TI 2023-069/I
Tinbergen Institute Discussion Paper

Individual oath-swearing and lying under peer pressure

Sorravich Kingsuwankul[1]

1 Vrije Universiteit Amsterdam and Tinbergen Institute

Individual oath-swearing and lying under peer pressure*

Sorravich Kingsuwankul†

October 13, 2023

Abstract

Previous studies have shown that an oath can reduce lying at an individual level. Can oaths reduce lying in groups, a context where the prevalence of lying is typically higher? Results from a lab experiment reveal that the impact of an oath on lying in a group context depends on the incentive structure. Oath reduces lying only when payoffs are independent. Evidence supports the notion that payoff interdependence creates pressure on individuals to conform to the group, crowding out the impact of oath on honest reporting. An implication is that to be effective in an organizational context, an oath intervention must be designed in strict connection with the incentive structures.

Keywords: Honesty oath, lying, group incentive, peer pressure, lab experiment

JEL: C92, D71, D91

*Acknowledgements: The author is grateful to B. Corgnet, L. Gangadharan, N. Jacquemet, C. Rott, V. Valero, M. C. Villeval, A. Zylbersztejn, participants at the Winter School on Credence Goods, Incentives and Behavior in Innsbruck, ASFEE in Lyon, and Experimental Seminar at SBE VU for their useful comments. The author is thankful to Q. Thevenet for his assistance in programming and to B. Monjour and A. Thampi for their research assistance. This research has benefited from the financial support of IDEXLYON from Université de Lyon (INDEPTH-IDEX/SBP/2018/03) within the Programme Investissements d'Avenir (ANR-16-IDEX-0005) operated by the French National Research Agency and GATE-LAB funding program. This research also benefited from the support of the LABEX CORTEX (ANR-11-LABX-0042) of the Université de Lyon, within the Programme Investissements d'Avenir (ANR-11-IDEX-007) operated by the French National Research Agency (ANR).

†School of Business and Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Tinbergen Institute, 1082 MS Amsterdam, The Netherlands. Email: s.kingsuwankul@vu.nl

1 Introduction

Frauds and misconducts often involve a group of individuals. The scandals at Enron, Wells Fargo and Madoff Ponzi scheme are some examples where a number of people were involved in the intricate webs of financial crimes, hindering public trust in the financial sector (Guiso et al., 2008; Sapienza and Zingales, 2012). To remedy these trust issues, professional oaths and codes of conducts have recently been introduced in many organizations and are becoming increasingly common (e.g., bankers' and accountants' oath in the Netherlands). In the non-financial world, academia has similarly seen a considerable share of scientific misconducts – from egregious ones like data fabrication to those in the grey area like HARKing and p-hacking. Many universities have thus put in place honor pledges to establish ethical standards and induce moral engagement among their graduates. An example is the widely known MBA oath, which is a 'Hippocratic oath for managers' initiated by the class of 2009 Harvard Business School with the aim of engaging graduates and alumni in being responsible and ethical actors in the society.[1] A similar commitment device has been recently adopted in France. Since December 2022, PhD graduates from French universities are required to take an ethics oath after having successfully defended their dissertations, where they pledge to maintain integrity in their research conduct.[2] While individuals can be asked to take an ethical pledge before the start of their jobs or at graduation, an open question is whether such an oath is able to effectively discourage dishonest behavior in a collective setting where people can interact with their group members.

This study investigates the impact of oath-swearing on lying behavior in a group setting. Evidence in psychology and behavioral economics have shown that oath reduces lying at an individual level (e.g., Jacquemet et al., 2019; Heinicke et al., 2019; Beck, 2021). However, there is still little evidence on the effectiveness of the oath against dishonesty in a group setting – an environment where the prevalence of lying is high (e.g., Weisel and Shalvi, 2015; Kocher et al., 2018). On the one hand, an oath increases the costs of lying and makes the honesty norm more salient. This can lead individuals to encourage ethical behavior within their group. On the other hand, group communication allows for an exchange of justifications and gives rise to group conformity. This may compromise the impact of oath.

The study further examines whether the impact of oath in a group setting depends on the type of incentive group members face. In particular, it explores the case of the payoff commonality rule (Kocher et al., 2018). Under this incentive scheme, group members are required to coordinate on an action, and a failure results in a loss for the whole group. Without this rule, however, the incentives of group members remain independent. Kocher et al. (2018) found that individuals in a group setting lied to a similar extent under the two incentive schemes because of peer conformity following group communication.

A dimension not discussed in Kocher et al. (2018) is the role of peer pressure under the payoff commonality rule (on top of peer conformity). Simply put, in a group setting where

[1] See https://mbaoath.org/ for more information about the MBA oath, accessed on 13th October 2023.

[2] Also see https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000046229036 for information on the the ethics oath for French doctorates, accessed on 13th October 2023.

payoffs are interdependent, an individual may lie not only out of peer conformity (i.e., I do what others are doing) but also out of peer pressure (i.e., I do it because others want me to). Therefore, it is interesting to assess the impact of oath under the two incentive schemes. As individuals face a trade-off between group incentive and the norm of honesty, lying out of peer pressure may crowd-out the impact of oath. One can thus expect oath to be more effective when group members do not earn a common payoff compared to when they do.

A laboratory experiment uses a variant of the observed cheating game developed by Gneezy et al. (2018) in a mixed design (similar to that of Kocher et al., 2018). The within-subjects variation consists of three parts. In part 1, participants play the game in an individual setting, without any social interaction. A video of a die outcome is shown to the participants via computer. They are asked to memorize and report it later to obtain a payoff. In this game, participants can misreport the outcome of the die and earn additional payoffs. Since the die outcome is observable by the experimenter, lying behavior can identified *ex-post* at an individual level. The decision in part 1 is used as a proxy for individual type (honest or dishonest). In part 2, participants play the same game but in a group setting. They are randomly matched to form a group of three. All the three group members first view the same video of a die outcome, after which they can anonymously chat via computer for five minutes. The chat is free-form and can be about anything, except revealing personal information. After the chat, group members report the number individually as in part 1. The decision in part 2 captures lying behavior in a group setting – the main outcome of interest in this study. Finally, in part 3, participants play the game in an individual setting as in part 1.

There are four (2x2) between-subjects treatments varying in the incentive schemes and the existence of oath. The between-subjects dimension is introduced at the beginning of part 2 of the experiment. This means that part 1 remains the same across treatments. In **BaseNoPC**, payoffs of the group members are independent. Participants are informed that the number they report will not affect the payoff of their group members, and vice versa. In **BasePC**, group members face payoff commonality. They are informed that all the three members must report the same number to earn the payoff associated with the number. Otherwise, they earn nothing. This variation in the incentive structures follow Kocher et al. (2018) and serve as the baseline conditions where individuals lie in a group setting under two group incentives. The two remaining treatments, **OathNoPC** and **OathPC**, follow the payoff commonality rule of their respective baseline, *except* that at the beginning of part 2 (before receiving the instructions), all participants individually swear an oath in which they commit to be honest and will tell the truth for the rest of the experiment. This is common knowledge.

The baseline conditions of the experiment replicate the *dishonesty shift in group* of Kocher et al. (2018). Moving from the individual setting (part 1) to the group setting (part 2), the fractions of liars in BaseNoPC and BasePC almost doubled . In addition, the lying rates in a group setting across two incentive structures are similar (91.7% in BaseNoPC vs. 93.5% in BasePC), replicating no relevance of other-regarding preferences in this setting.

The main finding of the experiment shows that oath reduces lying in group *only* when group member payoffs are independent. Given no payoff commonality, the fraction of liars reduces from 91.7% under no oath to 73.1% under oath. On the other hand, the reduction in lying caused by an oath under payoff commonality exists to a smaller extent (from 93.5% to 84.4%) and remains insignificant. Thus, the impact of oath against lying in group persists only when group members earn independent payoffs. This result is consistent with the mechanism of peer pressure under the payoff commonality rule in crowding out the impact of the oath on honest reporting.

Analyzing individual lying decision after group interaction (part 3) sheds light on the role of peer pressure under the payoff commonality rule. Those who lied in the group setting under the payoff commonality rule were less likely to lie later when they were alone compared to the treatments under no payoff commonality. This behavioral pattern is consistent with the peer pressure mechanism under payoff commonality. If individuals lied in a group setting out of pure conformity, one can expect them to continue to conform (i.e., to lie) when they were alone. However, if they lied out of peer pressure, one can expect less lying when they were alone because the prior action was inconsistent with what one would have wanted 'personally.' While peer conformity exists in both incentive structures, peer pressure exists only under payoff commonality.

These findings have implications for organizations that implement oaths or codes of conducts with the intention to promote and maintain professional integrity. In addition to individual's ability to update beliefs about normative expectations, which may lead to group conformity, individuals in a group may break the oath due to peer pressure caused by the incentive set by the organization. Thus, to avoid the crowding out effect, this 'soft' intervention must be carefully designed in strict connection with the incentive structure in the organization.

The remainder of this paper is organized as follows. Section 2 briefly reviews the related literature. Section 3 outlines the experimental design and procedure. Section 4 describes the behavioral conjectures. Section 5 reports the results of the experiment. Section 6 discusses these findings and concludes.

2 Related Literature

This study contributes to the literature on the effect of honesty oaths on lying. Most prior studies focused on an individual setting and showed that it induces truth-telling (Jacquemet et al., 2019; Beck, 2021). A general interpretation is that oath raises moral awareness and interferes with the justification process. Under oath, the costs of lying increase because doing so involves two moral transgressions (i.e., telling a lie and breaking an oath). This makes lying under oath less attractive.[3]

[3] A meta-analysis by Zickfeld et al. (2022) reported an overall negative impact of honesty oaths on lying. Recent studies show that the effect of oath may vary. Oath can have a limited impact on partial lies compared to obvious ones (Heinicke et al., 2019; Jacquemet et al., 2021), and on lies that are mutually beneficial

Two recent studies examined the impact of an honesty oath in a group setting. Dunaiev and Khadjavi (2021) tested the effectiveness of a moral nudge[4] (i.e., signing an honesty declaration vs. no declaration) on misreporting performances in a real-effort task in dyads. Their nudge intervention reduces cheating in dyads and is as effective as when it is implemented on an individual. Zickfeld et al. (2022) focused on the interplay of social and moral commitment in cheating games. They found that oath reduces dyad's misreporting. However, its impact is slightly smaller when there is social commitment (dyad setting) compared to no social commitment (individual setting). The present study differs in that it is only interested in the influence of oath on lying in a group context. In other words, it focuses on how a moral commitment affects the way individuals lie, keeping social commitment constant. Further, the contribution goes beyond that by investigating the role of peer pressure, induced by group incentive, in moderating the effect of oaths. This aspect has not been examined by the two studies.

