HAL
open science

Truth Telling Under Oath

Nicolas Jacquemet, Stephane Luchini, Julie Rosaz, Jason F Shogren

To cite this version:

Nicolas Jacquemet, Stephane Luchini, Julie Rosaz, Jason F Shogren. Truth Telling Under Oath. Management Science, 2019, 65 (1), pp.426-438. 10.1287/mnsc.2017.2892 . halshs-02018089

HAL Id: halshs-02018089

https://shs.hal.science/halshs-02018089

Submitted on 13 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Truth-telling under Oath*

Nicolas Jacquemet† Stéphane Luchini‡ Julie Rosaz § Jason F. Shogren¶

January 2019

Abstract

Oath-taking for senior executives has been promoted as a mean to enhance honesty within and towards organizations. Herein we explore whether people who voluntarily sign a solemn truth-telling oath are more committed to sincere behavior when offered the chance to lie. We design an experiment to test how the oath affects truth-telling in two contexts: a neutral context replicating the typical experiment in the literature, and a "loaded" context in which we remind subjects that "a lie is a lie." We consider four payoff configurations, with differential monetary incentives to lie, implemented as within-subjects treatment variables. The results are reinforced by robustness investigations in which each subject made only one lying decision. Our results show that the oath reduces lying, especially in the loaded environment—falsehoods are reduced by fifty percent. The oath, however, have a weaker effect on lying in the neutral environment. The oath did affect decision times in all instances: the average person takes significantly more time deciding whether to lie under oath.

Keywords: Deception; Lies; Truth-telling oath; Laboratory Experiment.

JEL Classification: C92, D03, D63.

*Published in Management Science vol. 65 (1), pp. 426-438. This paper is a revised version of CES Working Paper n°2015-68. We thank the reviewers, the editor, Theodore Alexopoulos, Marie-Pierre Fayant Robert-Vincent Joulé, Rosemarie Nagel and Marie-Claire Villeval for their helpful comments. We thank Justine Jouxtel for her efficient research assistance. Jacquemet gratefully acknowledges funding from the *Institut Universitaire de France* and ANR-10-LABX-93-01, part of this work has been completed while he was affiliated at Université de Lorraine (BETA). Shogren thanks the University of Alaska-Anchorage for the support and hospitality while working on this paper. Jacquemet and Luchini thank the Health Chair—a joint initiative by PSL, Université Paris Dauphine, ENSAE and MGEN under the aegis of the Fondation du Risque (FDR) for its support. Financial support from the Provence-Alpes-Côte-d'Azur Regional Council is gratefully acknowledged.

†Paris School of Economics and University Paris I Panthéon–Sorbonne. Centre d'Economie de la Sorbonne, 106 Bd. de l'Hôpital, 75013 Paris. Nicolas.Jacquemet@univ-paris1.fr

‡GREQAM, Aix-Marseille Univ. (Aix-Marseille School of Economics), CNRS, EHESS, Centrale Marseille, 5-9 Boulevard Bourdet, CS 50498 13205 Marseille Cedex 1, France. stephane.luchini@univ-amu.fr

§Université Montpellier 1, UMR5474 LAMETA, F-34000 Montpellier, France. rosaz@lameta.univ-montp1.fr

¶Department of Economics, University of Wyoming, Laramie, WY 82071-3985, United States. JRamses@uwyo.edu

"*And undoubtedly the evil in which oaths take their rise is the prevalence of falsehoods and wrong, and the consequent prevalence of suspicion and mistrust*" (Tyler, 1834, p.6).

1 Introduction

Sometimes people lie. Societies have responded to these falsehoods with ex post punishments and ex ante institutions designed to commit people to truth-telling. In Antiquity, a merchant who claimed to have been robbed during his journey was asked to take an oath to avoid retaliation from the persons who were supposed to receive those goods (Silver, 1995). Today's business and managerial world is no different. After the 2002 Enron and Worldcom scandals and the 2008 financial crisis, many in the management community grew concerned that business leaders had lost their legitimacy with the general public (see, for example, Khurana and Nohria, 2008; Podolny, 2009). These authorities proposed that private and public sector executives take an oath of honesty that commits them to have a positive role in society, similar in spirit to the Hippocratic oath used in medicine. The purpose of oath-taking as a management instrument is to set and commit to standards of behavior (honesty and integrity) to recapture trust with the general public. For instance, MBA students at Harvard were concerned enough about mistrust and the lack of self-policing that they created the MBA Oath (see Bishop and Green, 2011), which lead to the Oath Project created at the World Economic Forum in Davos.[1]

Despite this growing use of oaths in management practices, little is still known about whether such procedures would make any difference by actually changing behavior. Herein we explore this question: are people who voluntarily sign a solemn truth-telling oath more committed to sincere behavior when offered the chance to lie? Based on the existing literature on the non-monetary incentives to tell the truth, an oath can be expected to change behavior through two possible channels: a tighter link between words and actions, and a reminder of the moral issue of lying.[2] Our aim is to separate and identify these two channels. To that end, we use the social psychology theory of commitment to design a solemn oath of honesty that commit people to truth-telling. We examine how this oath affects truth-telling in two different contexts: a neutral context replicating the typical experiment in the literature, and a context in which subjects are reminded that a lie

[1] See Anderson and Escher (2010) who discuss the path leading to the creation of the MBA Oath. Beyond MBA programs, DeMartino (2011) for instance proposed an Economist's Oath. Also see for example Cohn, Fehr, and Maréchal (2014) for a longer discussion of dishonesty within the banking industry.

[2] In economics, the link between words and actions is the topic of studies on strategic communication, which are associated with two strands of literature (Kartik, 2009): games of persuasion, in which the information is verifiable (Milgrom, 1981; Grossman, 1981) and lying is infinitely costly, and "cheap talk" games in which information is not verifiable and can be manipulated at no cost (see, e.g., Crawford and Sobel, 1982). The recent literature in behavioral economics on the "intrinsic costs of lying" bridges the two strands of literature by putting forward the trade-off between the individual cost of lying and its material consequences (see, e.g, Boles, Croson, and Murnighan, 2000; Sutter, 2009).

is a lie. This treatment introduces moral reminders of ethical standards by explicitly labeling untruthful communication as a "lie" and truthful communication as "truth".

Within the framework of commitment theory, taking an oath works similar to foot-in-the-door experiments: it is a preliminary task that makes people more likely to behave truthfully in subsequent decision problems. In social psychology, commitment is defined as a "binding of the individual to behavioral acts". Commitment is obtained through preparatory actions, which are purposefully designed to induce a predictable change in subsequent decisions. In a classic field experiment, Harris (1972) asked people for the time before asking them for a dime. People were 4 times more likely to give him a dime if they were first asked to give away the "time" for free (44% time-dime vs 11% no time-dime). Harris found that asking for the time was a costless action that committed the other person to go for a similar and costly second action. People who gave the time away for free were more likely to give away a dime "for free". People complied freely to what was expected of them. These procedures have a common behavioral mechanism: compliance with a demanding request (e.g., tell the truth) is improved significantly by using a process that first gets a person to commit to being the type of person that does a certain action (e.g., signing an oath), as long as the course of action remains consistent and compliance to the preparatory action is free. Under these conditions, commitment induces strong and lasting changes in behavior (as shown in field experiments by, e.g., Freedman and Fraser, 1966; Joule, Girandola, and Bernard, 2007). The truth-telling oath as a commitment device has proved to be useful in triggering economic commitment in preference elicitation. The oath increases truth telling in hypothetical settings with no monetary incentives (Jacquemet, Joule, Luchini, and Shogren, 2013; Jacquemet, James, Luchini, and Shogren, 2016) and in coordination games with cheap talk communication (Jacquemet, Luchini, Shogren, and Zylbersztejn, 2017; Kataria and Winter, 2013).

