in a very suspicious way by always picking program 1 or program 2\(^\text{12}\) or by taking very little time to complete the survey in the lie detection condition. Among them, we checked the cardiac pulse and excluded participants from the lottery whose cardiac pulse significantly increases when responding the valuation question. We only excluded four participants because we applied a conservative rule: we excluded participants only when we had strong doubts. Another reason is that information on cardiac pulse could not be used for some respondents (e.g., the device was relatively sensitive).

6 Results

\begin{quote}
\textbf{Result 1:} The time to complete the 16 valuation tasks was higher in “lie detection” than in the two other conditions although the perceived effectiveness of the lie detection matters.
\end{quote}

\textit{Support:} In the baseline, lie detection and oath conditions, on average participants took 2min 47, 3min 07 and 2min 40, respectively. The differences are statistically significant between baseline and lie detector conditions (p-value = 0.0032) and between oath and lie detector conditions (p-value = 0.0001). However, the difference between baseline and oath conditions was not statistically significant (p-value = 0.223). Figure 3 indicates differences between conditions are relatively small for the first and last valuation tasks. On average it took 22 seconds to respond to the first valuation task in each condition, but only 10 seconds for the last valuation question. Although for some valuation questions, differences between conditions were more prominent (e.g., it took 13.97, 17.20 and 14.15 seconds to respond to the second valuation task in the baseline, lie detection and oath conditions, respectively), lines depicting average response time across all valuation tasks generally exhibited a very similar pattern.

\(^{12}\) Only one respondent always pick the same alternative (program 1).
Result 2: Responses are less “random” with a “lie detector”

Support: We analyzed the effects of the oath and lie detector condition on both scale and preference heterogeneity. We adopted the modelling strategy of Hess and Stathopoulos (2012) and estimated a scaled random parameter logit hybrid choice model. \( U_{int} \) represents the utility of alternative \( i \) for respondent \( n \) in choice situation \( t \) (\( t = 1, \ldots, 16 \)). \( U_{int} \) comprises a modelled component \( V_{int} \) and an extreme value distributed (type I) random error term \( \varepsilon_{int} \). As a result, we propose the following naive model specification (subscripts omitted for convenience):

\[
V = e^{(\beta_{\text{lie detector}} \cdot \text{lie detector} + \beta_{\text{oath}} \cdot \text{oath} + \beta_{\text{stress}} \cdot \text{stress} + \beta_{\text{credibility}} \cdot \text{credibility})}
\times (\beta_{\text{asc}} \cdot \text{asc} + \beta_{\text{online}} \cdot \text{online} + \beta_{\text{ecosystem}} \cdot \text{ecosystem}
+ \beta_{\text{country}} \cdot \text{country} + \beta_{\text{price}} \cdot \text{price} + \beta_{\text{price oath}} \cdot \text{price oath}
\times \text{oath} + \beta_{\text{price lie detector}} \cdot \text{price lie detector})
\]

Variables are described in Table 1 and Table 2. \( \text{asc} \) is an alternative-specific constant that takes the value one if the status quo option was selected and 0 otherwise. \( \text{Price\_lie\_detector} \) and \( \text{price\_oath} \) are interaction variables between the price attribute and the conditions. The three non-monetary attributes and the \( \text{asc} \) parameter are assumed to follow a normal distribution while the price attribute is held fixed. The random coefficients are correlated with each other.

The above specification may be prone to biases, and Hess and Stathopoulos (2012) showed that using attitudes as covariates in a choice model is not straightforward; it puts analysts at risk of measurement error and endogeneity bias. We circumvented these issues by using a hybrid model structure where stress and attitude toward the lie detector were indicators of survey engagement in a measurement model component, with survey engagement treated as a latent variable. In this paper, we chose to broadly describe the principles of the hybrid choice
model. To save space, we refer readers to consult Hess and Stathopoulo\(s\) (2012) for an in-depth description of the model.