In relation to the literature on lying in group, prior studies indicate that groups are more likely to lie than individuals (e.g., Weisel and Shalvi, 2015; Kocher et al., 2018).[5] Several mechanisms have been proposed. First, studies on group decision-making found that groups are more sophisticated and are more likely to make a self-interested choice compared to individuals (see a review by Charness and Sutter, 2012). Sutter (2009) reported that groups are more likely to use 'sophisticated deception' (i.e., telling the truth, but expecting not to be believed) than individuals do in a deception game. Second, the group setting allows for diffusion of responsibility (e.g., Falk and Szech, 2013; Falk et al., 2020). Individuals may perceive a lower sense of accountability when their role is less pivotal towards the group final outcome. This creates a moral wiggle room by hiding behind others (e.g., Bauer et al., 2021). Third, groups may lie more than individuals because doing so benefits others (e.g., Wiltermuth, 2011; Conrads et al., 2013). Lastly, individuals in a group tend to lie more because they can update the empirical information about honesty through observation (e.g., Diekmann et al., 2015; Bicchieri et al., 2022) or communication (Kocher et al., 2018). In other words, a high prevalence of lying in group results from an erosion of the honesty norm as individuals can learn from and conform to others.

By design, this study rules out sophistication and diffusion of responsibility and looks at the role of conformity and incentives. Studies on peer effects showed that people tend to misbehave after being exposed to the norm violations by peers (see Keizer et al., 2008; Fosgaard et al., 2013; Kroher and Wolbring, 2015), and the pattern of conformity tends to be asymmetric. That is, people become more dishonest after observing norm violations but

compared to self-serving ones (Jacquemet et al., 2021). Some studies reported null effects (Koessler et al., 2019; Prima et al., 2020; Cagala et al., 2023). Also see Jacquemet et al. (2019) for the (moral) framing effect, and Cagala et al. (2021) for the backfiring effect.

[4]Despite being called a nudge, their intervention is akin to a moral commitment device. Participants were presented with an honesty declaration before the opportunity to cheat.

[5]An exception is Castillo et al. (2022) who found that groups are not more dishonest than individuals when lying imposes a negative externality on a third-party.

do not become more honest after observing norm compliance.[6] Further, there are evidence that people prefer being paired with a peer of the same type to reduce the moral cost of misbehaving (e.g., Gross et al., 2018; Charroin et al., 2022). This study contributes by examining the effect of oath in a group setting where individuals can conform to others.

As for the role of incentives, this study examines the case of the payoff commonality rule (Kocher et al., 2018). Under this scheme, group members are required to coordinate on an action, and failing to do so results in a loss for all. The scheme reflects how individuals in a group can be interconnected via an incentive structure, as without it individual payoff becomes independent of others' actions. Kocher et al. (2018) varied the presence of payoff commonality across conditions and argued that they capture "the relevance of the other-regarding concerns argument in group decision making that involves a trade-off between payoff maximization and norm compliance." They documented similar lying rates across the two incentive schemes and argued that group communication allows for learning of the empirical norms, putting other-regarding preferences in the backseat.

Nonetheless, Kocher et al. (2018) is silent regarding the role of peer pressure under payoff commonality. When group member payoffs are interdependent, individuals may also lie out of pressure (i.e., I lie because others want me to). This implies that a person may lie not only because they simply follow others (conformity) but also because they succumb to pressure. Thus, while only conformity drives lying in a group under no payoff commonality, peer pressure is an additional force to induce lying when group members face the payoff commonality rule. This study will examine the two mechanisms under oath. As Kocher et al. (2018) remarked about the trade-off between payoff maximization and norm compliance, this tension can become more important under oath as the intervention reinforces the norm of honesty.[7] On the other hand, when payoffs are independent, there is no pressure to break the oath 'for the group.'

3 Experimental Design and Procedures

Design

Participants played a variant of the observed cheating game used in Kocher et al. (2018) (see Gneezy et al. (2018) for the original task). They observed the video of the outcome of a die roll (i.e., ⚀, ⚁, ⚂, ⚃, ⚄ or ⚅), which was randomly selected by the computer. Each face of the die was associated with different payoffs: each eye on the die represented one token, except for '⚅' which yielded zero tokens. Participants were asked to memorize the outcome

[6] This asymmetric conformity has been documented in a tax evasion game (Lefebvre et al., 2015) and a die-rolling paradigm (Colzani et al., 2023). Innes and Mitra (2013) reported symmetric conformity in a deception game with samples in India, but asymmetric conformity in the samples in the USA.

[7] This is consistent with the justified-ethicality perspective of Leib et al. (2021) which posits that the individual likelihood to engage in a collaborative dishonesty is driven by prosocial concerns towards group members and is attenuated by the honest-image concerns. The payoff commonality rule reinforces the former, while the oath reinforces the latter.

and report it later. They earned the payoff depending on their report and could misreport the die outcome to earn more tokens.

A mixed experimental design was used as in Kocher et al. (2018). The within-subjects dimension consisted of three parts and varied in whether the game was played in an individual or a group context. In part 1, the game was played individually as described, without any social interaction. The decision in part 1 serves as a control for individual moral type. In part 2, participants were randomly matched to form a group of three. The three members viewed the same video of the die outcome, after which they could chat anonymously for five minutes via computers. The chat allowed for a free-form communication apart from revealing identity. Messages were shown to all the three group members. They could leave the chat at any point of time. The chat ended either after all group members had left the chat or the time had elapsed. After the chat, each group member reported the number individually as in part 1. The decision in part 2 is the main outcome measure in this study. Finally, in part 3, participants played the game individually as in part 1. The decision in part 3 serves not only to replicate Kocher et al. (2018), but also will shed light on the role of peer pressure on the decision to lie in part 2. No feedback was given regarding the decisions of other participants.

There were four (2x2) between-subjects treatments varying in the group incentives and the existence of oath. The between-subjects dimension was introduced at the beginning of part 2. Thus, part 1 remained the same across treatments. In **BaseNoPC**, payoffs of the group members were independent. Participants were informed that the number they enter would not affect the payoff of their group members, and vice versa. In **BasePC**, group members faced payoff commonality. They were informed that all the three members must report the same number to earn the payoff associated with that number. Otherwise, they would earn nothing. BaseNoPC and BasePC serve as the baseline conditions where individuals lie in a group setting under two different incentive structures. The two treatments, **OathNoPC** and **OathPC**, followed the payoff commonality rule of their respective baselines, *except* that at the beginning of part 2 (before receiving the instructions on computer screens), all participants individually swore an oath by typing the text of the oath on their computers to commit to be honest and to tell the truth for the rest of the experiment. The oath-taking was common knowledge.[8]

The oath-taking procedure differs from how it was typically implemented in the literature in two ways. First, oath-taking was compulsory, and was made common knowledge with all participants in the session. This is a design choice to align it with how in reality individuals within the same organization or a profession do not have the option of not signing an oath.[9] For example, employees in the Dutch banking sector are required to swear bankers' oath before taking up the office. Making an oath compulsory also eliminates the uncertainty regarding

[8]The computer program checks whether the following words are typed correctly: '*I swear upon my honor*' '*honestly*' and '*truth.*' Checking what the participants typed after the experiment, all but one typed correctly with only minor typing errors that do not change the meaning of the oath. The one exception in OathNoPC added the word 'not' (i.e., 'I will not be honest'). As such, this observation and the two participants matched to it are excluded from the analysis to have a clean identification of the impact of the oath.

[9]Though participants could leave the experiment if they refused to sign the oath; none did.

whether the group members have signed it or not, which may be counterproductive (Davis and Jaber-Lopez, 2022). Second, the oath was introduced in the middle of the experimental session and not upon arriving at the lab. Specifically, participants swore the oath at the beginning of part 2 and before receiving the relevant instructions. There is a trade-off here. It was done to remain close to Kocher et al. (2018), while having the lying decision in part 1 as a control for individual's moral type before oath-taking. A drawback is the observed effect of an oath can be at the lower bound because they have been exposed to a similar task in part 1 (though they were not aware of the nature of the task in part 2 when they swore the oath).

Finally, at the end of the session, unannounced, participants answered three questions about their beliefs about the empirical norm of honesty for each part of the experiment in the session. More precisely, they had to guess the proportions of participants reporting each die number in each part. For each belief question, they were informed that they would receive 5 Euros minus any penalty (i.e., 0.04 Euros per percentage point deviation from the true value). To avoid negative earnings in case of extreme deviations, the minimum payment was fixed at 0.50 Euros. At the end of the experiment, the program randomly selected one of the three belief questions for an additional payment. Participants then answered standard socio-demographic questions and received feedback about their payoffs for each part and for the belief question selected by the program.

Procedures

The experiment was run in-person at GATE-Lab in Lyon, France, between November and December 2021. A total of 23 sessions were conducted: 5 sessions with 99 participants in BasePC, 6 sessions with 102 participants each for BaseNoPC, OathNoPC and OathPC.[10] All 405 participants (54.6% females) were recruited via HRoot (Bock et al., 2014), mainly from local engineering and business schools.[11] The experiment was programmed using z-Tree (Fischbacher, 2007). The videos of the die outcomes were shown on the computer screen. As in Kocher et al. (2018), the randomization of the videos of the die outcomes was conducted

[10] The number of observations is based on an ex-ante power calculation. Assuming a directional hypothesis of the effect of an oath on lying, the power is based on a one-tailed test. Assuming a Type-I error rate of 0.05, a power level of 0.8 and a medium to large effect size (Cohen's d = 0.65), the required number of observations to uncover the hypothesized effect between the treatments and the baseline conditions using One-tailed Mann-Whitney U test is 32 observations per treatment. Given that one group comprises of three individuals, this implies 96 individual observations per treatment. The actual number of observations is slightly higher than the initial power calculation to account for instances where participants saw a video of '⚄' in part 2, which left no room to observe upward lying.

[11] As discussed in this footnote 8, three observations in OathNoPC are excluded as one of them modified the text of the oath. Therefore, the final number of observations comes to 402 participants. See Table D1 in the Appendix for the summary statistics of the socio-demographic characteristics.

for the first two sessions and then used for the rest of the sessions to increase statistical power and simplify treatment comparisons.[12]

Upon arrival, participants randomly drew a ticket from an opaque bag which assigned them to computer terminals. A general instruction was given and read aloud at the beginning of the session. The remaining of the instructions (starting from part 1 of the experiment) were shown on computer screens at the beginning of each relevant part.[13] The average duration of the sessions was 50 minutes.

Participants' additional earnings consisted of their payoffs from one randomly selected part and one randomly selected belief question. Before the computer screen displayed the payoffs, the experimenter selected one participant at random to draw a raffle which indicated a participant ID who would roll a six-sided die to determine which part of the experiment was payoff relevant. After the die was rolled, the participant announced the die outcome and the relevant part for payment to the session. This procedure was described in the general instruction at the beginning of the experiment. The average payment was 17.32 Euros (SD = 2.71), including a fixed fee of 5 Euros.[14]

4 Conjectures

This section formulates behavioral conjectures regarding the impact of oath on lying in a group setting depending on the incentive structure.[15]

In part 2 of the experiment, participants anonymously formed a group of three and observed the same video of a die outcome. Before entering their number individually, they communicated via a computerized chat for five minutes. The baseline conditions differ in the incentive structure. In BaseNoPC, there was no payoff commonality rule. The number a group member entered did not affect the payoff of the other group members, and vice versa. In BasePC, group members faced payoff commonality where they needed to enter the same number to receive the associated payoff. If at least one of them deviated, all members earned zero tokens. Kocher et al. (2018) found that groups cheated to a similar extent irrespective of the payoff commonality. Since BaseNoPC and BasePC closely follow the treatment variations of Kocher et al. (2018), a similar pattern of behavior is expected.[16]

[12] Despite this plan, due to different turn-up rate across sessions, the videos shown are not perfectly balanced. Table D4 in the Appendix shows that the distributions of the videos displayed are not different across treatments. This implies that the results are free from any differences in the videos shown.