We study the two non-monetary devices in a 2*2 between-subjects experimental design that combines the oath and a loaded environment. We use the white lies experimental game from Erat and Gneezy (2012). This game provides incentivized observations of truth-telling through the choice of a signal sent to a receiver, which can be either consistent or inconsistent with the sender's private information. We consider several payoff configurations among the set introduced by Erat and Gneezy (2012), ranging from a purely selfish, self-beneficial lie occuring at the expense of others to situations in which the organization itself (i.e., both players in the game) benefits from the lie. While selfish lies are the primary target of the use of oaths in organizations, restored trust from the general public can only be achieved if oaths are efficient in a broader range of lying situations. We implemented the payoff configurations as within-subjects treatment variables—each subject made four decisions on whether to lie or not. We complement this evidence with a set of between-subjects robustness experiments in which each subject had one decision on whether to lie. We also control for decision times in all conditions to assess the extent of cognitive reasoning in decision making (Kahneman, 2003; Rubinstein, 2007). In typical promises experiments, subjects are committed endogenously to tell the truth by individually deciding whether to make a promise

or not. With the oath-procedure, subjects are made to freely commit to tell the truth, that is, they are committed exogenously to truth-telling. As we explain in the experimental design section, the oath is freely taken and signed but is designed in such a way that virtually all subjects comply. In our case, we had 96.5% (222 of 230 subjects in our within and between treatments) sign the oath. This allows measurement of the causal effect of commitment on truth-telling motivation unlike in promises experiments where the causal relationship between promises and cooperation can be both ways. The oath also implements commitment to truth telling before the game takes place, rather than in-game promises. Last, it is a real-world and easy to implement procedure, which provides guidance into the design of institutional devices aimed to foster truth-telling.

Our results provide a clear answer to our initial question: Does a solemn oath enhance sincere behavior in truth-telling situations? The answer is 'yes', especially if the social context is such that people are explicitly aware that their insincere behavior is lying: they reduce their lying by half relative to those who were not asked to sign a truth-telling oath. The answer is more mixed if people are not explicitly told they would be considered as "liars". Without an explicit reference to the social context of the decision, the oath significantly impacts the frequency of lies only in the between-subjects implementations of the game. The oath does, however, affect decision times when a lie is involved even in the neutral environment. After signing the oath, the average person in the neutral environment takes more time to decide when they lie to the other player, which indicates that the oath makes the person ponder their choice to lie a little longer. Subjects under oath in the neutral environment seem to be aware that they were lying, but less reluctantly do so without moral reminders.

2 Empirical strategy

Our baseline condition replicates the sender-receiver game originally proposed by Erat and Gneezy (2012). We study commitment-by-oath as a between-subjects treatment variable. One issue with the commitment interpretation is that the oath could affect behavior because it contains "moral reminders" and provides cues about the social context in which subjects are taking decisions (see, for example, Haley and Fessler, 2005; Bateson, Nettle, and Roberts, 2006; Rigdon, Ishii, Watabe, and Kitayama, 2009). To disentangle these two dimensions (real truth-telling versus the explicit reference to truth-telling) we consider an additional treatment variable that makes explicit the opportunities to lie in the game: the decisions are "loaded" according to what they actually are—truth-telling or lies. We combine the two treatments variables according to a 2*2 factorial design, resulting in four between-subjects treatments.

2.1 Design of the baseline experiment

Two players, a sender and a receiver (in the instructions, the sender is player A and the receiver is player B), are randomly matched. The computer makes a random draw by "rolling" a 6-sided

Table 1: Design of the baseline experiment

Payoff configurations				Sequences of plays (2 sessions each)		
Label	Lie	Receiver guesses correct	Receiver guesses wrong	Order 1	Order 2	Order 3
T[-5; 10]	Altruistic	(20;20)	(15;30)	1	2	3
T[10; 10]	Pareto	(20;20)	(30;30)	2	3	1
T[1; -5]	Selfish	(20;20)	(21;15)	3	1	4
T[-10;-10]	Spite	(30;30)	(20;20)	4	4	2

Note. The four payoff configurations presented in row are implemented together in all experimental sessions in varying order, as described in the right-hand side column. Option X describes the payoff implemented if the receiver choice matches the die roll—($sender; receiver$); otherwise subjects earn the payoff described in Option Y.

die, and only the sender is informed about the result. The sender is then asked to send a message to the receiver, which has to be one of the six available messages: "*The outcome of the roll of die was x*", in which x is any integer between 1 and 6. The sender faces no constraint on whether he or she sends the true outcome to the receiver. One message out of six (the one that matches the actual roll) is an experimental measure of truth-telling, the other five are observations of lies.

Once the message has been received, the receiver chooses a number in the set $[1, 2, 3, 4, 5, 6]$, which determines the payment of both subjects. There are two payment options, X and Y. Only the sender knows the actual payoffs generated by each option. If the number chosen by the receiver matches the die roll, both subjects are paid based on option X; otherwise, Y is implemented. This is common knowledge to all subjects.

To assess the robustness and sensitivity of our own treatment variables to monetary incentives, we focus on a subset of 4 payoff configuration—see Table 1. Each corresponds to a particular type of lie, two are presented in Erat and Gneezy (2012) and two only appear in the working paper version, Erat and Gneezy (2009). The label of each payoff configuration refers to the gain/loss of each player (the sender for the first number, the receiver for the second) when the receiver chooses a number other than the actual outcome of the roll of the die, as compared to their payoff when the receiver chooses the actual roll. In T[-5;10], if the sender chooses to lie and the receiver chooses a number that follows the sender's message, the lie is called *altruistic* since the receiver gains 10 from the lie but the sender loses 5. Following the same logic, T[1;-5] characterizes a *selfish* lie, whereas T[10;10] is a *Pareto* lie since both the sender and the receiver benefit from the sender sending a message that is not the real outcome of the die. The fourth payoff configuration acts as the control configuration since both players lose from the sender's lie.

We introduce two main changes in the original design. First, instead of paying only one pair of subjects out of 10 pairs, we pay each and every subject based on actual decisions. Second, we implement all payoff configurations one after the other in each session, as opposed to the original between-subjects implementations. This within-subjects design (also implemented as a robustness

treatment in Erat and Gneezy, 2012, Section 4) allows us to put most of the identification power of the experiment on the treatment variables. Since all results are derived from comparisons with the baseline, we control for these changes in our analysis of the data. We use a perfect stranger design—subjects are rematched across payoff configuration—to rule out any incentive to invest in partner-specific reputation. Such a within-subjects design reduces the noise due to individual heterogeneity in the observed treatment effects, even though its well-known drawback is to confound order and payoff configuration effects. We control for this dimension by implementing three different orders, chosen in such a way that each payoff configuration appears first in each order (except for the reverse pay-off, which is only used as a control configuration): (1) altruistic / Pareto / selfish / reverse , (2) selfish / altruistic / Pareto / reverse and (3) Pareto / reverse / altruistic / selfish. Individual earnings from the experiment are computed based on one randomly drawn decision out of the four, so as to avoid changes in expected wealth across payoff configurations. This core experimental game is played under two treatment variables, implemented according to a factorial design.

Except for these two changes, the wording and framing of the baseline experimental instructions closely follow Erat and Gneezy (2012). We label this condition the *neutral* treatment in that the game is described without using words such as "lie, truth, or (dis)honesty", as follows:

> *Player A is first informed of the payoffs for two possible options: option X and option Y. These payoffs change from one period to another. Player A must then roll a six sided die, and is the only participant to know the outcome of this roll. Once informed of the die rolls' outcome,* ***player A must choose a message to send to player B****.*

2.2 Treatment variable 1: loaded environment

Recent experimental findings suggest that social cues are not only a substitute for learning/experience (Cooper and Kagel, 2003), but also affect subjects' reasoning process (Cooper and Kagel, 2009). Self-awareness theory (Silvia and Duval, 2001) and self-concept maintenance theory (Mazar, Amir, and Ariely, 2008) put forward the idea that social cues in the form of reminders about what moral behavior is in the situation at stake make it harder for subjects to maintain a positive self-image while engaging in dishonest behavior. When a person learns that dishonesty "pays", he or she then has to balance the desire to enjoy the monetary gain and the desire to preserve his or her positive self-view, i.e. not to risk being a "liar"—a self-concept maintenance (see Mazar, Amir, and Ariely, 2008). In such a context, some have argued that the intrinsic rewards from honesty (and the intrinsic penalty from dishonesty) are likely to be more significant because people are induced to think deeper about honesty and dishonesty in the course of their economic decisions. The context helps people think about moral standards in solving the trade-off between profiting from cheating and allowing oneself to believe he or she is a honest person. As stressed by Bénabou and Tirole (2011) this strengthens "*the repudiation of intended untruthful actions from the point of view of the virtual judgments of imagined spectators*"—fostering the feeling of guilt experienced by the decision maker. Existing evidence also suggests that recalling moral

standards just before a subject makes a decision can significantly reduce dishonest behavior (see, e.g., Ariely, 2012, for a discussion). To test the effect of moral reminders in the context of our game and to distinguish them from the truth-telling oath, our first treatment variable creates a "loaded environment" in which we provide moral reminders. This is done by explicitly labeling untruthful communication as a "lie" and truthful communication as "truth" in the instructions:

> *Player A has a choice between: (1) being honest and telling the truth about the die roll's outcome; or (2) lying and sending a message that is different from the real outcome. Please note that player B is only shown player A's message: If player A choose to lie (choice nº 2), player B has no way of knowing it.*

This change in the wording of the experimental instructions is the only difference between the loaded treatment and the baseline neutral treatment.

