How to Improve Response Consistency in Discrete Choice Experiments? An Induced Values Investigation

N. Jacquemet[1] S. Luchini[2] J. Shogren[3] V. Watson[4]

Abstract

In this paper, we carry out induced value experiments that recreates the salient features of Discrete Choice Experiments (DCE) in an experimental economics laboratory setting. We first show that non payoff maximizing choices in a simple induced value DCE are numerous. We then study the potential of several simple devices to improve choices. Only a truth telling oath is effective in our setting whereas traditional devices such as monetary incentives or helping respondents (a calculator) have little or no effect on choices. The analysis of response time together with the implementation of the two oath procedures that target effort rather than truth-telling tell us that truth-telling in DCE is a serious concern. We suggest that this concern should be addressed in a systematic way in field surveys, by asking respondents to take an oath prior to being interviewed as we have done in this paper or by implementing weaker forms of commitment like a preliminary pledge or even a simple signed agreement to tell the truth.

Keywords: Experimental economics, Discrete choice experiment, Oath, Truth-telling, effort Demand revelation

JEL codes: D1, D6

Introduction

Discrete choice experiments (DCEs) are a survey-based tool used to elicit individuals' preferences for health and the provision of health care. Following Lancaster (1966)'s consumer theory, DCEs ask respondents to make hypothetical choices between alternative multi-attribute bundles of goods (health profiles, descriptions of health care provision). The application of DCEs in health economics and health services research is widespread and increasing (de Bekker Grob et al, 2014). The results from DCE studies have been used to construct the new value set for the EQ-5D patient reported outcome measure (PROM) (Oppe et al, 2014); as evidence about patient acceptable risk-benefit trade-offs and to value end-points in therapy. All these values may inform health technology reimbursement decisions. It is important, therefore, that the hypothetical DCE choices reliably measure individuals' preferences.

A direct test of DCE reliability is a comparison between the hypothetical DCE choices made in surveys and similar real world decisions. However, patients' choices are restricted in many health care systems and patients rarely bear the health care cost at the point of delivery. This means that researchers are seldom able to collect data about real world decisions that are comparable to the DCE choices. To date, only a few studies have compared individuals' DCE choices with their real-world health care actions (Mark and Swait, 2004, Ryan and Watson, 2009, Kesternich et al, 2014). In lieu many studies test the internal (theoretical) validity of DCE choices and propose survey modifications or econometric corrections for troublesome results. However, these tests are limited

[1]Université de Lorraine (BETA) and Paris School of Economics.
[2]Universit d'Aix-Marseille (Aix-Marseille School of Economics, CNRS & EHESS.
[3]Department of Economics and Finance, University of Wyoming.
[4]HERU, University of Aberdeen.

Token attributes	Level	£
Size	Small	0.50
	Medium	2.50
	Large	4.00
Colour	Red	1.00
	Yellow	1.50
	Blue	2.00
Shape	Circle	1.50
	Triangle	3.00
	Square	6.00
Cost		2.00
		3.00
		4.00

Table 1: Subjects' induced values for all treatments.

because respondents' true preferences are unknown and therefore the reliability of the DCE choices can not be measured. Instead, researchers can only make statements about the results' *plausability*.

An alternative approach to test the reliability of DCE choices is an experimental economics technique called an induced value experiment. In an induced value experiment the researcher uses (financial) rewards to "create" (or induce) individuals' preferences for goods (Smith, 1976). The researcher induces and, therefore, knows the individual's true preference, and can test if individuals reveal their induced value when asked to complete a decision task in a controlled laboratory setting. By using an induced value experiment we can measure the difference between 'true' induced preferences and the preferences revealed by the DCE choices.

Generic induced value DCE task

We develop an induced value experiment that recreates the salient features of a DCE in an experimental economics laboratory setting.[5] We induce preferences for a multi-attribute good that we call a token. A token has four attributes and each attribute has three levels: colour (red, yellow, blue); shape (circle, triangle, square); size (small, medium, large); and cost (see first and second column of Table 1). We then tell subjects that they can buy tokens during the experiment at the announced cost, and then sell the tokens back to us at the end of the experiment. We induce preferences by announcing the price we will pay for a token depends on its attribute levels. Specifically, the price we will pay for a token is a linear additive function of the attribute levels.[6]

We create a DCE to elicit subjects' induced preferences. In the DCE, tokens are arranged into choice sets with three alternative actions. The three alternatives are to buy one of two different tokens presented in the choice set at the posted price or to buy no token. Subjects are asked to complete nine choice sets and the order of the choice sets was randomised across subjects. The tokens included in the choices sets were chosen using a fractional factorial design (Louviere et al , 2000). The DCE task and tokens remain identical for all forthcoming treatments of this paper. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007).

[5]The baseline experiment is the wide hypothetical treatment of Luchini and Watson (2014).
[6]Linear additive utility is the utility typically assumed in DCE econometric models.

Subjects were recruited from students at the University of Aberdeen using Exlab and ORSEE software[7]. All subjects received a consent form, experiment instructions, and payment form before taking part in the experiment. Before the experiment started, the subjects read and signed the consent form and this was collected by the experimenter, then the experimenter read aloud the experiment instructions to the group and answered questions.

In homegrown preference settings, the observer cannot assess directly whether choices correspond to those that maximize utility or not. Preferences are unknown to the observer and the maximization of utility is an assumption made by the practitioner to econometrically estimate preferences over attributes. In an induced value design, one can assess directly whether participants make choices that maximise their payoff. In our set up, in each choice set subjects should buy the token with largest difference between its resale value and cost. We assess the reliability of the DCE method by calculating the proportion of payoff maximizing choices in the lab. We explore differences in the proportion of payoff maximizing choices across choice sets[8]. We consider the distribution of the total number of payoff maximizing choices made by an subject. We also record the time taken for a subject to make each choice to the nearest second. We consider the relationship between response time as a proxy for cognitive reasoning in the task and payoff maximizing choices. We assume that respondents who take longer to make their choices engage in more cognitive reasoning.

Experiment 1: Hypothetical DCE

In the experiment 1 choices are hypothetical as they are in DCE field surveys. In the rest of the paper, we take experiment 1 as our baseline treatment to which we compare all other treatments. Subjects are paid £12 for taking part in the experiment irrespective of the choices they make (tokens they buy). The experiment instructions use subjective language by asking subjects to "Put yourself in a situation where your account balance at the end of the experiment would depend on the choice you made..." (Taylor et. al., 2001).

Table 2 shows the subjects' payoff from each token in the nine choice sets (A - I), the payoff of each of the two tokens in the choice sets, the payoff difference between the two tokens, and the proportion subjects who chose the payoff maximizing token. There are two main results at the aggregate level. First, only a little over half of the choices are payoff maximizing (56.3%) with no significant pattern across rounds (see Appendix A). Second, the proportion of payoff maximizing choices varies greatly between choice sets, from 14.9% in choice sets A and C to 76.5% in choice set D.

Choice sets A, B and C have the worst choices, share the feature that while token A has a positive payoff, token B has the highest payoff. Of the three choice sets, choice set B has the highest proportion of payoff maximizing choices (38.3%) and the smallest payoff for token A. In contrast, in choice set D token B has a high payoff, but token A's has a negative payoff. The choice sets with the highest proportion of payoff maximizing choices have one token with zero or very small payoffs.

In Figure 1, we compute, for each subject, the percentage of payoff maximizing choices that subjects made in the 9 choice sets and we present its empirical distribution function (EDF) each bullet in the figure corresponds to a subject. No subject made 100% (or 9) payoff maximising choices. The highest percentage of payoff maximizing choices observed in the experiment is 77.7%, which corresponds to 7 choices out of 9. Most of the subjects (44.7%)

[7] When students log on to University computers they see a virtual notice board. This notice board includes adverts encouraging students to register to participate in economics experiments. Students who are registered receive an email notifying them of new experiments they can participate in. While many participants had participated in other experiments, no participant had participated in any similar experiment.