The latent attitude toward survey engagement for respondent \(n\), \(\alpha_n\), is defined by:

\[
\alpha_n = \gamma_{age} \cdot age + \gamma_{female} \cdot female + \gamma_{income} \cdot income + \gamma_{age^2} \cdot age^2 \\
+ \gamma_{income^2} \cdot income^2 + \eta_n
\]  

(2)

The variables are defined in Table 2; \(\eta_n\) is a random disturbance assumed to follow a Normal distribution \(g(\eta_n)\) across respondents, with a mean of zero and a standard deviation of one. Equation (1) is then replaced by:

\[
V = e^{((\beta_{lie detector} + \tau_{lie detector} \cdot alpha) \cdot lie detector + ((\beta_{oath} + \tau_{oath} \cdot alpha) \cdot oath)} \\
* (\beta_{asc} \cdot asc + \beta_{online} \cdot online + \beta_{ecosystem} \cdot ecosystem \\
+ \beta_{country} \cdot country + \beta_{price} \cdot price + (\beta_{price \cdot oath} \\
* \tau_{price \cdot oath}) \cdot price \cdot oath + (\beta_{price \cdot lie detector} \\
* \tau_{price \cdot lie detector} \cdot alpha) \cdot price \cdot lie detector)
\]  

(3)

where the \(\tau\) variables measure the impact of the latent variable \(\alpha\) on the scale of utility and the sensitivity to the price attribute. As noted earlier, lie detector stress and credibility were treated as indicators of survey engagement, allowing us to use additional information in these variables without risking measurement errors and endogeneity bias. Hence, we obtained two indicators, \(I_{stress}\) and \(I_{credibility}\).

We treated the two indicators as continuous. The likelihood for these indicators (we use \(I_{stress}\) as an illustration) is given by:

\[
L_{Istress} = \phi \left[ \frac{(stress - \zeta_{stress} \cdot alpha)}{\sigma_{stress}} \right]
\]  

(4)

where \(\sigma_{stress}\) and \(\zeta_{stress}\) are estimated. Finally, as demonstrated by Hess and Stathopoulo\(s\) (2012), estimating the model requires one to maximise the joint likelihood of the observed sequence of choices and the observed measures of stress and lie detection (coded as continuous variables). The log-likelihood (LL) corresponds to:

\[
LL = (\Omega, \gamma, \tau, \zeta, \zeta) \\
= \sum_{n=1}^{N} \ln \int_{\beta} L(y_{n} | \cdot) L_{Istress} \cdot L_{credibility} \cdot g(\eta) \cdot m(\beta | \Omega) \cdot d\eta \cdot d\beta
\]  

(5)

Equation (5) is integrated over the distribution of \(\eta\) and \(\beta_n \sim m(\beta_n | \Omega)\) where \(\Omega\) is a vector of parameters to be estimated. It is worth noting that \(L_{credibility}\) is fixed to 1 for respondents
who were not in the lie detector condition. Model results are in Table 3. Below (model was estimated with 500 Halton draws.)

Table 3 shows all attributes have the expected sign and are significant at usual levels. The variance-covariance matrix of the randomly distributed parameters indicates the presence of significant heterogeneity in preferences. The negative $asc$ indicates no status quo bias. Several results are worth noting. First, $βoath$ is positive and insignificant while $βlie_detector$ is negative and significant. This result indicates that responses from participants in the oath condition were less random than the responses of those in the lie detector condition. However, this does not indicate that the oath has a significant effect on scale and does not necessarily indicate the oath approach is superior compared to the lie detector. Indeed, one must look at the effect of the survey engagement latent variable.

