



HAL
open science

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

Nicolas Jacquemet, Stéphane Luchini, Jason Shogren, Adam Zylbersztejn

► **To cite this version:**

Nicolas Jacquemet, Stéphane Luchini, Jason Shogren, Adam Zylbersztejn. Commitment to the truth creates trust in market exchange: Experimental evidence. *Games and Economic Behavior*, In press, 10.1016/j.geb.2024.09.011 . halshs-04722343

HAL Id: halshs-04722343

<https://shs.hal.science/halshs-04722343v1>

Submitted on 4 Oct 2024

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.

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

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

October 4, 2024

Abstract

Under incomplete contracts, the mutual belief in reciprocity facilitates how traders create value through economic exchange. Creating such beliefs among strangers can be challenging even when they are allowed to communicate, because communication is cheap. In this paper, we first extend the literature showing that a truth-telling oath increases honesty to a sequential trust game with pre-play, fixed-form, and cheap-talk communication. Our results confirm that the oath creates more trust and cooperative behavior thanks to an improvement in communication; but we also show that the oath induces selection into communication — it makes people more wary of using communication, precisely because communication speaks louder under oath. We next designed additional treatments featuring mild and deterrent fines for deception to measure the monetary equivalent of the non-monetary incentives implemented by a truth-telling oath. We find that the oath is behaviorally equivalent to mild fines. The deterrent fine induces the highest level of cooperation. Altogether, these results confirm that allowing for interactions under oath within a trust game with communication creates significantly more economic value than the identical exchange institutions without the oath.

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

*Published in Games and Economics Behavior. Revised version of GATE WP n°2311. We wish to thank Sorravich Kingsuwankul, Rupert Sausgruber, Eli Spiegelmann, Jean-Robert Tyran, Jeroen van de Ven, Marie Claire Villeval, as well as several workshop and conference participants for helpful comments. We gratefully acknowledge the associate editor and the reviewers for their 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 U.) and from the Foundation Aix-Marseille U. (research program “Mutual trust and commitment”). We acknowledge financial support from both the French National Research Agency, through the LABEX CORTEX (ANR-11-LABX-0042, which is part of the program Investissements d’Avenir ANR-10-LABX-93-0), the EUR-OSE (ANR-17-EURE-0001 and ANR-17-EURE-0020), and the Excellence Initiative of Aix-Marseille U.-A*MIDEX. Adam Zylbersztejn acknowledges VISTULA Fellowship.

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

[‡]Aix Marseille Univ., CNRS, AMSE, Marseille, France.stephane.luchini@univ-amu.fr

[§]Department of Economics, U. Wyoming. JRamses@uwoyo.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.

“[...] 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 et al., 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, and cooperative behavior (see, e.g., Charness and Dufwenberg, 2006). Honesty oaths have played an important role in the development of commerce and economic exchanges in Western societies (see, e.g., Silver, 1995). In modern economic exchange, oaths are 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; Weitzel and Kirchler, 2023). Recent literature in both economics and psychology points to a particular channel through

¹Trust 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 Feltoich, 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 et al., 1992; Putnam, 2000). In an early contribution, Putnam et al. (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. In the same vein, Algan and Cahuc (2014) conclude in their survey article that (i) trust has a causal impact on economic development, through its channels of influence on the financial, product, and labor markets, and with a direct effect on total factor productivity and organization of firms; and that (ii) trust and institutions strongly interact, with causality running in both directions. Given the major role it plays in human interactions and organizations, trust constitutes a prominent research topic spanning multiple disciplines of social sciences (Rousseau et al., 1998).

which honesty oaths affect economic outcomes: taking a truth-telling oath creates commitment towards being more honest (Jacquemet et al., 2018a; Zickfeld et al., 2023, 2024), which in turn makes communication more efficient (see, e.g. Jacquemet et al., 2018b, for an application to a coordination game).

In this paper, we investigate whether the solemn truth-telling oath can be used to create more trust in a reciprocal exchange through commitment to truthful communication. We use the classic trust game (Berg et al., 1995) in which communication between the trustee and the trustor is considered as one the common tools to create trust within the pair (see, e.g., Bracht and Feltovich, 2009). From a standard theoretical point of view, however, communication is nothing but cheap talk without credibility unless it manages to change the underlying incentives to behave opportunistically (see, e.g., Farrell and Rabin, 1996; Bliege Bird et al., 2018; Crawford, 1998, for a review of cheap talk in games). In this case, 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.

Our objective is twofold. First, 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 — hence extending the existing oath literature to a trust game in which aligning beliefs with actual behavior is key to achieve efficiency.² To measure the causal effect of the oath on cooperation through enhancing the credibility of communication, we carry out the oath procedure put forward by Jacquemet et al. (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 intrinsic commitment to honesty.

Second, we use these changes in behavior under oath as a benchmark to measure the economic value of the oath in exchange institutions. To that end, we introduce monetary fines: if the trustee falsely communicates her intentions, whether it is to exploit or to cooperate with the trustor, she pays a monetary fine. Following Bohnet et al. (2001), the level of the fine is varied in a between-subject design: the mild fine reflects weak enforcement, while the deterrent fine wipes out all the monetary gains of deceptive behavior. Contrasting communication and cooperation behavior

²The truth-telling oath procedure we consider has been carried out in several settings giving rise to dishonesty like the sender-receiver game (Jacquemet et al., 2021b), a dice experiment (Beck et al., 2020), or a coin flipping task (Jacquemet et al., 2021a). All experiments support an economically substantial and long-lasting (Peer and Feldman, 2021) increase in the likelihood of truthful reporting of information. Several studies confirm that this change in behavior extends to, e.g., self-reported income in a tax evasion experiment both in the lab (Jacquemet et al., 2020) and in the field (Koessler et al., 2019), as well as to self-reported happiness (Carlsson and Kataria, 2018).

between fine and oath treatments hence provides a measure of the monetary equivalent of the non-monetary commitment achieved by the oath.

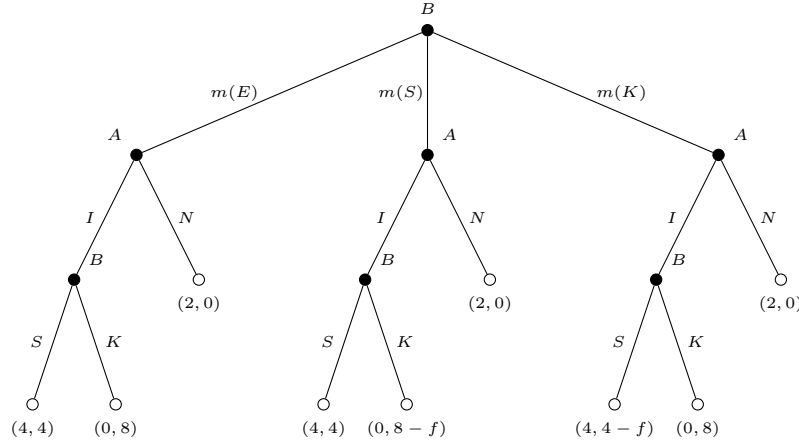
Our results on trust game played under oath are unambiguous: communication under oath helps create more trust and cooperative behavior. We first show that this improvement in efficiency is solely due to improved pre-play communication rather than to an effect of the truth-telling oath *per se* on cooperative behavior in the game: the oath has no behavioral effect in the absence of pre-play communication. Second, our results confirm that being exposed to the oath leads to a drastic increase in the willingness to tell the truth. Since words speak louder under oath, communication becomes a powerful tool to coordinate actions on welfare-improving outcomes. Third, this improvement arises from two countervailing effects: *(i)* the oath promotes truth-telling — subjects under oath keep their word significantly more than without the oath; but *(ii)* the oath also induces a selection effect: it makes people more wary of using communication altogether.