[13] Full instructions are provided in Section A of the Appendix.

[14] At the time, the project benefited from two sources of funding. Due to administrative issues, one funding had to be paid by bank transfer, while the other in cash. The author tried to balance the mode of payment across the treatments (see Table D5 in the Appendix).

[15] The design, conjectures and ex-ante power calculation were pre-registered at AsPredicted (#79968) prior to running the experiment. A clerical error was made where the belief elicitation stage was not mentioned in the pre-registration, despite having it in the design and the programme. The author decided to place the analysis of belief elicitations along with other exploratory analyses, which can be found in Section C of the Appendix.

[16] Kocher et al. (2018) elicited beliefs about the fraction of liars in a reference experiment (i.e., Fischbacher and Föllmi-Heusi, 2013) twice (before part 2 and after part 3). They were interested in how group interaction

If the honesty oath interferes with the justification process to lie by increasing the moral costs of lying, one can expect a reduction in lying, but how does that translate into a group context where communication is possible? On the one hand, oath may fail to promote honesty in group because individuals can exchange justifications to break the oath and conform to on another, which may attenuate the moral costs of lying heightened by the oath. On the other hand, as the honesty norm is made salient, it may shift the content of group communication towards an ethical one, and thereby promoting honesty in group. Whether the latter prevails the former is an empirical question to answer.

The first conjecture is formulated as follows:

Conjecture 1: *For a given payoff rule, an oath reduces lying in a group setting.*

As discussed earlier, the incentive structure may play an important role on the effect of honesty oath in a group context. The payoff commonality rule requires all the group members to coordinate on their decisions, which may create a pressure to conform to 'what the group wants,' and this can be either for or against honesty. There is, on the one hand, the honesty norm made salient by the oath. On the other hand, it can be impeded by other competing norms preferred by the group (e.g., efficiency, cooperation prosociality). Such a trade-off is consistent with the justified-ethicality perspective of Leib et al. (2021) as in this setting a group preference can be used to justify dishonesty. This is likely since lying under payoff commonality can be viewed as pareto-efficient (economically speaking), assuming that the majority of the group prefers to violate the honesty norm and thus the oath. Thus, an independence of decisions and payoffs allows each individual to respect the honesty norm (be it personal or induced by the oath) irrespective of whatever the others choose to do – though there is still room for conformity. As a result, one may expect that the payoff commonality rule to crowd out the impact of oath due to peer pressure.

The second conjecture is formulated as follows:

Conjecture 2: *The impact of an oath against lying is larger when group members do not face payoff commonality than when they do.*

5 Results

This section is organized as follows. Section 5.1 provides an evidence that oath impacts lying in a group setting but only when group members do not earn a common payoff. Section 5.2 shows that payoff commonality rule creates the pressure on group members to lie. Section 5.3 analyzes the chat data and explores how oath affects the content of communication.

shifted the beliefs about the empirical honesty in the reference experiment. As this study aims to test the impact of oath in a group setting, these belief elicitations of Kocher et al. (2018) were removed to avoid any confounding effect with the oath.

5.1 The Impact of Oath on Lying in Group

This section evaluates the impact of oath on lying in group (part 2 of the experiment). The participant is classified as a liar if the reported number of the die outcome does not match the one shown on the video. As participants could communicate via chat before making their decisions, treatment comparisons of the decisions in part 2 using non-parametric analyses are performed at the group level (collapsed at median and at mean). Unless otherwise indicated, two-sided Fisher exact tests are used. These analyses are confirmed with linear probability regressions (robust standard errors clustered at the group level) controlling for socio-demographics.[17]

Before formally testing the conjectures, an analysis is performed to check whether (i) lying behavior in part 1 is similar across four treatments, and (ii) the baseline conditions (BaseNoPC and BasePC) replicate the finding of Kocher et al. (2018).

Recall that in part 1 of all treatments individuals played the observed cheating game without any social interaction. The fractions of liars in part 1 range between 48% and 53% and there are no significant differences between them (the smallest p-value from the pairwise comparisons is 0.574).[18]

Next, do the baseline conditions replicate the dishonesty shift in group observed in Kocher et al. (2018)? A comparison is made between individual decisions in part 1 and 2 of the baseline conditions. Excluding groups that observed ⚄ in part 2,[19] there is a large significant dishonesty shift due to group communication. The fractions of liars increase from about 49% to 91.7% in BaseNoPC ($p < 0.001$, signrank test). The pattern is similar in BasePC (from 48% to 93.5%, $p < 0.001$, signrank test). In sum, the baseline conditions replicate the finding of Kocher et al. (2018) – a group context with communication pushes individuals to lie more and the payoff commonality rule does not play a role in the absence of an oath.

What is the impact of honesty oaths on the decision to lie in part 2? Looking at groups that had an opportunity to lie, the oath is impactful only when group members did not face payoff commonality. Compared to BaseNoPC, the fraction of liars in a group setting is lower in OathNoPC (91.7% vs. 73.1%, respectively). The difference is weakly significant at the median ($p = 0.082$) and significant at the mean ($p = 0.015$). The pattern of reduction is similar under payoff commonality. Compared to BasePC, the fraction of liars in a group setting is lower in OathPC (93.5% vs. 84.4%, respectively). The difference is, however,

[17] These specifications deviate from the pre-registered analysis plan (Mann Whitney U test and probit regression). Given that liars are identifiable at the individual level, a Fisher's exact test is more appropriate than Mann Whitney U test, which is ideal when lying is not observable ex-post. Linear probability regression is preferred due to its presentation and interpretation of the interaction term (for testing the second conjecture). The same conclusions can be drawn from the marginal effects of probit regressions.

[18] Excluding those who observed ⚄ in part 1 do not change this conclusion. The smallest p-value becomes 0.433.

[19] Individuals who observed ⚄ could not lie upward. In terms of analysis, there is a trade-off between excluding these data points and identification. The choice to exclude such observations only in part 2 is to have a clean identification for conjecture testing, while not losing power. Robustness tests by further exclusion (e.g., those observed ⚄ in part 1) are reported in the footnotes. Regression analyses include relevant control variables.

not statistically significant either at the median or mean (both $p = 0.426$). The fraction of liars in OathNoPC is significantly lower than that in OathPC at the mean ($p < 0.001$), but not significant at the median ($p = 0.536$). This analysis suggests that the power of oath in reducing lies in groups is dependent on the incentive structure – it can reduce lying in group when payoffs are independent, but it is insufficient to change behavior when payoffs are interdependent.

To confirm these results, Table 1 reports the coefficient estimations from the linear probability regressions. The dependent variable is a binary indicator whether the participant lied or not in part 2 (1 if liar and 0 otherwise). Groups that observed ⚄ in part 2 were excluded as they could not lie upward (no group lied downward). Model 1 includes in the independent variables a dummy variable 'PC' which takes the value of 1 if individuals faced the payoff commonality rule and 0 otherwise, a dummy variable 'Oath' which takes the value of 1 if individuals were under oath and 0 otherwise, and the interaction term between the two dummies ('PC X Oath'). Model 2 further controls for socio-demographics (age, gender and self-reported risk attitudes or SOEP, field of study in business and economics), the number of tokens observed in the video in part 2 (ranging from 0 to 4 tokens), and moral types based on decisions in part 1 (honest by choice or by design with liar as a reference category).

Table 1 is consistent the non-parametric analyses. Model 1 suggests that oath is only impactful when group members do not face the payoff commonality – swearing an oath reduces lying in a group context by 18.5 percentage point when payoffs are independent ($p = 0.008$). The coefficient of the interaction term is positive, indicating that the impact of oath under payoff commonality is smaller compared to under no payoff commonality. However, the effect of the oath under payoff commonality is not significant (Wald test, $p = 0.245$). Lastly, the coefficient of the interaction term is not significant ($p = 0.370$). Model 2 shows the robustness of these results – after additional control variables, lying in group reduces under oath by about 19 percentage point when the payoffs of group members are independent ($p = 0.004$). As in model 1, the impact of oath is less pronounced under payoff commonality (as suggested by a positive but insignificant coefficient of the interaction term, $p = 0.221$). Oath has no significant impact under payoff commonality (Wald test, $p = 0.313$). Control variables show no significant relationships between age, gender or risk attitudes and the decision to lie. Students with background in business or economics are about 10 percentage point more likely to lie compared to other fields. An additional token observed in the video reduces the likelihood to lie by about 5 percentage point, denoting pecuniary considerations in the decision to lie. Lastly, individual type (measured by the decision in part 1) matters: honest persons, either by choice or by design (i.e., observed ⚄), are 13-15 percentage point less likely to lie compared to those who did. The coefficients of being honest by choice and by design are not significantly different (Wald test, $p = 0.725$).

In all, these analyses indicate a partial support for conjecture 1 and no support for conjecture 2.

Table 1: The Impact of Oath on Lying in Group

	(1)		(2)	
	Coeff.	St.Err.	Coeff.	St.Err.
PC	0.019	(0.059)	0.002	(0.056)
Oath	-0.185***	(0.069)	-0.189***	(0.065)
PC X Oath	0.094	(0.104)	0.118	(0.096)
Age			-0.009	(0.007)
Male			-0.024	(0.037)
SOEP			0.009	(0.008)
Business and Economics			0.103***	(0.034)
Tokens observed in part 2			-0.049**	(0.019)
Honest in part 1			-0.127***	(0.043)
Observed '⚄' in part 1			-0.145***	(0.049)
Constants	0.917***	(0.039)	1.181***	(0.162)
N	378		378	
Cluster	126		126	
R-Squared	0.052		0.169	

Notes: This table reports coefficient estimations from the linear probability models. The dependent variable is a binary indicator whether the participant lied or not in part 2 (coded one for liar, zero otherwise). Groups that observed '⚄' in part 2 were excluded. Model 1 includes dummy variables PC (one if payoff commonality, zero otherwise), Oath (one if oath, zero otherwise) and their interaction term (PC X Oath). Model 2 further controls for socio-demographics, namely age (in years), gender (coded one for male, zero otherwise), self-reported risk attitudes (ranging from 0 to 10), a dummy indicating a field of study in business and economics, the number of tokens observed in part 2 (ranging from 0 to 4), and a categorical variable for individual moral types based on their decision in part 1, namely honest by choice and by design (liar as a reference category). Robust standard errors clustered at the group level in parentheses. ** $p < 0.05$, *** $p < 0.01$

Result 1. The impact of oath on lying extends to a group setting, but it depends on the incentive structure. Oath causes a reduction in lying *only* when group members do not earn a common payoff. Oath has no significant impact on lying when group member payoffs are interdependent. Conjecture 1 is partially supported.

Result 2. The impact of the oath is larger when group members do not earn a common payoff than when they do. Nonetheless, the difference is only marginal. Conjecture 2 is not supported.

5.2 Peer Pressure under Payoff Commonality

The previous section has shown that the impact of oath on lying in group persists only when payoffs of the group members are independent. This section sheds light on the role of peer pressure under payoff commonality in crowding out the impact of oath on honest reporting

by analyzing the individual decision to lie after group interaction (part 3) compared to the decision in the group setting (part 2).