[8] The computer based experiment means that the order of choice sets was randomised allowing us to separate choice sets effects from decision round effects

Table 2: Token values and the proportion of correct choices across choice sets in experiment 1 ($n = 47$)

Choice	Value A	Value B	Payoff Diff	Baseline N=47 Percentage
A	5.5	6.5	1	14.9
B	2.5	9.5	7	38.3
C	3.5	8	4.5	14.9
D	-0.5	7	7.5	76.5
E	8	3	5	72.3
F	4.5	3	1.5	72.3
G	6	4	2	74.4
H	3	0.5	2.5	68.1
I	8	1	7	74.4
Overall (%)				**56.26%**

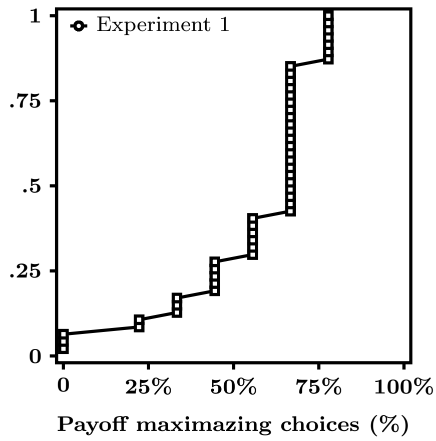


Figure 1: Empirical distribution function of percentage of payoff maximizing choices in experiment 1

made only 6 payoff maximizing choices in experiment 1. Many choices do not maximize induced utility in our simple DCE design.

One explanation for this result may be that subjects make mistakes when making choices. In particular, in choice sets A, B and C in which token A has a positive payoff but token B has the highest payoff. If subjects make mistakes, then we should observe significantly shorter response times for choices that are not payoff maximizing because subjects making mistakes engage less in cognitive reasoning (Rubinstein, 2007). This is not what results suggest. When all choice sets are considered together, median response time for non payoff maximizing choices is 15 seconds compared to 17 seconds for payoff maximizing choices. For choice sets A, B and C, median response time of payoff maximizing choices is shorter than non payoff maximizing choices (12 seconds compared to 17 seconds). The response time difference is even greater when only choice sets A and C are considered: median response time for payoff maximizing choices is 8 seconds compared to 17 seconds for non payoff maximizing choices. Bootstrap tests indicate that the increase in median response time is statistically significant for choice sets A and C ($p = .052$ and $p = .001$) whereas there is no significant difference in median response time for choice set B ($p = .358$). We observe an opposite pattern for choice sets D-I. Median response time is greater for payoff maximizing choices (17 seconds) than for non payoff maximizing choices (12.5 seconds). Bootstrap tests indicate that the difference is significant for choice sets D ($p = .024$), G ($p = .011$) and I ($p = .028$). Results on median response time are confirmed by Kolmogorov-Smirnov (KS) bootstrap distribution tests (see Appendix B for EDF graphs and KS tests).

From experiment 1, we observe: 1/ the proportion of payoff maximizing choices is very low and casts doubt on the use of DCE results for policy decisions. 2/ the response times indicate that non payoff maximizing choices take longer than payoff maximizing choices in those choice sets with the lowest proportion of payoff maximizing choices. Non payoff maximizing choices are not therefore “mere” mistakes based on intuitive responses rather than cognitive reasoning, but may involve more complex choice processes. Given the low reliability of DCE choices we aim to test a set of simple devices that can increase the proportion of payoff maximizing choices in the following experiments.

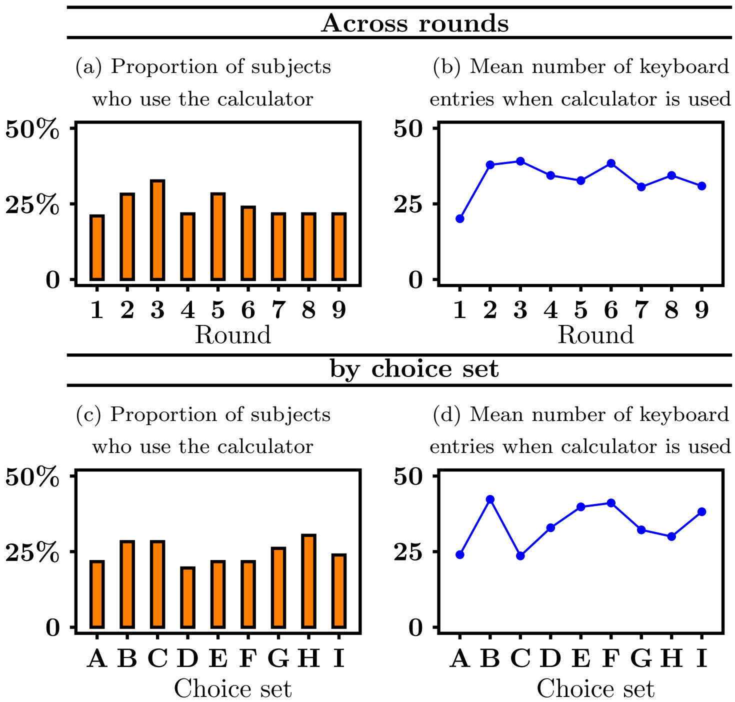


Figure 2: Use of the calculator across rounds and by choice set

Experiment 2: Making Choices with the Help of a Calculator

Several studies suggest that the DCE choice tasks presented in the field are too complex for individuals, and, as a consequence, observed choices may not maximize utility (DeShazo and Fermo, 2002; Mazzotta and Opaluch, 1995; Swait and Adamowicz, 2001). Choice complexity is higher when the difference in the utility between the alternatives in the choice set is small (Swait and Adamowicz, 2001b). Complexity is found to increase with the number of attributes and number of alternatives in a choice set, and with the number of choice sets individuals are asked to complete (DeShazo and Fermo, 2002; Mazzotta and Opaluch, 1995; Swait and Adamowicz, 2001). Although our induced valuation DCE task may seem to involve simple mathematics not all subjects may be able to complete this task.

In experiment 2, we provide subjects with a computerized calculator to help them make the calculations. We replicate experiment 1, but we add to each choice set screen a button, on which subjects can click, that gives access to Microsoft windows® calculator. Subjects calculator use is recorded throughout the experiment. We record if the calculator is activated by a subject in a given choice task and how many keyboard entries are made by the subject when the calculator is activated. One keyboard entry corresponds to a number, an operator, a decimal mark or a delete key. For instance, a subject who would calculate the value of a small yellow square token would type "`.5 + 1.5 + 6 =`" and this would be counted as 9 keyboard entries. In the experiment instructions, subjects are told how to access and use the calculator. Otherwise, experiment 2 is identical to experiment 1.

Figure 2 presents the proportion of subjects who activated the calculator and the mean number of keyboard entries across rounds and by choice set. The calculator was activated in 24.6% of the choice tasks. Fifty percent of subjects never activated the calculator, 19.5% activated it only once and 13% activated it in every round, the remaining subjects are equally distributed in between. Figure 2.a shows that the activation of the calculator is

Table 3: Token values and the proportion of correct choices across choice sets in experiment 2 ($n = 54$)

Choice	Value A	Value B	Payoff Diff	Calculator N=47 Percentage
A	5.5	6.5	1	4.3
B	2.5	9.5	7	36.9
C	3.5	8	4.5	13.0
D	-0.5	7	7.5	78.3
E	8	3	5	80.4
F	4.5	3	1.5	80.4
G	6	4	2	84.8
H	3	0.5	2.5	82.6
I	8	1	7	93.5
Overall (%)				**61.6%**

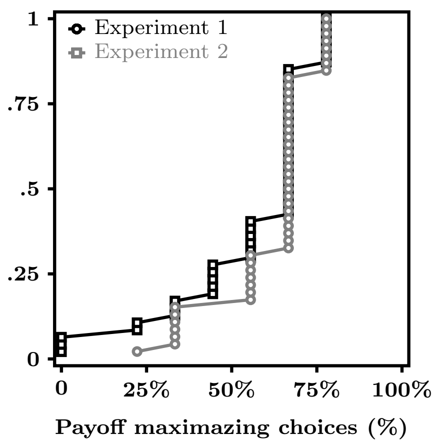


Figure 3: Empirical distribution function of percentage of payoff maximizing choices in experiments 1 and 2

relatively stable across rounds. There is no clear round effect, with only a small increase in activation in rounds 2 and 3 (28.2% and 32.6% respectively and 21.7% in round 1).

