Discrete Choice under Oaths

Nicolas Jacquemet, Stephane Luchini, Jason Shogren, Verity Watson

Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers

Discrete Choice under Oaths

Nicolas JACQUEMET, Stéphane LUCHINI
Jason F. SHOGREN, Verity WATSON

2019.07

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
https://centredeconomiesorbonne.univ-paris1.fr/
ISSN : 1955-611X

Discrete Choice under Oaths*

Nicolas Jacquemet[†] Stéphane Luchini[‡] Jason F. Shogren[§] Verity Watson[¶]

May 9, 2019

Abstract

Using discrete choices to elicit preferences is a major tool to help guide public policy. Although Discrete Choice Experiment (DCE) remains by far the most popular mechanism used to elicit preferences, its reliability still is questionable. Using an induced value experimental design, we show that standard benchmarks achieve no more than 56% (hypothetical answers with no monetary incentives) to 60% (real monetary incentives) of payoff maximizing choices. Herein we demonstrate that having respondents sign a the truth-telling oath reduces non-payoff maximizing choices by nearly 50% relative to these benchmarks. The explicit and voluntary commitment to honesty improved decisions. Further, we show that it is the explicit commitment to honesty induced by the truth-telling oath that improves choices, not just *any oath mechanism*, i.e., an oath to task or to duty did not improve choices.

Keywords: Discrete Choice Experiments, Stated Preferences, Oath, Truth-telling, External validity, Welfare.

JEL Codes: C9; H4; Q5.

*Financial support from the Health Chair, a joint initiative by PSL, Université Paris-Dauphine, ENSAE and MGEN under the aegis of the Fondation du Risque (FDR) is gratefully acknowledged. J. Shogren thanks the Rasmuson Chair at the University of Alaska-Anchorage for the support. The Health Economics Research Unit is funded by the Chief Scientist Office of the Scottish Government Health and Social Care Directorates.

†Paris School of Economics and University Paris 1 Panthéon-Sorbonne. CES-Centre d'Economie de la Sorbonne – Maison des Sciences Economiques 106-112 boulevard de l'Hôpital 75013 Paris

‡Aix-Marseille University (Aix-Marseille School of Economics), CNRS and EHESS, Centre de la Vieille Charité, 13236, Marseille Cedex 02, France.

§Corresponding author: Department of Economics, University of Wyoming, Laramie, WY, 82071-3985, USA. Tel: 307.766.5430. Email: JRamses@uwyo.edu

¶University of Aberdeen. Health Economics Research Unit University (HERU) of Aberdeen, Polwarth Building Foresterhill Aberdeen, AB25 2ZD. UK.

1 Introduction

Discrete choice experiments (DCE) have emerged as a go-to method to elicit preferences for new goods, services, and public policy options. Examples abound. DCEs have been used, for example, in employment searches, health care and medical options, better environmental quality policy, and new food products (see e.g., Mas and Pallais, 2017; Soekhai, de Bekker-Grob, Ellis, and Vass, 2018; Adamowicz, Glenk, and Meyerhoff, 2014; Vossler, Doyon, and Rondeau, 2012; Hensher, Rose, and Green, 2015). In the UK, the HM Treasury Green Book recommends using DCEs to elicit values for cost-benefit analysis (CBA) of new projects, such as infrastructure investments. Three main reasons help explain why DCEs are attractive: (1) binary choices are grounded in economic theory, although they are not necessarily incentive compatible;[1] (2) binary choices—between two goods differentiated by alternative levels of common attributes and cost[2]—are familiar and understandable to most people who make conscious and implicit multi-attribute decisions everyday; and (3) binary choices reveal the implicit marginal monetary values of attributes, which are relatively straightforward to estimate empirically.[3]

But as with any preference elicitation method, DCEs still face questions of reliability. Reliability is an issue because people misrepresent their preferences in DCEs either by mistake or strategic behavior (see, for example, Meginnis, Burton, Chan, and Rigby, 2018). Based on our previous work, we use laboratory experiments to explore whether a truth-telling oath can increase the reliability of the DCE by inducing a real commitment to honesty (Jacquemet, James, Luchini, and Shogren, 2011; Jacquemet, Joule, Luchini, and Shogren, 2013). We have observed that the truth-telling oath promotes more honesty in a variety of settings, including bidding in a second-price auction, voting for a referenda, strategic choices in coordination games, and decisions to tell white lies (Jacquemet, James, Luchini, and Shogren, 2017; Jacquemet, Luchini, Shogren, and Zylbersztejn, 2018; Jacquemet, Luchini, Rosaz, and Shogren, 2018). The truth-telling oath works as a "preparatory act" that commits a person to truth-telling in subsequent tasks—in this case stated preferences.[4] In terms of preference elicitation,Jacquemet, Joule, Luchini, and Shogren (2013) have shown that people bid more sincerely in both induced and home-grown auctions when under oath. The truth-telling oath can help commit the person to the choice at hand (also see

[1]See Manski and Lerman (1977); McFadden (2001), and Carson and Groves (2007).

[2]See the discussion in Freeman III, Herriges, and Kling (2014).

[3]See e.g., Train (2003); Adamowicz and Swait (2011).

[4]The literature on the compliance without pressure paradigm shows that, under certain conditions, the sequence of action matters: decisions made at one point in time have strong and long-lasting consequences on subsequent behavior. The literature has shown that volitional actions are much more likely to produce the target behavior than mandatory preliminary behavior (see, e.g., Beauvois and Joule, 2002). In a seminal experiment in this field, an experimenter stands at the corner of a street and asks people who pass-by for a dime so as to take the bus. While the share of people who comply with this request is around 1/3, it is twice as much when time has been asked before asking for a dime. Asking for time works as a preparatory act which produces significant changes in subsequent behavior. The oath procedure carried out herein follows that line of thought.

Table 1: Attributes' induced values

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

other citations). The truth-telling oath–similar to that taken by witnesses before giving evidence in a court of law–improves truthful revelation of preferences in both induced and home-grown value auctions (Jacquemet, Joule, Luchini, and Shogren, 2013) and home-grown value referenda (Jacquemet, James, Luchini, and Shogren, 2017). de Magistris and Pascucci (2014) have also shown in a home-grown DCE survey that a truth-telling oath can reduce the gap between hypothetical choices and real economic choices. In the absence of an induced-value design, this study however cannot contrast elicited choices with respondents' true preferences.

