GATE LYON SAINT-ÉTIENNE
UMR 5824

93, chemin des Mouilles
69130 Ecully - France

Maison de l'Université, Bâtiment B
10, rue Tréfilerie
42023 Saint-Etienne cedex 02 - France

http://www.gate.cnrs.fr
gate@gate.cnrs.fr

WP 2311 – July 2023

Commitment to the truth creates trust in market exchange: Experimental evidence

Nicolas Jacquemet, Stéphane Luchini, Jason F. Shogren, Adam Zylbersztejn

Abstract:

Social norms like the mutual belief in reciprocity facilitate economic exchange. But this reciprocity norm requires trust among traders, which can be challenging to create among strangers even with communication. The honesty oath is a time-honored mechanism that societies use to overcome this challenge – taking a solemn oath to tell the truth sends a trustworthy signal of real economic commitment given incomplete contracts. Herein we explore how the truth-telling oath creates trust within the sequential reciprocity trust game with pre-play, fixed-form, and cheap-talk communication. Four key results emerge: (1) communication under oath creates more trust and cooperative behavior; but (2) the oath induces a selection effect – it makes people more wary of using communication as a signal. (3) Although the overall net effect on cooperation is positive, the oath cannot reverse a general decay of cooperation over time. (4) By comparing the oath's performance to mild and deterrent fines for deception, we find that the oath is behaviorally equivalent to mild fines. The deterrent fine induces the highest level of cooperation.

Keywords:
Trust game; cooperation; communication; commitment; deception; fine; oath

JEL codes:
C72; D83

Working Paper

Commitment to the truth creates trust in market exchange: Experimental evidence*

Nicolas Jacquemet† Stéphane Luchini‡ Jason F. Shogren§ Adam Zylbersztejn¶

July 21, 2023

Abstract

Social norms like the mutual belief in reciprocity facilitate economic exchange. But this reciprocity norm requires trust among traders, which can be challenging to create among strangers even with communication. The honesty oath is a time-honored mechanism that societies use to overcome this challenge — taking a solemn oath to tell the truth sends a trustworthy signal of real economic commitment given incomplete contracts. Herein we explore how the truth-telling oath creates trust within the sequential reciprocity trust game with pre-play, fixed-form, and cheap-talk communication. Four key results emerge: (1) communication under oath creates more trust and cooperative behavior; but (2) the oath induces a selection effect — it makes people more wary of using communication as a signal. (3) Although the overall net effect on cooperation is positive, the oath cannot reverse a general decay of cooperation over time. (4) By comparing the oath's performance to mild and deterrent fines for deception, we find that the oath is behaviorally equivalent to mild fines. The deterrent fine induces the highest level of cooperation.

Keywords: Trust game; cooperation; communication; commitment; deception; fine; oath.

JEL Classification: C72; D83.

---

*We wish to thank Sorravich Kingsuwankul, Rupert Sausgruber, Eli Spiegelmann, Jean-Robert Tyran, Jeroen van de Ven, Marie Claire Villeval, and several workshop and conference participants for helpful comments. Sophie Cottet, Maxim Frolov, and Ivan Ouss provided valuable research assistance. This project has received funding from the chair "Economie Publique et Développement Durable" (Aix-Marseille University) and from the Foundation Aix-Marseille University (research program "Mutual trust and commitment"). We gratefully acknowledge financial support from the National Research Agency, through both the LABEX CORTEX, ANR-11-LABX-0042, which is part of the program Investissements d'Avenir ANR-10–LABX-93-0, and the EUR-OSE program, ANR-17-EURE-0001, ANR-17-EURE-0020, and from the Excellence Initiative of Aix-Marseille University-A*MIDEX. Adam Zylbersztejn acknowledges VISTULA Fellowship.

†Paris School of Economics and U. Paris 1 Panthéon-Sorbonne. nicolas.jacquemet@univ-paris1.fr

‡Aix-Marseille U., CNRS, EHESS, Centrale Marseille, Aix-Marseille School of Economics. stephane.luchini@univ-amu.fr

§Department of Economics, U. Wyoming. JRamses@uwyo.edu

¶Univ Lyon, Université Lyon 2, GATE UMR 5824, F-69130 Ecully, France; research fellow at Vistula University Warsaw (AFiBV), Warsaw, Poland. zylbersztejn@gate.cnrs.fr.

"Bare Virtue can't make Nations live In Splendor; they, that would revive A Golden Age, must be as free, For Acorns, as for Honesty. "

Bernard Mandeville. Knaves turn'd Honest, 1705.

"[...] there is no vengeance by men, and no available human witness. The man who has sworn is really face to face with nothing but his own sense of Aidôs [honor], plus a vague fear of gods and spirits... The thing that makes the perjury especially base... is precisely his security from danger. "

G. Murray. The Rise of the Greek Epic, 1934

1 Introduction

Creating economic value through cooperation without binding legal contracts depends on social norms — informal rules that create social beliefs about fair and honorable human behavior (see, for example, Elster, 1989, for a discussion of this vast literature). One classic social norm which helps facilitate cooperation is trust and sequential reciprocity, i.e., you scratch my back, I'll scratch yours (see e.g. Buckholtz and Marois, 2012; Duffy, Xie, and Lee, 2013; Dufwenberg and Kirchsteiger, 2004).[1] But as noted by Dasgupta (2009, p. 3308), "*[c]reating trust is no easy matter*". While communication between players can help by fostering common social beliefs when people differ by culture and clan, words have their limits. Familiar phrases such as "trust me" or "my word is my bond" may or may not be a true signal of real economic commitment to honest reciprocal cooperative behavior (see, e.g., Charness and Dufwenberg, 2006).

Herein we consider whether one historic mechanism — the solemn truth-telling oath — can be used to create more trust as a social norm between traders in a reciprocal exchange. We explore whether communication under oath becomes a key mechanism to unite people into a common belief of the social norm of fair sequential sharing. Such honesty oaths have played an important

[1] Trust is "the belief that others act in the interest of some measure of fairness or social welfare rather than their own self-interest" (Bracht and Feltovich, 2009, p. 1036). Trustworthiness, in turn, is the extent to which trust in a person is warranted. Trust and trustworthiness are considered as primary components of social capital and constitute a milestone of economic success at the organization-, community- and nation-wide level (Putnam, Leonardi, and Nanetti, 1992; Putnam, 2000). In an early contribution, Putnam, Leonardi, and Nanetti (1992) draw a link between macro-level outcomes — the political and economic success of large social collectivities and generalized trust and cooperation — to the micro-level phenomena such as networks, associations, and institutions: " *Voluntary cooperation is easier in a community that has inherited a substantial stock of social capital, in the form of norms of reciprocity and networks of civic engagement. Social capital here refers to features of social organizations, such as trust, norms and networks that can improve the efficiency of society by facilitating coordinated action*" (p. 167). An empirical study by Knack and Keefer (1997) corroborates this insight: their sample of 26 countries reveals a strong association between trust and economic growth. Given the major role it plays in human interactions and organizations, trust constitutes a prominent research topic spanning multiple disciplines of social sciences (Rousseau, Sitkin, Burt, and Camerer, 1998).

role in the development of commerce and economic exchanges in Western societies. An oath is a "sacrament" (whether it is religious or not) which aims to create a commitment between people or with a political body to a specific action, e.g., truth telling, duty, integrity (Prodi, 1992). The use of oaths to create and enhance trust in ancient economies characterized by high transactions costs is well documented (see, e.g., Silver, 1995). Oaths are also commonly used today in modern economic exchange. In business disputes and lawsuits when gathering information through legal discovery, witnesses typically answer questions under oath in their depositions and interrogatories. Oaths are also involved in the regulation of business conducts, as it were the case in the pre-modern world. Business professionals (such as bankers, managers, or pharmacists) can be asked to take oaths of integrity (de Bruin, 2016).[2]

We use the classic trust game (Berg, Dickhaut, and McCabe, 1995) to understand better how a truth-telling oath can affect trust and cooperative behavior. Communication between the trustee and the trustor is a classical tool to create trust within the pair (see, e.g., Bracht and Feltovich, 2009). But if communication does not change the underlying incentives to behave opportunistically, it is nothing more than cheap talk without credibility (Farrell and Rabin, 1996; Bliege Bird, Ready, and Power, 2018; Crawford, 1998, for a review of cheap talk in games). From this viewpoint, such communication may only give rise to strategic deception: opportunistic partners may use it to lure others and exploit their trust. As a consequence, whether partners have a chance to communicate or not prior to interacting, no cooperation arises in the subgame perfect Nash equilibrium. We examine whether the oath can foster trust and trustworthiness by promoting honest communication, i.e., transforming noisy "words" into credible signals of cooperative behavior. Our objective is to capture a "pure oath effect" rather than the effect of a subsequent monetary punishment (that is usually associated with the otherwise symbolic oath in real world institutions) if a person lies under oath. To that aim, we carry out the oath procedure put forward by Jacquemet, Joule, Luchini, and Shogren (2013) that borrows insights from the social psychology of commitment (see, e.g., Kiesler, 1971; Kiesler and Sakumura, 1966; Joule and Beauvois, 2010). The truth-telling oath is voluntary, made in written and in private, and has no payoff implications. As a result, truth-telling under oath is solely achieved by means of fostering one's own intrinsic commitment to honesty.