### Table 3. Hybrid choice model results

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimates</th>
<th>Robust standard errors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preferences</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$asc$</td>
<td>-5.2782</td>
<td>0.8464 ***</td>
</tr>
<tr>
<td>$βonline$</td>
<td>0.7684</td>
<td>0.0775 ***</td>
</tr>
<tr>
<td>$βecosystem$</td>
<td>-0.0549</td>
<td>0.0875</td>
</tr>
<tr>
<td>$βcountry$</td>
<td>0.0215</td>
<td>0.0546</td>
</tr>
<tr>
<td>$Bprice$</td>
<td>-0.1774</td>
<td>0.0215 ***</td>
</tr>
<tr>
<td>$βprice_oath$</td>
<td>-0.1341</td>
<td>0.0913</td>
</tr>
<tr>
<td>$τprice_oath$</td>
<td>0.0961</td>
<td>0.0476 **</td>
</tr>
<tr>
<td>$βprice_lie$</td>
<td>-0.1190</td>
<td>0.0377 ***</td>
</tr>
<tr>
<td>$τprice_lie_detector$</td>
<td>0.0452</td>
<td>0.0188 **</td>
</tr>
<tr>
<td><strong>Scale</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$βoath$</td>
<td>0.4681</td>
<td>0.5676</td>
</tr>
<tr>
<td>$βlie$</td>
<td>-0.7413</td>
<td>0.1911 ***</td>
</tr>
<tr>
<td>$τlie$</td>
<td>0.8908</td>
<td>0.3039 ***</td>
</tr>
<tr>
<td>$τoath$</td>
<td>-0.3184</td>
<td>0.3528</td>
</tr>
<tr>
<td><strong>Hybrid component</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$γage$</td>
<td>0.1471</td>
<td>0.0734 **</td>
</tr>
<tr>
<td>$γage^2$</td>
<td>0.0121</td>
<td>0.0041 ***</td>
</tr>
<tr>
<td>$γfemale$</td>
<td>1.0650</td>
<td>0.3544 ***</td>
</tr>
<tr>
<td>$γincome$</td>
<td>-1.6361</td>
<td>1.0105</td>
</tr>
<tr>
<td>$γincome^2$</td>
<td>5.9715</td>
<td>1.8707 ***</td>
</tr>
<tr>
<td>$ζstress$</td>
<td>0.1041</td>
<td>0.0871</td>
</tr>
<tr>
<td>$ζcredibility$</td>
<td>1.5307</td>
<td>0.2430 ***</td>
</tr>
<tr>
<td>$σstress$</td>
<td>1.7886</td>
<td>0.0710 ***</td>
</tr>
<tr>
<td>$Σcred$</td>
<td>3.0132</td>
<td>0.2873 ***</td>
</tr>
</tbody>
</table>
Random heterogeneity in preferences

Variance covariance matrix

<table>
<thead>
<tr>
<th></th>
<th>Online</th>
<th>ecosystem * online</th>
<th>ecosystem</th>
<th>country * online</th>
<th>country * ecosystem</th>
<th>Country</th>
<th>asc * online</th>
<th>asc * ecosystem</th>
<th>asc * country</th>
<th>Asc</th>
<th>Log-Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online</td>
<td>0.6355</td>
<td>0.1142</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-4517.971</td>
</tr>
<tr>
<td>ecosystem * online</td>
<td>0.0977</td>
<td>0.0526</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ecosystem</td>
<td>0.6605</td>
<td>0.0837</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>country * online</td>
<td>-0.0748</td>
<td>0.0859</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>country * ecosystem</td>
<td>-0.1879</td>
<td>0.0855</td>
<td>**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>0.4232</td>
<td>0.0785</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>asc * online</td>
<td>0.3148</td>
<td>0.9899</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>asc * ecosystem</td>
<td>0.2465</td>
<td>1.2889</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>asc * country</td>
<td>-0.3354</td>
<td>0.3121</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asc</td>
<td>3.2219</td>
<td>0.3266</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-Likelihood</td>
<td>-4517.971</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: ***, **, * indicate significance at 1%, 5% and 10% level, respectively. Income is divided by 1000.