We make four purposeful choices in our laboratory induced value DCE under oath.[5] First, we use induced values to control for preferences; second, we set our 2x2 benchmarks using hypothetical and paid choices with and without a calculator, which controls for both hypothetical bias and mistakes; third, we introduce the truth-telling oath to ask people to commit to honesty and forego strategic behavior; and fourth, we explore two oaths frames to understand better whether it is commitment to honesty or to the task or duty that leads more reliability in the DCE. Our results again suggest that the truth-telling oath improves preference elicitation–we find that DCE under oath reduces non-payoff maximizing choices by nearly 50%. Our results suggest that we can improve the reliability of preference elicitation with an explicit and voluntary commitment to honesty. Further, we find that the oath itself is not what matters–rather it is the commitment to the truth that matters. We use two alternative oath frames (oath on task or on duty) and find that people still use non-optimizing decision rules more frequently than under the truth-telling oath.

2 Experimental Design: DCE with Induced Values

Our induced value experimental design recreates the salient features of a stated choice study in an experimental economics laboratory setting–with induced (i.e., exogenous and perfectly observed) preferences. We follow the design of Luchini and Watson (2014, "wide hypothetical" treatment) and induce preferences for a multi-attribute good called a "token". Subjects may buy tokens

[5]For additional lab experiments that used an induced value approach to examine the reliability of DCEs without an oath, see e.g., Collins and Vossler (2009); Taylor, Morrison, and Boyle (2010); Carson, Groves, and List (2014); Interis, Xu, Petrolia, and Coatney (2016); Luchini and Watson (2014); Meginnis, Burton, Chan, and Rigby (2018).

during the experiment at the announced cost, which they will then sell back to the experimenter at the end of the experiment. Subject's preferences for the tokens are induced by announcing that the amount of money they will receive for a token depends on its attribute levels. A token has four attributes and each attribute can take three possible levels: colour (red, yellow, blue); shape (circle, triangle, square); size (small, medium, large); and cost. The levels of each of the attributes is associated with a monetary value, as shown in Table 1. The sum of the attribute levels determines the value of the token, i.e., the amount the monitor will pay to buy it back from the subject. This replicates the linear additive utility typically assumed in the estimation of preference parameters from DCE studies (Train, 2003).

Subjects are asked to complete nine choice tasks on a computer with one choice task per screen. In each task they are offered two tokens, and can choose either to buy one of the two tokens (and sell it back to the experimenter later), or to buy no token at all. The tokens included in the choice tasks were chosen using a fractional factorial design . The order in which choice tasks are presented is randomized at the individual level: for each subject, before each screen is displayed, we randomly draw without replacement which one of the remaining choice tasks will be displayed on the upcoming screen.

In this context, the payoff-maximizing choice amounts to buying the token with the highest profit in each choice set, i.e., the token in which the difference between the value (sum of attribute levels) and the cost is the greatest. The monetary value of attributes induces subject's preferences over the attributes in the choice sets—profits are the lab counterfactual of individual preferences in real life. In contrast with a home-grown preferences DCE study, in which preferences underlying elicited choices are respondent's private information, this induced value design allows us to assess whether subjects make the *best choice* for them, i.e., make choices that maximize their profit/experienced utility. It also allows us to assess how best choices vary across experimental treatments. The main outcome variable is the share of choices that coincide with the payoff-maximizing choice—which measures the preference revelation performance of the DCE.[6]

We also record response time—the time elapsed between the appearance of the choice of the screen and the decision. Assuming that subjects who respond more slowly engage in more cognitive reasoning, this outcome can be used as a proxy for cognitive reasoning in the task .[7]

[6] Subjects who take part in our experiments are students at the University of Aberdeen, who are recruited to the experiments using Exlab and ORSEE software (Greiner, 2015). 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. The experiment was programmed and conducted with the software z-Tree (Fischbacher, 2007). The computer based experiment means that the order of choice sets was randomized allowing us to separate choice sets effects from decision round effects.

[7] This relationship between cognitive effort and response time (RT) is confirmed by recent empirical evidence Krajbich, Bartling, Hare, and Fehr (2015): the more obvious a choice is, the lower is RT (Evans, Dillon, and Rand, 2015). It is worth highlighting we do not aim to perform reverse inference—i.e., identify the mental process at work

Table 2: Summary of the experimental treatments

Experimental treatments	Design	Hypothetical	Calculator	Oath
Treatment 1. *Baseline*	Each subject completes 9 choice tasks: (1 choice per screen). Task: Offered two tokens. He or she chooses to (1) buy one token to sell back, or (2) buy no token	Yes	No	No
Treatment 2. *Calculator*	Each subject has access to the Microsoft Windows ™ calculator	Yes	Yes	No
Treatment 3. *Paid*	We pay each subject based on his or her choices. Payoffs are determined by randomly selecting 1 round out of 9	No	No	No
Treatment 4. *Calc. + paid*	Combines experiments 2 and 3	No	Yes	No
Treatment 5. *Oath on truth*	"I, ..., the undersigned do solemnly swear that during the whole experiment, I will tell the truth and always provide honest answers"	Yes	No	Yes
Treatment 6. *Oath on task*	"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"	Yes	No	Yes
Treatment 7. *Oath on duty*	"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"	Yes	No	Yes

3 Benchmark DCE Treatments

Given our focus we want to define a set of benchmark behaviors that allows us to capture and control for hypothetical bias and mistakes. Our benchmark treatments consists of a 2×2 design: the experiment is either hypothetical or paid and subjects are provided with a calculator or not. All treatments are implemented between subjects—the same person participates in only one of the seven treatments (see Table 2 that summarizes all experimental treatments carried out in this paper).

Treatment 1 (*hypothetical*) mimics a typical DCE survey: subjects' choices are hypothetical. 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 individuals to "*put yourself in a situation where your account balance at the end of the experiment would depend*

in decision making (e.g., intuitive vs deliberative reasoning) based on RT variations, which requires one to control for unobserved components of decision-making (White, Curl, and Sloane, 2016).

on the choice you made..." (Taylor, McKee, Laury, and Cummings, 2001). The data for this treatment comes from Luchini and Watson (2014) "wide treatment".