With respect to keyboard entries (Figure 2.b), when the calculator was activated, again we observe no round effect (except in round 1). Figure 2.c and Figure 2.d present the activation of the calculator and mean number keyboard entries by choice sets. The proportion of activation does not depend on the choice set. For choice sets A, B and C, the choice sets which exhibit a low proportion of payoff maximizing choices in experiment 1, the proportion of activation is 26.1% whereas it is 23.9% for the remaining choice sets D to I. When the calculator was activated, the mean number of keyboard entries was also very similar across choice sets. Although, the two choice sets with the lowest mean number of keyboard entries were also those with the lowest proportion of payoff maximizing choices in experiment 1 (A and C).

Table 4 presents the proportion of payoff maximizing choices in experiment 2. Providing a calculator had a small, but statistically insignificant effect, on the proportion of payoff maximizing choices compared to experiment 1 (61.6% vs. 56.3%). The two proportions are compared using a two-sided bootstrap test of proportions that allows for within subject correlation, the p-value is $p = .298$. The proportion of payoff maximizing choices for choice sets D-I is 83.3% in experiment 2 compared to 73.0% in experiment 1 but the difference is not significant ($p = .176$). Similarly, the calculator does not significantly change the proportion of payoff maximizing choices in choice sets A, B and C compared to experiment 1. The proportion of payoff maximizing choices in these choice sets remains low.

At the individual level, we observe no improvement in the percentage of payoff maximizing choices a subject makes. Figure 3 presents the EDF of the percentage of payoff maximizing choices by subject in experiment 1 and 2. The EDF in experiment 2 is slightly to the right of the EDF in experiment 1 but first order dominance is not significant ($p = .192$).[9]

We test if subjects who use the calculator make better choices. We find that in choice sets D to I, subjects who activate the calculator made payoff maximizing choices 92.5% of the time compared to for subjects who did not activated the calculator who made payoff maximizing choices 80.5% of the time. In choice sets A, B and C, subjects who did not activate the calculator made payoff maximizing choices 18.6% of the time compared to

[9]We use a bootstrap version of the Kolmogorov-Smirnov test. The advantage of this test as compared to the standard KS test is to allows for ties and small sample size (see Abadie, 2002; Sekhon, J. , 2011).

Table 4: Token values and the proportion of correct choices across choice sets in experiment 3 ($n = 46$)

Choice	Value A	Value B	Payoff Diff	Mon. Incentives N=54 Percentage
A	5.5	6.5	1	5
B	2.5	9.5	7	33.3
C	3.5	8	4.5	27.7
D	-0.5	7	7.5	85.2
E	8	3	5	74.1
F	4.5	3	1.5	74.1
G	6	4	2	81.5
H	3	0.5	2.5	79.6
I	8	1	7	74.1
Overall (%)				**59.9%**

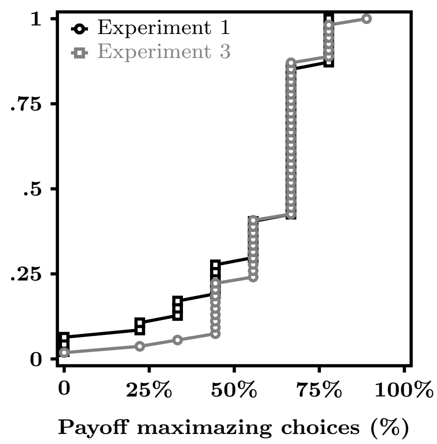


Figure 4: Empirical distribution function of percentage of payoff maximizing choices in experiments 1 and 3

16.7% when the calculator was activated. Pairwise correlation between the number of times the calculator was activated by a subject and the total number of payoff maximizing choices made is positive but not statistically significant for choice sets D to I: .271 with $p = .128$. Pairwise correlation is negative but not significant for choice sets A, B and C: -.140 with $p = .649$. In this experiment, providing a way to reduce the cognitive reasoning needed does not improve choices.

Experiment 3: Engaging Subjects With Monetary Incentives

In experiments 1 and 2, choices are not rewarded. Subjects are paid a fixed amount for taking part in the experiment irrespective of the choices that they make. One may question (economists in particular) the subjects' motivation to maximize their payoffs when those payoffs are hypothetical. Subjects' intrinsic motivation alone may not be enough to engage them in making the necessary cognitive effort to solve the task (see, e.g., Camerer and Hogarth, 1999, for a discussion of this issue in economic experiments). To test whether the low proportion of payoff maximizing choices observed in experiments 1 and 2 is due to a lack of monetary incentives, in experiment 3, we replicate experiment 1, but we add monetary incentives.

In experiment 3, subjects earnings are based on the payoff that they receive from selling purchased tokens at the end of the experiment. This means that subjects need to have money with which they can buy tokens. In this study, we provide subjects with an electronic account with £4 in it. All tokens offered for sale in the experiment cost less than £4. As in experiment 1, subjects complete nine choice sets, but their earnings depend on only one randomly chosen round.[10] Randomly selecting the round that is binding prevents subjects previous choices from influencing the amount of money subjects have to spend in each round. To ensure that subjects will not earn £0, all subjects are given £2 for showing up on time and participating. Experiment 3 is identical to experiment 1 and the experimental instructions are identical to those for experiment 1 except that they do not use subjective language.

Table 4 presents the proportion of payoff maximizing choices by choice set in experiment 3. Overall, we observe that 59.9% of choices are payoff maximizing when monetary incentives are at stake. This proportion is not

[10]Subjects were asked to consider a comparable hypothetical situation in the experiment 1

statistically different from that of experiment 1 ($p = .607$). The proportion of payoff maximizing choices in choice sets D to I is close to that observed in experiment 1: 78.1% with monetary incentives and 73.0% without. There are no differences for the most problematic choice sets A, B and C: 23.5% in experiment 2 and 22.7% in experiment 1. Figure 4, that plots the EDF of the percentage of payoff maximizing choices by subject in experiment 1 and 3, confirms aggregate findings at the individual level: the EDF are very much alike and although the EDF in experiment 3 is slightly to the right of the EDF in experiment 1, first order dominance is not significant ($p = .480$).

Subjects took more time to take their choices with monetary incentives than without. Median total response time (time taken by a subject to answer all nine choice sets) in experiment 3 is 197 seconds compared to 157 seconds in experiment 1, the increase is significant with $p = .050$ (median difference bootstrap test). The EDF of total response time shows that monetary incentives increased response time for subjects with the lowest response times (see figure in Appendix C). A KS bootstrap distribution test indicates that the EDF of response time in experiment 3 first order dominates the EDF of response time in experiment 1 ($p < .025$). Longer response times seem to be significantly associated with a higher proportion of payoff maximizing choices by subject: pairwise correlation is .349 with $p = .010$. In other words, subjects who made more cognitive effort were more successful at making payoff maximizing choices.

Experiment 4: Combining Monetary Incentives and the Calculator

Monetary incentives in experiment 3 did not improve subjects' choices overall, but they did encourage respondents to take longer to make their choices and response time was positively correlated with making payoff maximizing decisions. In experiment 4, we combine monetary incentives (as in experiment 3) with the provision of the calculator (as in experiment 2). Our hypothesis is that monetary incentives could foster cognitive reasoning and induce subjects to use the calculator more and this would lead, in turn, to a higher proportion of payoff maximizing choices. In other words, monetary incentives and the calculator could act as complements to improve choices.