ζ_{stress} is positive and not significant while ζ_{credibility} is positive and significant, indicating that participants with more positive value for the latent variable α_n tended to find the lie detector more credible but there is no effect on reported levels of stress. γ_{age} and γ_{age^2} are both positive and significant, indicating that α_n is increasingly higher for older respondents. α_n is higher for female respondents. γ_{income} is not significant but γ_{income^2} is significant and positive, indicating that α_n is increasingly higher as income increases.

τ_{lie detector} and τ_{oath} indicate that participants expressing higher survey engagement had a higher scale than those in oath condition. Indeed, τ_{lie detector} is positive and significant while τ_{oath} is not significant. This is an important result as it indicates that those in the lie detector condition provided more deterministic answers (i.e. less random answers) than those in the oath condition. The value of τ_{lie detector} shows that a unit increase in α_n increases τ_{lie detector} by 0.8908. The value of the γ coefficients indicates a higher scale for a vast majority of those in the lie detector condition.

Finally, β_{price oath} was not significant while β_{price lie detector} was negative and significant, indicating that participants in the lie detector condition generally had lower WTP. However, τ_{price lie detector} and τ_{price oath} were both positive and significant, indicating higher survey engagement increased WTP.

7 Discussion and further research

Response times now are commonly used in SP surveys to assess the validity of results based on the idea that people who take time to respond to valuation questions respond with more care (Campbell et al., 2012). We found monetary incentives increased time taken by people to make choices, which is consistent with a recent induced experiment (Jacquemet, Luchini, Shogren, & Watson, 2016). Each choice set was composed of 3 alternatives displaying several monetary amounts using tokens. One choice set was randomly chosen and amounts for the selected alternatives were added and offered to a respondent. The authors found that
participants took more time to make choices with monetary incentives than without, and people who took more time were more successful at making payoff maximizing choices. Similarly, individual response times were found associated with higher average pay-offs, higher levels of rationality and higher levels of honesty. Overall, our results suggest that monetary incentives increases reliability of SP data.

We also found that the variance of the error term decreased when lie detection was used, suggesting responses were less “random” or preferences less “fuzzy” when lie detection was used. This supports our theoretical model: when a survey is inconsequential, the best strategy for rational people is to provide random responses to minimize the level of effort invested in the valuation task. This result casts doubt on using inconsequential surveys in the literature; in particular, such surveys should be combined with a tool that increases the validity of results, otherwise results may not be fully valid.

More research is needed on the reward approach before using it for policy making purposes. A future challenge is to use lie detection in a field experiment, and one must be careful when choosing the lie detecting device (pre-test with focus groups, etc…). If some participants refuse to use a lie detector or think it is not effective at all, one can ask these participants to take an oath as the oath and lie detector can be viewed as complementary, not competitive. Also, some participants may be stressed, and studies have shown stress can negatively impact behavior (a recent study dealing with WTP is Maier and Wilken, 2014). The literature provides mixed results, with people who are stressed spending more time responding to the valuation task but making more mistakes (as suggested by the variance of the error term in the hybrid model; Appendix C reports self-report scores on a scale of 1 to 10). More research is needed to determine if the stress implied by lie detection can distort results, but use of the device may not necessarily lead to high levels of stress due to increasing personal use of connected tools (e.g., sport watch with a heart rate monitor).

Moreover, it would be interesting to test whether an incentive compatible survey leads to the same results as a non-incentive compatible survey that uses a lie detector. From the perspective of a researcher it would be valuable to know the true valuation, which depends on preferences and budget constraints (i.e., not conditional on other peoples’ responses). Finally, it is worth noting that our reward approach also can be used in consequential surveys, but will be more costly since rational people will trade-off the monetary reward and benefits from lying (i.e., a rational individual can prefer increasing the chance for the good/program to be implemented rather than getting a low monetary reward). Cognitive burden can be high if people try to manipulate responses, especially in DCEs; so the reward should compensate both the lack of benefit from deception and the efforts invested in the valuation task.