Several studies however suggest that the DCE tasks are too complicated for respondents, and, as a consequence, respondents might not choose what is best for them. Choices are more complicated when the multi-attribute goods included in the choice set are similar (Mazzotta and Opaluch, 1995; Swait and Adamowicz, 2001), when the goods are described by many attributes (Swait and Adamowicz, 2001; DeShazo and Fermo, 2002), when the choices sets include many alternatives (DeShazo and Fermo, 2002), or when individuals are asked to answer many choice tasks (DeShazo and Fermo, 2002). Although our induced value DCE tasks may seem to involve basic mathematics (addition and subtraction) not all subjects may be able to complete this task efficiently. To account for this issue in our benchmark treatments, subjects are provided a computerized calculator to help them make the calculations. Treatment 2 (*hypothetical + calc*) is identical to *hypothetical*, but with a button added to each choice set screen, by clicking on this button subjects can access the Microsoft windows^{TM} calculator. Subjects' calculator use is recorded throughout the experiment.

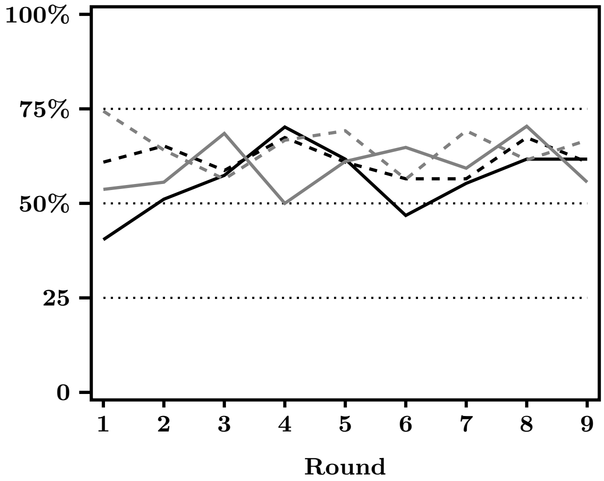
In Treatments 1 and 2, choices are hypothetical and do not affect how much subjects earn in the experiment. Critics of stated preference methods, and survey methods in general, question peoples' motivation to choose the best for them when answers are hypothetical (see, e.g., Olof and Henrik, 2008). Subjects' 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). In Treatment 3 (*paid*), we replicate Treatment 1, except that subjects are paid based on the choices that they make in the experiment.

To that end, each subject receives a £2 show-up fee plus an extra £4 in an account, which changes depending on earnings. We designed this payment such that the expected earnings is about £12, which is the same in expectation as the flat payment in Treatments 1 and 2. To avoid a binding budget constraint, all choice tokens cost less than £4. Subjects face the identical nine choices as in Treatment 1. At the end of Treatment 3, for each subject, the monitor selects 1 of the 9 choice tasks to be the binding choice: the balance of the account is updated with the cost of the token and the values of the attributes. This random selection of one binding choice follows standard protocol to avoid income effects (see for example Shogren, Shin, Hayes, and Kliebenstein, 1994). The experimental instructions are identical to Treatment 1, with one exception—there is no subjective language. The data for this treatment is the "wide-monetary" treatment in Luchini and Watson (2014).

Finally, in Treatment 4 (*paid+calculator*), monetary incentives are combined with a calculator. Our hypothesis is that, when choices affect earnings, cognitive reasoning could be fostered and this may encourage subjects to use the calculator more. This would, in turn, lead to a higher proportion of correct choices. In economic terms, monetary incentives and the calculator could

Figure 1: Proportion of correct choices by treatment and round

Treatment		n	
1: Hypothetical	—	47	56.3%
2: Hyp. + calculator	- - -	47	61.6%
3: Real	—	54	59.9%
4: Real + calculator	- - -	47	64.9%



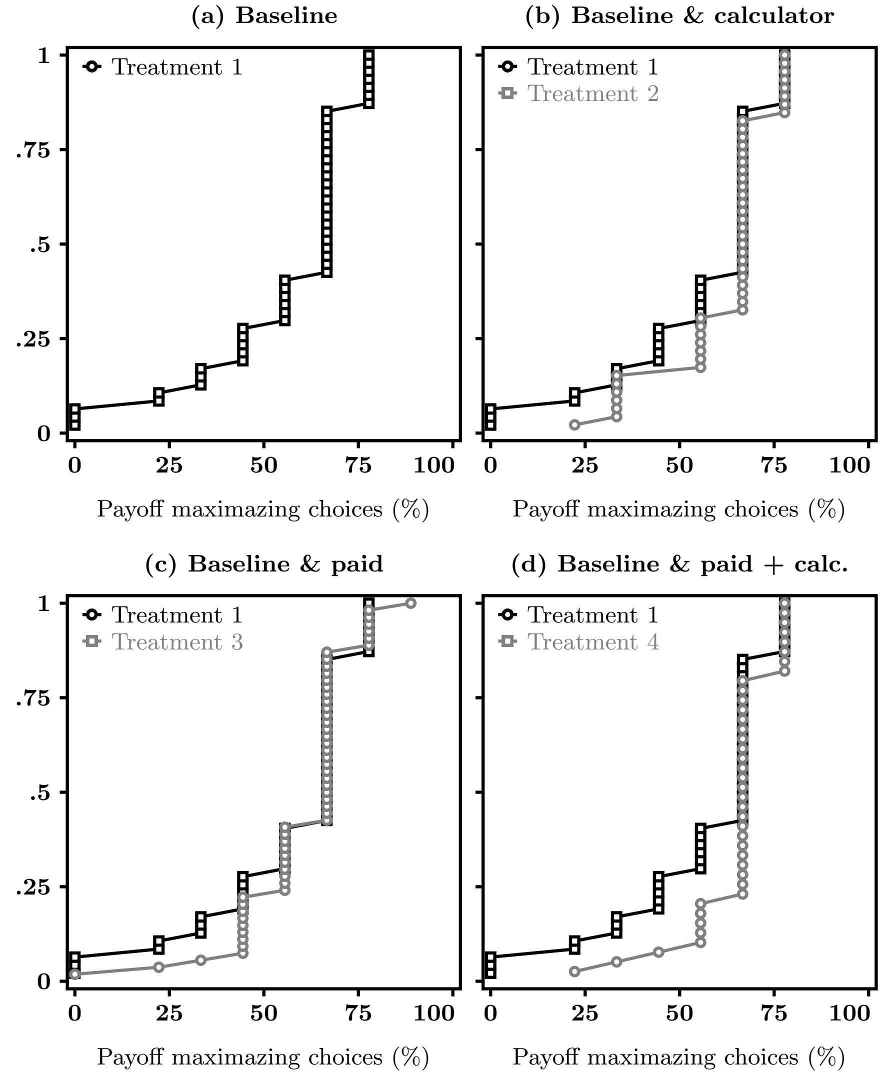
act as complements to improve decision making.