Figure 5 presents calculator use in experiment 4, across rounds and by choice set. Monetary incentives do not increase calculator use compared to experiment 2. The calculator was activated in 24.2% of the choice tasks in experiment 4 compared to 24.6% of the choices tasks in experiment 2. Again, nearly fifty percent of subjects (48.7%) never activated the calculator and 12.8% activated it in every choice set (13% in experiment 2). Figure 5 presents the activations of the calculator as well as the mean number of entries by period and by choice set. As in experiment 2, there little evidence of round or choice sets effects. Once the calculator has been activated, we do not find differences in the number of keyboard entries between experiment 4 and experiment 2. Overall, mean number of keyboard entries is 33.9 in experiment 4 and 33.6 in experiment 2.

Table 5 presents the proportion of payoff maximizing choices by choice set. Monetary incentives combined with the calculator leads to a small, but significant, increase in the proportion of payoff maximizing choices: 64.9% against 56.3% with $p = .037$. This improvement comes from better decisions in choice sets D to I: 89.8% of payoff maximizing choices in experiment 4 and 73.0% in experiment 1 ($p = .005$.). There is no significant change in the proportion of payoff maximizing choices in the most problematical choice sets A, B and C: 15.4% in experiment 4 and 22.7% in experiment 1 ($p = .328$).

Figure 6 presents the EDF of the percentage of payoff maximizing choices by subject in experiment 1 and experiment 4. The EDF in experiment 4 is shifted to the right for the lowest percentages (because of the improvements for choice sets D to I) whereas the upper part of the EDF remains similar to that of experiment

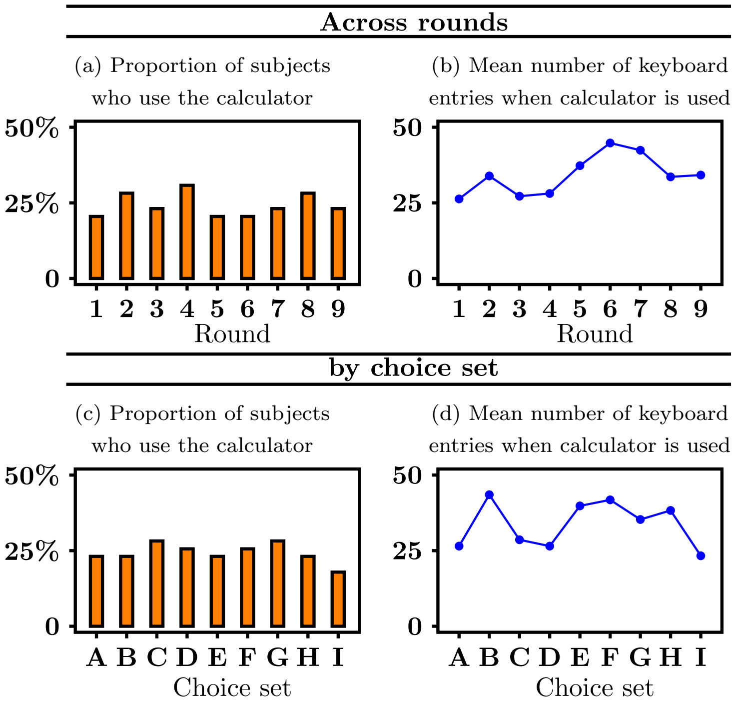


Figure 5: Use of the calculator across rounds and by choice set in experiment 4

1 (because of the unchanged proportions for choice sets A, B and C). The EDF in experiment 4 first order dominates the EDF in experiment 4 ($p = 0.015$).

In choice sets D to I, there is no difference in payoff maximizing choices between those who used the calculator (92.9%) and those who did not (88.8%). These proportions were respectively 92.5% and 80.5% in experiment 2. In choice sets A, B and C, 17.2% of choices made with the calculator are payoff maximizing compared to 14.7% of choices made without the calculator. The effect of the calculator provision and monetary incentives is additive, both improved the proportion of payoff maximizing choices in choice sets D to I but they are unsuccessful at improving choices in choice sets A, B and C. The proportion of payoff maximizing choices in these choice sets remains disappointing.

Experiment 5: DCE under Truth-telling Oath

In experiment 2, the potential of the calculator to improve choices was limited because too few subjects used it. In experiment 3, the use of monetary incentives was not sufficient to engage subjects. The combination of monetary incentives with a calculator in experiment 4 lead to small improvements, but not for the problematic choice sets. In experiment 5, we implement a non-priced commitment device –a truth-telling oath– in an hypothetical setting similar to of experiment 1.

The truth-telling oath has been found to improve bidding behavior in both induced and homegrown values auctions (Jacquemet et al, 2013). de Magistris and Pascucci (2014) have shown in a field DCE survey that a truth-telling oath can reduce hypothetical bias. The effect of the truth-telling oath can be explained as follows. Respondents' answers in hypothetical surveys are not binding. In the absence of a direct link between respondents and their

Table 5: Token values and the proportion of correct choices across choice sets in experiment 4 ($n = 39$)

Choice	Value A	Value B	Payoff Diff	Calc. & Inc. N=47 Percentage
A	5.5	6.5	1	5.2
B	2.5	9.5	7	30.8
C	3.5	8	4.5	10.3
D	-0.5	7	7.5	87.1
E	8	3	5	84.6
F	4.5	3	1.5	89.7
G	6	4	2	94.9
H	3	0.5	2.5	87.2
I	8	1	7	94.9
Overall (%)				**64.9%**

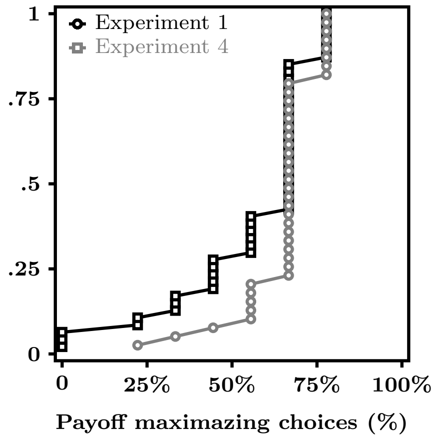


Figure 6: Empirical distribution function of percentage of payoff maximizing choices in experiments 1 and 4

declaration, respondents may lack the necessary commitment to provide reliable answers. Taking an oath restores this link by committing people to truth-telling (see Jacquemet et al, 2013, 2011, for more details)

The truth-telling oath procedure in experiment 5 follows that of Jacquemet et al (2013) and Jacquemet et al (2011). Subjects are presented with the oath upon entry into the lab, but after completing the consent form. Subjects are presented with an oath form at a private desk. The form is entitled "Solemn oath" and contains an unique sentence with a single prescription that reads "I, ..., the undersigned do solemnly swear that during the whole experiment, I will **tell the truth and always provide honest answers**" (the oath form is presented in Appendix D). Subjects are told that signing the form is voluntary and that neither their participation in the experiment nor their payoff depend on signing (see Jacquemet et al, 2011, who explain this choice by borrowing insights from the social psychology of commitment). The oath procedure was carried out by the same person for all subjects –she also ran experiments 1 to 4. All subjects but one signed the oath in experiment 5, selection is therefore not an issue in what follows.[11]

Table 6 presents the proportion of payoff maximizing choices in experiment 5. Overall, the results are unambiguous. The oath significantly increases the proportion of payoff maximizing choices compared to experiment 1: 78.3% compared to 56.3% ($p < .001$). The increase is due to improvements in choice sets A, B and C: the proportion of payoff maximizing choices increases by a factor 3, from 22.7% in experiment 1 to 76.5% in experiment 5 ($p < .001$). We observe no significant difference in choice sets D to I: 73% in experiment 1 and 78.8% when subjects are under oath ($p = .683$).