Turning to the monetary equivalent of the oath, our results show that the oath is behaviorally equivalent to a mild fine, but is dominated by a deterrent fine: while both types of fines shift up cooperation, only deterrent fine prevents it from eroding over time. Last, we 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 indicates 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 deterrent fine treatment. When compensation through redistribution is not feasible, the oath can still compare to the deterrent fine in terms of social welfare. This analysis confirms that allowing for economic exchange under oath creates significantly more value than an identical exchange institution without the oath. Altogether, this result documents that the oath – an enforcement mechanism employed in a variety of forms and contexts for thousands of years – indeed serves a valuable purpose to traders engaging in economic transactions under incomplete contracts.

2 Design of the experiment

Figure 1 presents the two-player trust game implemented in our experiment. 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

Figure 1: Trust game with prior communication and dishonesty fines



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). A risk-neutral and own-payoff-maximizing player A only invests if his subjective probability of facing a player B who decides to split is higher than 0.5.³ Before player A's decision to invest or not, player B can try to manipulate this belief by sending 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 €. ⁴ The game is repeated 5 times with constant roles, and pairs are randomly rematched (see Section 2.3 for further details).

³The investment multiplier of 4 has become standard in the experimental literature since the seminal study by Berg et al. (1995).

⁴In the written instructions (see Appendix A 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 whose aim is to turn 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 et al., 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 et al. (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 B); 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.

We investigate the effect of the oath on cooperation in the trust game by carrying out a between-subject setting that combines cheap talk communication ($f = 0$) and the truth-telling

⁵In the previous example describing 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).

oath according to a 2×2 factorial design with 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, 3 sessions, 60 subjects). This last treatment aims at investigating whether the behavioral effect of the oath is triggered by a change in communication, as we hypothesize, or whether the oath directly affects cooperation behavior.

The main treatment of interest combines cheap talk communication with an oath (OATH & COMMUNICATION, 4 sessions, 80 subjects). We observe a high oath acceptance rate in all treatments that include this treatment variable (97.3%).⁶ 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.

2.2 Fine equivalents of oath-based commitment

To measure the monetary equivalent of the commitment to truth-telling induced by the oath, we complement this 2×2 design with four treatments that introduce a positive dishonesty fine f .⁷ 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 $[f]$ € 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

⁶This rate includes all the oath-based conditions reported in the paper: OATH, OATH & COMMUNICATION, as well as OATH & COM. & $f = 4$ presented below, that combines the oath with a deterrent monetary fine.

⁷Our 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 to herself. 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: Overview of the experimental design

	Communication	Oath	f	Sessions	N	Cooperation (%)	Trust (%)	Trustworthiness (%)
BASELINE	No	No	0	3	60	16.0	60.7	26.4
COMMUNICATION	Yes	No	0	3	60	22.0	56.0	39.3
OATH	No	Yes	0	3	60	21.3	60.7	35.2
OATH & COMMUNICATION	Yes	Yes	0	4	80	34.5	65.5	52.7
Mild fine, $f = 1$	Yes	No	1	4	80	31.0	63.0	49.2
Mild fine, $f = 2$	Yes	No	2	4	80	37.5	62.0	60.5
Mild fine, $f = 3$	Yes	No	3	4	80	32.0	60.5	52.9
Deterrent fine, $f = 4$	Yes	No	4	4	80	46.0	64.5	71.3
OATH & COM. & $f = 4$	Yes	Yes	4	4	80	35.0	53.5	65.4

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 deterrent fine treatment. These four additional treatments are implemented between subjects (4 sessions, 80 subjects per treatment). In addition, our data also include four sessions (with a total of 80 subjects) in which the OATH & COMMUNICATION condition is combined with a deterrent fine (described in Section 5 below).

2.3 Experimental procedures

We run a total of 33 session in April 2014, March 2015 and June 2015. They all took place in the LEEP (Paris) and were computerized using REGATE (Zeiliger, 2000). Table 1 provides an overview of the experimental design, along with sample sizes and aggregate outcomes.

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 participants in the role of player A and 10 participant in the role of player B — 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.⁸ No subject participated in more than one experimental session. The instructions make clear that each pair of players interacts only once to prevent reputational concerns, and subjects are only informed that the experiment consists of multiple rounds (so as to rule out explicit end-game effects). The computer interface includes

⁸Appendices C and D provide a matching-group-level summary of the main outcomes of interest.

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: player A’s (player B’s) decision at each decision node is associated to a €2 (€4) variation in their own payoff for treatments without fines, which implies a €0.4 (€0.8) expected variation in their final earnings (taking the number of repetitions as given). Subjects additionally earn a show-up fee of €5. Subjects (660 in total, 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 €7.6.

3 Honesty, trust and trustworthiness under oath

The top part of Table 1 provides a summary of the key aggregate outcomes in the four treatments combining communication with the oath: 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. The top part of Table 2 disaggregates the cooperation data at the round level to highlight time trend differences across treatments. In BASELINE, cooperation starts at 23.3% in round 1 and ends up at 6.7% in round 5.⁹ The decrease between round 1 and round 5 is significant with $p = .044$ (one-sided Wilcoxon-Mann-Whitney ranksum paired-samples test).

Adding pre-play communication *per se* barely changes the outcomes. The overall rate of cooperation in COMMUNICATION is only 22% (which is not significantly higher than the baseline rate of 16%, $p = .357$, one-sided bootstrap proportion test).¹⁰ Table 2 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 increase in trustworthiness remains statistically insignificant ($p = .188$, one-sided bootstrap proportion test). 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.

⁹The round-level data on trust and trustworthiness are provided in Appendix E.

¹⁰The 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 within each matching group, but not between matching groups that can be assumed independent. We thus run all statistical tests at the matching-group-level (unless stated otherwise). Our bootstrap procedure accounts for within-cluster correlation of unknown form by bootstrapping with replacement on clusters rather than individual observations. 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. Unless stated otherwise, this procedure is used for all statistical tests reported in the paper.