Recall that in part 2, participants reported the die outcome shown on the video after an anonymous chat with group members via computer. The chat allowed group members to exchange justifications to lie, giving rise to *peer conformity* (i.e., I lie because others also lie). However, group incentive can further create *peer pressure* on individuals to lie (i.e., I lie because they want me to). Since group members facing the payoff commonality rule need to coordinate on their reports (else they earn nothing), an individual may be pressured to conform to what the group wants. In other words, when payoffs are interdependent a person might lie out of peer pressure at the cost of neglecting what they want as an individual. This conflict between group vs. individual interests does not exist under no payoff commonality as individual report does not affect others.

How would an individual who were pressured behave differently later compared to those who simply conformed? One might expect that people who lied out of peer pressure should be less likely to lie (again) when they were alone in part 3 because it contradicted what they wanted as an individual. On the other hand, if an individual lied out of conformity, they can be expected to continue to behave in a similar manner in part 3 because the action was not forced (i.e., they learned that lying was acceptable by others). If this is so, the fraction of liars in an individual setting after group interaction is expected to be *lower* in treatments where group members earn a common payoff compared to when they do not.

Looking only at individuals who lied in part 2 and had the opportunity to lie in part 3, the fractions of liars in part 3 in the treatments with no payoff commonality (97.3% in BaseNoPC and 96.5% in OathNoPC) are higher than in the treatments with payoff commonality (90.8% in BasePC and 86.1% in OathPC).[20] In addition, the difference in the fractions in part 3 from part 2 is not significantly different for BaseNoPC and OathNoPC (both $p = 0.500$, signrank tests) but is significantly different for BasePC and OathPC ($p = 0.016$ and $p = 0.002$ respectively, signrank tests). Overall, the pattern in the data is consistent the mechanisms of conformity and pressure under the two incentive structures. Liars who conformed to the group are more likely to lie later in an individual setting, while liars who succumbed to pressure were less likely to lie.[21]

To confirm this analysis, Table 2 reports the coefficient estimations from the linear probability regressions. The dependent variable is a binary indicator whether the participant lied or not in part 3 (1 if liar and 0 otherwise). Individuals that observed '⚄' in part 2 and 3 were

[20] Though they are qualitatively consistent with the pressure mechanism, the differences are insignificant between BaseNoPC and BasePC ($p = 0.167$), and only marginal between OathNoPC and OathPC ($p = 0.065$). If one pools the treatments by the incentive structure (PC vs. NoPC), the difference is significant ($p = 0.011$). In addition, using the same exclusion criteria, the fractions of liars in part 3 from Kocher et al. (2018) exhibit a similar pattern: 90.1% under no payoff commonality and 85.7% under payoff commonality ($p = 0.396$).

[21] Interestingly, looking at those who had the opportunity to lie in part 1 and 3, the fraction of liars in part 3 is significantly higher than that in part 1 for all treatments ($p < 0.001$ in both BaseNoPC and BasePC; $p = 0.002$ in both OathNoPC and OathPC, signrank tests). The same conclusion is reached if groups that observed '⚄' are excluded. This implies that effect of individual oath-swearing does not persist after group interaction.

Table 2: Peer pressure under payoff commonality

	(1)		(2)	
	Coeff.	St.Err.	Coeff.	St.Err.
PC	-0.065*	(0.037)	-0.069**	(0.032)
Oath	-0.008	(0.031)	-0.020	(0.031)
PC X Oath	-0.039	(0.055)	-0.044	(0.053)
Age			-0.025**	(0.010)
Male			-0.006	(0.031)
SOEP			-0.011*	(0.006)
Tokens observed in part 3			-0.029***	(0.011)
Business and Economics			0.009	(0.029)
Honest in part 1			-0.174***	(0.040)
Observed '⚄' in part 1			-0.106**	(0.043)
Constants	0.973***	(0.019)	1.708***	(0.210)
N	278		278	
Cluster	114		114	
R-Squared	0.030		0.218	

Notes: This table reports coefficient estimations from the linear probability models. The dependent variable is a binary indicator whether the participant lied or not in part 3 (coded one for liar, zero otherwise). Individuals that observed '⚄' in part 2 and 3 were excluded. Furthermore, those who were honest in part 2 were excluded to focus only on liars. Model 1 includes dummy variables PC (one if payoff commonality, zero otherwise), Oath (one if oath, zero otherwise) and their interaction term (PC X Oath). Model 2 further controls for socio-demographics, namely age (in years), gender (coded one for male, zero otherwise), self-reported risk attitudes (ranging from 0 to 10), a dummy indicating a field of study in business and economics, the number of tokens observed in part 2 (ranging from 0 to 4), and a categorical variable for individual moral types based on their decision in part 1, namely honest by choice and by design (liar as a reference category). Robust standard errors clustered at the group level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

excluded as they could not lie upward (no one lied downward). Further, as the mechanism concerns lying out of pressure vs. conformity, those who were honest in part 2 were also excluded. Model specifications are the same as in Table 1. The result of model 1 suggests that liars who faced the payoff commonality rule in part 2 are about 7 percentage point less likely to lie than those who did under no payoff commonality (though the coefficient is only weakly significant, $p = 0.079$).[22] The negative sign of the interaction term is consistent with oath increasing the pressure but the coefficient is far from being significant. Model 2 suggests that after accounting for the control variables, liars under the payoff commonality rule were 7 percentage point less likely to lie in part 3 compared to liars who did not face the payoff commonality rule. This main effect of pressure under payoff commonality is significant

[22]Model 1 included the interaction term to keep the presentation of regressions uniform with other tables. However, if one considers a model without an interaction term, the coefficient of PC returns -0.083 ($p = 0.003$). This is consistent with the non-parametric analysis.

(p = 0.031), but oath does not significantly enhance the pressure. Lastly, those who observed higher tokens, and were honest in part 1 (both by choice or by design) were less likely to lie in part 3.

In all, these analyses support the notion that the payoff commonality rule creates pressure on individuals to lie for the group. Individuals who lie under pressure are less likely to lie later compared to those who simply conform to peer misbehavior. This explains the main result where the impact of oath on lying in group only persists when payoffs are not common as peer pressure under payoff commonality crowds-out the impact of oath.

These analyses support the following result:

Result 3. Liars under payoff commonality are less likely to lie later compared to liars under no payoff commonality. This is consistent with the role of peer pressure under payoff commonality, resulting in the crowding out of the impact of oath on honest reporting under payoff commonality.

5.3 Chat Analysis

This section provides additional insights into the content of group communication in part 2. This helps to understand whether or not oath changes the way group members communicate and how communication affects the decision to lie in part 2. It also sheds light on the tension between individuals face between behaving prosocially towards group members and behaving in line with the norm of honesty.

Messages sent during chat were categorized by two independent research assistants who were blind to the objective of the experiment. More specifically, the two coders classified the messages into pre-defined categories based on the argument explicitly made (i.e., whether the argument was made for or against honesty *and* its relation to money, honesty, insecurity, rules, others, consequences or oath). These categories were taken from Kocher et al. (2018), except the ones about the oath. A message may belong to more than one category. Thus, for each group chat, the coders counted the number of times messages with explicit arguments were sent for each category.[23]

This section proceeds by presenting the descriptive statistics of the chat then explores its content by the types of argument (honest vs. dishonest) used by the group members. Finally, the analysis examines the themes of arguments (i.e., whether the argument was made related to money or morality).[24]

[23] As coders might interpret the messages differently, their counts were averaged, which to some extent minimize the bias in case of extreme opinions. The code book used by the two coders is reported in Section B in the Appendix.

[24] Of 134 groups, two groups per treatment observed '⚄' and were excluded from this analysis. In addition, chat data are missing for one group in OathNoPC as none participated in the chat stage. One subject in BaseNoPC did not contribute to the chat. Therefore, there are 125 groups for this content analysis.

Descriptive Statistics of Chat

The average chat duration is 179 seconds (SD = 89.7) during which group members exchanged about 19 messages (SD = 10.4). Table 3 shows the average chat duration and volume of messages by treatment. Interestingly, oath does not impact the length of the conversation nor the volume of messages. As oath does not make a difference to chat duration and volume of messages, pooling the data together based on the incentive structure reveals that the payoff commonality rule reduces both the duration and volume (p = 0.001 and p = 0.032, respectively, ranksum tests). Thus, it is the group incentive, and not the oath, that determines the length and frequency of group communication.

Table 3: Chat Duration and Volume of Messages

	(1)	(2)	(3)	(4)	(1-2)	(1-3)	(2-4)	(3-4)
	BaseNoPC	BasePC	OathNoPC	OathPC		p-value		
Duration (in sec.)	204.25 (91.40)	158 (89.62)	210.33 (82.40)	145 (80.51)	0.047**	0.848	0.559	0.005***
Volume of messages	22.86 (11.70)	17.77 (11.74)	18.27 (9.30)	16.19 (7.49)	0.024**	0.155	0.899	0.387
N	32	31	30	32				

Notes: This table presents average chat duration (in seconds) and the volume of messages by treatment. Standard deviations in parentheses. Figures are reported at the group level. p-values reported using ranksum tests. ** $p < 0.05$ *** $p < 0.01$

Turning to the content of communication, the majority of groups (66.4%) used dishonest arguments at least once, while honest arguments were made at least once in only 33.6% of the groups ($p < 0.001$).[25] Defining participant's moral type based on their decision in part 1, the first proposition about which number to report in part 2 was made more by dishonest individuals than honest individuals (58.5% vs. 41.5%, p = 0.034).[26] Overall, the average share of dishonest arguments (54.1%) is higher than honest arguments (16.3%, $p < 0.001$, within-subjects comparison at the group level using signrank test).[27] This pattern holds for all treatments. However, it is noteworthy that the difference between the shares of dishonest and honest arguments is the least in OathPC (43.5% vs. 22.1%, p = 0.095, signrank test), while the gap between the shares are significantly larger in all the other treatments (all $p < 0.001$, signrank tests). Lastly, oath does not reduce the share of dishonest arguments (58% in pooled baselines vs. 50% in pooled oath treatments, p = 0.264, ranksum test), but

[25] See Table D2 in the Appendix for the number of groups by types of arguments in which honest and/or dishonest argument is mentioned at least once.

[26] The same conclusion holds after excluding those who observed '⚄' in part 1. The proportions become 68% vs. 32%, respectively (p = 0.028). In addition, there are no differences in the proportion of dishonest or honest proposers across treatments (the smallest p-value is 0.197).

[27] Shares of honest and dishonest messages are calculated at the group level (i.e., the number of times messages encouraging honesty were sent divided by the total number of times messages encouraging honesty and dishonesty were sent in a group). This makes the sum of shares of honest and dishonest messages equal 100% at the group level, but not at the treatment level or the global level.

increase the share of honest arguments (12% in pooled baselines vs. 21% in pooled oath treatments, $p = 0.040$, ranksum test).[28]

In sum, the oath does not affect the duration of the chat nor the volume of the messages. Group members used dishonest arguments more often than honest ones in all treatments. Introducing oath increases the share of messages encouraging honesty, but does not reduce the share of messages encouraging dishonesty.