The EDF of the percentage of payoff maximizing choices a subject makes is plotted in Figure 7. Now, 40.9% of subjects make payoff maximizing choices in all 9 choice sets and 54.5% make at most one non payoff maximizing choice. Consequently, the EDF in experiment 5 first order dominates significantly the EDF in experiment 1 ($p < .001$).

Response time in experiment 5 is significantly longer compared to experiment 1. The response time increase is comparable to the increase observed with monetary incentives in experiment 3. Median total response time is 208 seconds in experiment 5 compared to 157 seconds in experiment 1 ($p = .008$) and 197 seconds in experiment 3 ($p = .400$). The EDF of response time in experiment 5 first order dominates that of experiment 1 ($p = .004$,

[11]Choices of the subject who did not signed are not dropped from the statistical analysis, i.e. we adopt an intention to treat strategy.

Table 6: Token values and the proportion of correct choices across choice sets in experiment 5 ($n = 44$)

Choice	Value A	Value B	Payoff Diff	Truth-Tell. oath N=44 Percentage
A	5.5	6.5	1	59.1
B	2.5	9.5	7	86.4
C	3.5	8	4.5	84.1
D	-0.5	7	7.5	90.9
E	8	3	5	77.3
F	4.5	3	1.5	65.9
G	6	4	2	81.8
H	3	0.5	2.5	77.3
I	8	1	7	79.5
Overall (%)				**78.3%**

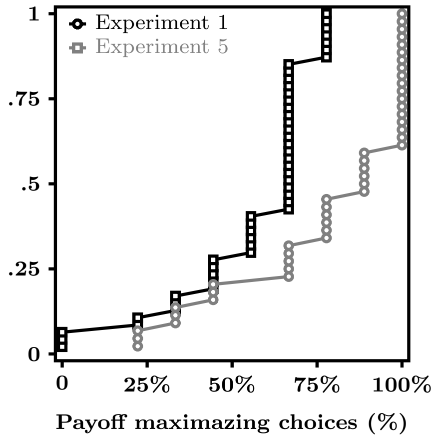


Figure 7: Empirical distribution function of percentage of payoff maximizing choices in experiments 1 and 5

see EDF of response time in experiments 1, 3 and 5 in appendix E). The longer response time in experiment 5 has two sources. First, subjects making a payoff maximizing choices in choice sets A, B and C take longer in experiment 5 than in experiment 1: a median response time is 21 seconds in experiment 5 compared to 12 seconds in experiment 1. In experiments 1 and 3, payoff maximizing choices were made quicker than non payoff maximizing choices: median response time for payoff maximizing choices was 14 seconds compared to 21 seconds for non payoff maximizing choices. This is no longer the case in experiment 5: median response time for non payoff maximizing choices is now 19 seconds. Second, response times are longer for both payoff maximizing and non payoff maximizing choices in choice sets D to I. Median response time is increased by 6 seconds to 18 seconds for non payoff maximizing choices (12 seconds in experiment 1 and 16 seconds in experiment 3) and and by 4 seconds to 21 seconds for payoff maximizing choices (17 seconds in experiment 1 and 20 seconds in experiment 3).

The truth-telling oath induced a large and significant improvement in choices. This improvement is due to a dramatic improvement in choice sets A, B and C whereas the proportion of these payoff maximizing decisions in choice sets D to I remain unchanged. The oath is also associated with an increase in response time of payoff maximizing choices, to a level similar to that of non payoff maximizing choices when compared to response times observed experiment 1. This phenomenon was not observed when monetary incentives were introduced.

Experiment 6: Fostering Effort with a Task Oath

The response time in experiment 5 shows that the truth-telling oath increased cognitive reasoning. One explanation is that subjects under oath dedicate more effort to making accurate choices. However, monetary incentives also increase cognitive reasoning, but without improving choices in choice sets A, B, and C. In experiment 6, we explore whether the explanation that the oath improves choices because it foster cognitive reasoning is reasonable. Experiment 6 replicates experiment 5, but with a modified oath form that explicitly targets cognitive effort without referring to truth-telling behavior. The oath form now reads "I, ..., the undersigned do solemnly swear that during the entire experiment, I will **faithfully and conscientiously fulfil the tasks that I am asked to complete to the best of my skill and knowledge**". Otherwise, the oath form and the oath procedure are identical to that of experiment 5. All subjects agreed to sign the oath in experiment 6.

Table 7: Token values and the proportion of correct choices across choice sets in experiment 6 ($n = 37$)

Choice	Value A	Value B	Payoff Diff	Task oath N=44 Percentage
A	5.5	6.5	1	10.8
B	2.5	9.5	7	35.1
C	3.5	8	4.5	18.9
D	-0.5	7	7.5	89.2
E	8	3	5	83.8
F	4.5	3	1.5	75.7
G	6	4	2	86.5
H	3	0.5	2.5	83.8
I	8	1	7	89.2
Overall (%)				**63.7%**

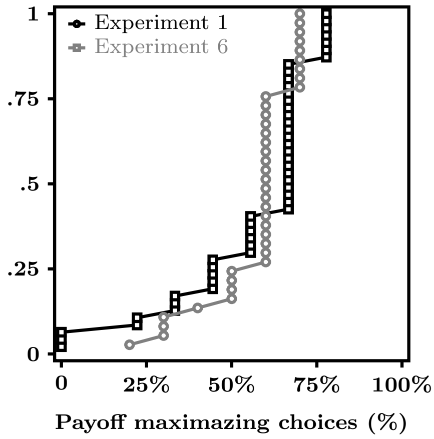


Figure 8: Empirical distribution function of percentage of payoff maximizing choices in experiments 1 and 6

The proportion of payoff maximizing choices presented in Table 7 shows that the task oath had only a small positive effect on choices compared to experiment 1: 63.7% compared to 56.26% ($p = .074$). This improvement is because of improvements in choice sets D to I (84.5% in experiment 6 vs. 73% experiment 1, $p = .071$). There is no effect of the task oath on choices in choice sets A, B and C (21.6% in experiment 6 vs. 22.7% in experiment 1, $p = .321$). This result is confirmed at the subject level by the comparison of the EDF in experiments 1 and 6 (Figure 8): the EDF in experiment 6 does not first order dominates that in experiment 1 ($p = .536$).

Response times show that subjects took the task oath seriously. There is a significant increase in response time compared to experiment 1: median total response time is 237 seconds in experiment 6 compared to 157 seconds in experiment 1, $p = .012$. The EDF of response time in experiment 6 first order dominates that of experiment 1 with $p < .001$. The response time increase is similar to that for the truth-telling oath (experiment 5): median total response time was 208 seconds in experiment 5 (a response time of 237 seconds is not significantly different from 208, $p = .176$).

In experiment 6, payoff maximizing choices were make quicker than non-payoff maximising choices. Median response time of payoff maximizing choices in choice set A, B and C is 13.5s, 18s and 5s respectively. Median response time of non payoff maximizing choices in choice sets A, B and C is 19s, 37.5s and 26s respectively. This was not the case on experiment 5 in which payoff maximizing choices took longer, but was observed in experiment 1 and experiment 3 (when monetary incentives were at stake). In choice sets D to I, subjects took more time to make payoff maximizing choices (25s) than non payoff maximizing choices (14.5s) – 17s and 12.5s in experiment 1 and 22s and 18s in experiment 5.

Experiment 7: The Oath of Office

Response time in experiment 6 showed that the task oath fostered cognitive reasoning which improved choices made in choice sets D to I, but had no impact on choices in choice sets A, B and C. The task oath may appear too abstract and singular, whereas the truth-telling oath is a real world institution with a moral content. In other words, the task oath was too far from what one would encounter in the field and, consequently, it did not succeed in increasing the proportion of payoff maximizing choices.