Table 2: Cooperation rates by round (%)

Treatment	Round				
	1	2	3	4	5
BASELINE	23.3	26.7	10.0	13.3	6.7
COMMUNICATION	33.3	23.3	23.3	13.3	16.7
OATH	36.7	23.3	16.7	16.7	13.3
OATH & COMMUNICATION	45.0	32.5	37.5	32.5	25.0
Mild fine, $f = 1$	45.0	25.0	27.5	35.0	22.5
Mild fine, $f = 2$	57.5	32.5	37.5	35.0	25.0
Mild fine, $f = 3$	45.0	27.5	25.0	27.5	35.0
Deterrent fine, $f = 4$	42.5	40.0	52.5	45.0	50.0
OATH & COM. & $f = 4$	30.0	30.0	37.5	35.0	42.5

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

In contrast with these benchmarks, the oath coupled with pre-play communication increases the aggregate rate of cooperation (see Table 1). 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, $p = .078$, and COMMUNICATION, $p = .076$, one-sided bootstrap proportion tests). Moreover, this upward shift is persistent across rounds (see Table 2): 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 is mainly driven by enhanced trustworthiness — the average level of trustworthiness is higher when communication and oath are combined ($p = .003$, one-sided bootstrap proportion test), whereas trust only changes marginally and insignificantly ($p = .290$). The increase in trustworthiness falls short of achieving statistical significance when comparisons are made with that observed in COMMUNICATION ($p = .127$, one-sided bootstrap test) but is significant when compared to OATH ($p = .032$, one-sided bootstrap test).

The outcomes observed in OATH (which does not allow for pre-play communication) confirm that the oath only impacts behavior through communication: the level of cooperation resembles the one observed in the two previous treatments (21.3%, see Table 1; which is not significantly higher than either BASELINE, $p = .305$, or COMMUNICATION, $p = .940$, one-sided bootstrap proportion tests). 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, and is coupled with a small and statistically insignificant increase in trustworthiness from player Bs ($p = .202$, one-sided bootstrap proportion tests). This result echoes previous evidence showing that the oath has no effect in a coordination game with Pareto-rankable equilibria in the absence of pre-play communication (Jacquemet et al.,

Table 3: Communication behavior, trust, and trustworthiness

	Messages (%)		Outcomes (%)		
	“ <i>I will split</i> ”	“ <i>I will keep</i> ”	Cooperation	Trust	Trustworthiness
BASELINE	–	–	16.0	60.7	26.4
	Conditional on “ <i>I will split</i> ”				
COMMUNICATION	86.4	4.7	24.0	61.2	39.2
OATH & COMMUNICATION	74.0	10.0	43.2	75.7	57.1

2018b).

Result 2 *The oath increases cooperation when combined with communication.*

Higher cooperation when oath and communication are combined results from significant changes in the use of pre-play communication. Table 3 summarizes the communication behavior observed in the two communication treatments, with and without an oath.¹¹ This comparison highlights two countervailing effects of the oath on communication. First, consistent with the hypothesis that words speak louder under oath, senders under oath are more reluctant to signal their willingness to split the resources: the message “*I will split*” becomes significantly less frequent under oath (from 86.4% to 74.0%, $p = .036$, one-sided bootstrap proportion tests).

Second, receivers respond to the content of the message only when it has been delivered under oath. As expected, 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 ($p = .211$ for trustworthiness, $p = .946$ for trust and $p = .198$ for cooperation, one-sided bootstrap proportion tests). This confirms that communication alone turns out to be too cheap to be interpreted as an informative signal of intended behavior. This stands in sharp contrast with the consequences of the message when delivered under oath. When communication occurs in the oath treatment, both trust and trustworthiness conditional on the message “*I will split*” significantly increase as compared to BASELINE ($p = .023$ and $p = .002$, one-sided bootstrap proportion tests). Cooperation is also significantly higher ($p = .002$, one-sided bootstrap proportion tests).

To further investigate the effect of the oath on communication, Figure 2 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 2a, summarizes the informational content

¹¹Communication behavior exhibits no clear pattern across rounds, see Appendix F.

of the messages sent, while the trusted ratio, presented in Figure 2b, measures the response to this information. Distributions of both ratios exhibit the same shift to the right 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 of 40) in OATH & COMMUNICATION ($p = .034$, one-sided bootstrap proportion test). 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 & COMMUNICATION ($p = .199$, one-sided bootstrap proportion test).

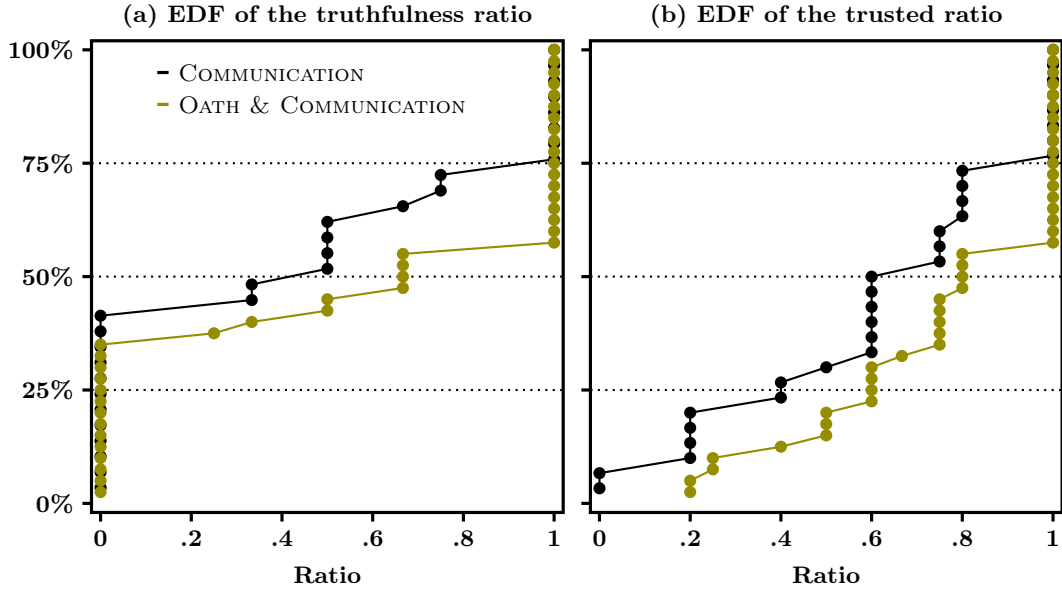
To sum up, improved cooperation under oath occurs thanks to more efficient communication: the trustees' messages are more credible, and also perceived as such by the trustors. At the same time, the truth-telling oath makes subjects more wary of announcing cooperation, suggesting that talk is no longer cheap under oath. Given this smaller share of subjects who do make use of communication, the effect of the oath on communication is however not large enough for the increase in trust and trustworthiness to be clearly statistically significant when compared to COMMUNICATION ($p = .072$ and $= .084$, one-sided bootstrap proportion tests). However, these improvements are insufficient to induce a statistically significant increase in cooperation after the message “*I will split*” compared to COMMUNICATION ($p = .056$, one-sided bootstrap proportion test).