Our results on trust are unambiguous: communication under oath helps create more trust and cooperative behavior. We document that this improvement in efficiency arises from two

[2] Today we have also seen a resurgence of oaths to restore or create trust for environmental or ethical issues. For instance, green oaths commit people to take actions for environmental issues, e.g., "Ocean oath" (`https://oceanconservancy.org/blog/2017/01/22/take-the-oath/`) or the "Pledge for our Planet" (`https://support.worldwildlife.org/site/Advocacy?cmd=display&page=UserAction&id=885`). Geophysicists have introduced an ethical code of conduct named the "Geoethical Promise" (`http://www.geoethics.org/geopromise`). New ethical initiatives in digital technologies promote the use of "Tech oaths" (see, e.g. `https://www.doteveryone.org.uk/2018/03/oaths-pledges-and-manifestos-a-master-list-of-ethical-tech-values/` inspired by the Hippocratic Oath.

countervailing effects: *(i)* the oath promotes truth-telling — subjects under oath keep their word significantly more than without the oath (also see Jacquemet, Luchini, Shogren, and Zylbersztejn, 2018; Jacquemet, Luchini, Rosaz, and Shogren, 2018); but *(ii)* the oath also induces a selection effect: it makes people more wary of using communication altogether. We moreover show the overall net improvement in efficiency is due to better pre-play communication rather than to an effect of the oath *per se* on the cooperative behavior in the game: the oath has no behavioral effect in the absence of pre-play communication.

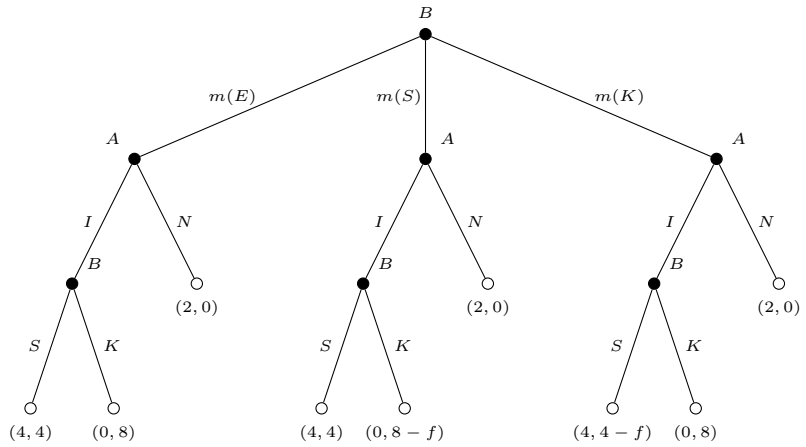
Finally, we assess the economic value of the oath using a two-steps process. First, we introduce monetary fines into the experiment: if the trustee falsely communicates her intentions, whether it is to cooperate with or to exploit the trustor, she pays a monetary fine. Following Bohnet, Frey, and Huck (2001), the level of the fine is varied in a between-subject design: the Mild fine reflects weak enforcement, while the Heavy fine is a deterrent fine that wipes out all the monetary gains of deceptive behavior. Our results confirm that both types of fines shift up cooperation, but only heavy fine prevents it from eroding over time. The oath is behaviorally equivalent to a mild fine. In the second step, we then assess the economic value of the oath by comparing the distribution of earnings between treatments and show that the oath induces a small but significant welfare improvement relative to the baseline without the oath. A first analysis of earnings in the fines treatments indicate that when the investor is sanctioned for lying but the trustor receives no monetary compensation for being deceived, the fine treatments are dominated by both the baseline and the oath treatments. Our second simulation counterfactually compensates the trustor for deceptive communication: we redistribute the garnered fines to the trustor, holding behavior constant. Fine treatments then unambiguously dominate both the baseline and the oath treatments. Since the legal enforcement of contracts is usually costly, we estimate the efficiency ratios that equalize social welfare in the oath and the fines treatments with redistribution. These ratios are all positive indicating that there is room for the oath to operate when contract enforcement is incomplete. This holds even for the heavy fine treatment. When compensation through redistribution is infeasible, the oath can still compare to the heavy fine in terms of social welfare.

2 Design of the experiment

Our design relies on the two-players trust game with pre-play communication and dishonesty fines presented in Figure 1. The core game is a classical trust game: player A holds an initial endowment of 2 and can either maintain the status quo (N) or trust player B by investing the entire amount (I). If player A invests, the amount quadruples and player B can act in a trustworthy manner by making a 50-50 split with player A (S) or keep the money for herself (K).[3] However, before

[3] The investment multiplier of 4 has become standard in the experimental literature since the seminal study by Berg, Dickhaut, and McCabe (1995).

Figure 1: Trust game with prior communication and dishonesty fines



player A's decision to invest or not, player B can send a message to player A. The message can be either $m(K)$ — "*I will keep*", $m(S)$ —"*I will split*", or an empty message $m(E)$ — "*I will either keep or split*". We allow for different communication technologies: cheap talk ($f = 0$) or binding communication where honesty is incentivized through monetary fines paid by player B when the message (provided it is not empty) does not coincide with the action undertaken in the trust game ($f > 0$). The game boils down to a classical trust game when the prior communication phase is not carried out.

In all treatments, subjects are split into two groups at the beginning of each session — player As and player Bs. Subjects then play the experimental game with all payoffs given in EUR.[4] The game is repeated 5 times with constant roles. Pairs are rematched using a round-robin procedure: each 20-subjects session is split into 2 matching groups, resulting in two independent clusters of observations per session. Subjects are only informed that the experiment consists of multiple rounds so as to rule out explicit end-game effects. Moreover, the instructions make clear that each pair of players interacts only once to prevent reputational concerns. Further implementation details are provided in Appendix A.

[4] In the written instructions (see Appendix B, for an English translation of the original instructions in French), we use a neutral left-right (*gauche/droite* in French) framing of players' action spaces. More specifically, decision G (g) corresponds to Not investing (Keeping), and D (d) to Investing (Splitting). Hence, the exact wording of the non-empty messages is either *I will make the decision g* or *I will make the decision d*. For the sake of simplicity, throughout the paper we use the traditional trust game nomenclature *(i.e.*, investing or not, keeping or splitting) when referring to players' actions or messages.

2.1 Non-monetary commitment: Truth-telling oath

Our main treatment variable is a truth-telling oath, a non-monetary commitment device aimed at transforming cheap talk into binding communication. Our oath procedure borrows insights from the theory of commitment in social psychology (Kiesler, 1971), which investigates institutions that induce people to "comply without pressure" (see the reviews in Joule, Girandola, and Bernard, 2007; Cialdini and Sagarin, 2005). Experiments on the social psychology of commitment have shown that decisions made in the past, even when they appear innocuous, can commit people to a particular course of action in the future (Freedman and Fraser, 1966). For instance, a person may be more likely to help a stranger asking for some change to take the bus after having agreed to answer a preliminary request from that stranger about telling the time. The experimental literature in social psychology does not only document a significant commitment effect of past decisions, but more importantly shows that the magnitude of the behavioral change is large, long-lasting and can have strong attitudinal consequences.[5] Accumulated evidence in social psychology suggests that commitment is likely to be stronger when it is freely made, signed and/or publicly expressed.

We apply these insights to the truth-telling oath procedure implemented in the experiment, following the design introduced in Jacquemet, Joule, Luchini, and Shogren (2013). Before entering the lab (and learning about the content and proceedings of the experiment), each subject is invited to a separate room adjoining the laboratory where she is welcomed by one of the monitors. The monitor offers each subject a form to sign entitled "solemn oath" (see Appendix C); the word oath is written on the form and read by the subject, but never said aloud. The monitor explicitly points out before the subject reads the form that she is free to sign the oath or not, that participation and earnings in the experiment are not conditional on signing the oath, and that whether she signs the oath or not would be private information that would not be revealed to anyone else within or outside the experiment. 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). Regardless of whether the subject signs the oath, she is thanked and invited to enter the lab. The exact wording used by the monitors to offer the oath to respondents was scripted to standardize the phrasing of the oath. The monitor did not leave the room at any time, and another monitor remained 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 at the oath-desk. This additional stage during which subjects are exposed to voluntarily signing the oath is the only difference between oath and no-oath treatments.

[5] In the previous example, describing the design of a typical commitment experiment in social psychology, the compliance rate with the target request jumps on average from 1/3 to 2/3 (see Joule and Beauvois, 2010, for a detailed discussion).