Table 8: Token values and the proportion of correct choices across choice sets in experiment 7 ($n = 37$)

Choice	Value A	Value B	Payoff Diff	Oath of office N=44 Percentage
A	5.5	6.5	1	0.0
B	2.5	9.5	7	24.3
C	3.5	8	4.5	8.1
D	-0.5	7	7.5	91.9
E	8	3	5	94.6
F	4.5	3	1.5	75.7
G	6	4	2	91.9
H	3	0.5	2.5	75.7
I	8	1	7	91.9
Overall (%)				**61.6%**

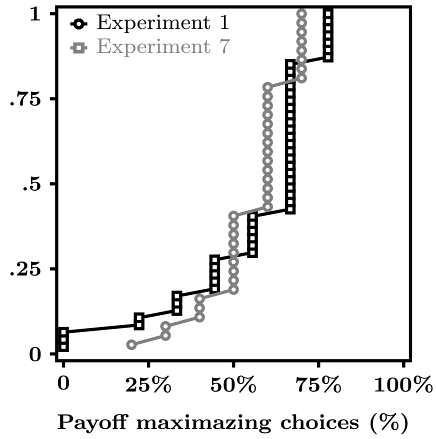


Figure 9: Empirical distribution function of percentage of payoff maximizing choices in experiments 1 and 7

In experiment 7, we replicate experiment 6 but bringing into the lab a real world oath that also targets effort to perform one's assigned task but with a moral content: the *oath of office*. In experiment 7, we carry out the same oath procedure than in experiment 5 and 6, using the same oath form but that now reads "I, ..., the undersigned do solemnly swear that during the whole experiment, I will **faithfully and conscientiously fulfil my duties to the best of my skill and knowledge**". The oath of office also targets effort but contains the necessary moral reminders that one would encounter in the field if she had, for instance, to take an oath before undertaking the duties of a public office. Providing ethical standards to people, which the truth-telling oath and the oath of office do, has been shown to have significant effect on behavior (Mazar et al, 2008). All subjects, but one, signed the oath in experiment 7. As in experiment 5, statistical analysis is carried out without dropping observations.

Empirical results indicate that the oath of office has no overall effect on behavior. Table 8 shows that 61.6% of choices are payoff maximizing, which is not statistically different than experiment 1 ($p = .185$). As with the task oath, we observe, a small but significant increase in payoff maximizing choices in choices sets *D-I* (84.7% in experiment 7 vs. 73.0% in experiment 1, $p = .068$). The oath of office has no effect in choice sets A, B and C (21.6% in experiment 7 vs. 22.7% in experiment 1, $p = .834$). At the individual level, the comparison of the EDF of the percentage of payoff maximizing choices in experiment 1 and 7 presented in Figure 9 confirms that the oath of office is not effective at improving choices in our setting.

Comparing the oath of office to the truth-telling and the task oaths, we find that response time data shows that subjects take the oath of office seriously. Median total response time in experiment 7 is 213s, significantly greater than in experiment 1 (157s) and similar to both other oath experiments. The EDF of total response time for each subject presented in appendix G shows that the EDF in experiment 7 first order dominates the EDF in experiment 1 (KS bootstrap test, $p = .009$) and that it is very similar to that of experiment 6. Response times for payoff maximising choices are longer than non payoff maximising choices with the oath of office, as with the truth-telling oath. Subjects maximizing payoffs took longer to choose (25s) than those who did not payoff maximise (21s). Still, subjects did often not maximize their hypothetical monetary payoff in choice sets A, B and C.

Concluding remarks

In the present paper, we adopt the induced values design of Luchini and Watson (2014) that shows that non payoff maximizing choices in a simple induced value DCE are numerous. In this article, we study the potential of several simple devices to improve choices. Only the truth telling oath was effective in our setting. Traditional devices such as monetary incentives or helping respondents (a calculator) or the combination of both had little or no effect on choices. The analysis of response time tells us that one cannot consider decisions in choice sets A, B and C as bare mistakes for at least two reasons: 1/ response time in choice sets A, B and C are larger for non payoff maximizing decisions and 2/ because devices that target an increase of cognitive effort show that people spend more time to decide, which indicates that they produce more effort (but without improving decisions in choice sets A, B and C) . This suggests that people may deliberately choosing token A in choice sets A, B and C after careful thinking, because in the truth-telling oath experiment, payoff maximizing choices take more time than non payoff maximizing choices (given that we also overserved in experiment 3 a general increase in response time).

The analysis of response time together with the implementation of the two oath procedures that target effort rather than truth-telling tell us that truth-telling in DCE is a serious concern. We suggest that this concern should be addressed in a systematic way in field surveys, by asking respondents to take an oath prior to being interviewed as we have done in this paper or by implementing weaker forms of commitment like a preliminary pledge or even a simple signed agreement to tell the truth.

One may argue that our induced values design –however a now standard tool in experimental economics to testbed economic institutions– has an obvious limitation: it may seem too abstract to many. Respondents in typical DCE questionnaires do not solve simple mathematical problems but rather make choices between real world alternatives, and the benchmark for hypothetical/choice modeling should rather be real choices in the field. This reasoning has at least two shortcomings. First, it supposes that real choices exist and are observable. However, in many situations, like in health care or environmental preservation, the economic assessment is often carried out ex-ante and no markets in which one could observe choices and competitive prices are available (this is however less clear in marketing research). Second, it assumes that real choices in the field fulfil the necessary rational requirements so that welfare estimates are meaningful for policy purposes. In other words, context may well matter in the field but one may question the idea that it systematically leads to welfare maximizing decisions, that would actually reveal meaningful preferences. For these reasons, we believe that our design (that replicates very closely the new consumer theory put forward by Lancaster) is a good starting point for exploring further the extent to which preferences revealed in field DCE surveys are relevant for collective decision making. We believe that in this article we open a promising avenue for future research in addition to what is currently done in DCE research.

References

Abadie, A. Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Model *Journal of the American Statistical Association*, **97**, 284–292.

de Bekker-Grob, E. Ryan, M. and Gerard, K. (2012). Discrete choice experiments in health economics: A review of the literature. *Health Economics*, **20**, 145–172.

Camerer, C. and Hogarth, R (1999). The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework *Journal of Risk and Uncertainty*, **19**, 7–42.

Carlsson, F. and Martinsson, P. (2001). Do hypothetical and actual marginal willingness to pay differ in choice experiments? Application to the valuation of the environment. *Journal of Environmental Economics and Management*, **41**, 179–192.

Carlsson, F. Fryblom, P. and Lagervist, C.J. (2005). Do hypothetical and actual marginal willingness to pay differ in choice experiments? Application to the valuation of the environment. *Journal of Environmental Economics and Management*, **41**, 179–192.

Carson, R.T. and T. Groves (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics*, **37**, 181–210.

DeShazo, J.R. and Fermo G. (2002). Designing choice sets for stated preference methods: The effects of complexity on choice consistency. *Journal of Environmental Economics and Management*, **44**, 123–43.

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

Hole, A. R. (2011) A discrete choice model with endogenous attribute attendance. *Economics Letters* , **110**, 203–205.

Jacquemet, N., Joulé, R.-V., Luchini, S. and J.S. Shogren (2013). Preference Elicitation under Oath *Journal of Environmental Economics and Management* , **65**, 110-132

Jacquemet, N., A. James, S. Luchini, and J. Shogren (2011). Social psychology and environmental economics: a new look at ex ante corrections of biased preference evaluation *Environmental & Resource Economics* , **48**, 411–433.

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

Kanninen, B. (2006) *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice.* Springer, Dordrecht, The Netherlands.

Lancaster K. (1966) A new approach to consumer theory. *Journal of Political Economy*, **74**, 132-157.

Louviere, J.J., Pihlensa, D. and R. Carson (2011) Design of Discrete Choice Experiments: A Discussion of Issues That Matter in Future Applied Research *Journal of Choice Modelling*, **4**,1-8.

Louviere, J.J., Islam, T., Wasi, N., Street, D. and L. Burgess (2008). Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price? *Journal of Consumer Research*, **35**, 360–375.

Louviere, J.J., Hensher, D.A. and J. D. Swait (2000) *Stated Choice Methods: Analysis and Applications* Cambridge University Press: Cambridge, UK.