Result 3 *The positive impact of communication under oath on cooperation 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.*

4 The monetary equivalent of non-monetary commitment

We now turn to an explicit comparison between the truth-telling oath and incentivized communication. The bottom part of Table 1 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 to €3) 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 $f = 3$, one-sided bootstrap proportion tests), and are not significantly different from OATH & COMMUNICATION ($p = .978$ for $f = 1$, $p = .844$ for $f = 2$ and $p = .777$ for $f = 3$, two-sided bootstrap proportion tests). Trust and trustworthiness are also very much alike in the mild fines to those observed in OATH & COMMUNICATION. In mild fine treatments, trust ranges from 60.5% to

Figure 2: Empirical Distribution Functions of individual behavior by treatment



63.0% (with $p = .779$ for $f = 1$, $p = .973$ for $f = 2$ and $p = .868$ for $f = 3$ when compared to OATH & COMMUNICATION, two-sided bootstrap proportion tests) and trustworthiness ranges from 49.2% to 60.5% ($p = .724$ for $f = 1$, $p = .419$ for $f = 2$ and $p = .956$ for $f = 3$, two-sided bootstrap proportion tests).

The comparison of OATH & COMMUNICATION with the deterrent fine treatment leads to a different conclusion. A deterrent fine induces a higher rate of cooperation than the oath ($p = .036$, one-sided bootstrap proportion test). Since the level of trust is similar in the two conditions, this difference in cooperation is mainly driven by higher trustworthiness (although the difference is on the fence in purely statistical terms with $p = .105$, one-sided bootstrap proportion test). An important driving force of the observed aggregate difference between the deterrent fine and the other treatments is related to the (lack of) decay of cooperation over time. The bottom part of Table 2 presents the dynamics of cooperation across rounds for the different fine treatments. Under mild fines, we once again observe a decreasing trend in cooperation which is similar to the one observed in OATH & COMMUNICATION — although the trend seems to bend upwards in round 5 for $f = 3$. The effect of the deterrent fine on the dynamics of cooperation, by contrast, is stronger: the deterrent 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 & COMMUNICATION, 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%).

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

	Messages (%)		Conditional on “ <i>I will split</i> ” (%)		
	“ <i>I will split</i> ”	“ <i>I will keep</i> ”	Cooperation	Trust	Trustworthiness
COMMUNICATION	86.4	4.7	24.0	61.2	39.2
OATH & COMMUNICATION	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
Deterrent fine, $f = 4$	65.5	14.0	67.9	84.0	80.9

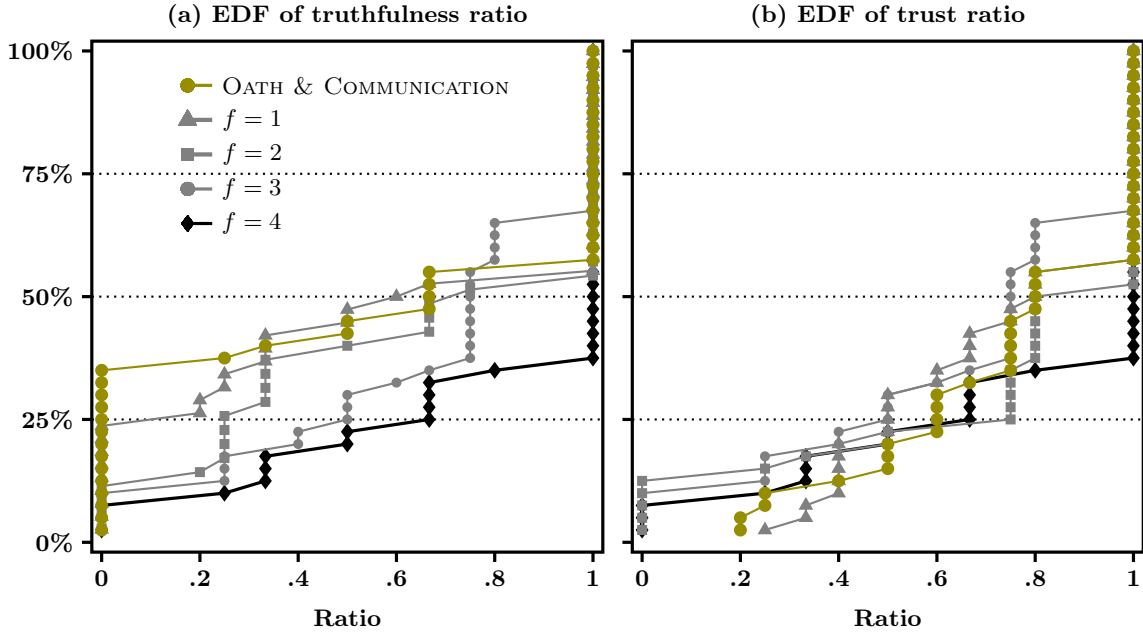
This negative effect of fines on the rate of message “*I will split*” is strongly significant when all fine treatments are pooled and compared to COMMUNICATION ($p = .013$, one-sided bootstrap proportion test). This change in the use of communication is even stronger with a deterrent fine: the use of the message “*I will split*” drops from 74.2% under mild fines (combined data) to 65.5% ($p = .078$, one-sided bootstrap proportion test).

Second, we also observe that communication with fines induces a higher rate of messages “*I will keep*” as compared to COMMUNICATION ($p = .052$, one-sided bootstrap proportion test). The rate jumps to 14.0% when communication is combined with a deterrent fine, but this raise is not statistically significant against the mild fines ($p = .137$ for deterrent fines against mild fines combined). The effect of OATH & COMMUNICATION 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 to a similar extent as compared to the benchmark situation with communication only. A deterrent fine has an even larger effect than both the oath and mild fines on both cooperation ($p = .007$ and $p = .002$ respectively, one-sided bootstrap proportion tests) and trustworthiness ($p = .005$ and $p = .006$ respectively, one-sided bootstrap proportion tests). While the increase in trust is significant as compared to mild fine treatments ($p = .079$, one-sided bootstrap proportion tests), it is not statistically significant when compared to subjects taking an oath ($p = .141$, one-sided bootstrap proportion tests).

Figure 3 further substantiates the use of communication by comparing the EDF of the truthfulness and the trust ratios in each fine treatment to OATH & COMMUNICATION. Figure 3a 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 fine treatments is 49.9% as compared to 45.0% in the OATH & COMMUNICATION ($p = .112$,