Themes of Arguments

This section further explores the categories of explicit arguments made in the messages that were identified by the two independent coders. The categories were taken from Kocher et al. (2018), in addition to two categories relating to an oath.[29] Messages can be used to encourage (dis)honesty in relation to **Money** (*e.g., 'Reporting 4 is not so bad, it gives us 8 Euros', 'We should report 5 to maximize our gains'*), **Morality** (*e.g., 'We should be honest', 'Respect the oath', 'Lying is ok'*), **Instructions** (*e.g., 'We should stick to the rule/ what the instruction says', 'It is not mentioned that we cannot lie'*) or **Empirical Norms** (*e.g., 'I think other people are honest, 'Many people will lie'*). The analysis focuses on the arguments related to Money and Morality, as they occur most frequently.[30]

Figure 1 displays the frequencies of arguments related to Money and Morality, where the upper panel reports for dishonest use, and the lower panel reports for honest use.[31] Participants frequently used 'money' to encourage group members to lie irrespective of the treatments.[32] Arguments related to money in favor of honesty are rare except OathPC where the frequencies increase but are not different from BasePC ($p = 0.532$). This means that dishonest arguments related to money remain unimpeded by the oath.

On the other hand, participants under oath frequently made morality related arguments in favor of truth-telling than those in the baseline conditions (lower panel; BaseNoPC vs. OathNoPC, $p = 0.060$; BasePC vs. OathPC, $p = 0.057$; pooled baselines vs. pooled oath treatments, $p = 0.006$). As for the dishonest use (upper panel), oath does not reduce the frequencies of immoral arguments to encourage lying under no payoff commonality (BaseNoPC vs. OathNoPC, $p = 0.849$). If anything, immoral arguments increase under oath when group members face payoff commonality (BasePC vs. OathPC, $p = 0.022$). The increase in the frequencies of immoral arguments under oath with payoff commonality may relate to the

[28] See Table D3 in the Appendix for shares of dishonest and honest messages by treatment and pairwise treatment comparisons.

[29] Given a rather large number of categories to begin with, they are afterwards combined for ease of analysis and presentation. The category 'Morality' contains both arguments about 'Honesty' and 'Oath'. The category 'Instructions' contains both arguments about 'Rules' and 'Insecurity.'

[30] The shares of arguments in favor of dishonesty: 57% are related to Money, followed by Morality (24%), Instructions (14%) and Empirical Norms (5%). The shares of arguments in favor of honesty: 59% are related to Morality, followed by Money (20%), Instructions (20%) and Empirical Norms (1%).

[31] See Figure E1 in the Appendix for the frequencies of all types of arguments.

[32] No significant differences across treatments using Fisher's exact tests. Comparing pooled baselines and pooled oath treatments yields ($p = 0.370$).

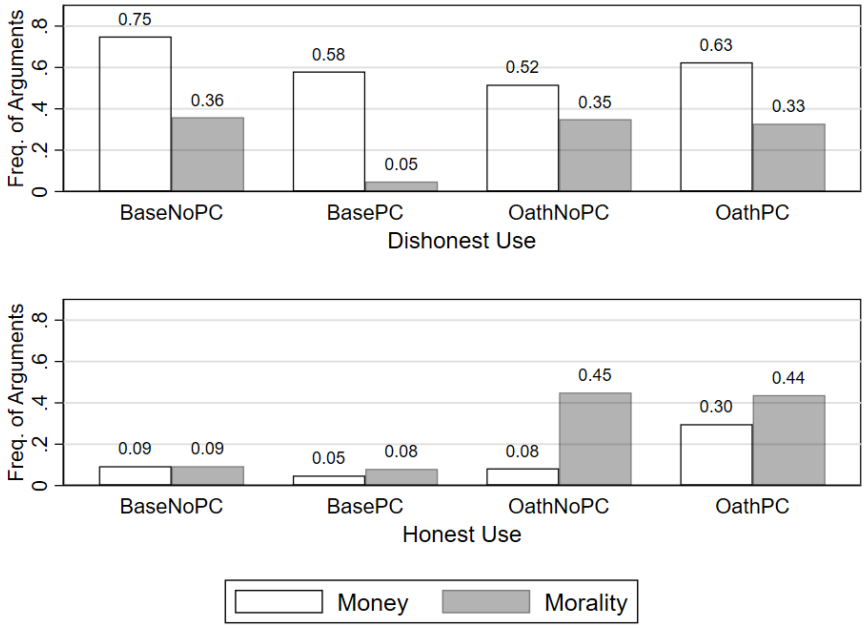

Figure 1: Relative Frequencies of Arguments Related to Money and Morality

Notes: This figure displays the frequencies of arguments related to Money and Morality. Dishonest use (Upper Panel) Honest Use (Lower Panel).

peer pressure mechanism. That is, the payoff commonality rule may push group members to convince the others to lie by justifying that lying is ok or downplaying the oath.

In sum, the chat analysis shows that group members favor lying to maximize their gains (as documented in Kocher et al., 2018) and this still remains a core discussion even when they are under oath. An increase in morality related arguments favoring honesty reflects a more salient norm of honesty under an oath. Overall, the dynamic of group communication reflect the tension group members face between prosociality towards the group and the norm of honesty, as Leib et al. (2021) posited.

Finally, to examine the impact of arguments on the decision to lie in the group setting, Table 4 reports the coefficient estimations from the linear probability model where the dependent variable is a binary indicator whether the participant lied or not in part 2 (coded one for liar, zero otherwise). Model 1 simply replicates the initial result shown in Model 2 of Table 1. Model 2 further includes the arguments used in the chat (i.e., money or morality related, favoring honesty or dishonesty). It can be seen that dishonest arguments referring to money increase the likelihood to lie, while morality related arguments favoring honesty decrease it, which explain most of the treatment effects.

These analyses support the following result:

Result 4. Oath does not impact the duration of communication nor the volume of messages. The content of communication changes under oath as people make more moral arguments to encourage truth-telling. However, immoral arguments to encourage lying by downplaying morality or favoring profit maximization remain high. This dynamic is consistent with the tension between prosociality towards the group and the honesty norm.

Table 4: Impact of arguments on the decision to lie in part 2

Dep. Var.	(1)		(2)	
Lying in part 2	Coeff.	St.Err.	Coeff.	St.Err.
PC	0.002	(0.056)	0.031	(0.054)
Oath	-0.189***	(0.065)	-0.079	(0.066)
PC X Oath	0.118	(0.096)	0.087	(0.094)
Age	-0.009	(0.007)	-0.002	(0.004)
Male	-0.024	(0.037)	-0.020	(0.033)
SOEP	0.009	(0.008)	0.003	(0.006)
Business and Economics	0.103***	(0.034)	0.089***	(0.032)
Tokens observed in part 2	-0.049**	(0.019)	-0.023	(0.016)
Honest in part 1	-0.127***	(0.043)	-0.110***	(0.041)
Observed '⚄' in part 1	-0.145***	(0.049)	-0.116**	(0.045)
Money related & dishonest			0.057**	(0.024)
Money related & honest			-0.064	(0.085)
Morality related & dishonest			0.060	(0.040)
Morality related & honest			-0.247***	(0.070)
Constants	1.181***	(0.162)	0.982***	(0.117)
N	378		375	
Cluster	126		125	
R-Squared	0.169		0.331	

Notes: This table reports coefficient estimations from the linear probability models. The dependent variable is a binary indicator whether the participant lied or not in part 2 (coded one for liar, zero otherwise). Individuals that observed '⚄' in part 2 were excluded. Model 1 shows the initial result shown in Model 2 of Table 1. Model 2 further includes the relative frequencies of arguments made in chat (related to Money or Morality *and* in favor of (dis)honesty). One group in OathPC did not engage in chat, thus excluded from model 2. Robust standard errors clustered at the group level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6 Discussion and Conclusion

To restore and maintain professional integrity, ethical oaths and codes of conduct have been recently introduced in various professions such as bankers, accountants and even academia. While a moral commitment can be demanded from and made by an individual, how effective is such an intervention in reducing dishonesty when individuals are in a group setting? Results from a laboratory experiment reveal that the impact of oath on lying in a group setting depends on the incentive structure – an honesty oath reduces lying in group *only* when the incentives of group members are independent. Interdependency of group member payoffs can crowd-out the effectiveness of oath.

In a group setting, an individual faces a tension between behaving in line with what the group wants vs. what they personally want. The chat analysis shows that individuals

under oath mostly use arguments related to money and morality, favoring both honesty and dishonesty. Thus, the tension within group is between profit maximization (as most of the liars lied to the full extent) and respecting the oath. In such a context where group members earn a common payoff, an individual (who initially may intend to respect the oath) might consider to lie out of pressure to satisfy the group. Thereby, they may end up neglecting their own preference for honesty by favoring the choice of the group. On the other hand, when payoffs of group members are independent, individuals do not face this conflict between behaving prosocially (towards the group at least) and abiding by the norm of honesty. This explains why the incentive structure can modulate the impact of oath in a group setting.

An additional reason why liars in the group setting with payoff commonality were less likely to later lie in an individual setting may also arise from an image concern. If participants believe that they have 'lied a lot' with the group, they may be more reluctant to lie again to dissociate themselves from bad apples and to restore their image with the experimenter (or even in their own eyes). This is consistent with moral balancing (see Monin and Miller, 2001; Ploner and Regner, 2013; Rahwan et al., 2018). Such motives may co-exist in parallel with peer pressure.

An interesting point is that oath loses its power after group interaction as evident from a higher fraction of liars in part 3 compared to part 1. This implies that even though the oath has a desirable effect on lying in group when members earn independent payoffs, individuals might have been exposed to the justifications of others during the group communication, which could dampen the effect of oath on individual lying behavior overall. This is in contrast to the findings of Peer and Feldman (2021) and Kingsuwankul et al. (2023) – both of which used individual lying games. Nonetheless, participants in these two studies could not observe behaviors of others nor could they communicate with their peers. These channels of norm erosion exist in the setting of the current study. Future research can explore this avenue.

In all, the findings of this study suggest that while an honesty oath can foster honesty of individuals in a group setting, the desired impact becomes negligible when group members earn a common payoff due to peer pressure from group members. While one still needs to remain cautious about the external validity of these findings in a more complex environment like that of a real organizational setting, they highlight how practitioners must pay close attention to the way an oath or a compulsory code of conduct is introduced within an organizational context as it must be designed in close connection with the incentive schemes. In addition, it suggests that oath may not be a *be-all-end-all* policy to globally limit dishonesty in a collective setting. Thus, firms may benefit from optimizing this 'soft' intervention, for example, by incorporating the social identity dimension such as proximity, which has been shown to halt norm erosion (Bicchieri et al., 2022), or implementing it in tandem with 'hard' interventions, such as internal auditing.

References

Bauer, M., J. Cahlíková, D. Celik-Katreniak, J. Chytilová, L. Cingl, and T. Želinsk (2021). Nastiness in Groups. *CERGE-EI Working Paper Series*.

Beck, T. (2021, October). How the honesty oath works: Quick, intuitive truth telling under oath. *Journal of Behavioral and Experimental Economics 94*, 101728.

Bicchieri, C., E. Dimant, S. Gächter, and D. Nosenzo (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior 132*, 59–72.