Two additional characteristics of the oath procedure are worth noting. First, subjects face no penalty (either in terms of monetary sanctions in the lab or negative reputational effects during the experimental session, or hampered participation in the future experiments) for not signing the oath, or for violating the oath they had signed (e.g., by lying under oath). Compliance with the oath is not obtained by external social pressure from the monitor or other subjects. Subjects do not know whether others have been proposed or signed an oath at any stage of the experiment. Any change in behavior from subjects who signed the oath hence solely relies on one's own intrinsic commitment to honesty.

Second, the oath is generic and minimalistic in the sense that it only targets truth-telling. Our purpose in this paper is to assess the sole effect of committing subjects to truth-telling by an oath in economic exchanges.[6] We thus depart from other types of oaths such as, for instance, the commonplace banker's oath (see, e.g., Weitzel and Kirchler, 2023) that contains multiple clauses with loaded wording that make explicit the reference to the general importance of honesty to the society as a whole "*I will keep confidential that which has been entrusted to me*", "*I will not abuse my knowledge*", and "*I will act openly and accountably, knowing my responsibility to society*". We exclude such moral reminders from the oath procedure and thus elicit rather conservative estimates of its effect on behavior. The efficiency of this intervention is substantiated by a large amount of both lab and field experimental evidence (in contrast with other kinds of social commitment devices, see Zickfeld, Karg, Engen, Gonzalez, Michael, and Mitkidis, 2022).

2.2 Main experimental treatments

We investigate the effect of the oath on cooperation in the trust game by carrying out a between-subject design that combines cheap talk communication ($f = 0$) and the truth-telling oath according to a 2×2 factorial design that consists of three benchmark conditions: trust game without communication (Baseline, 3 sessions, 60 subjects), trust game with cheap talk pre-play communication (Communication, 3 sessions, 60 subjects) and trust game under oath without communication (Oath only, 3 sessions, 60 subjects). The main treatment of interest combines cheap talk communication with an oath (Oath, 4 sessions, 80 subjects). We observe a high oath acceptance rate in all treatments under oath (97.3%).[7] This guarantees that there is no selection effect in our data. Consequently, we do not distinguish oath compliers and non-compliers in the statistical analysis and we apply an intention to treat analysis of the data — i.e., all the analyses rely on comparisons between treatments.

[6] This historically resonates with the three fundamental precepts of roman law (*tria praecepta iuris*): *honeste vivere, alterum non laedere, suum cuique tribuere* ("to live honestly, not to harm others, to render to each his own") — which are the basis of contemporary legal theory and practice in the law of delict and contract law (Petrak, 2014).

[7] This rate includes all the oath-based conditions reported in the paper: Oath only, Oath, as well as an additional condition combining the oath and a heavy monetary fine.

2.3 Fine-equivalents of oath-based commitment

We complement this 2×2 design with four treatments that introduce a positive dishonesty fine f.[8] The fine is paid when there is a mismatch between the announcement (provided it is informative) and the subsequent decision, as shown in Figure 1. This allows us to compare the effect of the oath with the effect of dishonesty fines, and deduce the monetary equivalent of the truth-telling oath from comparisons with our main treatment. Dishonesty fines are explained to the subjects in the instructions by adding the following statement:

> "*Moreover,* ***had the player A chosen I, if the decision made by the player B does not match the decision announced in her message, then the amount of [monetary value of f] EUR shall be subtracted from her gains in a given period.*** *This procedure does not apply if the participant B did not announce her decision.*"

We consider four values of the fine, $f = \{1, 2, 3, 4\}$. The first three are mild, since announcing splitting and then deciding to keep (conditional on player A's trust) remains beneficial to player B in purely pecuniary terms. In contrast, $f = 4$ is considered deterrent, since it washes out all the monetary gains of deceptive behavior. We refer to this treatment as the heavy fine treatment. These four additional treatments are implemented between subjects (4 sessions, 80 subjects for each one of them). In addition, our data also include four sessions (with a total of 80 subjects from 4 sessions) in which the OATH condition is combined with a heavy fine.

3 Honesty, trust and trustworthiness under oath

Table 1 provides a summary of the key aggregate outcomes by treatment: the rate of cooperation (player A invests and player B splits), the rate of trust (player A invests), and the rate of trustworthiness (player B splits provided that player A invested). The baseline rate of cooperation is low and equals 16%. This arises in spite of a relatively high level of trust (60.7%) and mainly because of a relatively low level of trustworthiness — only one player B in four splits following player A's investment. Figure 2.a presents the rate of cooperation over rounds. Cooperation starts at 23.3% in round 1 and ends up at 6.7% in round 5 (recall that subjects do not know when the game ends).

[8] Our dishonesty fine echoes the escrow procedure proposed by Bracht and Feltovich (2008) for the trust game played without communication. In their design, the second mover places an amount into escrow prior to the first mover's choice. The escrow is forfeited if the second mover keeps the proceeds of investment for himself. Bracht and Feltovich (2008) report the following behavioral properties of their escrow procedure. First, a deterrent escrow — i.e., one that makes splitting more beneficial than keeping for the second mover — almost universally leads to the cooperative outcome. The effect of a mild escrow, in turn, is also positive, yet weaker. Second, these effects do not depend on whether escrow choices are made exogenously (i.e., randomly determined by the experimenter) or endogenously (i.e., voluntarily chosen by the subject). See also Tyran and Feld (2006) for related experimental evidence on introducing the cost of opportunistic behavior in either a "mild" or a "deterrent" way, this time in the context of sanctions for free riding in the public goods game.

Table 1: Cooperation, trust and trustworthiness

	Cooperation (%)	Trust (%)	Trustworthiness (%)
Baseline	16.0	60.7	26.4
Communication	22.0	56.0	39.3
Oath only	21.3	60.7	35.2
Oath	34.5	65.5	52.7

The decrease between round 1 and round 5 is significant with $p = .044$ (Wilcoxon-Mann-Whitney rank sum paired samples test).

Adding pre-play communication *per se* barely changes the outcomes. The overall rate of cooperation in the communication treatment is only 22% (which is not significantly higher than the baseline rate of 16%, $p = .357$, bootstrap proportion test).[9] Figure 2.b shows that cooperation also exhibits a similar decreasing pattern as in Baseline. As shown in Table 1, there is hardly any between-treatment variation in the aggregate trust rate, while the small increase in trustworthiness remains statistically insignificant ($p = .415$). These findings replicate existing experimental results (see, e.g., Bracht and Feltovich, 2009): simple communication neither fosters cooperation, nor can it serve as a remedy against its decay over time.

Result 1 *Communication alone has no significant effect on cooperation, trust and trustworthiness.*

The overall rate of cooperation in the Oath only treatment (which does not allow for pre-play communication), in turn, resembles the one observed in the two previous treatments (21.3%, which is not significantly higher than either the baseline, $p = .305$, or the plain communication treatment, $p = .940$). This similarity in outcomes between treatments results from comparable levels of trust and trustworthiness: as shown by the summary statistics in Table 1, the level of trust exhibited by player As is similar in Baseline and Oath only, and is coupled with a small and statistically insignificant increase in trustworthiness from player Bs ($p = .307$). This result echoes previous evidence showing that the oath has no effect in a coordination game with Pareto-rankable equilibria without pre-play communication (Jacquemet, Luchini, Shogren, and Zylbersztejn, 2018).

Result 2 *The oath alone has no significant effect on cooperation, trust or trustworthiness.*

[9] The game is repeated over five rounds and pairs of players are being rematched through the round-robin scheme. This implies that there might be correlation between individual observations. Our bootstrap procedure accounts for within cluster correlation by bootstrapping with replacement on clusters rather than individual observations. This testing procedure accounts for a within-cluster correlation of unknown form. In practice, the test is based on a standard bootstrap procedure with 9999 draws that yields an empirical bootstrap distribution of players' sets of choices. This is testing procedure that we generally adopt in the following. When we rely on another testing procedure, it will be indicated.