Figure 3: Empirical Distribution Functions of individual behavior by treatment



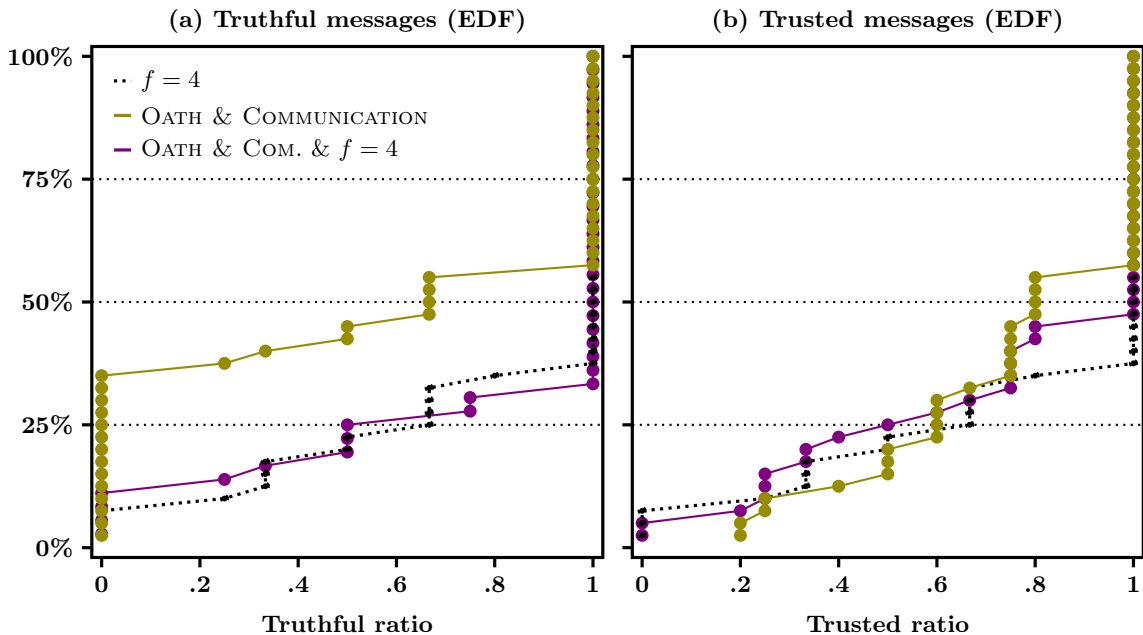
two-sided bootstrap proportion test). On the lower end, mild fines seem to mitigate lying behavior more efficiently than the oath: the proportion of subjects who lie at every occasion is 19.6% when facing mild fines and 35.0% under oath ($p = .021$, one sided bootstrap test). The effect of deterrent fine is present on both the upper- and lower-end of the EDF distribution than with mild fines, and thus dominates the one induced by the oath. 65.0% of our subjects always tell the truth under a deterrent fine ($p = .068$ when compared to the OATH & COMMUNICATION, one-sided bootstrap proportion test) and 7.5% always lie ($p = .003$ when compared to the OATH & COMMUNICATION, one sided bootstrap proportion test). Finally, Figure 3b 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 deterrent fines (60%) than with the oath (45%) but the increase is not statistically significant ($p = .205$, one-sided bootstrap proportion test).

Result 4 *In our setting, the oath turns out to be behaviorally equivalent to a mild fine. Commitment 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 the deterrent fine ($f = 4$) is implemented

Figure 4: Empirical Distribution Functions of individual behavior when the deterrent fine and the oath are combined



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 deterrent fine treatment.

The magnitude of cooperation in this new treatment is 35.0% (see Table 1), a level comparable with the one observed in OATH & COMMUNICATION (34.5%) but lower than that in the deterrent fine alone treatment (46.0%, although this decrease is not statistically significant with $p = .193$, one-sided bootstrap proportion test). Table 2 shows that the evolution of cooperation is very similar under deterrent fine alone and under the deterrent fine combined with the oath treatment, but happens at a lower level. The overall rate of trust is at 53.5%, not statistically different to the level observed in the deterrent fine treatment (65.4%, $p = .282$, two-sided bootstrap proportion test), whereas trustworthiness is at 65.4% (vs. 71.3% in the deterrent fine treatment, $p = .722$, two-sided bootstrap proportion test). These results indicate that the combination of the deterrent 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 combining the oath and the deterrent fine on the rate of messages “*I will split*” is further exaggerated: 56.0% as compared to 65.5% in the deterrent fine treatment ($p = .139$, one-sided bootstrap proportion test) and 74.0% in OATH & COMMUNICATION ($p = .014$, one-sided bootstrap proportion test). The positive effect on the rate of messages “*I will keep*” also increases: 24.5% as compared to

14.0% in the deterrent fine alone treatment ($p = .077$, one-sided bootstrap proportion test) and to 10.0% in OATH & COMMUNICATION ($p = .031$, one-sided bootstrap proportion test). Second, the examination of the EDF in Figures 4a and 4b indicates that combining devices has no impact on the truth and trust ratios. EDF are close to those observed in the deterrent 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 deterrent fine treatment (61.7%, $p = .997$, two-sided bootstrap proportion test). This is also true for the level of trust (75.9%, $p = .601$ two-sided bootstrap proportion test) and the level of trustworthiness (81.2%, $p = .981$, two-sided bootstrap proportion test). Altogether, these results indicate that weakened cooperation when the deterrent fine is combined with the oath occurs due to subjects being less willing to announce cooperation with no improvement in the truthfulness of, and the trust in, messages.

Result 5 *The combination of the deterrent fine with the oath leads to a decline in the use of cooperative messages with no improvement in truthfulness and trust.*