2.3 Treatment variable 2: commitment through a truth-telling oath

Our second and main treatment variable is a truth-telling oath. Our oath procedure, first designed by Jacquemet, Joule, Luchini, and Shogren (2013), is based on findings in social psychology. The social-psychological theory of commitment investigates several institutions that induce people to "comply without pressure" (see the review in Joule, Girandola, and Bernard, 2007; Cialdini and Sagarin, 2005). Social psychologists have observed that commitment is stronger if it has been made freely, and/or publicly expressed, or signed. The truth-telling oath procedure is fulfilling these conditions: the oath is taken freely and signed in the presence of the monitor. We consider the slightest possible change to the baseline game so as to ease comparison with the original design of Erat and Gneezy (2012). Only senders are offered to sign the oath. Receivers in this game play a rather passive role (very much like receivers in dictator games), and only senders are given an opportunity to cheat. Subjects are offered to sign a truth-telling oath before entering the lab. At this time, they know nothing about the details of the forthcoming experiment.

Each subject enters alone and is directed to a monitor at the front of the laboratory. Each subject that will play as the sender in the experiment is then offered a form to sign entitled "solemn oath" (see Section A in the Appendix). The monitor explicitly points out to the subject before she reads the form that she is free to sign the oath or not, and that participation and earnings are not conditional on signing the oath. Importantly, subjects are not informed about the topic of the experiment when asked to take the oath. The subject reads the form, which asks whether she agrees "to swear upon my honor that, during the whole experiment, I will **tell the truth and always provide honest answers**" (in bold in the original form). The word "oath" is written on the form and read by the subject, but never said aloud. Regardless of whether the subject signs the oath, he is thanked and invited to enter the lab. The exact wording used by the monitor to offer the oath to respondents was scripted to standardize the phrasing of the oath. One monitor stayed in the lab until all subjects had been presented with the oath, to avoid communication prior to the experiment. Subjects waiting their turn could neither see nor hear what was happening

Table 2: Random effect panel Logit regressions: baseline *versus* loaded treatment

	Baseline treatment (n = 240, ♯ subjects = 60)			**Baseline & Loaded treatments (n = 480, ♯ subjects = 120)**		
	Estimated Parameter	Odds-ratio	p-value	Estimated Parameter	Odds-ratio	p-value
Constant term	-.587	-	.339	-.294	-	.521
Payoff dummies						
Altruistic lie	-1.615	0.198	0.001	-1.910	0.148	0.000
Pareto lie	1.188	3.283	0.003	1.137	3.120	0.000
Spite lie	-1.615	0.198	0.001	-2.134	0.118	0.000
Order effect dummies						
Pareto first	1.084	2.957	0.012	0.882	2.417	0.005
Altruistic first	0.843	2.323	0.040	0.323	1.382	0.281
Treatment effect						
Loaded treatment	-	-	-	-0.486	0.615	0.047

Note. The endogenous variable is whether sender i sends a message which does not match the actual roll of the die in the tth game. An estimated odds ratio of 1.5 corresponds to a 50% increase in odds. Dummy variables control for order effect, payoff configurations, and demographics (not reported in the table, we include gender, marital status and religion); the reference person is an atheist single female in a selfish lie payoff configuration.

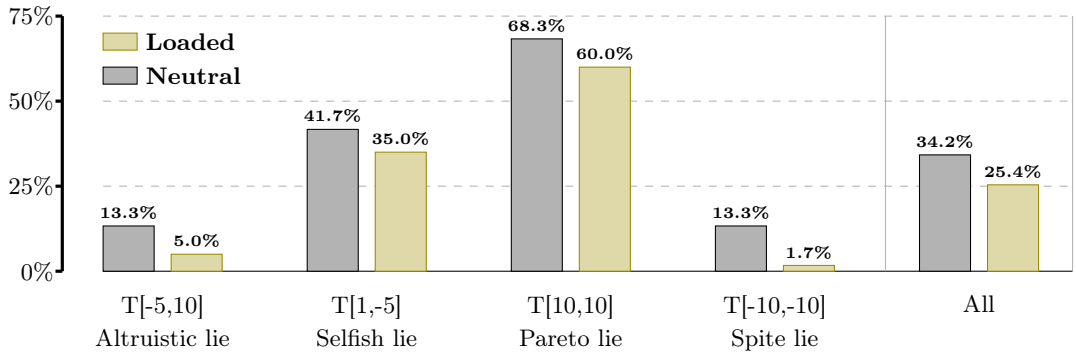
at the oath-desk. No information is given to the subjects about *(i)* whether other subjects were exposed to the oath procedure, or *(ii)* whether any one else decided to sign the oath or not. Again, adding the oath procedure before entering the laboratory is the only difference with the baseline experiment. To be sure that it is the oath that produces a change in behavior, it was important that almost all subjects sign the oath, which is also what we observed. In all experimental sessions reported on in Section 3, the compliance rate with the oath is 95.8% (115 among 120 subjects signed the oath). This less-than-100% rate suggests the subjects knew they had the option to not sign the oath. The Appendix (Section B) presents the detailed procedures of the experiment.

3 Results

For the sake of statistical analysis, we define lying as any decision in which player A deceives player B by sending a message which does not match the roll of the die. We first benchmark the internal validity of our study by comparing the results to Erat and Gneezy (2009, 2012). We then move to our main treatment effects: loaded environment and commitment.

In the Appendix (Section C), we provide a benchmark comparison of observed behavior in our baseline treatment (neutral, no oath) with previous evidence. The main qualitative difference is that our study produces less altruistic lies than previous experiments. The left-hand side of Table 2 provides the results of a panel Logit regression with random effects that controls for payoff configuration, order effects and demographics. The dependent variable is whether the subject lies and the three demographics considered are: gender, living as a couple and being religious. The

Figure 1: Lying behavior in neutral and loaded environments



reference person is an atheist single female observed in the selfish-lie payoff configuration (since demographics are only included as control variables, to downplay the risk of under-specification, the results are not reported in the tables). First, conditional statistics confirm that lying is significantly less likely for altruistic and spite lies relative to selfish lies. Second, subjects are more likely to make Pareto lies. According to odds ratios, the probability of making a Pareto lie is three times greater than making a selfish lie. The probability of making an altruistic and a spite lie are 20% lower than a selfish lie. All three dummy variables are significant at a 1% threshold. Third, order effects are present in the baseline: lying is fostered by facing a Pareto or an altruistic lie first in the sequence.

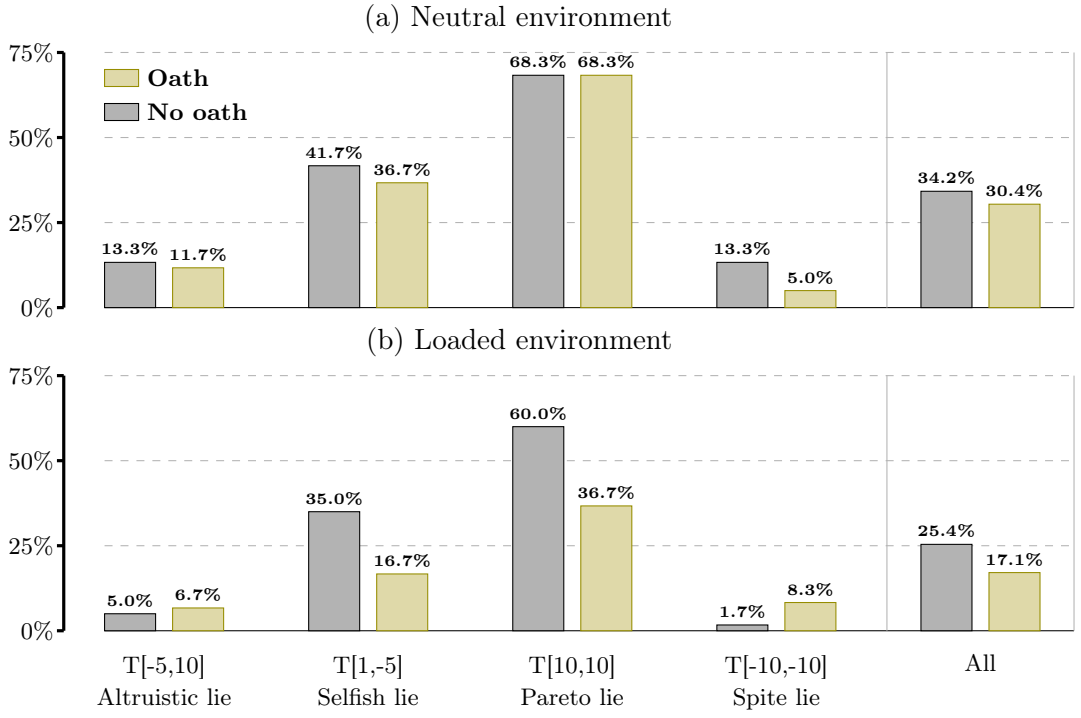
3.1 Truth-telling in a loaded environment