Figure 2: Rate of cooperation by round and treatment

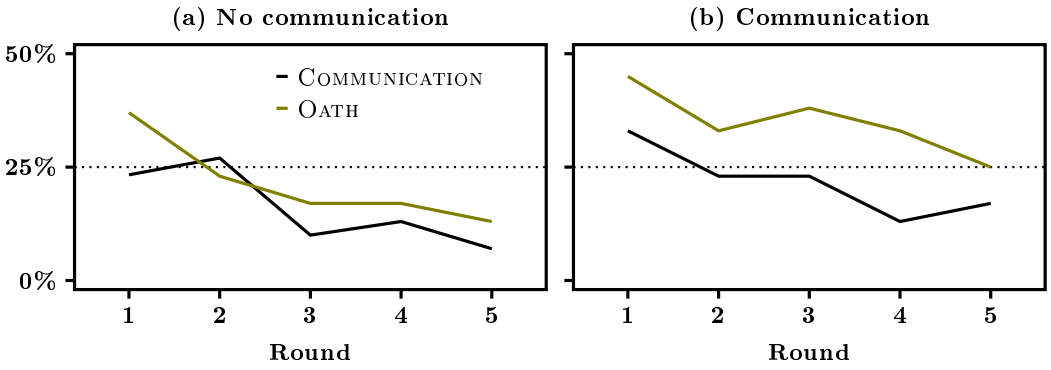


Table 2: Communication behavior, trust and trustworthiness

	Messages (%)		Outcomes (%)		
	"*I will split*"	"*I will keep*"	Cooperation	Trust	Trustworthiness
Baseline	–	–	16.0	60.7	26.4
			Conditional on "*I will split*"		
Communication	86.4	4.7	24.0	61.2	39.2
Oath	74.0	10.0	43.2	75.7	57.1

In contrast with these benchmarks, the oath coupled with pre-play communication increases the aggregate rate of cooperation, as shown in Figure 2.b. The oath induces a significant upward shift in cooperation to 34.5% (which is significantly different from both Baseline, $p = .004$, and, to a lesser extent, from Oath only, $p = .078$, and Communication, $p = .076$). Moreover, this upward shift is persistent across rounds: as compared to Baseline, the rate of cooperation is higher both in round 1 (21.7%, $p = .066$) and in round 5 (18.3%, $p = .043$). The oath falls short of mitigating the decay of cooperation over time—cooperation falls from 45.0% in round 1 to 25.0% in round 5. As compared to Baseline, the observed improvement in cooperation mainly comes from enhanced trustworthiness — the average level of trustworthiness is higher when communication and oath are combined ($p = .024$, bootstrap proportion test), whereas trust only marginally and not significantly changes ($p = .290$). The increase in trustworthiness falls short of achieving statistical significance when comparisons are made with that observed in either Oath only or Communication ($p = .381$ and $p = .372$).

Result 3 *The oath combined with communication increases cooperation.*

Higher cooperation when oath and communication are combined results from significant changes in the use of pre-play communication. Table 2 summarizes the communication behavior observed

in the two communication treatments, with and without an oath. First, we observe that the message "*I will split*" becomes significantly less frequent under oath (from 86.4% to 74.0%, $p = .036$).[10] Second, the content of the message does not change behavior when communication is implemented alone: as compared to BASELINE, cooperation, trust and trustworthiness are not different after receiving the message "*I will Split*" in COMMUNICATION ($p = .429$ for trustworthiness, $p = .945$ for trust and $p = .498$ for cooperation). Third, when communication occurs in the oath treatment, both trust and trustworthiness conditional on the message "*I will Split*" significantly increase as compared to BASELINE ($p = .063$ and $p = .018$). Cooperation is also significantly higher ($p = .011$). The effect of the oath on communication is however not large enough for the increase in trust and trustworthiness to be statistically significant when compared to COMMUNICATION ($p = .218$ and $= .288$). The same is true for cooperation ($p = .184$).

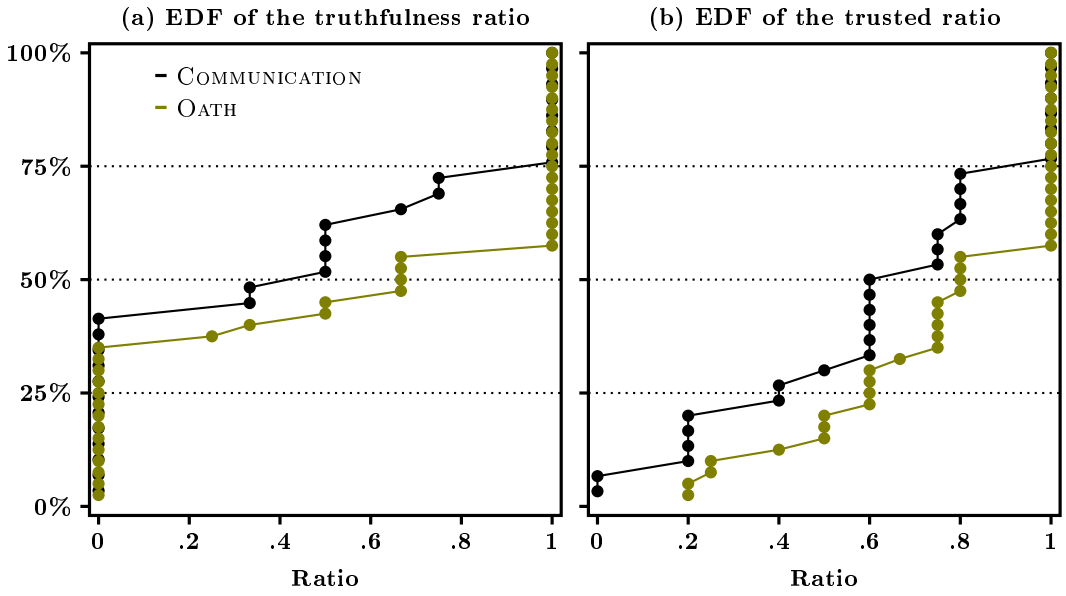
To further investigate the effect of the oath on communication, Figure 3 reorganizes the data at the individual level (each plot mark corresponds to a subject). For each treatment, we compute the truthfulness ratio, defined as the proportion of decisions from a trustee which are consistent with the message sent (given the message is informative, i.e., split after "*I will split*" and keep after "*I will keep*"), and the trusted ratio, defined as the proportion of decisions from a trustor that are consistent with the message received (i.e., invest after "*I will split*" and not invest after "*I will keep*"). Thus, the truthfulness ratio, presented in Figure 3.a, summarizes the informational content of the messages sent, while the trusted ratio, presented in Figure 3.b, measures the response to this information. Both ratios exhibit the same shift to the left when subjects are under oath. Informative messages sent by subjects who communicate under oath are more truthful, mainly due to an increase in the number of those subjects who always truthfully reveal their intentions: they represent 26.7% of player Bs (8 out of 30 subjects) in COMMUNICATION and 45% (18 out 40) in OATH ($p = .034$). Receivers also consider messages as more credible when they are sent under oath, in no small part due to an increase in the number of subjects who always follow the informative signal they received: they account for 26.7% of player As (8 out of 30) in COMMUNICATION and 45% (18 out of 40) in OATH ($p = .199$).

Result 4 *The positive impact of oath on cooperation in the communication condition is driven by two opposite effects. First, trustees are less likely to announce cooperation under oath. Second, messages announcing cooperation become more truthful and are trusted more often under oath.*

To sum up, improved cooperation under oath occurs thanks to more efficient communication: the trustees' messages are more credible, and also perceived as such by the trustors. Moreover, the truth-telling oath makes subjects more wary of announcing cooperation, suggesting that talk is no longer cheap under oath.

[10] Communication behavior exhibits no clear pattern across rounds, see Appendix D.

Figure 3: Empirical Distribution Functions of individual behavior by treatment



4 How does taking an oath compare to receiving fines?

We now turn to an explicit comparison between the truth-telling oath and incentivized communication. Table 3 presents aggregate statistics on cooperation, trust and trustworthiness for the fine treatments together with a reminder of the oath with communication treatment. The effect of mild fines (i.e., fines lower than or equal 3 EUR) echoes the effect of the oath. Each of the three mild fine treatments induces a significant increase in the rate of cooperation in comparison to Baseline, ($p = .023$ for $f = 1$, $p = .008$ for $f = 2$ and $p = .011$ for $f = 3$), and are not significantly different from Oath ($p = .978$ for $f = 1$, $p = .271$ for $f = 2$ and $p = .777$ for $f = 3$). Trust and trustworthiness are also very much alike in the mild fines to those observed in Oath. In mild fines treatments, trust ranges from 60.5% to 63.0% (with $p = .779$ for $f = 1$, $p = .973$ for $f = 2$ and $p = .868$ for $f = 3$ when compared to Oath) and trustworthiness ranges from 49.2%

Table 3: Cooperation, trust and trustworthiness with pre-play communication

	Cooperation (%)	Trust (%)	Trustworthiness (%)
Oath	34.5	65.5	52.7
Mild fine, $f = 1$	31.0	63.0	49.2
Mild fine, $f = 2$	37.5	62.0	60.5
Mild fine, $f = 3$	32.0	60.5	52.9
Heavy fine, $f = 4$	46.0	64.5	71.3

Figure 4: Rate of cooperation by round: Comparison between treatments with pre-play communication