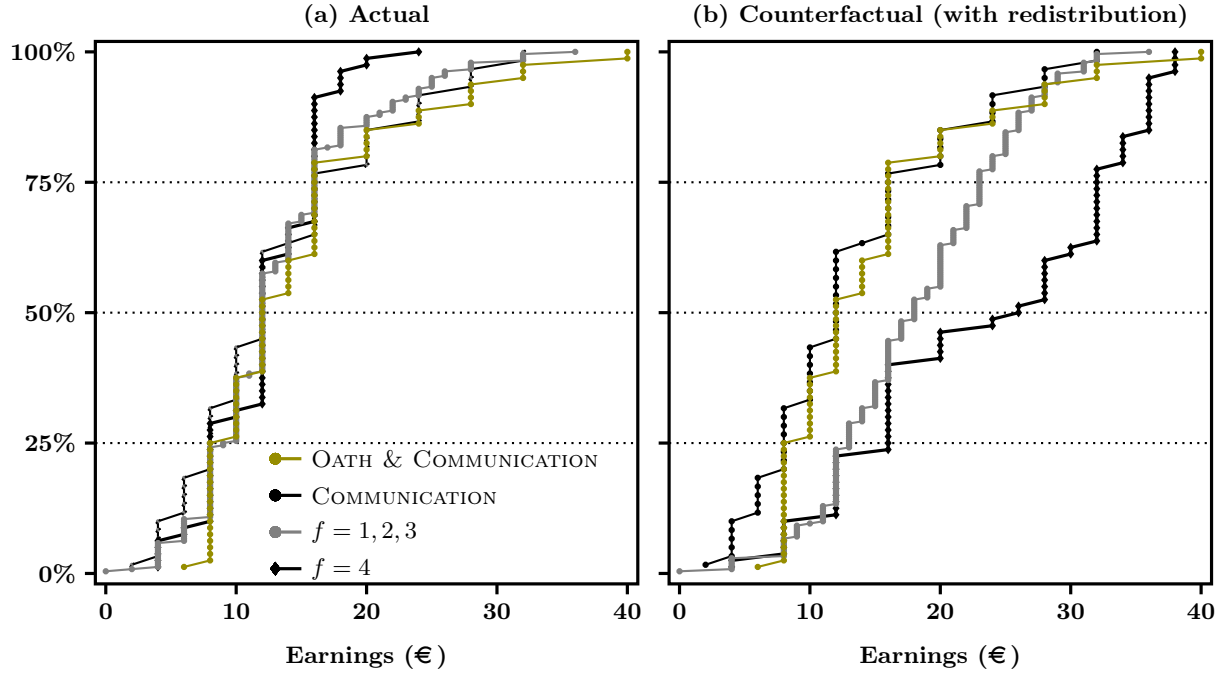
6 The economic value of the oath

Overall, the oath proves to be behaviorally equivalent to mild fines, while the deterrent fine is more effective in restoring cooperation. Should we conclude that deterrent fines are socially preferable to mild fines and non-monetary commitment devices such as the truth-telling oath as way to foster trust and cooperation through communication? 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 sum of their payoff over all rounds of play (see Appendix G for detailed statistics). The mean of feasible earnings in OATH & COMMUNICATION is €14.8 whereas it is €13.4 in COMMUNICATION. The increase is small in size, but significant ($p = .065$, based on a one-sided 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 in COMMUNICATION to €10.3 in OATH & COMMUNICATION, and from €18.0 to €19.3 among player Bs. Figure 5a 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 with an oath and €4.8 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 with an oath and €29.6 without).

¹²The test operates by bootstrapping on clusters in the same way as the bootstrap proportion test presented in footnote 10.

Figure 5: Empirical Distribution Functions of feasible earnings under communication



Welfare (as measured by feasible earnings) is not improved by the fines. This is particularly true for the deterrent 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 ($\text{€}13.2$, $p = .017$, one-sided bootstrap mean test) and the deterrent fine treatment ($\text{€}13.6$, $p = .063$) give rise to lower feasible earnings than OATH & COMMUNICATION. Fines seem to punish dishonest trustees at the upper end of the distribution without benefiting trustors at the lower 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. This amounts to factoring the enforcement authority’s revenue into the computation, and can be seen 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 5b 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 deterrent fine treatment.

Here, mean feasible earnings are now €23.6, as compared to €18.2 with the mild fine treatments (bunched together). Redistribution largely benefits the trustors: their mean feasible earnings are now €32.8, to be compared with €10.4 in OATH & COMMUNICATION when there is no redistribution, and €12.8 in the deterrent fine treatment.

In practice, fines and redistribution are however costly, leading to a marginal cost of public funds (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 deterrent fine obviously dominates all other conditions in this case, with a coefficient of 8. For the lowest level of fine ($f = 1$), the coefficient is 0.93. It means the oath dominates what can be achieved under this level of 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. A classical tool to restore trust is communication between traders, but it often fails because communication is cheap and people can easily be duplicitous. How can we commit people to tell the truth in such strategic contexts? Using a mechanism that fosters a commitment to the truth becomes essential.

Herein we use the lab to explore how a non-monetary mechanism — a solemn truth-telling oath — can improve the trust-enhancing performance of communication through commitment to the truth, and measure the monetary equivalent of this non-monetary commitment device. Using a classic trust game experimental design, our results extend the existing evidence about the economic consequences of the commitment to truth-telling achieved by an oath to the context of communication in a trust game. Although communication alone fails to promote fair economic exchange, we find that communication under oath can create more trust and cooperative behavior. This overall increase in cooperation results from two countervailing effects of the oath on communication. While messages announcing cooperation become more truthful and are trusted more often under oath, this improvement in the communication technology comes at a cost on the intensive margin of communication: trustees are less likely to announce cooperation under oath.

A truth-telling oath appears to be a powerful behavioral intervention when communication is

¹³The value of this ratio 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 et al. (2008) obtain an estimated value equal to 1.16 for the UK.

a key requirement to reach desirable economic outcomes. Our second contribution is to investigate the monetary value of this non-monetary intervention through a direct comparison with the monetary commitment implemented by mild and deterrent fines for deception. We find that the oath is behaviorally equivalent to mild fines, but is dominated by deterrent fines. Similar to mild fines, the oath fails to reverse a general decay of cooperation over time. Importantly, however, the shift in cooperation induced by the oath seems to persist over time. An obvious avenue for further research would be to assess the robustness of this important result with a larger number of repetitions.

The deterrent fine, in contrast with both the oath and mild fines, manage to maintain cooperation at a high level with no significant decrease over time. While one may be tempted to strengthen the incentives further by combining the truth-telling oath with the deterrent fines, we show this may be counterproductive: combining the oath with the deterrent fine exacerbates the reluctance to announce cooperation and comes with no clear improvement in the efficiency of communication, resulting in a significantly lower level of cooperation.

References

- Algan, Y., Cahuc, P., 2014. Trust, growth, and well-being: New evidence and policy implications, in: *Handbook of Economic Growth*. Elsevier. volume 2, pp. 49–120.
- Beck, T., Bühren, C., Frank, B., Khachatryan, E., 2020. Can Honesty Oaths, Peer Interaction, or Monitoring Mitigate Lying? *Journal of Business Ethics* 163, 467–484.
- Belan, P., Gauthier, S., Laroque, G., 2008. Optimal grouping of commodities for indirect taxation. *Journal of Public Economics* 92, 1738–1750.
- Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. *Games and Economic Behavior* 10, 122–142.
- Bliege Bird, R., Ready, E., Power, E., 2018. The social significance of subtle signals. *Nature Human Behavior* 2, 452–457.
- Bohnet, I., Frey, B., Huck, S., 2001. More order with less law: On contract enforcement, trust, and crowding. *American Political Science Review* 95, 131–144.
- Bracht, J., Feltovich, N., 2008. Efficiency in the trust game: an experimental study of precommitment. *International Journal of Game Theory* 37, 39–72.
- Bracht, J., Feltovich, N., 2009. Whatever you say, your reputation precedes you: Observation and cheap talk in the trust game. *Journal of Public Economics* 93, 1036–1044.
- de Bruin, B., 2016. Pledging integrity: Oaths as forms of business ethics management. *Journal of Business Ethics* 136, 23–42.