Figure 1 illustrates the effect of our first treatment variable through the fraction of subjects who lie by payoff configuration in both the baseline and the loaded treatments. Overall, lying occurs in 25.4% of observed decisions, which is a 25% decrease in lying behavior in comparison to the baseline treatment. We test for the effect of loading the game by computing a bootstrap unconditional proportion test that accounts for within correlation.[3] According to this testing procedure, the decrease in lying behavior is significant, with $p = 0.022$.

Result 1 *Lying is less likely in the loaded environment.*

[3]The correlation between decisions comes from the within subject implementation of the payoff configurations—uncorrected p-values mistakenly treat each individual decision as independent. The bootstrap procedure is implemented by bootstrapping subjects and their 4 behaviors in the sample rather than observations. We also allow for asymmetry in the empirical distribution of lying behavior (Davidson and MacKinnon, 2006). The procedure is accurate only when applied to pooled data. Due to within subject correlation, it is impossible to test the proportion of lies for each payoff configuration using this test. We rather use conditional tests on parameters from econometric regressions which account for order effects.

Figure 2: Changes in lying behavior due to signing a truth-telling oath



Conditional regressions confirm the overall decrease in lying behavior. On the right-hand side of Table 2, we report the results from a panel Logit regression that pools the data from baseline and loaded treatments. We introduce a dummy variable into the regression to control for the loaded environment. The parameter associated with this variable is negative and significant at a 5% threshold. The odds ratio indicates that the probability to lie in the baseline, after conditioning, decreases by 61.5% when subjects face loaded instructions with explicit wording.

3.2 Truth-telling under oath

Figure 2 contrasts observed behavior in both the neutral and loaded environment according to whether subjects are offered to sign a truth-telling oath or not. The upper part of the figure shows the effect of the oath when the environment is neutral. The oath only marginally decreases lying behavior in the altruistic, selfish and spite payoff conditions, whereas it has no effect on Pareto lies. The bootstrap proportion test cannot reject the null of no difference in overall lying with and without the oath ($p = .631$) in the neutral treatment. Table 3 presents conditional regressions that pool data without and with oath. As in the regressions presented earlier, we control for demographics, payoff configuration and order effects. The results presented on the left-hand side, pooling data from the neutral environment and the neutral environment with oath, confirm that

Table 3: Random effect panel Logit regressions: subjects under oath

	Neutral & neutral oath treatments (n = 480, ♯ subjects = 120)			**Loaded & Loaded oath treatments (n = 480, ♯ subjects = 120)**		
	Estimated Parameter	Odds-ratio	p-value	Estimated Parameter	Odds-ratio	p-value
Constant term	-0.429	-	0.352	-1.136	-	0.034
Payoff dummies						
Altruistic lie	-1.594	0.202	0.000	-1.882	0.152	0.000
Pareto lie	1.301	3.675	0.000	1.140	3.127	0.000
Spite lie	-1.954	0.141	0.000	-2.050	0.128	0.000
Order effect dummies						
Pareto first	0.495	1.641	0.116	0.290	1.336	0.423
Altruistic first	0.538	1.713	0.088	-0.255	0.774	0.494
Treatment effects						
Oath	-0.291	0.747	0.263	-0.707	0.492	0.022

Note. The endogenous variable is whether sender i sends a message which does not match the actual roll of the die in the tth game. An estimated odds ratio of 1.5 corresponds to a 50% increase in odds. Dummy variables control for order effect, payoff configurations and demographics (not reported in the table, we include gender, marital status and religion); the reference person is an atheist single female in selfish lie payoff configuration).

lying behavior is the same with and without oath in this context ($p = .263$).

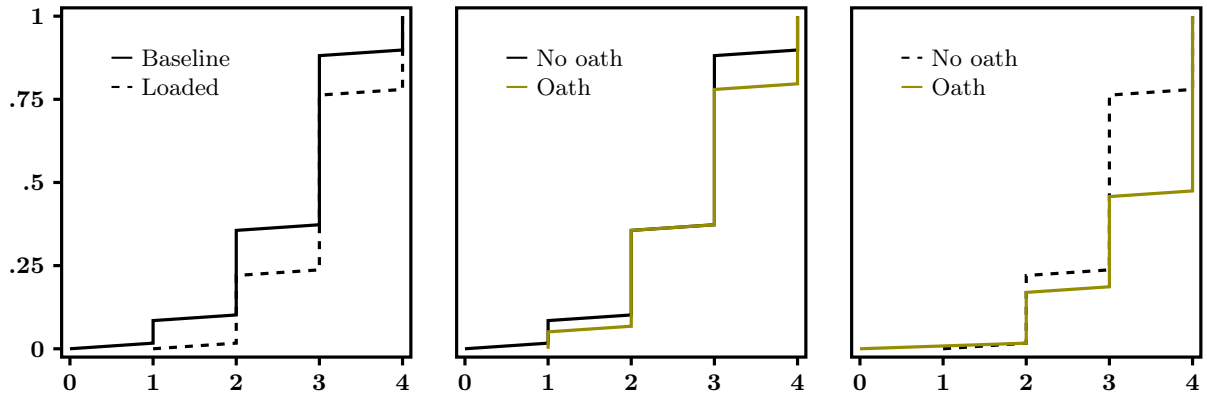
Result 2 *Having subjects sign a truth-telling oath before participating in the neutrally framed lying game leaves truth-telling behavior unchanged.*

We now turn to the comparison of loaded treatments with and without oath. Overall, lying amounts for 17.1% of all messages sent when the oath is combined with a loaded environment. This corresponds to a 33% decrease in lying behavior relative to the loaded-no oath context. The decrease is significant according to a bootstrap proportion test ($p = .008$). We observe that altruistic lies decrease by 52.2% (from 35% to 16.7%) and Pareto lies decrease by 38.3% (from 60% to 36.7%). Selfish lies are slightly increased from 5.0% to 6.7%. Spite lies are more frequent without the oath, but remain at a low level. A bootstrap proportion test indicates that the difference in overall lying between neutral and loaded under oath is statistically significant ($p < .001$).

Result 3 *The oath decreases lying when lies are made explicit in the loaded environment.*

The Logit regression presented in the right-hand side of Table 3 pools data from the loaded environment and the loaded environment with oath. We again control for the lie type, order effect and demographics, and we include a dummy variable that equals 1 if the subjects were offered to take an oath. The estimated parameter associated with this oath-dummy variable is significant at a 5% threshold ($p = .022$). It is highly significant ($p < .001$) when we restrict the sample to the oath treatments only.

Figure 3: Truth ratio by treatment



(a) No oath (b) Neutral (c) Loaded environment

3.3 Treatment effects at the individual level

We further explore the effect of the treatment on lying behavior by looking more closely at individual level data. First, we compute a truth ratio for each subject, i.e., the number of times a subject tells the truth out of the four decisions. Figure 3 provides 2×2 comparisons of the Empirical Distribution Functions (EDF) of the truth ratio across the four treatments. Figure 3.a plots the EDF of the baseline along with the loaded framing (both with no oath). The EDF of the loaded treatment first order dominates the EDF of the baseline at a 10% threshold, $p = .097$.[4] The two subsequent figures contrast behavior under oath with the no oath situation in the neutral and loaded treatment. In the neutral treatment (Figure 3.b), the only visible difference between the two curves is a small increase of truth ratios that equal 4, i.e., subjects who always tell the truth. This increase is however not significant ($p = 0.269$). Consequently, the EDF from the oath(-neutral) treatment does not first order dominate the EDF from the baseline treatment ($p = 0.549$). In contrast, in Figure 3.c, the EDF from the oath loaded environment unambiguously first order dominates the EDF from the no oath loaded environment—the difference is highly significant, with $p < .001$. Truth ratios are also highly significantly different in the two oath treatments: the EDF of the truth ratio when instructions are loaded first order dominates the EDF observed with neutral neutral wording ($p < .001$).