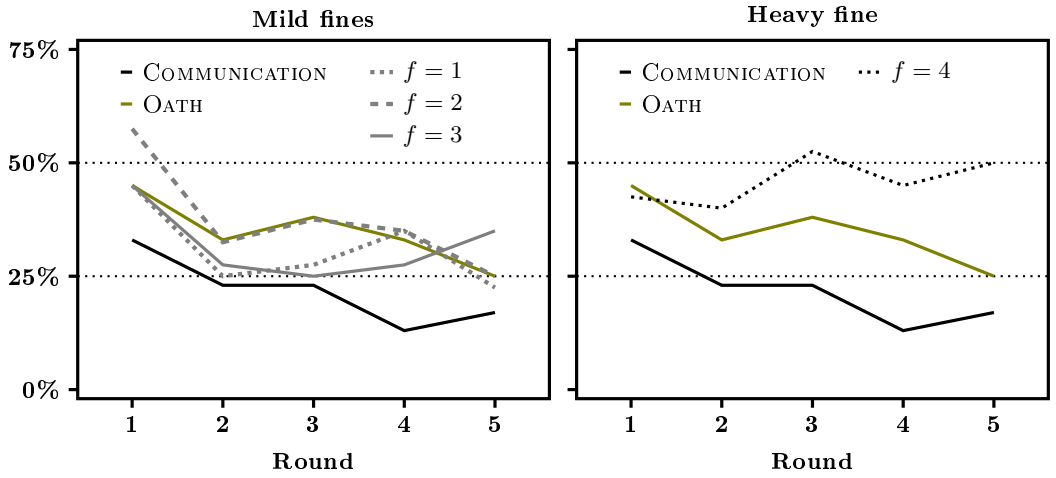


Table 4: Communication, trust and trustworthiness in mild and heavy fines treatments

	Messages (%)		Conditional on "*I will split*" (%)		
	"*I will split*"	"*I will keep*"	Cooperation	Trust	Trustworthiness
Oath	74.0	10.0	43.2	75.7	57.1
Mild fine, $f = 1$	75.0	10.5	40.0	74.7	53.6
Mild fine, $f = 2$	74.5	9.5	47.6	74.5	63.9
Mild fine, $f = 3$	73.0	9.0	43.2	68.5	63.0
Heavy fine, $f = 4$	65.5	14.0	67.9	84.0	80.9

to 60.5% ($p = .465$ for $f = 1$, $p = .747$ for $f = 2$ and $p = .653$ for $f = 3$).

The comparison of Oath with the heavy fine treatment leads to a different conclusion. An heavy fine induces a higher rate of cooperation than the oath ($p = .036$). Since the level of trust is similar in the two conditions, this difference in cooperation mainly arises as the result of a higher trustworthiness (although the difference is on the fence in purely statistical terms with $p = .105$). An important driving force of the observed aggregate difference between the heavy fine and the other treatments is related to the (lack of) decay of cooperation over time. Figure 4.a presents the dynamics of cooperation across rounds for the mild fines treatments together with Oath and Communication. Once again, it points to a decreasing trend in cooperation, similar to the one observed in Oath — although the trend seems to bend upwards in round 5 for $f = 3$. The effect of the heavy fine on the dynamics of cooperation presented in Figure 4.b, by contrast, is stronger: the heavy fine mitigates the decay of cooperation and even reverses it, thus clearly dominating the oath with that respect.

Table 4 provides evidence on the use and consequences of communication in the fine treatments.

First, echoing the previous results on OATH, fines are found to reduce the frequency of the message "*I will split*". Subjects send this message in 74.2% of the cases when they face a mild fine for lying, a rate comparable to the one observed in the oath treatment (74%). This negative effect of fines on the rate of message "*I will split*" is strongly significant when all fine treatments are pooled together and compared to COMMUNICATION ($p = .013$). This change in the use of communication is much stronger with heavy fines: the use of the message "*I will split*" drops from 74.2% under mild fines (combined data) to 65.5% ($p = .078$).

Second, we also observe that communication with fines induces a higher rate of messages "*I will keep*" as compared to COMMUNICATION ($p = .052$). The rate seems to increase with the fine, although the marginal effect of the heavy fine is not significant against the mild fines ($p = .136$ for heavy fines against mild fines combined). The effect of OATH on this rate is equivalent to that of mild fines. One possibility is that this shift results from a more honest signaling of non-cooperative intentions by trustees. However, given that investors show strong distrust towards the trustees who send this message (which means no investment is made and thus the trustee has no decision to make in the game), the data is too scarce to assess whether player Bs truly announce their non-cooperative intentions.

Third, both the oath and the mild fines have similar effects on behavior conditional on the message "*I will split*": cooperation, trust and trustworthiness all increase in the same manner as compared to the benchmark situation with communication only. A heavy fine has an even larger effect than both the oath and mild fines on cooperation ($p = .074$ and $p = .069$ respectively), but the effect on both trust and trustworthiness is not statistically significant when compared to subjects taking an oath ($p = .332$ and $p = .118$).

Figure 5 further substantiates the use of communication by comparing the EDF of the truthfulness and the trust ratios in each fine treatment to OATH. Figure 5.a shows that the effect of mild fines on the upper-end of the EDF of the truthfulness ratio is comparable to the one induced by the oath. The overall proportion of subjects always telling the truth in the mild fines treatments is 49.9% as compared to 45.0% in the OATH ($p = .947$). On the lower end, mild fines seem to mitigate lying behavior more efficiently than the oath: the proportion of subjects who lie at every occasion is 19.4% when facing mild fines and 35% under oath ($p = .021$). The deterrence of lying induced by the heavy fine is stronger on both the upper- and lower-end of the EDF distribution than with mild fines, and thus dominates the one induced by the oath. 65% of our subjects always tell the truth under a heavy fine ($p = .068$ when compared to the oath) and 7.5% always lie ($p = .003$ when compared to the oath). Figure 5.b presents the EDF of individual trust ratios by treatment and shows that the effect of the fines and the oath are very much alike. The proportion of subjects who always trust the message they receive is higher with the heavy fines (60%) than with the oath (45%) but the increase is not statistically significant ($p = .205$).

Result 5 *In our setting, the oath turns out to be behaviorally equivalent to a mild fine. Com-*

Figure 5: Empirical Distribution Functions of individual behavior by treatment

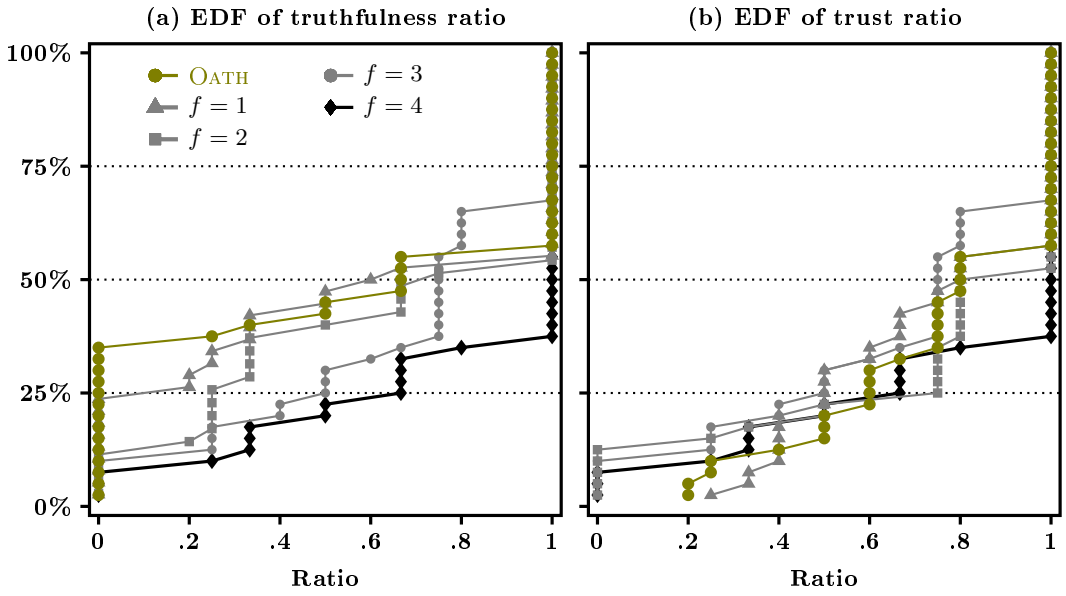
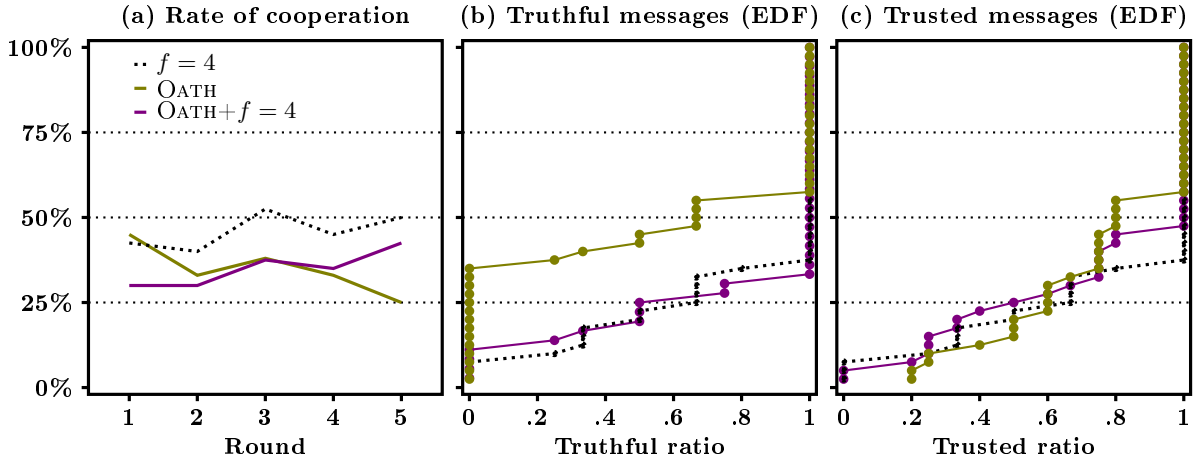