- Buckholtz, J.W., Marois, R., 2012. The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience* 15, 655–661.
- Carlsson, F., Kataria, M., 2018. Do people exaggerate how happy they are? Using a promise to induce truth-telling. *Oxford Economic Papers* 70, 784–798.
- Charness, G., Dufwenberg, M., 2006. Promises and partnership. *Econometrica* 74, 1579–1601.
- Cialdini, R., Sagarin, B., 2005. Interpersonal influence, in: Brock, T., Green, M. (Eds.), *Persuasion: Psychological insights and perspectives*. Sage Press, Newbury Park (CA), pp. 143–169.
- 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, 3301–3309.
- Duffy, J., Xie, H., Lee, Y.J., 2013. Social norms, information, and trust among strangers: Theory and evidence. *Economic Theory* 52, 669–708.
- Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. *Games and Economic Behavior* 47, 268–298.
- Elster, J., 1989. Social norms and economic theory. *Journal of Economic Perspectives* 3, 99–117.
- Farrell, J., Rabin, M., 1996. Cheap talk. *Journal of Economic Perspectives* 10, 103–118.
- Freedman, J.L., Fraser, S., 1966. Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology* 4, 195–202.
- Gauthier, S., 2013. Optimal tax base with administrative fixed costs. *International Tax and Public Finance* 20, 961–973.
- Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association* 1, 114–125.
- Jacquemet, N., James, A.G., Luchini, S., Murphy, J.J., Shogren, J.F., 2021a. Do truth-telling oaths improve honesty in crowd-working? *PloS ONE* 16, e0244958.
- Jacquemet, N., Joule, R.V., Luchini, S., Shogren, J.F., 2013. Preference elicitation under oath. *Journal of Environmental Economics and Management* 65, 110–132.
- Jacquemet, N., Luchini, S., Malézieux, A., Shogren, J., 2020. Who'll stop lying under oath? Experimental evidence from tax evasion games. *European Economic Review* 20, 103369.
- Jacquemet, N., Luchini, S., Rosaz, J., Shogren, J.F., 2018a. Truth-telling under oath. *Management Science* 65, 426–438.
- Jacquemet, N., Luchini, S., Rosaz, J., Shogren, J.F., 2021b. Can We Commit Future Managers to Honesty? *Frontiers in Psychology* 12, 2785.

- Jacquemet, N., Luchini, S., Shogren, J.F., Zylbersztejn, A., 2018b. Coordination with communication under oath. *Experimental Economics* 21, 627–649.
- Joule, R.V., Beauvois, J.L., 2010. *La soumission librement consentie*. 6th ed., Presses Universitaires de France, Paris.
- Joule, R.V., Girandola, F., Bernard, F., 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, 493–505.
- Kiesler, C., 1971. *The psychology of commitment. Experiments linking behavior to belief*. Academic Press, New York.
- Kiesler, C., Sakumura, J., 1966. A test of a model for commitment. *Journal of Personality and Social Psychology* 3, 349–353.
- Knack, S., Keefer, P., 1997. Does social capital have an economic payoff? A cross-country investigation. *Quarterly Journal of Economics* 112, 1251–1288.
- Koessler, A.K., Torgler, B., Feld, L.P., Frey, B.S., 2019. Commitment to pay taxes: Results from field and laboratory experiments. *European Economic Review* 115, 78–98.
- Peer, E., Feldman, Y., 2021. Honesty pledges for the behaviorally-based regulation of dishonesty. *Journal of European Public Policy* 28, 761–781.
- Putnam, R.D., 2000. *Bowling Alone: The Collapse and Revival of American Community*. Simon and schuster.
- Putnam, R.D., Leonardi, R., Nanetti, R.Y., 1992. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton university press.
- Rousseau, D.M., Sitkin, S.B., Burt, R.S., Camerer, C., 1998. Not so different after all: A cross-discipline view of trust. *Academy of Management Review* 23, 393–404.
- Silver, M., 1995. *Economic Structures of Antiquity*. Greenwood Press, Westport, CT.
- Tyran, J.R., Feld, L.P., 2006. Achieving compliance when legal sanctions are non-deterrent. *Scandinavian Journal of Economics* 108, 135–156.
- Weitzel, U., Kirchler, M., 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/> URL: <http://www.gate.cnrs.fr/~zeiliger/regate/RegateIntro.ppt>.
- Zickfeld, J.H., Karg, S.T.S., Engen, S.S., Gonzalez, A.S.R., Michael, J., Mitkidis, P., 2024. Committed (dis)honesty: A systematic meta-analytic review of the divergent effects of social commitment to individuals or honesty oaths on dishonest behavior. *Psychological Bulletin* 150, 586–620.

Zickfeld, J.H., Ścigala, K.A., Weiss, A., Michael, J., Mitkidis, P., 2023. Commitment to honesty oaths decreases dishonesty, but commitment to another individual does not affect dishonesty. *Communications Psychology* 1, 1–18.

Appendix

A Written 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 experiment 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 being matched into pairs: if you are participant A, then a participant B is randomly selected to your complete pair; analogously, if you 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.

B 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.

C Main aggregate outcomes: matching-group-level means

Treatment	Group	Outcomes (%)		
		Coop.	Trust	Trustworth.
BASELINE	1	28.0	64.0	43.8
	2	16.0	68.0	23.5
	3	8.0	60.0	13.3
	4	20.0	68.0	29.4
	5	8.0	52.0	15.4
	6	16.0	52.0	30.8
COMMUNICATION	1	24.0	48.0	50.0
	2	52.0	76.0	68.4
	3	12.0	52.0	23.1
	4	12.0	60.0	20.0
	5	20.0	48.0	41.7
	6	12.0	52.0	23.1
OATH	1	32.0	64.0	50.0
	2	12.0	44.0	27.3
	3	20.0	60.0	33.3
	4	36.0	80.0	45.0
	5	20.0	68.0	29.4
	6	8.0	48.0	16.7
COMMUNICATION & OATH	1	36.0	72.0	50.0
	2	32.0	56.0	57.1
	3	44.0	68.0	64.7
	4	44.0	84.0	52.4
	5	28.0	64.0	43.8
	6	16.0	52.0	30.8
	7	32.0	64.0	50.0
	8	44.0	64.0	68.8
$f = 1$	1	8.0	44.0	18.2
	2	20.0	68.0	29.4
	3	36.0	52.0	69.2
	4	44.0	64.0	68.8
	5	28.0	76.0	36.8
	6	40.0	60.0	66.7
	7	44.0	72.0	61.1
	8	28.0	68.0	41.2

Treatment	Group	Outcomes (%)		
		Coop.	Trust	Trustworth.
$f = 2$	1	48.0	68.0	70.6
	2	40.0	64.0	62.5
	3	4.0	36.0	11.1
	4	36.0	68.0	52.9
	5	40.0	72.0	55.6
	6	44.0	60.0	73.3
	7	48.0	76.0	63.2
	8	40.0	52.0	76.9
$f = 3$	1	32.0	68.0	47.1
	2	28.0	56.0	50.0
	3	32.0	72.0	44.4
	4	24.0	44.0	54.5
	5	16.0	64.0	25.0
	6	40.0	56.0	71.4
	7	32.0	60.0	53.3
	8	52.0	64.0	81.3
$f = 4$	1	48.0	68.0	70.6
	2	68.0	80.0	85.0
	3	40.0	56.0	71.4
	4	36.0	56.0	64.3
	5	52.0	72.0	72.2
	6	40.0	60.0	66.7
	7	44.0	64.0	68.8
	8	40.0	60.0	66.7
COMMUNICATION & OATH & $f = 4$	1	20.0	48.0	41.7
	2	24.0	48.0	50.0
	3	44.0	64.0	68.8
	4	72.0	84.0	85.7
	5	60.0	72.0	83.3
	6	36.0	52.0	69.2
	7	12.0	32.0	37.5
	8	12.0	28.0	42.9