The truth ratios inform about the intensity of lying at the individual level, but do not account for the types of lie in each payoff configuration. To better investigate possible variations in that regard, Table 4 reorganizes observed behavior on an individual basis according to the lying pattern over all four treatments. For each treatment, the table provides the number of subjects whom

[4]This result comes from a bootstrap version of the univariate Kolmogorov-Smirnov test. This modified test provides correct coverage even when the distributions being compared are not entirely continuous and, unlike the traditional Kolmogorov-Smirnov test, allows for ties (see Abadie, 2002; Sekhon, 2011).

Table 4: Lying patterns by treatment

Lying patterns																
	← Nothing but the truth															All lies →
Altruist	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Selfish	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
Pareto	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
Spite	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
Number of subjects																
Neutral	7	0	25	1	4	1	14	0	2	0	0	0	0	5	0	1
Neutral oath	13	0	19	1	3	0	16	1	3	0	1	1	0	0	2	0
Loaded	14	0	23	0	8	0	11	1	1	0	1	0	1	0	0	0
Loaded oath	32	0	12	2	3	1	5	0	1	0	2	0	0	0	0	1

Note. The upper part of the table describes all possible lying patterns in the experiment The bottom part describes for each treatment in row the number of subjects for whom the corresponding lying pattern has been observed.

decisions match each of the 16 potential combinations of lies. In the baseline treatment, we observe three main lying patterns: *(i)* No lie (7 subjects, 11.7%), *(ii)* Pareto lie only (25 subjects, 41.7%) and *(iii)* Selfish and Pareto lie (14 subjects, 23.3%). In the loaded environment, the number of subjects who engage in "nothing-but-the-truth" behavior, i.e., who never lie, happens to double. They are now 23.3% in comparison to 11.7% in the baseline. A bootstrap proportion test indicates that this change is however not significant, with $p = 0.189$. We observe smaller differences for subjects who choose to make a Pareto lie only or a selfish and Pareto lie. The main lesson from this comparison is that the overall decrease in lying observed in Result 1 cannot be attributed to a decrease in only one specific type of liar.

The comparison between oath and no oath treatments when instructions are neutral confirm that Result 2 applies to all lying patterns. The lying patterns are similar under both conditions, the dominant patterns under oath being: no lie at all, only Pareto lie and selfish and Pareto lies only. Again, this is in sharp contrast with the variation in lying patterns when lies are made explicit in the instructions. When this is combined with an oath, 32 subjects engage in truthful behavior for all four payoff configurations. Nothing-but-the-truth behavior equals 53.3% of the sample (this increase is significant relative to the loaded environment without the oath, $p = .003$). The decrease in Pareto lie only is significant ($p = .039$). We also observe less selfish-lie-only behavior: 8.3% of subjects make selfish and Pareto lies under oath, whereas 18.3% do so without the oath—the decrease, however, is not significant, $p = .231$.

As a summary of the main results, we pool all four treatments to jointly test all results. We estimate a random effect panel Logit regression based on the 960 observations from the four treatments. We introduce three dummy variables—one for each result. The first dummy, *loaded treatment*, provides a robustness test of Result 1 to further conditioning. A second dummy, *oath*

Table 5: Joint test - Random effect panel Logit regression

	All treatments (n = **960, ♯ subjects = 240**)		
	Estimated Parameter	Odds-ratio	p-value
Constant term	-0.520	-	0.168
Payoff dummies			
Altruistic lie	-1.712	0.180	0.000
Pareto lie	1.222	3.396	0.000
Spite lie	-2.007	0.134	0.000
Order effect dummies			
Pareto first	0.405	1.500	0.091
Altruistic first	0.141	1.152	0.555
Treatment effects			
Loaded treatment	-0.545	0.579	0.044
Neutral oath	-0.215	0.805	0.420
Oath-loaded treatment	-0.715	0.488	0.014

Note. The endogenous variable is whether sender i sends a message which does not match the actual roll of the die in the tth game. An estimated odds ratio of 1.5 corresponds to a 50% increase in odds. Dummy variables control for order effect, payoff configurations and demographics (not reported in the table, we include gender, marital status and religion); the reference person is an atheist single female in a selfish lie payoff configuration.

neutral, provides a test of Result 2. The third dummy, *oath-loaded treatment*, tests Result 3. Table 5 presents the estimations and confirms that the three key results jointly hold when data are pooled: the parameter associated with the dummy variable *oath neutral* is not significant ($p = .420$) while parameters associated with the dummies *loaded environment* and *oath-loaded environment* are both statistically significant at a 5% level ($p = .044$ and $p = .014$).

3.4 Why the oath (sometimes) works: evidence from decision times

Our main findings are twofold. First, we find that "loading" the decision problem with an explicit mention of its truth-telling content drastically affects a subject's willingness to lie: half of the original lies disappear. Second, a truth-telling oath strongly reinforces the non-monetary incentives to tell the truth: another third of the lies are eliminated under oath. There are good reasons to expect an effect of each of the two non monetary devices on truth-telling behavior based on the literature in social psychology. It is more surprising, and puzzling, that the oath works only in the loaded environment. We now discuss possible reasons for these results given evidence on observed decision times.

Two hypothesis are consistent with the observed asymmetric effect of the oath across experimental environments: that subjects either do not think that they lie when they send a false message, or they do but they perceive lying as reasonable in the neutral environment. The first hypothesis corresponds to a well-grounded argument within the social psychology of commitment

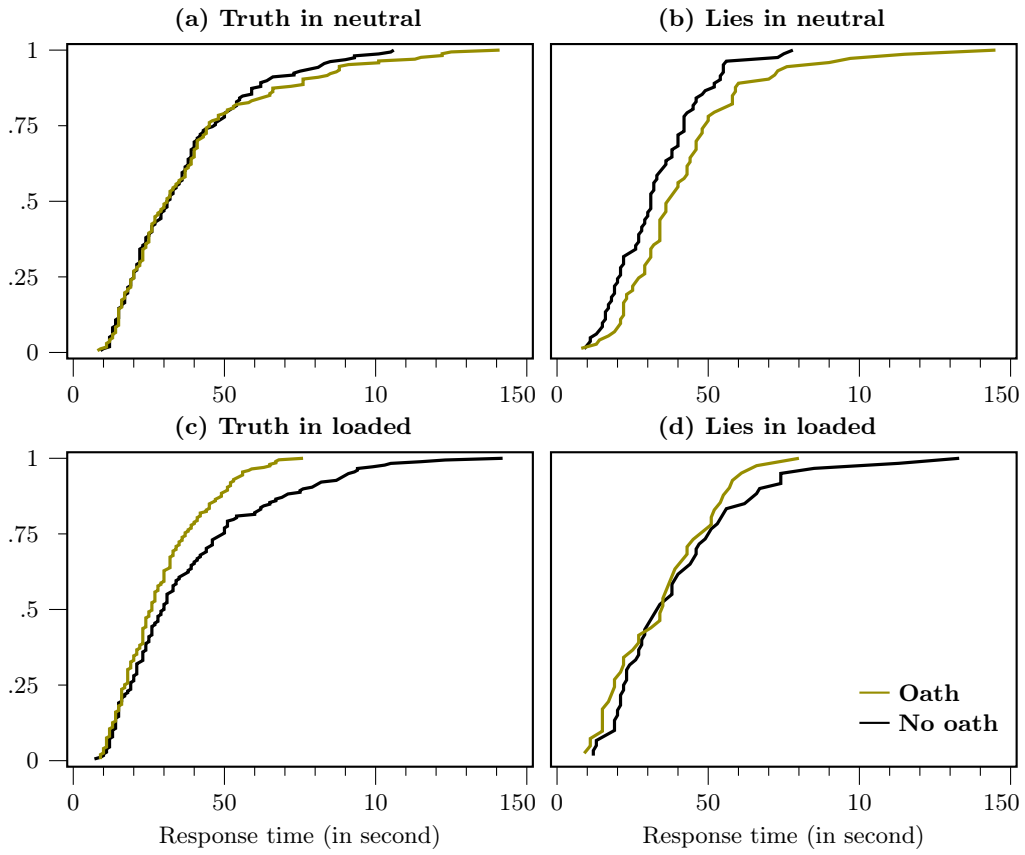
and action identification theory. Action identification theory posits that the way people label their behavior while they are performing an action (e.g., "I am driving", "I am preparing dinner", "I am lying") matters for decision-making (see e.g., Vallacher, Wegner, and Somoza, 1989; Vallacher and Wegner, 1989). People behave in accordance with what they think they are doing. A decision maker can associate his or her action with a different identification, which will result in different choices of actions. This phenomenon has been shown to matter for commitment because some identifications are more appropriate than others for maintaining the action, or for the emergence of new actions. For commitment to induce a significant change in behavior, preliminary behavior and target behavior should share the same level of action identification, see e.g., Joule and Beauvois (1998); Wegner, Vallacher, Macomber, Wood, and Arps (1984). Commitment binds the person to a behavioral act, but it must also belong to the same category of identification. Accordingly, a requirement for the oath to be effective at mitigating dishonest behavior is that subjects think that what they are doing is "lying" when they send a false message. If subjects do not realize they are lying in the neutral condition, then consistent with the social psychology of commitment, the oath would have no or weak effect on behavior because the truth-telling prescription from the oath is meaningless.