Luchini, S. and V. Watson (2014) Are choice experiments reliable ? Evidence from the lab *Economics Letters*, **124**, 9–13.

Lusk, L.J. and Schroeder, T.C. (2004) Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks *American Journal of Agricultural Economics*, **86**, 467–482.

de Magistris, T. and S. Pascucci (2014) The effect of the solemn oath script in hypothetical choice experiment survey: A pilot study *Economics Letters*, **123**, 252–255.

Mazar, N., On, A. and D. Ariely (2008). The Dishonesty of Honest People: A Theory of Self-concept Maintenance *Journal of Marketing Research* , **45**, 633-644.

Mazzotta, M. and Opaluch, J. (1995). Decision making when choices are complex: A test of Heiner's hypothesis. *Land Economics*, **71**, 500–515.

Mitani, and Flores, N., (2009). Demand revelation, hypothetical bias, and threshold pubic goods provision. *Environmental and Resource Economics*, , 231–243.

Rubinstein, A. (2007). Instinctive and Cognitive Reasoning: A Study of Response Times *Economic Journal*, **117**, 1243–1259.

Sekhon, J. (2011). Multivariate and Propensity Score Matching Software with Automated Balance Optimization *Journal of Statistical Software*, **42**, 1–53.

Smith, V.L. (1976). Experimental Economics: Induced Value Theory. *The American Economic Review* , **66**, 274–279.

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

Swait, J. and Adamowicz, W. (2001). Choice environment, market complexity and consumer behaviour: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behavior and Human Decision Processes.*, **86**, 141–167.

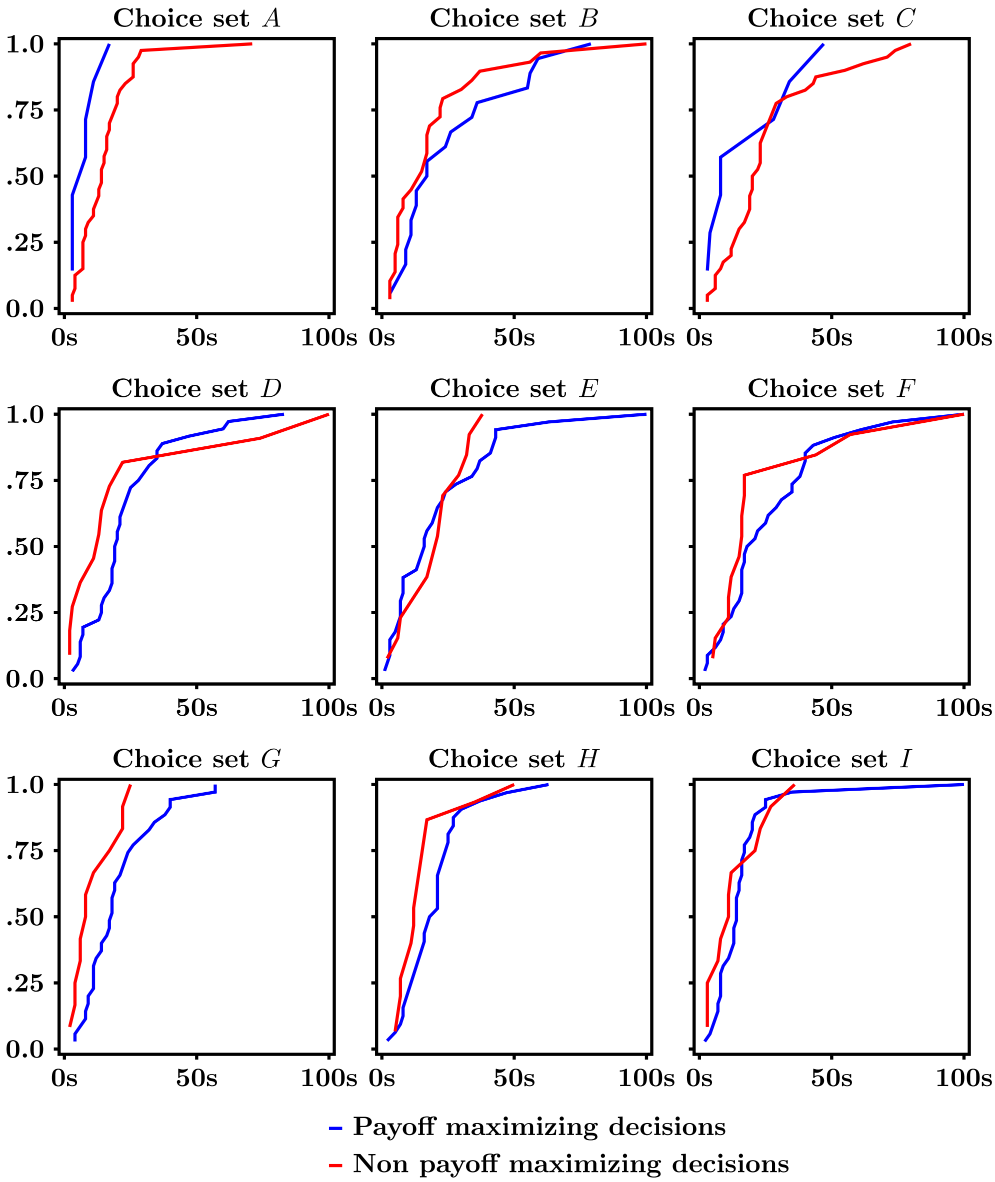
Taylor, L.O., and McKee, M., Laury, S.K. and Cummings, R.G. (2001). Induced value tests of the referendum voting mechanism. *Economic Letters* **71**, 61–65.

A Proportion of payoff maximizing decision by round and experiments

Round	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Baseline	42.5%	51.1%	57.4%	70.2%	61.7%	46.8%	55.3%	61.7%	61.7%
Calculator	60.9%	65.2%	58.7%	67.4%	60.8%	56.5%	56.5%	67.4%	60.8%
Money	53.7%	55.5%	68.5%	50.0%	61.1%	64.8%	59.3%	70.4%	55.5%
Money & calculator	74.4%	64.1%	56.4%	66.7%	69.2%	56.4%	69.2%	61.5%	66.7%
Oath	72.7%	77.3%	84.0%	75.0%	77.3%	79.5%	77.3%	77.3%	81.8%

B Response time by choice set in experiment 1

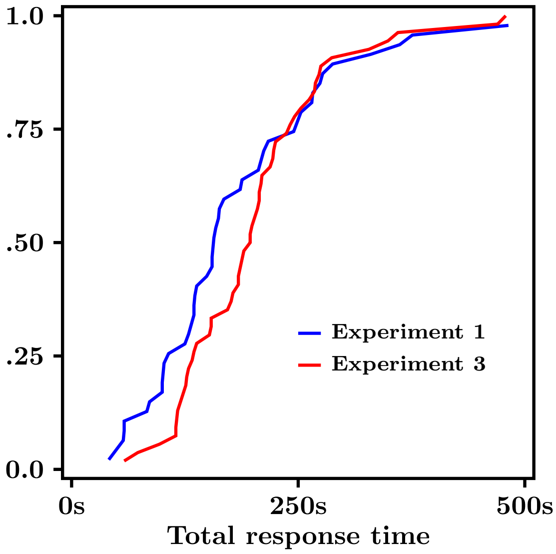
Empirical distribution functions or response time in experiment 1 are plotted by choice set In the figure below. The x-axis is labelled in seconds and ranges from 0 second to 100 seconds. For the sake of vizualisation, 6 observations –1.4% of the sample– greater than 100 seconds were dropped. The red line corresponds to decision time of non payoff maximizing decisions and the blue line to payoff maximizing decisions. In choice sets A and C, the red line appears to the right of the blue line: EDF of response time of non payoff maximizing decisions first order dominate (FOD) the EDF of response time of payoff maximizing decisions. That is, a non payoff maximizing decision take more time than a payoff maximizing decision. For all other choice sets, response time EDF are either similar or the EDF of non payoff maximizing decisions appear to the left of that of payoff maximizing decisions.