D Communication, trust, and trustworthiness: matching-group-level means

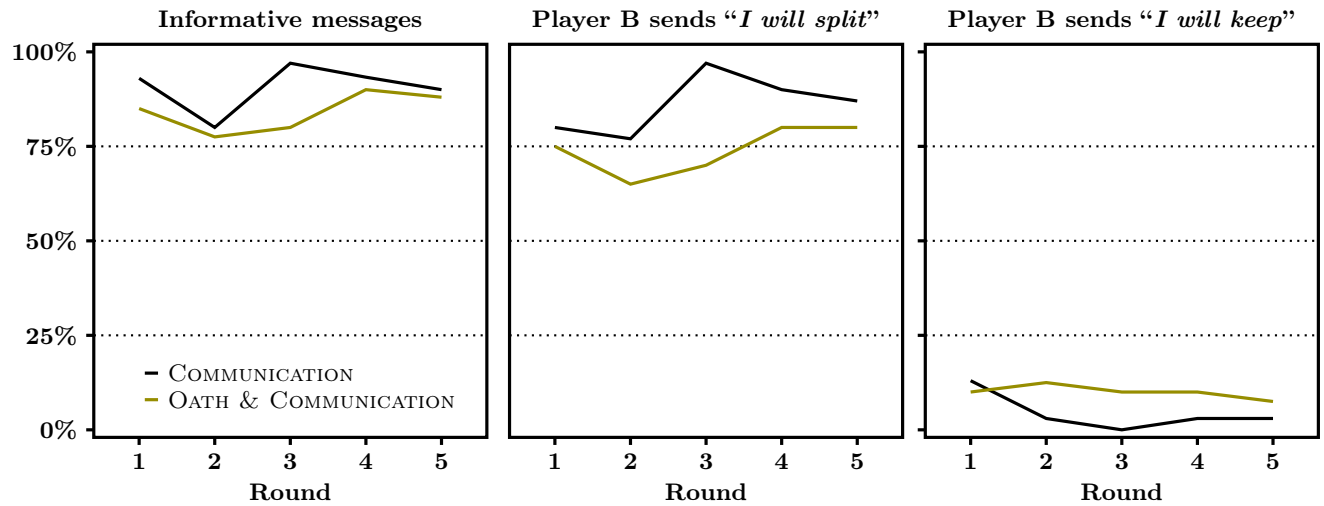
Treatment	Group	Messages (%)		Outcomes cond. on “ <i>I will split</i> ” (%)	
		“ <i>I will split</i> ”	“ <i>I will keep</i> ”	Trust	Trustworthiness
COMMUNICATION	1	92.0	4.0	52.2	50.0
	2	80.0	4.0	90.0	72.2
	3	100.0	0.0	52.0	23.1
	4	84.0	8.0	66.7	21.4
	5	80.0	4.0	50.0	30.0
	6	80.0	8.0	60.0	25.0
OATH & COMMUNICATION	1	84.0	4.0	85.7	50.0
	2	72.0	8.0	66.7	66.7
	3	72.0	8.0	83.3	73.3
	4	76.0	4.0	94.7	55.6
	5	80.0	4.0	65.0	38.5
	6	72.0	8.0	55.6	30.0
	7	52.0	32.0	76.9	70.0
	8	84.0	12.0	76.2	68.8
Mild fine, $f = 1$	1	80.0	8.0	50.0	20.0
	2	60.0	20.0	86.7	38.5
	3	68.0	16.0	76.5	69.2
	4	76.0	8.0	78.9	66.7
	5	68.0	16.0	82.4	42.9
	6	64.0	12.0	81.3	76.9
	7	88.0	4.0	77.3	64.7
	8	96.0	0.0	70.8	41.2
Mild fine, $f = 2$	1	96.0	0.0	70.8	70.6
	2	88.0	4.0	72.7	62.5
	3	68.0	12.0	47.1	12.5
	4	64.0	12.0	75.0	75.0
	5	88.0	4.0	77.3	52.9
	6	64.0	8.0	75.0	91.7
	7	76.0	4.0	58.5	58.8
	8	52.0	32.0	92.3	75.0
Mild fine, $f = 3$	1	68.0	12.0	64.7	72.7
	2	80.0	12.0	60.0	58.3
	3	88.0	0.0	77.3	47.1
	4	64.0	8.0	62.5	60.0
	5	76.0	12.0	73.7	28.6
	6	84.0	4.0	61.9	69.2
	7	56.0	8.0	64.3	88.9
	8	68.0	16.0	82.4	92.9
Deterrent fine, $f = 4$	1	72.0	4.0	77.8	78.6
	2	84.0	8.0	81.0	100.0
	3	48.0	24.0	100.0	83.3
	4	56.0	16.0	85.7	75.0
	5	64.0	12.0	93.8	80.0
	6	68.0	12.0	82.4	71.4
	7	64.0	12.0	87.5	78.6
	8	68.0	24.0	70.6	75.0
OATH & COM. & $f = 4$	1	40.0	8.0	90.0	55.6
	2	44.0	24.0	81.8	66.7
	3	60.0	36.0	66.7	100.0
	4	72.0	4.0	100.0	100.0
	5	76.0	20.0	94.7	83.3
	6	60.0	36.0	73.3	81.8
	7	36.0	48.0	44.4	75.0
	8	60.0	20.0	40.0	50.0

E Trust and trustworthiness by round

Table 5: Trust and trustworthiness rates by round (%)

Treatment	Trust					Trustworthiness				
	Round					Round				
	1	2	3	4	5	1	2	3	4	5
BASELINE	56.7	80.0	66.7	53.3	46.7	41.2	33.3	15.0	25.0	14.3
COMMUNICATION	73.3	46.7	63.3	50.0	46.7	45.5	50.0	36.8	26.7	35.7
OATH	63.3	70.0	63.3	60.0	46.7	57.9	33.3	26.3	27.8	28.6
OATH & COMMUNICATION	65.0	72.5	67.5	60.0	62.5	69.2	44.8	55.6	54.2	40.0
Mild fine, $f = 1$	70.0	65.0	52.5	70.0	57.5	64.3	38.5	52.4	50.0	39.1
Mild fine, $f = 2$	70.0	57.5	60.0	57.5	65.0	82.1	56.5	62.5	60.9	38.5
Mild fine, $f = 3$	72.5	62.5	52.5	52.5	62.5	62.1	44.0	47.6	52.4	56.0
Deterrent fine, $f = 4$	55.0	62.5	72.5	65.0	67.5	77.3	64.0	72.4	69.2	74.1
OATH & COM. & $f = 4$	47.5	47.5	60.0	52.5	60.0	63.2	63.2	62.5	66.7	70.8

F Communication behavior by round and treatment



G 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 & COMMUNICATION	6.0	8.0	10.0	10.3	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
Deterrent fine, $f = 4$	6.0	11.5	12.0	12.8	16.0	18.0

Player Bs' earnings						
(accumulated over 5 rounds)						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
COMMUNICATION	4.0	12.0	16.0	18.0	24.0	32.0
OATH & COMMUNICATION	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
Deterrent fine, $f = 4$	4.0	8.0	12.0	11.8	16.0	24.0