4 How well do subjects perform the DCE task in benchmark treatments?

Overall, our benchmark results reveal that people use non-optimizing choices in nearly 35-45% of all decisions, regardless of whether the DCE decision is hypothetical or paid, with or without a calculator. Given that the induced valuation task was relatively straightforward, this non-profit maximizing choices are substantial and consistent across these standard benchmark treatments. Neither raising the stakes with paid monetary incentives nor reducing complexity with a calculator improved choices. This result matters because it sets the comparative stage to determine whether the truth-telling oath will work in the context of DCE.

We now consider the set of benchmark results in detail. Figure 1 presents the proportion of correct choices (i.e., tokens that are payoff-maximizing among each pair) by treatment and round. In **Treatment 1** (*baseline*), there are two main results at the aggregate level. First, only a little over half of the choices are payoff-maximizing (56.3%). Second, there is no significant difference across rounds except for the first round for which the proportion of payoff-maximizing choices is a little lower than 50% (42.1%). Median response time for one choice decision is 16 seconds and median total response time for all 9 choices is 157s in this *hypothetical* treatment. In Figure 2, we compute, for each subject, the percentage of correct choices 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) correct choices. The highest percentage of correct choices

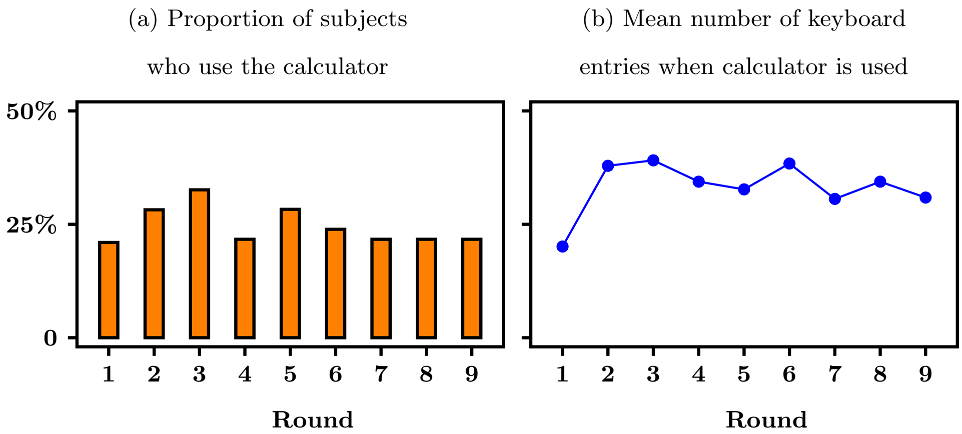
Figure 2: Empirical distribution function (EDF) of proportion of payoff-maximizing choices by subject



observed in the *hypothetical* treatment is 77.7%, which corresponds to 7 choices out of 9. Most of the subjects (44.7%) made only 6 correct choices in Treatment 1. In many choice sets, subjects do not choose what is profit-maximizing even in this relatively straightforward DCE design.

The proportion of correct choices is relatively low. One potential explanation might be that subjects make mistakes when making choices. Providing subjects with the windows calculator should help them making better decisions (**Treatment 2**, *calculator*). We record if the calculator is used in a choice task, and how many keyboard entries are made 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`

Figure 3: Use of the calculator across rounds in Treatment 2



`+ 1.5 + 6 =`" and this would be counted as 9 keyboard entries. Figure 3 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. 50% 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 3.a shows that the activation of the calculator is 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% and 21.7% in round 1). We do not observe a round effect in number of keyboard entries (Figure 3.b), when the calculator was activated (except in round 1).

As shown in Figure 1, providing a calculator has a small but statistically insignificant effect on the percentage of correct choices compared to Treatment 1 (61.6% vs. 56.3%). Comparing the two proportions based on a two-sided bootstrap test of proportions that allows for within subject correlation,[8] the p-value is $p = .298$. At the subject level, we observe no improvement in the percentage of correct choices a subject makes. Figure 2.b presents the EDF of the percentage of correct choices by subjects in Treatments 1 and 2. The EDF in Treatment 2 is slightly to the right of the EDF in Treatment 1 but first order dominance is not significant ($p = .192$).[9] Furthermore, we examine if subjects who use the calculator make better choices. At the choice level, the pairwise correlation between the activation of the calculator and the choice being payoff-maximizing is minimal (equal to .048). The correlation is a little larger for the the number of entries but still small (.097). At the subject level, pairwise correlation between the number of

[8]Our bootstrap procedure consists of bootstrapping on subjects rather than choices. This allows us to control for within-subject correlation.

[9]We use a bootstrap version of the Kolmogorov-Smirnov (KS) 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, 2011).

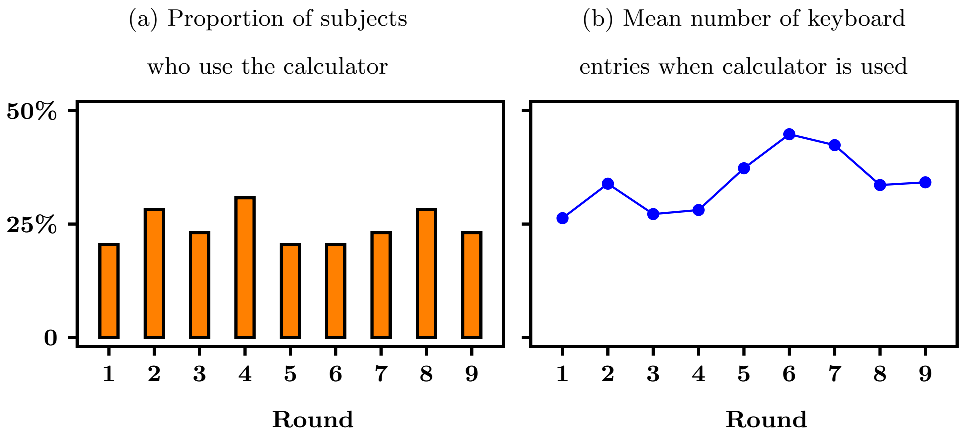
times the calculator is used by subject and the total number of payoff-maximizing choices made is positive (.196) but the coefficient is not significantly different from zero ($p = .193$). The same is true for the correlation between the total number of entries and the number of correct choices by subject (with a correlation coefficient equal to .207, $p = .167$). When choices are hypothetical, providing help to improve cognitive reasoning does not improve choices in this induced values setting.