The second explanation is illustrated by the famous *Stanford jail experiment* (Zimbardo, 2007). At Stanford in 1971, students were randomly selected to be either guards or prisoners in what students knew to be a fake jail. After only a few days, Zimbardo and his colleagues stopped the experiment before the end because the guards were adopting dangerous and psychologically harming behavior. One of Zimbardo's conclusions from this failure is that situational forces are crucial to generate detrimental behavior. One important driving force is that a legitimate authority explicitly banishes, or implicitly allows, certain types of behavior. Because the neutral condition does not mention either truth-telling or honesty, the subject may consider that lying is implicitly "allowed"—i.e., sending a false message to obtain more money is not perceived to go against any norm, at least in the laboratory. The neutral environment seems to give subjects more "wiggle" room to rationalize lying behavior.

The two explanations are consistent with the oath affecting behavior in the loaded environment while remaining unchanged in the neutral context. They are observationally identical in the game. One important difference between the two, though, is whether subjects recognize their actions as lies when they decide on a message. According to the first explanation, subjects do not realize that some actions in the game are "lies": the oath is ineffective because it is irrelevant. According to the second explanation, they realize what actions mean in terms of honesty, but they view lying *in that way* as something allowed by the experimenter, which would make the oath ineffective.

This difference has testable consequences on the time subjects take to choose their message. Our working hypothesis is the following: if the first explanation holds, response times in the neutral environment should be of comparable magnitude whether subjects are under oath or not. The oath should make no difference because subjects would not consider untruthful communication as

Figure 4: Empirical Distribution Function of Response times by treatment



"lies" in the neutral environment. In contrast, longer response times under oath would favor the second explanation: subjects under oath face a greater moral dilemma.[5]

Following a standard definition in the literature, we measure response times as the time elapsed from the moment the choice is presented to the subject on the computer screen until the choice is made. In Figure 4, we report the EDF of response times by treatment and type of behavior. The main observations are two-fold. First, the top panel reports the behavior of subjects in the neutral condition. While the curves almost perfectly coincide on the left-hand side figure (there is statistically no change in response time for honest decisions, $p = .544$) the right-hand side shows a sharp difference: subjects take significantly more time to decide to send an untruthful message when under oath. The EDF of response time under oath first order dominates that of response

[5]This interpretation holds only if the oath has no direct effect on response times. Otherwise, the correlation of response times with lying behavior could be a spurious consequence of the simultaneous effect of the oath on both lying and response times. To provide evidence on this interpretation, we carried out two separate between-subjects treatments (oath, $n = 66$, and no oath, $n = 63$) of an earned money task which consists of sorting numbers. We find no significant difference in response times between the two treatments.

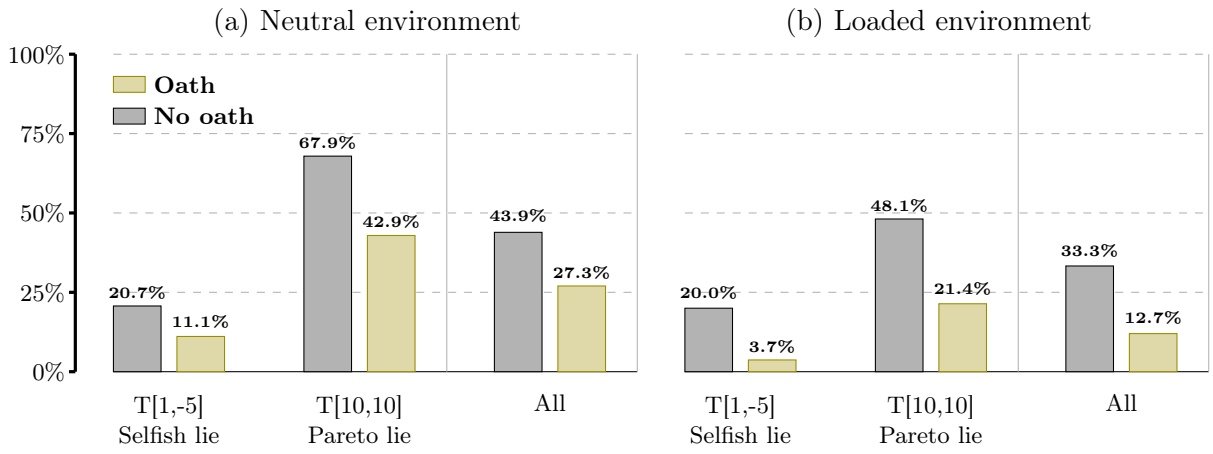
times in neutral without oath ($p = .011$). The increase in median response times amounts to six seconds: the decision to send a lie in the neutral-no oath condition takes 31 seconds whereas it takes 37 seconds under oath, a 20% increase in response time. Under the assumption that higher response times are associated with strengthened moral dilemmas, this suggests that the truth-telling prescription of the oath is meaningful for subjects who decide to send an untruthful message in the neutral condition, even though the oath has no behavioral consequence. Subjects have to solve a stronger moral dilemma in the game when they previously complied with a truth-telling oath, as compared to no oath, which seems to generate greater difficulties to overcome the oath prescription. This tends to rule out the possibility of a mismatch between the oath behavioral prescription and the behavior emphasized in the game. Rather, this result favors the second interpretation: the oath is meaningful to subjects, but it is ineffective due to the norms suggested by the neutral condition.

This is reinforced by the pattern of response times in the loaded environment, reported in the bottom panel. While response times of liars are both observationally and statistically ($p = .610$) the same, honest subjects in this loaded condition take less time to answer when under oath. The EDF of response times in the loaded environment-no oath treatment first order dominates that of response times in the loaded environment-oath treatment ($p = .002$). The difference is also visible in median response times: it is equal to 30 seconds in the loaded environment-no oath condition, while it is 26 seconds in the oath condition—a 13.3% decrease. Both results still hold when the comparison is performed on pooled data. Again, under the assumption of a positive correlation between response times and the intensity of the dilemmas subjects have to solve, this tends to suggest that loading the instructions with explicit wording removes any ambiguity as to what is the norm in the game, and indicates that honesty matters. Such correlations are consistent with the second interpretation: the instructions in the neutral conditions suggests that lying *in that way* is allowed in the lab and, although surprising at first glance, the oath is ineffective on lying. In the loaded treatment, by contrast, the oath eases the decision to tell the truth because subjects face a lower moral dilemma in the course of sending a truthful message.

4 Robustness check: truth-telling under oath in between-subjects design

Subjects have some moral wiggle room to rationalize lying in the within-subject design since only one decision out of four is binding. In their classroom experiment, for instance, Erat and Gneezy (2012) observe a lower share of lying in between-subjects implementation of the treatments as compared to a within-subjects design. We address this issue by rerunning our experiment using a between-subject design with only binding choices. This between-subject design is identical to our within-subject design with one exception: we focus strictly on the two payoffs that generate the most lying—Selfish and Pareto lies treatments, without/with a loaded environment and with-

Figure 5: Changes in lying behavior due to signing a truth-telling oath



out/with an oath. Now each subject faces only one payoff configuration and takes only one decision (the Appendix, Section D, details the procedures of this experiment). Among participants who were offered to sign the oath, 97.3% (107 subjects out of 110) complied with the request. Participants average age is about 23 years old. No subject participated in more than one experimental session and no subject had participated in any of our other oath experiments.