Bock, O., I. Baetge, and A. Nicklisch (2014). hroot: Hamburg Registration and Organization Online Tool. *European Economic Review 71*, 117–120.

Cagala, T., U. Glogowsky, and J. Rincke (2021). Detecting and Preventing Cheating in Exams: Evidence from a Field Experiment. *Journal of Human Resources*, 0620. Publisher: University of Wisconsin Press.

Cagala, T., U. Glogowsky, J. Rincke, and S. Schudy (2023). Commitment Requests Do Not Affect Truth-Telling in Laboratory and Online Experiments.

Castillo, G., L. Choo, and V. Grimm (2022). Are groups always more dishonest than individuals? The case of salient negative externalities. *Journal of Economic Behavior & Organization 198*, 598–611.

Charness, G. and M. Sutter (2012). Groups Make Better Self-Interested Decisions. *Journal of Economic Perspectives 26*(3), 157–176.

Charroin, L., B. Fortin, and M. C. Villeval (2022, August). Peer effects, self-selection and dishonesty. *Journal of Economic Behavior & Organization 200*, 618–637.

Colzani, P., G. Michailidou, and L. Santos-Pinto (2023). Experimental evidence on the transmission of honesty and dishonesty: A stairway to heaven and a highway to hell. *Economics Letters 231*, 111257.

Conrads, J., B. Irlenbusch, R. M. Rilke, and G. Walkowitz (2013). Lying and team incentives. *Journal of Economic Psychology 34*, 1–7.

Davis, B. J. and T. Jaber-Lopez (2022). Do voluntary commitment mechanisms improve welfare? The effect of mandatory and voluntary oaths in a social dilemma. *Bulletin of Economic Research*.

Diekmann, A., W. Przepiorka, and H. Rauhut (2015). Lifting the veil of ignorance: An experiment on the contagiousness of norm violations. *Rationality and Society 27*(3), 309–333.

Dunaiev, Y. and M. Khadjavi (2021). Collective Honesty? Experimental Evidence on the Effectiveness of Honesty Nudging for Teams. *Frontiers in Psychology 12*.

Falk, A., T. Neuber, and N. Szech (2020). Diffusion of Being Pivotal and Immoral Outcomes. *The Review of Economic Studies 87*(5), 2205–2229.

Falk, A. and N. Szech (2013). Morals and Markets. *Science 340*(6133), 707–711.

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

Fischbacher, U. and F. Föllmi-Heusi (2013). Lies in Disguise—an Experimental Study on Cheating. *Journal of the European Economic Association 11*(3), 525–547.

Fosgaard, T. R., L. G. Hansen, and M. Piovesan (2013). Separating Will from Grace: An experiment on conformity and awareness in cheating. *Journal of Economic Behavior & Organization 93*, 279–284.

Gneezy, U., A. Kajackaite, and J. Sobel (2018). Lying Aversion and the Size of the Lie. *American Economic Review 108*(2), 419–453.

Gross, J., M. Leib, T. Offerman, and S. Shalvi (2018). Ethical Free Riding: When Honest People Find Dishonest Partners. *Psychological Science 29*(12), 1956–1968.

Guiso, L., P. Sapienza, and L. Zingales (2008). Trusting the Stock Market. *The Journal of Finance 63*(6), 2557–2600.

Heinicke, F., S. Rosenkranz, and U. Weitzel (2019). The effect of pledges on the distribution of lying behavior: An online experiment. *Journal of Economic Psychology 73*, 136–151.

Innes, R. and A. Mitra (2013). Is dishonesty contagious? *Economic Inquiry 51*(1), 722–734.

Jacquemet, N., A. James, S. Luchini, J. Murphy, and J. Shogren (2021). Do truth-telling oaths improve honesty in crowd-working? *PLoS ONE 16*.

Jacquemet, N., S. Luchini, J. Rosaz, and J. F. Shogren (2019). Truth Telling Under Oath. *Management Science 65*(1), 426–438.

Jacquemet, N., S. Luchini, J. Rosaz, and J. F. Shogren (2021). Can we commit future managers to honesty? *Frontiers in Psychology*, 2785.

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

Keizer, K., S. Lindenberg, and L. Steg (2008). The spreading of disorder. *Science 322*(5908), 1681–1685.

Kingsuwankul, S., C. Tergiman, and M. C. Villeval (2023). Why do oaths work? Image concerns and credibility in promise keeping.

Kocher, M. G., S. Schudy, and L. Spantig (2018). I Lie? We Lie! Why? Experimental Evidence on a Dishonesty Shift in Groups. *Management Science 64*(9), 3995–4008.

Koessler, A.-K., B. Torgler, L. P. Feld, and B. S. Frey (2019). Commitment to pay taxes: Results from field and laboratory experiments. *European Economic Review 115*, 78–98.

Kroher, M. and T. Wolbring (2015). Social control, social learning, and cheating: Evidence from lab and online experiments on dishonesty. *Social Science Research 53*, 311–324.

Köbis, N. C., B. Verschuere, Y. Bereby-Meyer, D. Rand, and S. Shalvi (2019). Intuitive Honesty Versus Dishonesty: Meta-Analytic Evidence. *Perspectives on Psychological Science 14*(5), 778–796.

Lefebvre, M., P. Pestieau, A. Riedl, and M. C. Villeval (2015). Tax evasion and social information: an experiment in Belgium, France, and the Netherlands. *International Tax and Public Finance 22*(3), 401–425.

Leib, M., N. Köbis, I. Soraperra, O. Weisel, and S. Shalvi (2021). Collaborative dishonesty: A meta-analytic review. *Psychological Bulletin 147*(12), 1241–1268.

Monin, B. and D. T. Miller (2001). Moral credentials and the expression of prejudice. *Journal of personality and social psychology 81*(1), 33.

Peer, E. and Y. Feldman (2021). Honesty pledges for the behaviorally-based regulation of dishonesty. *Journal of European Public Policy 0*(0), 1–21.

Ploner, M. and T. Regner (2013). Self-image and moral balancing: An experimental analysis. *Journal of Economic Behavior & Organization 93*, 374–383.

Prima, R. A., S. Feeny, R. Hoffmann, and E. Satriawan (2020). THREATS, PLEDGES, AND ASSET MISREPORTING: A FRAMED FIELD EXPERIMENT IN INDONESIA.

Rahwan, Z., O. P. Hauser, E. Kochanowska, and B. Fasolo (2018). High stakes: A little more cheating, a lot less charity. *Journal of Economic Behavior & Organization 152*, 276–295.

Sapienza, P. and L. Zingales (2012). A Trust Crisis. *International Review of Finance 12*(2), 123–131.

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

Weisel, O. and S. Shalvi (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences 112*(34), 10651–10656.

Wiltermuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes 115*(2), 157–168.

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.

Zickfeld, J., K. A. Ścigała, A. Weiss, J. Michael, and P. Mitkidis (2022). Investigating the Impact of Social Commitment on Dishonest Behavior.

Appendix A Instructions (Translated from French)

The general instruction was given in print and read aloud at the beginning of the session. Then, the instruction for each part of the experiment was given via the computer screen. Instructions, both general and on-screen, and comprehension questions are kept as close as possible to those used in Kocher et al. (2018).

Instructions

Hello and welcome to this experiment in decision making. Please turn off your phone and put it away. You are not allowed to communicate with other participants during the experiment, unless invited to do so by the experimenters. You may be excluded from the session and gains. During this session, you can earn money. The amount you will earn depend on your decisions and the decisions of other participants in the session. Therefore, please read the instructions carefully.

Your earnings

During the session, we will not talk in Euros but in tokens. The conversion rate is as follows:

1 Token = 2 Euros

This session consists of three successive and independent parts. One of the three parts will be randomly selected to determine your earnings at the end of the session. At the end of the session, a randomly selected participant will roll a die. If the die displays ⚀ or ⚁, your gains in the first part will determine your earnings for the session. If the die displays ⚂ or ⚃, your gains in the second part will determine your earnings for the session. And if the die displays ⚄ or ⚅, your gains in the third part will determine your earnings for the session. Since all the parts have the same probability of being selected for payment, it is in your best interest to make your decision in each part as if that part were the one that counts towards your earnings.

The amount earned at the end of this session is the sum of your tokens in the selected part and your participation fee of 5 Euros. No one will know how much you have won, nor will you be informed of the earnings of the other participants.

You will be paid by transfer by the CNRS to your bank account. To do this, at the end of the session of the session we will ask for your IBAN. We will also ask you to to send us at the end of the session a bank statement (containing IBAN) in pdf format with the title "your first name-your last name" to the following email address: gatelab[at]gate.cnrs.fr.

Given the administrative and banking delays, the transfer may take up to two weeks. We commit ourselves that your banking information will be used only for this transfer. This information will be kept separate from the files containing your decisions and the researchers who will process your decisions will not have access to your banking information.

Anonymity of your decisions

All your decisions and responses will be anonymous. We will never link your name with the data generated in the experiment. You will not know the identity of the other participants, either before or after the experiment. The other participants will not know your identity either.

At the beginning of each part, you will receive detailed instructions on your screen. If you have any questions about the instructions or during the experiment, please raise your hand or press the red button on the side of your desk. We will come and answer your questions immediately in private.

-END OF GENERAL INSTRUCTIONS-

-BEFORE INSTRUCTIONS OF PART 1-

[All instructions after this point were displayed on the screen. After the welcome page, participants were asked two questions using SAM (Self-Assessment Manikin) to measure valence and arousal at the beginning and the end of the experiment.]

(Valence question)
Please choose an image that best describes your current feeling.

For example, if you feel deeply happy or satisfied, you can indicate this by choosing the image located on the right-most part of the screen (number 7). Conversely, if you feel deeply unhappy or dissatisfied, you can also indicate this by choosing the image located in the leftmost part of the screen (number 1). These images also allow you to describe a feeling of intermediate happiness by selecting one of them.

[Seven images of valence question displayed here]

(Arousal question on new screen)
Please choose an image that best describes your current feeling.

For example, if you are feeling extremely excited or tense, you can indicate this by choosing the rightmost image (number 7). Conversely, if you feel completely relaxed or calm, you can indicate this by choosing the leftmost image (number 1). These images also allow you to describe an intermediate feeling of excitement by selecting one of them.

[Seven images of arousal question displayed here]

-Instructions for Part 1-
(Same for all treatments)

The die roll
In part 1, you will see a video of a die roll, showing (⚀, ⚁, ⚂, ⚃, ⚄ or ⚅).
The video you will see is chosen randomly by the program and each of the 6 videos is chosen with the same probability. *(For session 3 and onward: These draws have been conducted before the start of the session.)*

Your task
Your task is to memorize the outcome of the die roll shown in the video and enter it later.

Your gains for this task The number you enter determines your gains in tokens as follows:

Number Entered	'1'	'2'	'3'	'4'	'5'	'6'
Tokens Gained	1	2	3	4	5	0

Comprehension Questions

1. What is your task in this part?

   - Enter the number displayed that you have memorized
   - Enter the number different from the one seen and memorized
   - Enter a random number

2. Suppose that you saw ⚂ and you enter "3", how many tokens do you receive?
3. Suppose that you saw ⚂ and you enter "2", how many tokens do you receive?
4. Suppose that you saw ⚂ and you enter "4", how many tokens do you receive?