Treatment 3 (*paid*), in which subjects are paid based on their decisions, aims to test whether such behavior is due to the lack of monetary incentives on elicited choices. Figure 1 presents the percentage of correct choices across rounds. Overall, we observe that 59.9% of choices are correct when money is at stake. This is not statistically different from that of Treatment 1 (*hypothetical*, $p = .607$). Again, the figure shows that no round effects are present in the data: the proportion of payoff-maximizing choices is stable across rounds, at a level close to that in Treatments 1 and 2 (*hypothetical* and *hypothetical + calc*). Figure 2.c, plots the EDF of the percentage of correct choices by subjects in Treatments 1 and 3. The comparison confirms aggregate findings at the subject level: the EDF are nearly identical and although the EDF in Treatment 3 is slightly to the right of the EDF in Treatment 1, first order dominance is not significant ($p = .480$). Although we do not find any difference in choices, subjects do react to monetary incentives. They pay more attention to their decisions: subjects take more time to make their choices in Treatment 3. The median total response time (the time taken by a subject to answer all nine choice sets) is 197 seconds in Treatment 3 compared to 157 seconds in Treatment 1, the increase is significant with $p = .050$ (based on a median difference bootstrap test). The EDF of total response time shows that being paid reduces the number of subjects with rapid response times. A KS bootstrap distribution test indicates that the EDF of response time in Treatment 3 first order dominates the EDF of response time in Treatment 1 ($p < .025$).

Treatment 4 (*paid+calculator*) combines the two devices studied separately in Treatment 2 and 3: subjects are paid based on their choices and can use a calculator to help them make choices. Figure 4.b reports the use of the calculator in Treatment 4 across rounds. Being paid based on choices does not increase calculator use as compared to Treatment 2. The calculator was activated in 24.2% of the choice tasks in Treatment 4 compared to 24.6% of the choices tasks in Treatment 2. Again, nearly 50% of subjects (48.7%) never use the calculator and 12.8% use it in every choice set (13% in Treatment 2). Among those subjects who actually use the calculator, we do not find differences in the number of keyboard entries between Treatment 4 and Treatment 2. Overall, the mean number of keyboard entries is 33.9 in Treatment 4 and 33.6 in Treatment 2.

The overall percentage of payoff-maximizing choices in Treatment 4, presented in Figure 1, shows that being paid in combination with the calculator leads to a small, but significant, increase in the percentage of correct choices: 64.9% compared to 56.3% ($p = .037$). Again, no round effect seem to be present in the data. Figure 2.d presents the EDF of the percentage of correct choices

Figure 4: Use of the calculator across rounds in Treatment 4



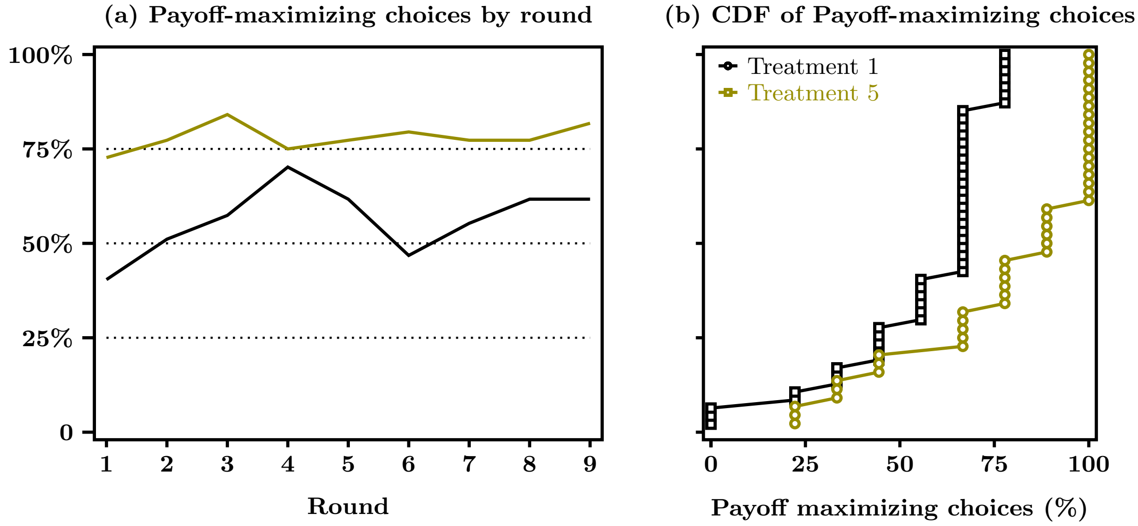
by individual in Treatments 1 and 4. The EDF in Treatment 4 is shifted to the right: there are fewer subjects with a low number of payoff-maximizing choices. The upper part of the EDF remains similar to that of Treatment 1. The EDF in Treatment 4 first order dominates the EDF in Treatment 4 ($p = 0.015$). The pairwise correlation between payoff-maximizing decision and both the activation of the calculator (.165, $p = .318$) and the number of entries in the calculator (.125, $p = .448$) are however small. These results suggest that the effect of the calculator and of monetary incentives add up—they both improve the proportion of payoff-maximizing; although a large proportion of dominated choices remain.

Result 1 *DCE choices with induced values are not profit maximizing for about 35-44% of the choices, regardless of whether these choices are hypothetical or real, with or without a calculator. The standard DCE method of hypothetical choices do as well as the typically proposed solutions, e.g., real monetary payments or tools to reduce computation costs (in our case, a calculator).*

5 Does the honesty Oath improve DCE choices?

The results obtained in the benchmark treatments are striking. The proportion of payoff-maximizing decisions is low. Monetary incentives combined with the help of a calculator increase the proportion of payoff-maximizing decisions, but only by a small amount. In **Treatment 5** (*truth-telling oath*), we implement a non-monetary commitment device—a truth-telling oath—before Treatment 1. The truth-telling oath procedure follows that of Jacquemet, Joule, Luchini, and Shogren (2013). A monitor presents each subject with the oath at a private desk upon entry into the lab, after completing the consent form. The form is entitled "Solemn oath" and contains an unique

Figure 5: Empirical distribution function (EDF) of proportion of payoff-maximizing choices by subject



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**". Subjects are told that signing the form is voluntary and that neither their participation in the experiment nor their earnings depend on signing. Honesty relies only on the goodwill of our subjects. All subjects except one signed the oath.[10]

Our key result now emerges. We find that the truth-telling oath reduces non-payoff maximizing choices by nearly 50% relative to the standard benchmarks. The explicit and voluntary commitment to honesty improved decisions. The results, presented in Figures 5.a and 5.b, are unambiguous. The oath significantly increases the percentage of correct choices compared to Treatment 1: 78.3% compared to 56.3% ($p < .001$). This increase is observed for all decision rounds: the truth-telling oath induces an upward shift in each and every round, above the 75% mark. Figure 5.b presents the percentage of correct choices a subject makes. Under a truth-telling oath, 40.9% of subjects make payoff-maximizing choices in all 9 choice sets (whereas only one subject did so over all benchmark treatments) and 54.5% make at most one dominated choice. The EDF of correct choices in Treatment 5 significantly first order dominates the EDF in Treatment 1 ($p < .001$).