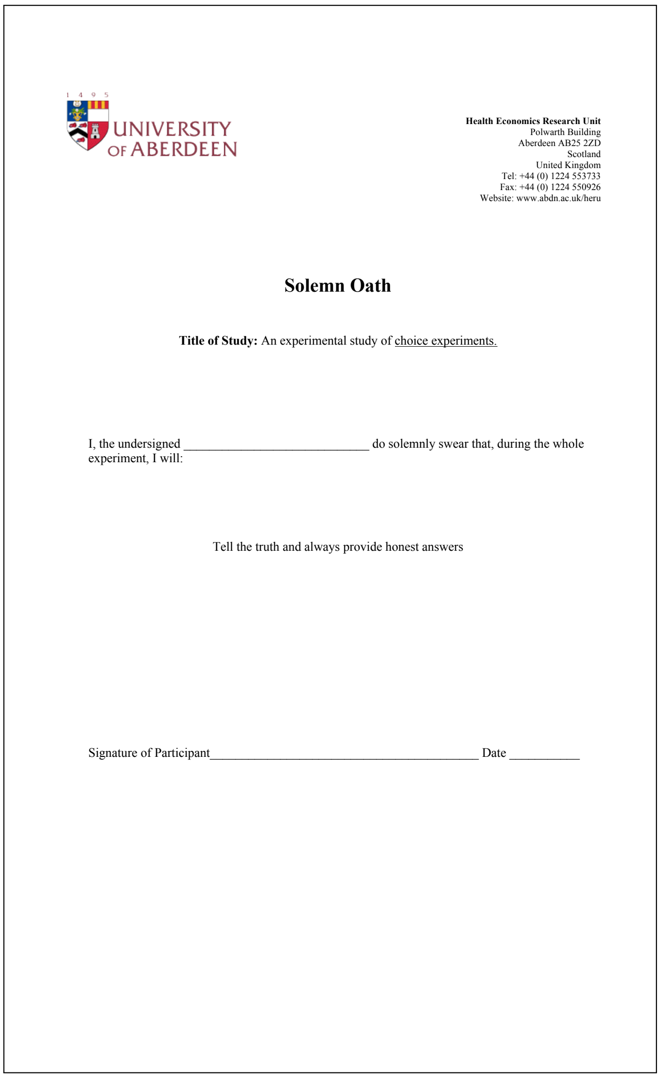
For each choice set, we peformed a Kolmogorov-Smirnov bootstrap distribution test to assess whether first order stochastic dominance (in one way or the other) was statistically significant or not. KS bootstrap tests indicate that FOD of response time of non payoff maximizing decisions over reponse time of payoff maximizing decisions is statistically significant for choice set A and C with $p = .032$ and $p = .069$ respectively. FOD of response time of payoff maximizing decisions over non payoff maximizing decisions is statistically significant for choice sets B, D, G and H with $p = .075$, $p = .041$, $p = .013$ and $p = .021$ repectively. The null of non FOD cannot be rejected for choice sets E ($p = .391$), F ($p = .114$) and I ($p = .127$).

C Response time in experiment 1 and experiment 3

In the Figure below, we computed for each subject the time she took to answer all 9 choice sets (total response time) and plotted the EDF of total response time in experiment 1 and experiment 3. Note that we dropped one subject in experiment 1 with a response time of 765 seconds to make the Figure easier to read. The KS bootstrap test presented in the text is carried out without dropping this subject.



D Oath form used in experiment 5



1495
UNIVERSITY OF ABERDEEN

Health Economics Research Unit
Polwarth Building
Aberdeen AB25 2ZD
Scotland
United Kingdom
Tel: +44 (0) 1224 553733
Fax: +44 (0) 1224 550926
Website: www.abdn.ac.uk/heru

Solemn Oath

Title of Study: An experimental study of choice experiments.

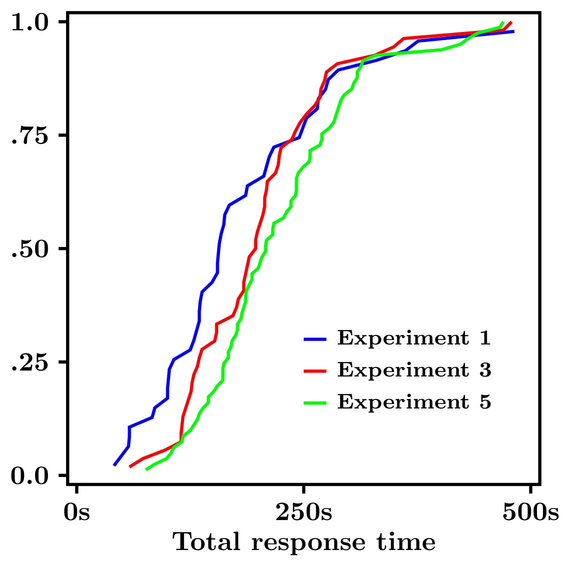
I, the undersigned ___________________________ do solemnly swear that, during the whole experiment, I will:

Tell the truth and always provide honest answers

Signature of Participant_______________________________________ Date __________

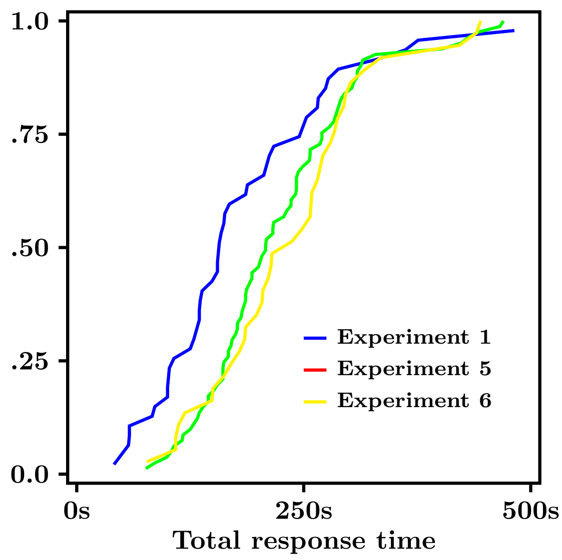
E Response time in experiments 1, 3 and 5

In the Figure below, we computed for each subject the time she took to answer all 9 choice sets (total response time) and plotted the EDF of total response time in experiments 1, 3 and 5.



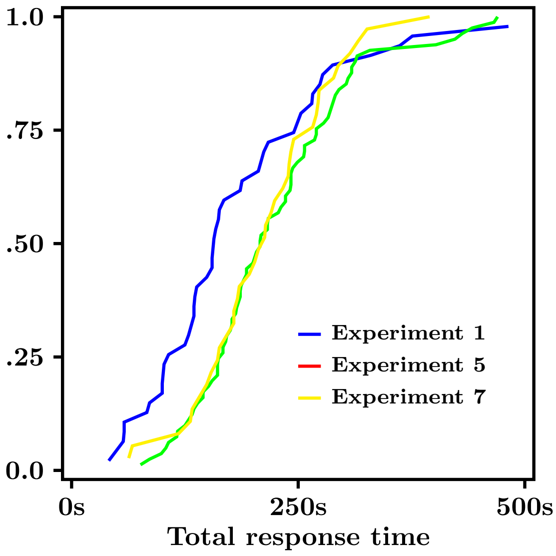
F Response time in experiments 1, 5 and 6

In the Figure below, we computed for each subject the time she took to answer all 9 choice sets (total response time) and plotted the EDF of total response time in experiments 1, 5 and 6.



G Response time in experiments 1, 5 and 7

In the Figure below, we computed for each subject the time she took to answer all 9 choice sets (total response time) and plotted the EDF of total response time in experiments 1, 5 and 7.


1.0
.75
.50
.25
0.0
0s
250s
500s
Total response time
Experiment 1
Experiment 5
Experiment 7