Figure 5 reports the main results from the experiment. In line with existing evidence, we observe a lower rate of lies in the between-subjects treatment. The ranking between payoff configurations remains unchanged. Overall, the results support the findings from the within-subjects design. If anything, the new results make a stronger case for using the oath to create more commitment to truth. This is confirmed with both unconditional and conditional empirical tests. We find two key results. First and most important, the oath reduces lying significantly in both treatments within the loaded environment. Selfish lies decrease to 3.7% from 20.0% ($p = .071$, using a classic proportion test) and Pareto lies decrease to 21.4% from 48.1% ($p = .035$). Together, lies in the loaded environment are around 2.5 times less likely when subjects are under oath compared to no-oath (33.3% in loaded-no oath vs. 12.7% under oath, $p = .009$). Second, the oath reduces lying in the neutral Pareto lie environment, albeit with a bit less influence: Pareto lies are 67.9% without vs. 42.9% under oath, with $p = .053$. The drop in selfish lies in this condition is statistically not significant ($p = .271$).

The econometric regressions presented in Table 6 confirm the overall effectiveness of the oath in both neutral and loaded environments once payoff configurations and socio-demographics are accounted for. In both environments, the parameter associated with the oath dummy is statistically significant. The oath-enhances truth-telling given all choices are binding. This gives less room for subjects to rationalize lying under oath.

Table 6: Probit regressions: subjects under oath

	Neutral & neutral oath treatments (♯ subjects = 112)			**Loaded & Loaded oath treatments (♯ subjects = 102)**		
	Estimated Parameter	Odds-ratio	*p*-value	Estimated Parameter	Odds-ratio	*p*-value
Constant term	-.238	-	0.5476	-0.586	-	0.143
Payoff dummies						
Pareto lie	1.122	3.173	0.000	0.921	5.053	0.003
Treatment effects						
Oath	-0.535	0.179	0.045	-0.746	0.270	0.022

Note. The endogenous variable is whether sender i sends a message which does not match the actual roll of the die. An estimated odds ratio of 1.5 corresponds to a 50% increase in odds. Dummy variables control for payoff configurations and demographics (not reported in the table, we include gender, marital status and religion); the reference person is an atheist single female in a selfish lie payoff configuration. The left-hand side regression is performed on pooled data from neutral treatments, the right-hand side regression is performed on pooled data from loaded treatments (one session is omitted in the right-hand side regression due to missing information on demographics, 10 observations are dropped).

5 Conclusion

Current management practices have increasingly turned to the oath as a device to enhance truth-telling and foster honesty. Herein we explore the behavioral consequences of a truth-telling oath and disentangle between its two possible components: commitment and moral reminders. We implement the two non-monetary incentives to tell the truth one by one in a 2*2 design. The first treatment variable is a truth-telling oath designed so as to elicit full compliance—hence providing identification of its causal effect on truth-telling behavior. The second treatment variable introduces moral reminders through a loaded environment. We consider different lying situations that we may find in organizations: from purely selfish ones (lying benefits the liars at the expense of other members) to efficient ones (all members of the organization benefit from lying). Both devices appear influential on the willingness to tell the truth. When we describe the game to be explicit about truth-telling/lies, we observe 50 percent fewer lies. In a neutral framing, we do not observe any improvement in truth telling after having the subject sign an oath to tell the truth before entering the laboratory. It is only when this commitment device is associated with reminders of ethical standards that the oath drastically improves truth-telling behavior—with a further one third drop in the share of liars. We confirmed these results by retesting behavior in a between-sample treatment in which each subject makes only one lying decision, giving subjects less room to rationalize lying.

Our results complement the growing literature on lying behavior in strategic communication and the ways to undermine it. Gneezy, Meier, and Rey-Biel (2011) illustrate how non-monetary incentives can work to contradict monetary incentives. Our results suggest that the social context of non-monetary commitment can make decisions to lie more difficult. These results are also in line with recent findings from psychology. First, in accordance with Mazar, Amir, and Ariely

(2008) self-concept maintenance theory, we find a strong effect of internal rewards to honesty when people are induced to think in terms of dis/honesty in making their decisions. To strengthen the interpretation of these results we provide evidence based on response times, showing that subjects decide more quickly to *(i)* tell the truth under oath in the loaded environment and *(ii)* to lie with no oath in the neutral environment. Under the assumption that higher response times are associated with more intense moral dilemmas, these correlations suggest that the oath is ineffective in the neutral condition not because it is misaligned with the game, but rather because the neutral condition suggests that lying does not contradict the norms implemented in the laboratory. Non-monetary incentives implemented through an oath do have a strong impact on lying behavior, but social context matters for commitment. Our next step is to explore whether and to what degree an oath of honesty can create, restore, and maintain trust within alternative exchange institutions.

References

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

Anderson, M., and P. Escher (2010): *The MBA oath: Setting a higher standard for business leaders.* Penguin.

Ariely, D. (2012): *The Honest Truth About Dishonesty: How We Lie to Everyone–Especially Ourselves.* HarperCollins Publishers, New York.

Bateson, M., D. Nettle, and G. Roberts (2006): "Cues of being watched enhance cooperation in a real-world setting," *Biology Letters*, 2(3), 412–414.

Bishop, M., and M. Green (2011): *The road from ruin: How to revive capitalism and put America back on top.* Crown Pub.

Boles, T. L., R. T. A. Croson, and J. K. Murnighan (2000): "Deception and Retribution in Repeated Ultimatum Bargaining," *Organizational Behavior and Human Decision Processes*, 83(2), 235–259.

Bénabou, R., and J. Tirole (2011): "Identity, Morals, and Taboos: Beliefs as Assets," *Quarterly Journal of Economics*, 126(2), 805–855.

Cialdini, R., and B. Sagarin (2005): "Interpersonal influence," in *Persuasion: Psychological insights and perspectives*, ed. by T. Brock, and M. Green, pp. 143–169. Sage Press, Newbury Park (CA).

Cohn, A., E. Fehr, and M. A. Maréchal (2014): "Business culture and dishonesty in the banking industry," *Nature*, 516(7529), 86–89.

Cooper, D. J., and J. H. Kagel (2003): "The impact of meaningful context on strategic play in signaling games," *Journal of Economic Behavior & Organization*, 50(3), 311–337.

Cooper, D. J., and J. H. Kagel (2009): "The role of context and team play in cross-game learning," *Journal of the European Economic Association*, 7(5), 1101–1139.

Crawford, V. P., and J. Sobel (1982): "Strategic Information Transmission," *Econometrica*, 50(6), 1431–1451.

Davidson, R., and J. G. MacKinnon (2006): "Bootstrap methods in econometrics," in *Palgrave Handbooks of Econometrics*, ed. by T. C. Mills, and K. D. Patterson, vol. 1, chap. 23, pp. 812–838. Palgrave Macmillan, Basingstone.

DeMartino, G. F. (2011): *The economist's oath: On the need for and content of professional economic ethics*. Oxford University Press.

Erat, S., and U. Gneezy (2009): "White Lies," *Rady School of Management Working paper*.

——— (2012): "White Lies," *Management Science*, 58(4), 723–733.

Freedman, J. L., and S. Fraser (1966): "Compliance without pressure: the foot-in-the-door technique," *Journal of Personality and Social Psychology*, 4(2), 195–202.

Gneezy, U., S. Meier, and P. Rey-Biel (2011): "When and Why Incentives (Don't) Work to Modify Behavior," *Journal of Economic Perspectives*, 25(4), 191–210.

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

Grossman, S. (1981): "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law & Economics*, 24(3), 461–83.

Haley, K. J., and D. M. T. Fessler (2005): "Nobody's watching? Subtle cues affect generosity in an anonymous economic game," *Evolution and Human Behavior*, 26(3), 245–256.

Harris, M. B. (1972): "The Effects of Performing One Altruistic Act on the Likelihood of Performing Another," *Journal of Social Psychology*, 88(1), 65–73.

Jacquemet, N., A. James, S. Luchini, and J. Shogren (2016): "Referenda under oath," *Environmental and Resource Economics*, Forthcoming.