Figure 6: Rate of cooperation and Empirical Distribution Functions of individual behavior in the combined treatment



mitment to truthful communication induced by an oath is equivalent to a drastic decrease in the monetary benefits from lying, but remains weaker than a deterrent fine.

5 Deterrent fine and truth-telling oath combined

In this section, we investigate the behavioral consequences of combining monetary and non-monetary devices based on a new treatment in which a heavy fine ($f = 4$) is implemented after the

truth-telling oath procedure. That is, subjects are asked to take an oath, similar to the previous oath treatments, prior to participating to the heavy fine treatment.[11]

The magnitude of cooperation in this new treatment is 35.0%, a level comparable with the one observed in OATH (34.5%) but lower than that in the heavy fine alone treatment (46.0%); although the decrease is not statistically significant ($p = .193$). Figure 6.a shows that the evolution of cooperation is very similar in the heavy fine alone and the heavy fine combined with the oath treatment, but shifted down. The overall rate of trust is at 53.5%, not statistically different to the level observed in the heavy fine treatment (65.4%, $p = .138$) whereas trustworthiness is at 65.4% (71.3% in the heavy fine treatment, $p = .484$). These results indicate that the combination of the heavy fine and the truth-telling oath does not further improve cooperation, rather pointing to a possible setback of coupling these two commitment devices.

A closer examination of the structure of messages and the EDF of truth and trust ratios help understand the source of this outcome. First, the negative effect of the oath and the fines on the rate of messages "*I will split*" is further exaggerated: 56% as compared to 65.5% in the heavy fine treatment ($p = .139$) and 74.0% in OATH ($p = .014$). The positive effect on the rate of messages "*I will keep*" also increases: 24.5% as compared to 14.0% in the heavy fine alone treatment ($p = .077$) and to 10.0% in OATH ($p = .031$). Second, the examination of the EDF in Figures 6.a and 6.b indicates that combining devices has no impact on the truth and trust ratios. EDF are close to those observed in the heavy fine treatment. This is consistent with the observed rates of cooperation conditional on the message "*I will split*", which is close to that in the heavy fine treatment (61.7%, $p = .493$). This is also true for the level of trust (75.9%, $p = .574$) and the level of trustworthiness (81.2%, $p = .984$). All together, these results indicate that weakened cooperation when the heavy fine is combined with the oath occurs due to subjects being less willing to announce cooperation with no improvement in truthfulness of, and trust in, the messages.

Result 6 *The combination of the heavy fine with the oath leads to a decline in the use of cooperative messages with no improvement in truthfulness and trust.*

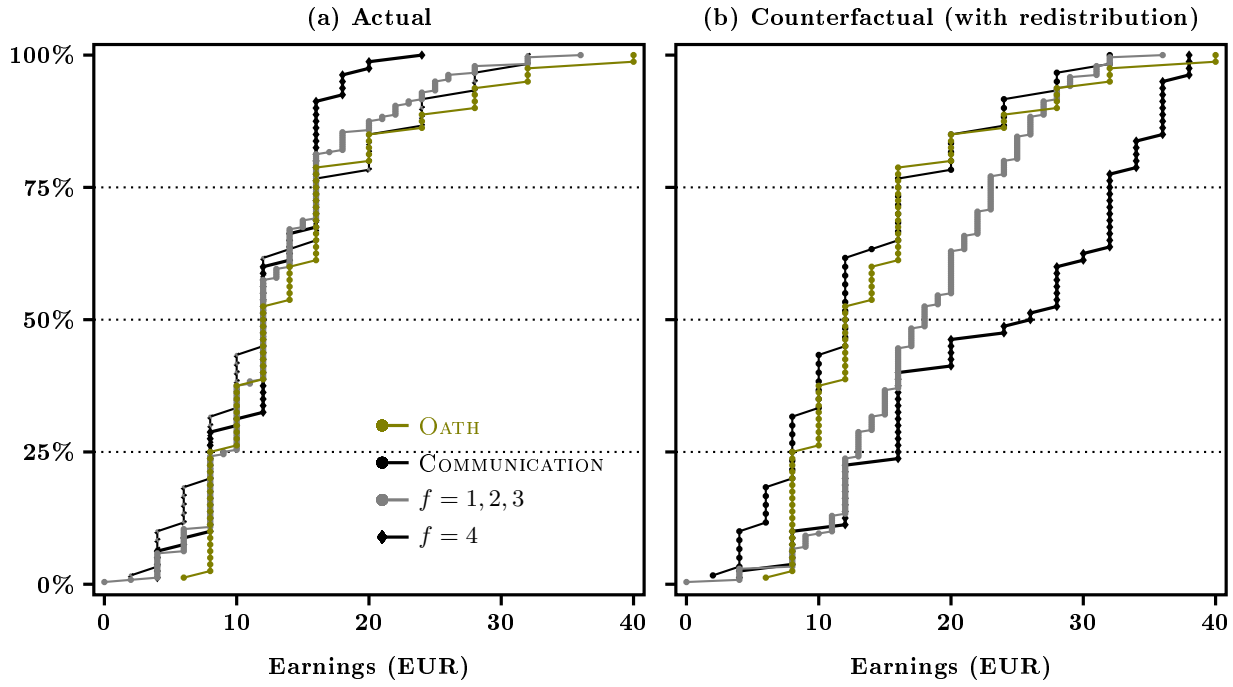
6 The economic value of the oath

Overall, the oath proves to be behaviorally equivalent to mild fines, while the heavy fine is more effective in restoring cooperation. Should we conclude that heavy fines are socially preferable to mild fines and non-monetary commitment devices such as the oath? To answer this question, we examine welfare outcomes in the experiment.

As a benchmark, we first compute the feasible monetary earnings of subjects in all conditions that include pre-play communication — i.e., how much subjects would earn would they get the

[11] For similar designs in non-strategic settings involving dishonesty in self-reporting, see Prima, Feeny, Hoffmann, and Satriawan (2020); Peer and Feldman (2021).

Figure 7: Empirical Distribution Functions of feasible earnings under communication



sum of their payoff over all rounds of play (see Appendix E for detailed statistics). The mean of feasible earnings in Oath is 14.8 EUR whereas it is 13.4 EUR in Communication. The increase is small in size, but significant ($p = .065$, based on a bootstrap mean test).[12] Both player As and player Bs benefit from this small increase in feasible earnings: the mean among player As raises from 8.8 EUR in Communication to 10.3 EUR in Oath, and from 18.0 EUR to 19.3 EUR among player Bs. Figure 7.a provides the EDF of feasible earnings by treatment. Subjects are almost always better off with the oath than without. A decomposition by roles shows that the oath is more likely to benefit player As at the low end of the distribution of feasible earnings (the mean feasible earnings among player As at the 25% bottom end of the feasible earnings distribution is 7.9 EUR with an oath and 4.8 EUR without), and player Bs at the upper-end of the distribution (the mean feasible earnings among player Bs at the 25% upper end of the feasible earnings distribution is 32.0 EUR with an oath and 29.6 EUR without).

This figure also clearly shows that welfare (as measured by feasible earnings) is not improved by the fines. This is particularly true for the heavy fine treatment that has a strong impact on the upper end of the feasible earnings distribution. On average, both the mild fine treatments taken together (13.2 EUR, $p = .017$) and the heavy fine treatment (13.6 EUR, $p = .063$) give rise to

[12] The test operates by bootstrapping on clusters in the same way as the bootstrap proportion test presented in Footnote 9.

lower feasible earnings than OATH ONLY. Fines seem to punish dishonest trustees on the upper end of the distribution without benefiting trustors on the low end of the distribution.