Result 2 *The honesty oath significantly improves standard DCE choices—incorrect choices were cut in half, they fall to around 20%.*

[10]Our statistical analysis includes the choices of the subject who did not sign, i.e., we adopt an intention to treat strategy to avoid selection effects.

Response time in Treatment 5 is significantly longer than in Treatment 1. The response time increase is comparable to the increase observed in Treatment 3 (*paid*). The median total response time is 208 seconds in Treatment 5 compared to 157 seconds in Treatment 1 ($p = .008$) and 197 seconds in Treatment 3 ($p = .400$). The EDF of response time in Treatment 5 first order dominates that of Treatment 1 ($p = .004$). The response time in Treatment 5 shows that the truth-telling oath increases the time spent by subjects to take their decision, to an extent that is similar to the response time observed in the *paid* treatment. Combined with the increase of payoff maximizing decisions, two explanations are consistent with these results. The first explanation is that the oath fosters truth-telling, which requires more time from subjects to identify the payoff-maxmizing option. An alternative explanation is that engaging subjects by an external non-monetary commitment device is more efficient than money (Treatment 3) at fostering one's cognitive effort to perform the task accurately. This is the hypothesis we test in the next section.

6 Does the truth-telling oath just foster cognitive reasoning?

We now explore whether it is the explicit commitment to honesty induced by the truth-telling oath that improves choices or if any alternative oath mechanism works just as well. We explore two alternative oath frames—an oath on task and an oath on duty—to determine if the basic idea of taking an oath causes people to use fewer non-optimizing decision rules.

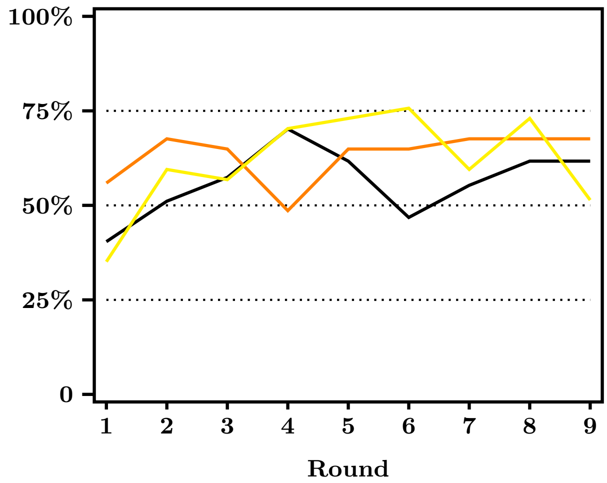
We carry out two alternative oaths treatments. In the first (**Treatment 6**, *oath on task*), we implement a modified oath that directly targets cognitive effort—a "task-oath". Treatment 6 replicates Treatment 5, but with a modified oath form that explicitly targets cognitive effort without referring 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 Treatment 5. All subjects accepted to take the *oath on task*.

The oath on task only targets cognitive effort while the truth-telling oath in Treatment 5 also had a moral connotation. Providing ethical standards to people has been shown to have significant effect on behavior (Mazar, Amir, and Ariely, 2008). In **Treatment 7** (*oath on duty*), we implement a second modified oath that targets cognitive effort, but with a moral component. To that end, we adapt a real world oath, in this case one that targets effort to perform one's assigned task with the moral reminders that one would encounter in the field if taking an oath before beginning the duties of a public office: the *oath of office*. The oath form 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**".[11] All subjects

[11] All oath procedures in Treatments 5, 6 and 7 were carried out by the same person for all subjects—who also

Figure 6: Proportion of correct choices by treatment and round

Treatment		n	
1: Hypothetical		47	56.3%
6: Oath on task		37	63.7%
7: Oath on duty		37	61.6%



except one accepted to take the *oath on duty*.[12]

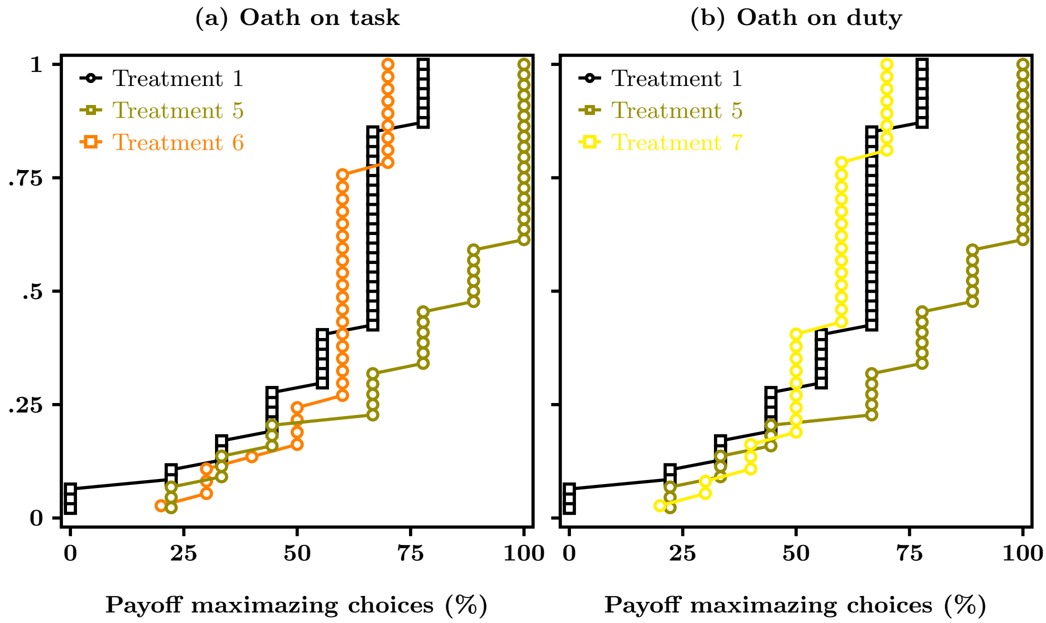
The results from the three oath treatments, along with the benchmark situation (Treatment 1), are presented in Figure 6. Comparing **Treatment 6** to Treatment 1, the task-oath has a small positive effect on choices: 63.7% (as compared to 56.26%, $p = .074$) of observed choices are payoff-maximizing. This increase is, however, significantly lower than the improvement observed with the *truth-telling oath*, which always dominates that observed with the *oath on task*. This result is confirmed at the subject level from the comparison of the EDF in Treatments 1 and 6 (Figure 7.a): the EDF in Treatment 6 does not first order dominates the one observed in Treatment 1 ($p = .536$).