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. Shogren, and A. Zylbersztejn (2017): "Coordination with Communication under Oath," *Experimental Economics*, Forthcoming.

Joule, R., and J. Beauvois (1998): *La soumission librement consentie*. Presses Universitaires de France, Paris.

Joule, R.-V., F. Girandola, and F. Bernard (2007): "How Can People Be Induced to Willingly Change Their Behavior? The Path from Persuasive Communication to Binding Communication," *Social and Personality Psychology Compass*, 1(1), 493–505.

Kahneman, D. (2003): "Maps of Bounded Rationality: Psychology for Behavioral Economics," *American Economic Review*, 93(5), 1449–1475.

Kartik, N. (2009): "Strategic signaling with lying costs," *Review of Economic Studies*, 76(4), 1359–1395.

Kataria, M., and F. Winter (2013): "Third Party Assessments in Trust Problems with Conflict of Interest: An Experiment on the Effects of Promises," *Economics Letters*, 120(1), 53–56.

Khurana, R., and N. Nohria (2008): "It's time to make management a true profession," *Harvard Business Review*, 86(10), 70–77.

Mazar, N., O. Amir, and D. Ariely (2008): "Identity, Morals, and Taboos: Beliefs as Assets," *Journal of Marketing Research*, 45(6), 633–644.

Milgrom, P. R. (1981): "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics*, 12(2), 380–391.

Podolny, J. M. (2009): "The buck stops (and starts) at business school," *Harvard Business Review*, 87(6), 62–67.

Rigdon, M., K. Ishii, M. Watabe, and S. Kitayama (2009): "Minimal social cues in the dictator game," *Journal of Economic Psychology*, 30(3), 358–367.

Rubinstein, A. (2007): "Instinctive and Cognitive Reasoning: A Study of Response Times," *Economic Journal*, 117(523), 1243–1259.

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

Silver, M. (1995): *Economic Structures of Antiquity*. Greenwood Press, Westport, CT.

Silvia, P. J., and T. S. Duval (2001): "Objective Self-Awareness Theory: Recent Progress and Enduring Problems," *Personality and Social Psychology Review*, 5(3), 230–241.

Sutter, M. (2009): "Deception Through Telling the Truth?! Experimental Evidence From Individuals and Teams," *Economic Journal*, 119(534), 47–60.

Tyler, J. E. (1834): *Oaths; Their Origins, Nature, and History*. J.W. Parker, London.

Vallacher, R. R., and D. M. Wegner (1989): "Levels of personal agency: Individual variation in action identification," *Journal of Personality and Social Psychology*, 57(4), 660.

Vallacher, R. R., D. M. Wegner, and M. P. Somoza (1989): "That's easy for you to say: action identification and speech fluency," *Journal of Personality and Social Psychology*, 56(2), 199.

Wegner, D. M., R. R. Vallacher, G. Macomber, R. Wood, and K. Arps (1984): "The emergence of action," *Journal of Personality and Social Psychology*, 46(2), 269.

Zimbardo, P. (2007): *The Lucifer effect. Understanding How Good People Turn Evil*. Random House, New York.

Appendix: experiment material

A Oath form used in the experiment

FACULTÉ D'ÉCONOMIE
université montpellier 1

Université Montpellier 1

SOLEMN OATH

I undersigned ..... swear upon my honor that, during the whole experiment, I will:

Tell the truth and always provide honest answers.

Montpellier, ... Signature.....

Avenue Raymond DUGRAND - Espace Richter—C.S. 79606—34960 —Montpellier cedex 2—Tel. 04.67.15.84.50—Fax. 04.67.15.84.86

B Experimental procedures for the within-subjects experiment

In the experiments reported in Section 3, subjects randomly draw a tag from a bag upon arrival, assigning them to a role: sender or receiver. This draw is performed in a separate room before entering to the laboratory. After drawing their tag, receivers are asked to enter the lab and sit in front of the computer indicated on their tag. Only the senders are asked to sign a solemn oath (in the oath treatments) and then sit in the lab.

Table 7: Comparison with Erat and Gneezy experimental results

(*Setting*) *Matching scheme* Payoffs	Type	Erat and Gneezy (2009) *(Internet)* *Between*	Erat and Gneezy (2012) *(Classroom)* *Between*	*Within*	This paper *(Laboratory)* *Within*	*Between* (Section 4)
T[-1;10]	Altruistic	33%	33%	43%	–	–
T[-5;10]	Altruistic	29%	–	–	13%	–
T[1;10]	Pareto	46%	49%	66%	–	–
T[10;10]	Pareto	61%	65%	76%	68%	68%
T[1; 0]	Pareto	41%	–	–	–	–
T[1;-1]	Selfish	32%	–	–	–	–
T[1;-5]	Selfish	20%	37%	52%	42%	21%
T[10; 0]	Pareto	47%	52%	-	-	–
Reverse	Spite	6%	–	–	13%	–

Note. For each payoff configuration in row, the Table reports the share of senders who decide to lie in each experiment presented in column. The last column reports the results from the baseline treatment (neutral, no oath).

Once all subjects are seated in the room, written instructions are distributed and read aloud. Subjects then observe their type displayed on the screen, A (sender) or B (receiver), which remains the same for the whole experiment. The experiment then consists of four periods. The only difference between the periods is the payoff configuration used. At the beginning of each period, the subjects are paired (one player A, one player B in each pair) and they are informed that they will never play twice with the same partner. The experiment ends with a demographic questionnaire. One of the four decision rounds is randomly drawn and each player receives the amount in Euros corresponding to her gains in that round, plus a show-up fee equal to 5 Euros.

All sessions took place in the experimental lab of LAMETA[6], (France) between March and June 2012. For each treatment, we ran six 20-subjects sessions (two for each order defined in Table 1): this generates 480 observations of sender's decisions from 120 different individuals. Overall 480 subjects participated in the experiment, 220 males and 260 females. 91.67% were students, among which 38.86% were likely to have some background in game theory due to their field of study.[7] Participants' average age is about 24 years old. No subject participated in more than one experimental session. Each session lasted about 50 minutes, with an average payoff of 11.70€ (including a 5€ show-up fee).

C Comparison of the results with existing evidence

Table 7 provides a benchmark comparison of observed behavior in our baseline treatment (neutral, no oath) with previous evidence. Pooling all payoff configurations, we observe that a significant fraction of our subjects, 34.2%, are willing to lie. Lying is the more likely when *(i)* it is in the subject's monetary interest to lie (selfish lie = 41.7%) and *(ii)* when the other player also gains from the lie (Pareto lie =

[6]Laboratoire Montpellierain d'Economie Théorique et Appliquée, `http://leem.lameta.univ-montp1.fr/`. The recruitment uses Orsee (Greiner, 2015).

[7]Disciplines such as economics, engineering, management, political science, psychology, mathematics applied in social science, mathematics, computer science, sociology.

68.3%). These results are in line with those obtained in both a classroom setting and over the internet by Erat and Gneezy (2009, 2012). The order in terms of intensity of lying is also preserved in all instances: lies are more likely in the Pareto payoff configuration, lower in the selfish condition, and (slightly) lower in the altruistic condition. The main qualitative difference is that our study produces less altruistic lies than previous experiments. In Erat and Gneezy (2009, 2012), 33% of subjects engage in altruistic lying in the between-subjects-implementations (subjects face only one payoff configuration) and 43% in the within-subjects-implementation (subjects face four payoff configurations). These results are obtained for an altruistic lie configuration with a low cost-benefit ratio, as it costs 1 to the sender for the receiver to gain 10. Our altruistic payoff configuration has only been implemented in Erat and Gneezy (2009) in a between subjects context and over the Internet, in which it results in a 29% share of liars—this is lower than the between-subjects-implementation, but larger than the percentage of liars that we observe.

D Experimental procedures for the between-subjects experiment

All sessions fo the experiment reported in Section 4 took place in the same experimental lab (at LAMETA, France) between March and October 2016. The same monitors who ran the within-treatments sessions ran these new between-treatment sessions. The same monitor administered the oath to all subjects. For each treatment and pay-off configuration, we run three 20-subjects sessions. In total, 448 subjects (224 senders) participated in the between-subject design treatments (249 males/199 females)—due to technical problems with computers, some sessions were carried out with less then 20 subjects. Each session lasted about 50 minutes, with an average payoff of 10.8 Euros (including a 5 Euros show-up fee).