The fact that fines are lost is likely to matter for the functioning of the economy implemented in our lab experiment. To examine this issue, we conduct a simulation exercise whereby fines are redistributed (at no cost) to those trustors that were exploited by the trustees — hence excluding trustees from the redistribution We consider this as a conservative way to build our comparison since we hold constant both players' behavior in the game. Note that this assumption defines a lower bound of the actual social welfare in such circumstances given that the trustors would also have a higher willingness to invest in a counter-factual setting in which they would be compensated for the dishonesty of the trustees. Without benefiting from redistribution, the incentives faced by these trustees would remain in line with the original game.

Figure 7.b reports the results from this simulation. Re-injecting the fines into the economy increases the welfare achieved with fines. This is particularly clear for the heavy fine treatment. Here, mean feasible earnings are now 23.6 EUR, as compared to 18.2 EUR with the mild fine treatments (bunched together). Redistribution largely benefits the trustors: their mean feasible earnings is now 32.8 EUR, to be compared with 10.4 EUR in OATH when there is no redistribution, and 12.8 EUR in the heavy fine treatment.

In practice, fines and redistribution are however costly, leading to a marginal cost of public funds (MCF, the ratio between the change in net wealth induced by increasing a tax rate and the amount of spendings it permits) which is typically higher than 1.[13] As a metric to compare fines to our non-monetary commitment device, we compute the efficiency ratio that would equalize social welfare across devices with redistribution. The heavy fine obviously dominates all other conditions in this case, with a coefficient of 8. For the lowest level of the fine ($f = 1$), the coefficient is 0.93. It means the oath dominates what can be achieved under this level of the fine under all possible values of the ratio (unless imposing fines improves efficiency). The coefficients increase for higher levels of mild fines ($f = 2, 3$) to 2.53 and 3.06.

7 Conclusion

In a world of incomplete contracts, social norms matter to promote market exchange. The social norms of mutual beliefs in reciprocity and trust create economic value by facilitating mutually beneficial trade that could not be sustained otherwise. Creating trust among traders can be challenging especially when dealing with strangers and outsiders. Even with communication, people can be duplicitous. How can we commit people to tell the truth in mutually beneficial exchange? Using a mechanism that fosters a commitment to the truth becomes essential.

[13] The value of the MCF is highly heterogeneous across countries, sectors and even goods as it depends in particular on the elasticity of demand for the taxed product (see, e.g., Gauthier, 2013, for a discussion). As an example, Belan, Gauthier, and Laroque (2008) obtain an estimated value equal to 1.16 for the UK.

Herein we use the lab to explore how a timeless non-market mechanism — the solemn truth-telling oath — can create this commitment to the truth such that even cheap talk communication becomes a trustworthy signal of reciprocity. Using a classic trust game experimental design, our results corroborate the intuition of our ancestors: honesty oaths help promote fair economic exchange. First, communication under oath can create more trust and cooperative behavior. Second, the oath does not solve all challenges because it induces a selection effect — it makes people more wary of using communication as a signal. Third, although the overall net effect on cooperation is positive, the oath cannot reverse a general decay of cooperation over time. Fourth, we also compare the oath's performance to mild and deterrent fines for deception. Here we find that the oath is behaviorally equivalent to mild fines. While one may be tempted to strengthen the incentives further by combining the truth-telling oath with the deterrent fines, this may be counterproductive if we see crowding-out, i.e., monetary incentives and non-monetary incentives acting as substitutes, not complements. The extent to which this holds remains to be established in further work on the oath.

References

Belan, P., S. Gauthier, and G. Laroque (2008): "Optimal Grouping of Commodities for Indirect Taxation," *Journal of Public Economics*, 92(7), 1738–1750.

Berg, J., J. Dickhaut, and K. McCabe (1995): "Trust, Reciprocity, and Social History," *Games and Economic Behavior*, 10(1), 122–142.

Bliege Bird, R., E. Ready, and E. Power (2018): "The social significance of subtle signals," *Nature Human Behavior*, 2, 452–457.

Bohnet, I., B. Frey, and S. Huck (2001): "More Order with Less Law: On Contract Enforcement, Trust, and Crowding," *American Political Science Review*, 95(1), 131–144.

Bracht, J., and N. Feltovich (2008): "Efficiency in the trust game: an experimental study of precommitment," *International Journal of Game Theory*, 37(1), 39–72.

——— (2009): "Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game," *Journal of Public Economics*, 93(9-10), 1036–1044.

Buckholtz, J. W., and R. Marois (2012): "The Roots of Modern Justice: Cognitive and Neural Foundations of Social Norms and Their Enforcement," *Nature Neuroscience*, 15(5), 655–661.

Charness, G., and M. Dufwenberg (2006): "Promises and Partnership," *Econometrica*, 74(6), 1579–1601.

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).

Crawford, V. (1998): "A Survey of Experiments on Communication via Cheap Talk," *Journal of Economic Theory*, 78, 286–298.

Dasgupta, P. (2009): "Trust and Cooperation among Economic Agents," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1533), 3301–3309.

de Bruin, B. (2016): "Pledging Integrity: Oaths as Forms of Business Ethics Management," *Journal of Business Ethics*, 136, 23–42.

Duffy, J., H. Xie, and Y.-J. Lee (2013): "Social Norms, Information, and Trust among Strangers: Theory and Evidence," *Economic Theory*, 52(2), 669–708.

Dufwenberg, M., and G. Kirchsteiger (2004): "A Theory of Sequential Reciprocity," *Games and Economic Behavior*, 47(2), 268–298.

Elster, J. (1989): "Social Norms and Economic Theory," *The Journal of Economic Perspectives*, 3(4), 99–117.

Farrell, J., and M. Rabin (1996): "Cheap Talk," *Journal of Economic Perspectives*, 10(3), 103–118.

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.

Gauthier, S. (2013): "Optimal Tax Base with Administrative Fixed Costs," *International Tax and Public Finance*, 20(6), 961–973.

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

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. F. Shogren (2018): "Truth-Telling under oath," *Management Science*, 65(1), 426–438.

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

Joule, R.-V., and J.-L. Beauvois (2010): *La soumission librement consentie*. Presses Universitaires de France, Paris, 6th edn.

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.

Kiesler, C. (1971): *The psychology of commitment. Experiments liking behavior to belief*. Academic Press, New York.

Kiesler, C., and J. Sakumura (1966): "A test of a model for commitment," *Journal of Personality and Social Psychology*, 3(3), 349–353.

Knack, S., and P. Keefer (1997): "Does social capital have an economic payoff? A cross-country investigation," *Quarterly Journal of Economics*, 112(4), 1251–1288.

Peer, E., and Y. Feldman (2021): "Honesty pledges for the behaviorally-based regulation of dishonesty," *Journal of European Public Policy*, 28(5), 761–781.

Petrak, M. (2014): "Plato and Ulpian's Praecepta Iuris," *Fundamina: A Journal of Legal History*, 2014(si-2), 694–701.

Prima, R. A., S. Feeny, R. Hoffmann, and E. Satriawan (2020): "Threats, pledges, and asset misreporting: A framed field experiment in Indonesia," .

Prodi, P. (1992): *Il sacramento del potere: il giuramento politico nella storia costituzionale dell'Occidente*. Il Mulino, Bologna.

Putnam, R. D. (2000): *Bowling Alone: The Collapse and Revival of American Community*. Simon and schuster.

Putnam, R. D., R. Leonardi, and R. Y. Nanetti (1992): *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton university press.

Rousseau, D. M., S. B. Sitkin, R. S. Burt, and C. Camerer (1998): "Not so different after all: A cross-discipline view of trust," *Academy of management review*, 23(3), 393–404.

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

Tyran, J.-R., and L. P. Feld (2006): "Achieving Compliance when Legal Sanctions are Non-deterrent," *Scandinavian Journal of Economics*, 108(1), 135–156.

Weitzel, U., and M. Kirchler (2023): "The banker's oath and financial advice," *Journal of Banking & Finance*, 148, 106750.

Zeiliger, R. (2000): "A presentation of Regate, Internet based Software for Experimental Economics," *http://regate-ng.gate.cnrs.fr/ sferriol/*.

Zickfeld, J., S. T. Karg, S. S. Engen, A. S. R. Gonzalez, J. Michael, and P. Mitkidis (2022): "Committed Dishonesty: A Systematic Meta-Analysis of the Effect of Social Commitment on Dishonest Behavior," .

Appendix

A Experimental procedures

We run a total of 33 session in April 2014, March 2015 and June 2015. They all took place in the LEEP laboratory in Paris and were computerized using REGATE (Zeiliger, 2000). For our three control

treatments – baseline, baseline with oath, and simple communication – we run 3 sessions per treatment. For each of the six remaining treatments, we run 4 sessions.