-Before Instructions for Part 2-
(OathNoPC and OathPC Only)

Before continuing the experiment, we ask all participants to take an oath.

With this oath, each participant swears that for the remainder of the experiment, he or she will be honest and will always tell the truth.

Please click "Next" to take the oath.
(Next Button)

You must now swear the following oath: "I swear upon my honor that for the rest of the experiment, I will be honest and that I will always tell the truth."

Please write this oath in the box below, then click the "Next" button.
(Next Button)

-Instructions for Part 2-
(Same for all treatments, except sentence about the payoff rule)

Part 2 of the experiment is similar to part 1, except that you will now decide in group. You will be randomly matched with two other participants to form a group of three.

The die roll
In part 2, you will see a video of a die roll, showing (⚀, ⚁, ⚂, ⚃, ⚄ or ⚅).

The video you will see is chosen randomly by the program and each of the 6 videos is chosen with the same probability. *(For session 3 and onward: These draws have been conducted before the start of the session.)* All the three members in your group will see the same video chosen at random.

Your task
Your task is to memorize the outcome of the die roll shown in the video and enter it later.

Your gains for this task The number you enter determines your gains in tokens as follows:

Number Entered	'1'	'2'	'3'	'4'	'5'	'6'
Tokens Gained	1	2	3	4	5	0

(The sentence for BaseNoPC & OathNoPC)
The numbers entered by the other two group members do not affect your gains in tokens in this part. Similarly, the number you enter does not affect the gains of the other members of your group.

(The sentence for BasePC & OathPC)
The numbers entered by the other two group members affect your gains in tokens in this part. If the numbers entered by all group members are not the same, all members of your group receive zero tokens in this part.

Comprehension Questions

1. What is your task in this part?

   - Enter the number displayed that you have memorized
   - Enter the number different from the one seen and memorized
   - Enter a random number

2. Suppose that you saw ⚂ and you enter "3", how many tokens do you receive, if ...

- the members in your group enter the same number?
- the members in your group enter a different number?

3. Suppose that you saw ⚂ and you enter "2", how many tokens do you receive, if ...
   - the members in your group enter the same number?
   - the members in your group enter a different number?
4. Suppose that you saw ⚃ and you enter "4", how many tokens do you receive, if ...
   - the members in your group enter the same number?
   - the members in your group enter a different number?

-Chat Stage in Part 2-
(Same for all treatments)

You now have the opportunity to chat with the other two members of the group to clarify about the number that each group member must enter.

You have 5 minutes to exchange information. The group chat ends after 5 minutes or as soon as all the 3 group members have pressed the button "Leave chat". If only 1 or 2 group members press the button, the chat will continue either until all of them press the button or until the time is up. If you have pressed the button "Leave Chat" but do not want to leave the chat, you can press the button "Back". After the group chat, each member of the group enters the number of the die roll.

You are free to choose the content of the communication, but it is not allowed to mention any personal information (such as name, age, gender, school, field of study or any other identifying details, e.g., seat number). In addition, you are not allowed to agree to any side payment outside of the experiment. If you break any of these rules, you will be excluded from the experiment and will not receive any payment.

During the chat, each group member can send as many messages as they want to other group members. Each of your messages will automatically appear on the screen of the other two group members. It is not possible to send messages to only one person.

The chat screen looks like this

(Screenshot of Chat Stage appears here)

To write a message, click on the purple tab, type your message and press the enter key on your keyboard. Your message will appear in the gray box above. Other participants in your group will not see your message until you press "Enter".

Comprehension Questions
Suppose that you have pressed the button "Leave Chat", when do you leave the chat stage?

- Immediately
- When you press the button "Back"
- When all the members of your group have pressed the button "Leave Chat" or when the time limit expires

-Instructions for Part 3-
(Same as in part 1 and in all treatments)

-Belief Elicitation Screens-
(Same for all treatments)

Question about part 1

Please answer as accurately as possible.
In part 1, each participant saw his or her video of the die outcome and reported the number.

(Actual distribution of the video of die outcomes displayed here)

Please estimate the percentages of participants who entered each of the numbers in Part 1. If at the end of the session, this question is selected by the program, you will receive 5 euros. For each percentage point by which your answers differ from the correct values, your payout will be reduced by 4 cents. The smallest possible amount for this question is 50 cents.

- What is the percentage of participants that entered the number "1"?
- What is the percentage of participants that entered the number "2"?
- What is the percentage of participants that entered the number "3"?
- What is the percentage of participants that entered the number "4"?
- What is the percentage of participants that entered the number "5"?
- What is the percentage of participants that entered the number "6"?

Question about part 2

Please answer as accurately as possible.
In Part 2, each participant was assigned to a group of three, watched the same video of the die roll result, and reported the number after the group discussion.

(Actual distribution of the video of die outcomes displayed here)

Please estimate the percentages of participants who entered each of the numbers in Part 2. If at the end of the session, this question is selected by the program, you will receive 5 euros. For each percentage point by which your answers differ from the correct values, your payout will be reduced by 4 cents. The smallest possible amount for this question is 50 cents.

- What is the percentage of participants that entered the number "1"?
- What is the percentage of participants that entered the number "2"?
- What is the percentage of participants that entered the number "3"?
- What is the percentage of participants that entered the number "4"?
- What is the percentage of participants that entered the number "5"?
- What is the percentage of participants that entered the number "6"?

-Belief Elicitation of Part 3-
(Same as the belief elicitation of part 1)

-AFTER BELIEF ELICITATION OF PART 3 AND BEFORE QUESTIONNAIRE-

[Before answering questionnaire, participants were asked the same Self-Assessment Manikin questions to measure valence and arousal. After the two SAM questions, participants respond to the questionnaire.]

- Age
- Gender
- Status
- School
- Self-reported risk attitude (SOEP): How do you see yourself: are you generally a person who is very willing to take risks or do you try to avoid taking risks? From 0 'Not at all' to 10 'Very ready to take risk.'
- In Part 2, you were able to communicate with your group members before reporting your number. Please explain the reasons for your decision in Part 2.
- Did you feel pressure from others to make a decision during Part 2? Please give an answer choosing between the numbers 1 and 7, with 1 if you felt no pressure and 7 if you felt extreme pressure.

[After all participants completed the questionnaire, the experimenter approached one participant at random who then drew a raffle indicating a seat number. The participant seated at that computer then threw a six-sided die to determine the part relevant for the bonus payment. The participant then announced the die outcome and the relevant part. After that participants saw on the computer screen the payoff for each part and for the belief question selected by the programme.]

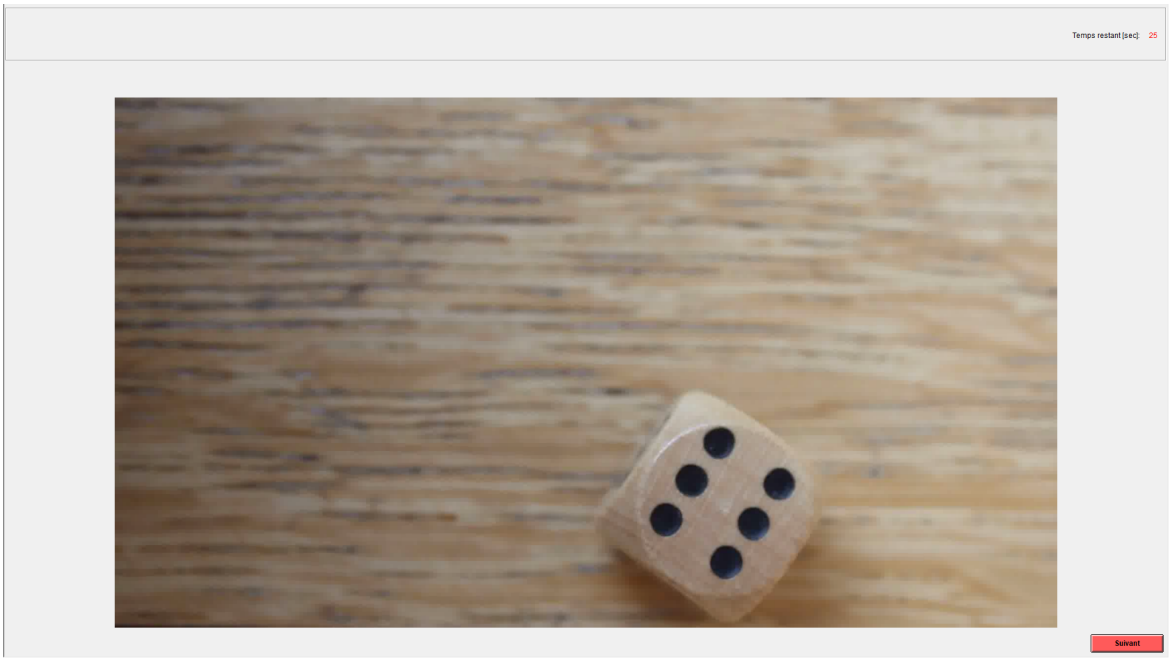

Figure A.1: Sample Video of Die Outcome

Note: This figure displays the computer screen in which a die outcome is shown through a video downloaded from the website of one of the authors of Kocher et al. (2018).

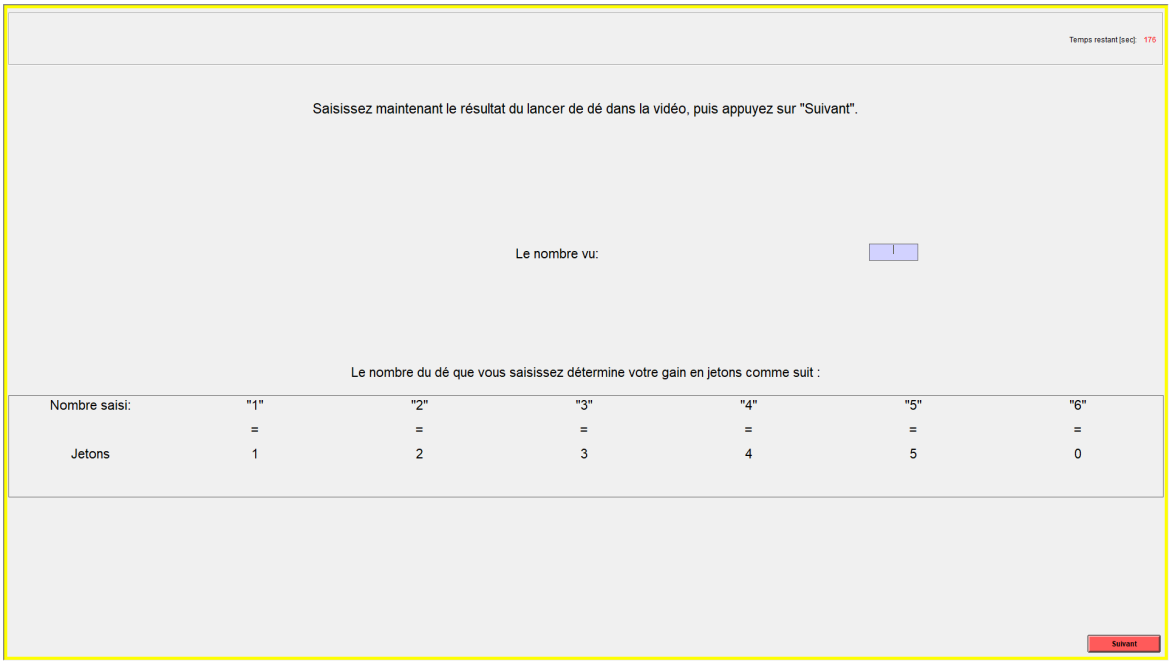

Figure A.2: Decision Screen in Part 1

Note: This figure displays the decision screen in part 1 of the experiment.