8 Conclusion

Surveys are seldom incentive compatible because their conditions are restrictive and increase the research costs. Ex-ante approaches have been proposed to enhance SP validity; we proposed a new approach that potentially can be used in any survey. It involves rewarding truth telling and using lie detection. Our approach has several potentially appealing aspects. (a) It is consistent with standard economic theory that emphasizes the probability of being caught lying and the magnitude of the reward as key ways to overcome dishonesty. (b) Lie detection can help to obtain truthful responses for any type of WTP questions (close-ended, open-ended question (…) and non-WTP questions. (c) Research in lie detection technology is on-going, which could benefit the approach we proposed: detecting deception is not easy for
researchers, nor one that can yet be done with certainty, but recent advances may make it easier and improve actual and perceived effectiveness. More research is needed before our approach can be used for policy making purpose.
References


Appendix A. Monetary payoff does not linearly increase the respondent’s utility

The assumption that the monetary payoff, $m$, increases a respondent’s utility linearly can be relaxed. It may be the case that a respondent’s utility is given by a function $U(m) = m^\alpha$, where $\alpha > 0$ (we note that the case of $\alpha = 1$ corresponds to previously considered examples). Again, we successively study the two possible behaviors in a more formal way. (i) Responding truthfully requires no effort, so people are indifferent between telling the truth and not telling the truth (ii) Responding truthfully requires an effort. Hence, people prefer not telling the truth (i.e., both answers imply no benefit but responding truthfully implies a cost).

- Responding truthfully requires no effort.

A respondent’s expected utility from lying is given by $EU_L = P_c x^\alpha + (1 - P_c)(x + a)^\alpha$, and their expected utility from not lying is given by $EU_{NL} = P_wc x^\alpha + (1 - P_wc)(x + a)^\alpha$. Given a respondent’s rationality, thy will answer sincerely (will not lie) upon $EU_L < EU_U$, which is equivalent to $P_c x^\alpha + (1 - P_c)(x + a)^\alpha < P_wc x^\alpha + (1 - P_wc)(x + a)^\alpha$. For $a > 0$ and $P_c > P_wc$, the condition will be fulfilled. However, the last condition reveals that the efficiency of the reward approach depends not only on the amount of the monetary reward for being truthful, $a$, but also on how the monetary payoff is translated into the respondent’s utility; in turn, this is tied to a respondent’s risk attitude. The lower the $\alpha$, which is related to stronger risk aversion, the more negligible the impact of $a$. On the other hand, the higher the $\alpha$, which is related to risk-seeking behavior, the more prominent the impact of $a$.

If the lie detector is perceived as 0% effective, $P_c = P_wc = P$, where $P$ can take any value between 0 and 1, it follows that $EU_{NL} = EU_L$ since $Px^\alpha + (1 - P)(x + a)^\alpha = Px^\alpha + (1 - P)(x + a)^\alpha$ and a rational individual will be indifferent between lying and not lying. If the lie detector is perceived as 100% effective, we obtain $P_wc = 0$ and $P_c = 1$ and it follows that $EU_{NL} > EU_L$ since $(x + a)^\alpha > x^\alpha$ and the best strategy is to tell the truth.

- Responding truthfully requires an effort.

A respondent’s expected utility from lying does not change ($EU_L = P_c x^\alpha + (1 - P_c)(x + a)^\alpha$) and their expected utility from not lying becomes $EU_{NL} = P_wc(x - c)^\alpha + (1 - P_wc)(x - c + a)^\alpha$. So, the expected gain from lying decreases. Therefore, the higher $c$ the higher $a$ must be for the lie detection approach to be effective. Also, the lower $\alpha$ the more negligible the impact of $a$. 
Appendix B. Histogram of the self-report level of lies detection effectiveness (credibility) in the condition group lies detection

![Histogram of self-report level of lies detection effectiveness](image)

Appendix C. Average score of stress on a scale ranging from 1 to 10 for each of the different conditions.

![Average stress score for different conditions](image)