Observed response times however confirm that subjects do take the task-oath seriously. There is a significant increase in response time compared to Treatment 1: the median total response time is 237 seconds in Treatment 6 compared to 157 seconds in Treatment 1 ($p = .012$). The EDF of response time in Treatment 6 first order dominates that of Treatment 1 ($p < .001$). The response time increase is similar to that observed when a truth-telling oath is implemented (Treatment 5), with a median total response time of 208 seconds (this difference in median response time is not statistically significant, $p = .176$). This change in response time shows that the task-oath fosters cognitive reasoning; it has, however, little to no impact on choices as compared to our benchmark treatments. This suggests that it is not by fostering cognitive reasoning that the truth-telling oath improves decision making. The task-oath might however appear too abstract and singular, whereas the truth-telling oath is a real world institution with a moral content. From this point of view, the task-oath would not succeed in increasing the proportion of payoff-maximizing choices

ran Treatments 1 to 4.

[12] As in Treatment 5, statistical analysis is carried out without dropping observations.

Figure 7: Empirical distribution function (EDF) of the % of payoff-maximizing choices by subject in Treatments 6 and 7



because it might be too far from what one would encounter in the field.

The duty-oath, implemented in **Treatment 7** aims to address this issue. Figure 6 shows that 61.6% of choices are payoff-maximizing in this treatment, which is not statistically different from Treatment 1 ($p = .185$). At the subject level, the comparison of the EDF of the percentage of payoff-maximizing choices in Treatments 1 and 7, presented in Figure 7.b, confirms that the duty-oath does not improve choices. Response time data in the duty-oath, compared to both the task-oath and the truth-telling oath, confirm that subjects take the duty-oath seriously. The median total response time in Treatment 7 is 213s—significantly greater than in Treatment 1 (157s) and similar to both Treatment 5 and 6. The EDF of total response time for each subject shows that the EDF in Treatment 7 first order dominates the EDF in Treatment 1 (the p-value from a KS bootstrap test is $p = .009$) and is similar to the ones observed in Treatment 6. Last, similar to what has been observed with the truth-telling oath, the response times associated with payoff-maximizing choices are longer (25s) than for non-payoff-maximizing choices (21s).

In sum, the two alternative oaths directly targeting cognitive effort do not translate into higher proportions of payoff-maximizing choices, in sharp contrast with the truth-telling oath. The analysis of response time suggests that subjects do engage into more cognitive reasoning, to a level comparable to those observed in Treatments 3 (*paid*) and 5 (*truth-telling oath*). This suggests that people who tell the truth when facing a truth-telling oath, deliberately choose the wrong token, even after careful thinking, when oaths fostering cognitive effort are implemented instead.

Result 3 *The notion of an oath in-and-of-itself is not enough. Neither the Oath on Duty nor the Oath on Task significantly improve standard DCE choices.*

7 Conclusion

Can a truth-telling oath increase the reliability of the Discrete Choice Experiment (DCE), a popular elicitation mechanism used to reveal preferences? Using an induced value experimental design, our results suggest the answer is yes. Relative to the standard benchmark treatments (hypothetical or paid choices, with or without a calculator), we find that the truth-telling oath reduces non-payoff maximizing choices by nearly 50%. Similar to our previous work on second-price auctions and referenda under oath, we find that an explicit and voluntary commitment to honesty can improve preference elicitation. Further, we show that is not the oath itself that matters but rather the commitment to the truth that matters. Using two alternative oath frame, an oath on task and an oath on duty, we still observe people using non-optimizing decision rules significantly more often than under the truth-telling oath.

Given that DCEs are a popular tool to elicit preferences for multi-attribute goods like health care, improved environmental quality, and new consumer products, our results suggest that the truth-telling oath should play a more significant role outside the lab. Strategic behavior reduces the reliability of preference elicitation methods, which has implications for the popular binary DCE and beyond, i.e., managing incomplete contracts. One path to address such behavior requires that we identify motives and target specific incentives or nudges, which can imply that we have to create a rather complex mechanism design to realign incentives. For example, creating a notion of consequentially in survey respondents requires a well-focused and refined version of cheap talk as the means to make a respondent believe their response matters (e.g., short versus long scripts, choice of wording). But our results suggest another more straightforward path—commitment to the truth through the honesty oath, which reduces the tendency to be strategic. People seem more prone to do what is right, which in our case was also more profitable. This suggests we should invest more effort to understand the simplicity of the oath to create reliable commitment to the truth relative to trying to disentangle the complexity of common information and the many motives, actions, beliefs, counter-actions, and so on that drive strategic behavior. With the oath, we do not have to try to outguess each respondent's strategic gaming: we do not have to know how they might misrepresent their preferences strategically from an infinitely large set so that we can design a behavioral antidote –a potentially effective cheap talk. We can take a behavioral shortcut and we ask them to swear to a solemn oath beforehand.

Truth-telling should be explicitly addressed and not implicitly assumed in field applications. Researchers could ask respondents to take a truth-telling oath prior to being interviewed (see, e.g., Jacquemet, James, Luchini, and Shogren, 2017). Alternatively, future research might consider whether weaker forms of commitment would be successful such as a preliminary pledge or even a signed agreement to tell the truth. One might challenge our recommendation by arguing our induced value experimental design was too abstract. Basic math problems do not reflect real world goods and services. While we appreciate this reasoning, we defend our approach based on two comments. First, this reasoning supposes that real choices exist and are observable. But in many situations, such as health care or environmental preservation, the DCE must be carried out before any concrete policy option exists, and no markets exist in which one could observe choices and competitive prices. Second, and most importantly, this reasoning presumes that real choices in the field fulfill the necessary rational requirements that underpin meaningful welfare estimates. As we have seen, in our basic task, monetary incentives alone do not guarantee that people choose their best option. Context matters. We believe our design is a necessary starting point to understand better whether preferences revealed in field DCE surveys are accurate enough for collective decision making.