At the beginning of each session, subjects fill in a short questionnaire on basic socio-demographic characteristics (age, gender, occupation, previous participation in experiments, etc.). Then, the experimental instructions are read aloud by the experimenter, followed by a short quiz of comprehension filled in by each subject. Finally, all the questions and answers to the quiz are read aloud by the experimenter. Once all the remaining questions (if any) are addressed, the experimental game starts.

Each session involves 20 subjects—10 participant As and 10 participant Bs—with fixed roles. Matching is anonymous and based on a round-robin scheme carried out over 5 periods within two separate subgroups of 10 players (5 player As and 5 player Bs) which yields two independent clusters of observations per session. Due to the between-subject nature of our design, no subject participated in more than one experimental session. The instructions state that the experiment consists of several rounds. The computer interface includes a short summary of all completed rounds – own choice and the resulting payoff, coupled with the message transmitted between players in treatments involving communication. Furthermore, subjects are informed that a single round will be drawn at random for payoff at the end of the experiment, and that there will be an additional show-up fee of 5 EUR.

Our 660 subjects (286 males) were recruited using ORSEE (Greiner, 2015). Participants' average age was 27. Most of them (65%) were students, and 74% participated in an experiment before. No subject participated in more than one experimental session. Sessions lasted between 30 and 60 minutes, with an average payoff of 7.60 EUR.

B Instructions

GENERIC INSTRUCTIONS INDICATING TREATMENT-SPECIFIC PARTS

You are about to take part in an experiment in which you can earn money. The amount of your gains will depend on your decisions, as well as on decisions made by other participants.

Before starting, we would like to ask you to answer a few standard questions (concerning your age, education, profession, ...) which will help us to get to know you better. **This information, as well as the amount of your gains from this experiment, will remain strictly confidential and anonymous.**

Please, fill in the questionnaire using the interface on your computer screen, which is divided into three parts:

- In the *top* section, you will find information that might help you in making decisions.
- In the *middle* section, you will submit your decisions by clicking on a relevant button.
- In the *bottom* section, you will see all your decisions and gains from previous rounds of the experiment.

Thank you.

The experiment

The experiments consists of several identical rounds. In each round, participants are divided by groups of two. Each pair consists of one participant A and one participant B. You will be randomly assigned to your role — participant A or participant B — at the beginning of the experiment, and retain it throughout the experimental session. A message on your computer screen will inform you about your role. **Your role will remain unchanged throughout the entire experiment.**

What happens in each round

At the beginning of each round, participants are be matched into pairs: if your are participant A, then a participant B is randomly selected to your complete pair; analogously, if your are participant B, then a participant A is randomly selected to complete your pair. Your pair will **change after each round**, and two participants in opposite roles **may interact at most once during the experiment.**

(ONLY IN TREATMENTS WITH COMMUNICATION) In each round of the experiment, participant B can send a message to participant A before the latter makes a decision. If she wishes to do so, participant B may announce the decision he will make in that round.

(ONLY IN TREATMENTS WITHOUT COMMUNICATION) Each round consists of **4 stages**. *(ONLY IN TREATMENTS WITH COMMUNICATION)* Each round consists of **6 stages**.

Stage 1. At the beginning, a participant is randomly matched to your group.

Stage 2. *(ONLY IN TREATMENTS WITH COMMUNICATION)* participant B is asked **to send a message to participant A** by choosing one of the options displayed on his computer screen and submitting it by clicking 'OK'. **This message does not affect neither participants' earnings.**

Stage 3. *(ONLY IN TREATMENTS WITH COMMUNICATION)* participant A reads the message from participant B and then clicks 'OK' in order to proceed to the next stage.

Stage 4. participant A chooses between L and R by clicking on a relevant button on his computer screen.

Stage 5. If participant A chose L in the previous stage, then participant B has no decision to make.
If participant A chose R in the previous stage, then participant B chooses between l and r by clicking on a relevant button on his computer screen.

Stage 6. End of the round and each participant is informed about his earnings:

- If **participant A chose** L, then:
  - participant A earns 2 € in this round;
  - participant B earns 0 € in this round;
- If **participant A chose** R and **participant B chose** l:
  - participant A earns 0 € in this round;
  - participant B earns 8 € in this round;
- If **participant A chose** R and **participant B chose** r:
  - participant A earns 4 € in this round;
  - participant B 4 € in this round;

(ONLY IN TREATMENTS WITH MONETARY FINE) Moreover, **had the participant A chosen R, if the decision made by the participant B does not match the decision announced in his message, then the amount of [monetary value of fine] € shall be subtracted from his gains in a given period.** This procedure does apply to a participant B who did not announce a decision in his message.

At the end of each round, a message on your computer screen will inform you that either a new round is about to start, or that the experiment ends.

Payment of your earnings

At the end of the experiment, **one round is picked at random**. Each participant receives a sum in EUR corresponding to the amount he earned in this round, plus a bonus of 5 € for completing the experiment. Payments are made individually and in cash.

For obvious reasons, **you are not allowed to talk during the experiment**. Participants who violate this rule will be excluded from the experiment and all payments. It is crucial that you understand perfectly the rules of this experiment. Should you have any questions to ask, please raise your hand.

Thank you for your participation.

C Oath form used in the experiment

PARIS SCHOOL OF ECONOMICS
ÉCOLE D'ÉCONOMIE DE PARIS

SOLEMN OATH

Topic: "JZ"; Research number 1842A

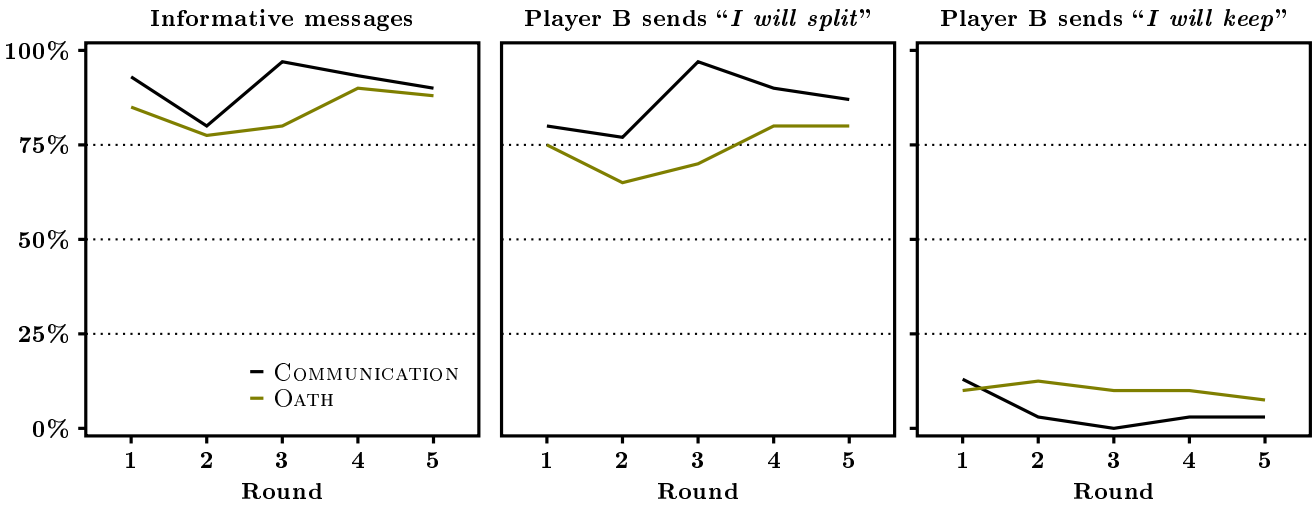
I undersigned ..... swear upon my honour that, during the whole experiment, I will:

Tell the truth and always provide honest answers.

Paris, ... Signature.....

Paris School of Economics, 48 Boulevard Jourdan 75014 Paris - France.

D Communication behavior by round and treatment



E Earnings: Summary statistics

Player As' earnings
(accumulated over 5 rounds)

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Communication	2.0	6.0	10.0	8.8	12.0	16.0
Oath	6.0	8.0	10.0	10.35	12.0	16.0
Mild fine, $f = 1$	4.0	7.5	10.0	9.9	12.0	18.0
Mild fine, $f = 2$	4.0	10.0	10.0	11.3	14.0	18.0
Mild fine, $f = 3$	2.0	8.0	10.0	10.4	12.5	16.0
Heavy fine, $f = 4$	6.0	11.5	12.0	12.8	16.0	18.0

Player Bs' earnings
(accumulated over 5 rounds)

Communication	4.0	12.0	16.0	18.0	24.0	32.0
Oath	8.0	16.0	16.0	19.3	24.0	40.0
Mild fine, $f = 1$	0.0	11.8	15.5	15.9	21.8	32.0
Mild fine, $f = 2$	0.0	12.0	15.0	13.6	18.0	28.0
Mild fine, $f = 3$	0.0	8.0	12.0	11.9	15.0	25.0
Heavy fine, $f = 4$	4.0	8.0	12.0	11.8	16.0	24.0