References

Abadie, A. (2002): "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Model," *Journal of the American Statistical Association*, 97(457), 284–292.

Adamowicz, W., K. Glenk, and J. Meyerhoff (2014): "Choice modelling research in environmental and resource economics," in *Handbook of Choice Modelling*, ed. by A. D. S. Hess. Edward Elgar Publishing, Cheltenham, UK.

Adamowicz, W., and J. Swait (2011): "Discrete Choice Theory and Modeling," in *Oxford Handbook of the Economics of Food Consumption and Policy*, ed. by J. Lusk, J. Roosen, and J. Shogren. Oxford University Press, NY.

Beauvois, J.-L., and R.-V. Joule (2002): *Petit traité de manipulation à l'usage des honnêtes gens*. Presse Universitaires de Grenoble.

Camerer, C. F., and R. M. Hogarth (1999): "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor-Production Framework," *Journal of Risk and Uncertainty*, 19(1), 7–42.

Carson, R., T. Groves, and J. List (2014): "Consequentiality: a theoretical and experimental exploration of a single binary choice," *Journal of the Association of Environmental and Resource Economists*, 1, 171–207.

Carson, R. T., and T. Groves (2007): "Incentive and informational properties of preference questions," *Environmental and Resource Economics*, 37(1), 181–210.

Collins, P., and A. Vossler (2009): "Incentive compatibility tests of choice experiment value elicitation questions," *Journal of Environmental Economics and Management*, 58(2), 226–235.

de Magistris, T., and S. Pascucci (2014): "Does "solemn oath" mitigate the hypothetical bias in choice experiment? A pilot Study," *Economics Letters*, 123(2), 252–255.

DeShazo, J., and G. Fermo (2002): "Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency," *Journal of Environmental Economics and Management*, 44(1), 123–143.

Evans, A. M., K. D. Dillon, and D. G. Rand (2015): "Fast but Not Intuitive, Slow but Not Reflective: Decision Conflict Drives Reaction Times in Social Dilemmas," *Journal of Experimental Psychology: General*, 144(5), 951–966.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10(2), 171–178.

Freeman III, A. M., J. Herriges, and C. Kling (2014): *The Measurement of Environmental and Resource Values*. RFF Press and Routledge, Washington, D.C. and NY.

Greiner, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1(1), 114–125.

Hensher, D., J. Rose, and W. Green (2015): *Applied choice analysis*. Cambridge University Press, Cambridge (UK).

Interis, M., C. Xu, D. Petrolia, and K. Coatney (2016): "Examining unconditional preference revelation in choice experiments: a voting game approach," *Journal of Environmental Economics and Policy*, 5(1), 125–142.

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(3), 411–433.

Jacquemet, N., A. James, S. Luchini, and J. F. Shogren (2017): "Referenda Under Oath," *Environmental and Resource Economics*, 67(3), 479–504.

Jacquemet, N., R.-V. Joule, S. Luchini, and J. F. Shogren (2013): "Preference Elicitation under Oath," *Journal of Environmental Economics and Management*, 65(1), 110–132.

Jacquemet, N., S. Luchini, J. Rosaz, and J. Shogren (2018): "Truth-Telling under Oath," *Management Science, In press.*

Jacquemet, N., S. Luchini, J. Shogren, and A. Zylbersztejn (2018): "Coordination with Communication under Oath," *Experimental Economics*, 21(3), 627–649.

Krajbich, I., B. Bartling, T. Hare, and E. Fehr (2015): "Rethinking fast and slow based on a critique of reaction-time reverse inference," *Nature Communications*, 6.

Luchini, S., and V. Watson (2014): “Are choice experiments reliable? Evidence from the lab,” *Economics Letters*, 124(1), 9–13.

Manski, C., and S. Lerman (1977): “The estimation of choice probabilities from choice based samples,” *Econometrica*, 45(8), 1977–88.

Mas, A., and A. Pallais (2017): “Valuing Alternative Work Arrangements,” *American Economic Review*, 107(12), 3722–59.

Mazar, N., O. Amir, and D. Ariely (2008): “The dishonesty of honest people: A theory of self-concept maintenance,” *Journal of Marketing Research*, 45(6), 633–644.

Mazzotta, M. J., and J. J. Opaluch (1995): “Decision Making When Choices Are Complex: A Test of Heiner’s Hypothesis,” *Land Economics*, 71(4), 500–515.

McFadden, D. (2001): “Economic Choices,” *American Economic Review*, 91(3), 351–378.

Meginnis, K., M. Burton, R. Chan, and D. Rigby (2018): “Strategic bias in discrete choice experiments,” *Journal of Environmental Economics and Management*, Forthcoming.

Olof, J.-S., and S. Henrik (2008): “Measuring Hypothetical Bias in Choice Experiments: The Importance of Cognitive Consistency,” *The B.E. Journal of Economic Analysis & Policy*, 8(1).

Sekhon, J. (2011): “Multivariate and Propensity Score Matching Software with Automated Balance Optimization,” *Journal of Statistical Software*, 42(7), 1–52.

Shogren, J. F., S. Y. Shin, D. J. Hayes, and J. B. Kliebenstein (1994): “Resolving Differences in Willingness to Pay and Willingness to Accept,” *American Economic Review*, 84(1), 255–270.

Soekhai, V., E. de Bekker-Grob, A. Ellis, and C. Vass (2018): “Discrete Choice Experiments in Health Economics: Past, Present and Future,” *Pharmacoeconomics*, Forthcoming.

Swait, J., and W. Adamowicz (2001): “Choice Environment, Market Complexity, and Consumer Behavior: A Theoretical and Empirical Approach for Incorporating Decision Complexity into Models of Consumer Choice,” *Organizational Behavior and Human Decision Processes*, 86(2), 141–167.

Taylor, L., M. Morrison, and K. Boyle (2010): “Exchange rules and the incentive compatibility of choice experiments,” *Environmental and Resource Economics*, 47(2), 197–220.

Taylor, L. O., M. McKee, S. K. Laury, and R. G. Cummings (2001): “Induced-value tests of the referendum voting mechanism,” *Economics Letters*, 71(1), 61–65.

Train, K. (2003): *Discrete choice methods with simulation.* Cambridge University Press, Cambridge (U.S.).

Vossler, C., M. Doyon, and D. Rondeau (2012): “Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments,” *American Economic Journal-Microeconomics*, 4(4), 145–171.

White, C. N., R. A. Curl, and J. F. Sloane (2016): "Using decision models to enhance investigations of individual differences in cognitive neuroscience," *Frontiers in Psychology*, 7.