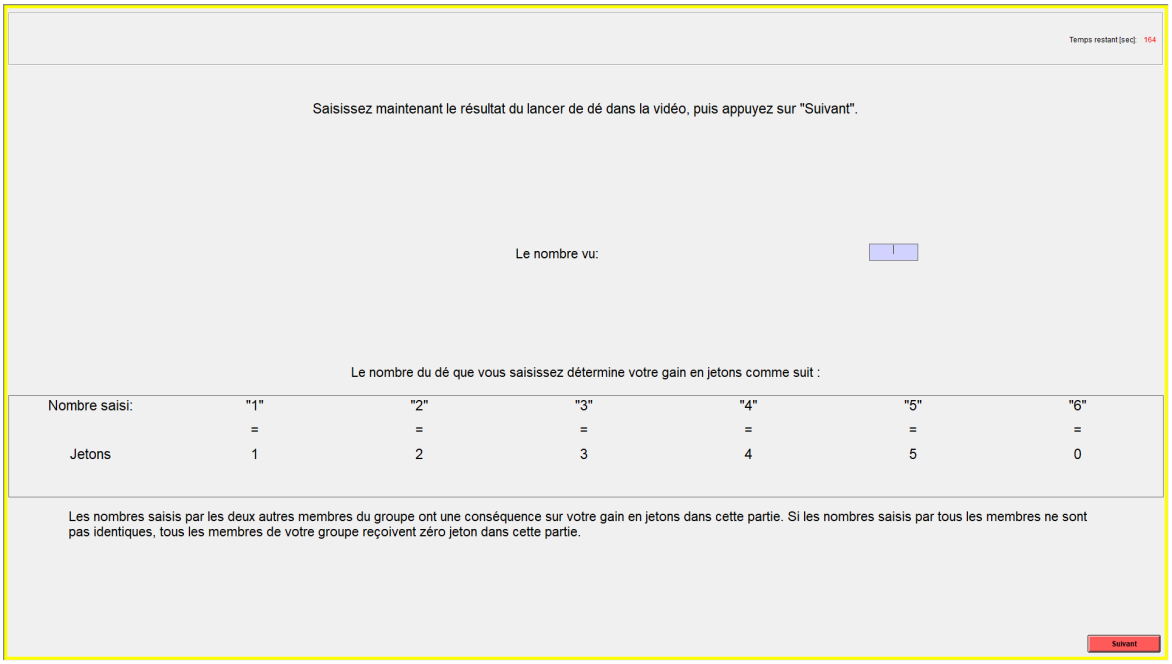

Figure A.3: Decision Screen in Part 2

Note: This figure displays the decision screen in part 2 of the experiment. The description below the payoff table varies according to the payoff commonality rule.

Figure A.4: Sample Chat Stage in Part 2

Note: This figure displays the chat stage in part 2 of the experiment. Participants can type the messages on the purple bar below. After pressing “Enter”, the message are sent to the whole group, appearing in the grey box above.

Appendix B Code Book for Content Analysis (Translated from French)

Instructions

Your job is to code messages exchanged by participants in an experiment.

Context of the task: You will read excerpts of conversations exchanged during an experiment that has already taken place. They are messages from a chat where participants communicate with other members of their group. More precisely, there were three members in each group (A, B and C).

Before they could interact with each other, they each saw a video on their screen displaying a roll of the die (⚀, ⚁, ⚂, ⚃, ⚄ or ⚅) that was randomly chosen by the program. Each video had the same probability of being chosen. All three members saw the same video. Their task was to memorize the number displayed on the die so that they could enter it later on their computer. The money they earned depended on the number they entered and was determined as follows:

Number Entered	'1'	'2'	'3'	'4'	'5'	'6'
Tokens Gained	1	2	3	4	5	0

One token was exchangeable for two Euros at the end of the experiment. Before entering their numbers individually, group members were given the opportunity to chat for 5 minutes. Although some of the messages are neutral, they can help you understand the discussion between the group members.
Data: For each conversation excerpt, you will have the following information (variable name in parentheses):

1. Group identifier (Group_id)
2. Subject identifier (Subject_id)
3. Name in chat (NameInChat → A, B or C)
4. Message sent by the participant to two other members of his/her group (Messages)
5. Time (in seconds), indicating the second the message was sent (TimeStampInChat)
6. Number of the die viewed in the video in part 2 (RandNumOfGroup)

Coding: Your job is to sort the messages into different categories based on the arguments participants use to persuade others in the group about the number to enter. You will see examples associated with each category.

IMPORTANT: Only count arguments that are EXPLICITLY mentioned. For example: if "⚂" was shown and the participant said. "I think we should be honest and enter 3" honesty=1 ; "I think we should enter 3" honesty=0.

List of categories:

1. **Money_dishon:** Discussion about money in favor of lying (# of times money/points/chips are mentioned). Ex: *"We should enter 5 because that will maximize our winnings."*
2. **Money_hon:** Discussion about money in favor of honesty (# of times money/points/chips are mentioned). Ex: *"We should enter 4, that will still give us 8 euros."*
3. **Honesty_dishon:** Discussion about honesty in favor of lying (# of times mentioned). Ex: *"There is no need to be honest; it's okay to lie."*
4. **Honesty_hon:** Discussion about honesty in favor of honesty (# of times mentioned). Ex: *"Let's be honest and report the number we saw."*

5. **Insecurity_dishon:** Discussion of uncertainty in favor of lying (# of times uncertainty or doubt is mentioned). Ex: *"I don't understand, I thought we could enter any number."*

6. **Insecurity_hon:** Discussion of uncertainty in favor of honesty (# of times uncertainty or doubt is mentioned). Ex: *"I don't understand, I thought we had to enter the number we saw earlier."*

7. **Rules_dishon:** Discussion of rules in favor of lying (# of times rules/instructions are referred to). Ex: *"We don't have to follow the rules; it is not mentioned in the instructions that we can't lie."*

8. **Rules_hon:** Discussion of rules in favor of honesty (# of times rules/instructions are referred to). Ex: *"We have to follow the rules; the instructions say to enter the number seen in the video."*

9. **Others_dishon:** Referring to the dishonesty of others outside the group (e.g., the experimenter, people in general, other participants, etc.) (# of times mentioned). Ex: *"Everyone lies; I don't believe they (the experimenters) showed us a random video."*

10. **Others_hon:** Refer to the honesty of others outside the group (e.g., the experimenter, people in general, other participants, etc.) (# of times mentioned) Ex: *"I think the other participants would be honest."*

11. **Neg_consq:** Discussion of negative consequences (# of times they talk about hurting conscience, breaking the oath, fear of punishment). Ex: *"It hurts your conscience if you lie; will we be punished if we break the oath/pledge?"*

12. **Pos_consq:** Discussion of positive consequences (# of times they talk about how lying could have a positive outcome). Ex: *"The group gains more if we all lie."*

13. **Oath_dishon:** Discussion of the oath in favor of lying (# of times mentioned). Ex: *"The oath is useless."*

14. **Oath_hon:** Discussion of the oath in favor of honesty (# of times mentioned). Ex: *"Respect the oath."*

A sentence can belong to several categories. Ex: *"Let's be honest and put 3. We will earn 6 euros anyway" Money_hon =1 AND Honesty_hon =1*

For each group, you must also note the 'Subject_id' of the first subject who proposed the number to be entered.

Appendix C Additional Analyses

This section reports additional analyses not included in the pre-registration. Section C.1 looks at group coordination on (dis)honest reports. Section C.2 examines the beliefs about the empirical norms of honesty. Section C.3 examines the response times under oath.

C.1 Group Coordination

This section explores the coordination rates in part 2 of the experiment. Kocher et al. (2018) analyzed group coordination under the two incentive schemes. In addition, Jacquemet et al. (2018) found in the setting of a coordination game that a truth-telling oath increased coordination rates by nearly 50%, resulting in an increase in efficiency because senders' message became more truthful while followers were more willing to trust.

Table C.1 displays the coordination rates at the group level by treatment. The rates in BaseNoPC and BasePC are high (84.4% and 100%, respectively, p = 0.053, Fisher's exact test).[33] Introducing oath reduces coordination rates, but only when there is no payoff commonality (84.4% in BaseNoPC vs. 54.8% in OathNoPC, p = 0.014, Fisher's exact test).

Table C.1: Coordination Among Group Members by Treatment

Coordination	Nb. Dishonest	BaseNoPC	BasePC	OathNoPC	OathPC
Yes (all honest)	0	1	2	2	5
No	1	1	0	5	0
No	2	3	0	9	0
No (all dishonest)	3	1	0	0	0
Yes (all dishonest)	3	26	29	15	27
Total Nb. of Groups		32	31	31	32
Coordination rates		84.4%	100%	54.8%	100%

Notes: This table presents coordination rates of group members across treatments. Figures are collapsed at the group level. Groups that observed '⚄' in part 2 were excluded.

Looking at whether groups coordinated on dishonest or honest reports (or remained uncoordinated) yield interesting findings. When group members do not face payoff commonality, oath reduces coordination rates on dishonest reports (81.3% in BaseNoPC vs. 48.4% in OathNoPC, p = 0.008), but does not increase coordination rates on honest reports (3.1% in BaseNoPC vs. 6.3% in OathNoPC, p = 0.613, Fisher's exact tests). This means that when payoffs are independent, introducing oath moves people away from dishonest coordination towards respecting their individual honesty oath. The rates at which groups remained uncoordinated (i.e., 1 or 2 group members being honest) increase from 12.5% in BaseNoPC to 45.2% in OathNoPC (p = 0.005, Fisher's exact test). On the other hand, under payoff commonality, oath does not significantly reduce coordination rates on dishonest reports (93.5% in BasePC vs. 84.4% in OathPC, p = 0.426) nor does it increase honest coordinations (6.5% in BasePC vs. 15.6% in OathPC, p = 0.426, Fisher's exact tests).

In sum, an honesty oath breaks down coordination on dishonest reporting when groups do not face payoff commonality – a setting where only peer conformity plays a role. On the other hand, the oath has no impact on coordination rates under payoff commonality – a setting where both conformity and pressure persist.

These analyses lead to the following observation:

Observation 1. Coordination rates in the baseline conditions are close to those reported in Kocher et al. (2018). An honesty oath breaks down group coordination on dishonest reporting when group member do not

[33] These rates are very close to those reported in Kocher et al. (2018): 84.6% in treatment without payoff commonality and 100% under payoff commonality.

earn a common payoff as people are likely to remain honest independently of others. However, when group member payoffs are interdependent, oath has no impact on coordination rates.

C.2 Belief about Others' (Dis)honesty

This section reports the analysis of individuals' beliefs about the empirical norms. Beliefs were elicited at the end of the experiment (before receiving feedback about the payoff). More precisely, for each part of the experiment, participants had to guess the proportion of individuals who reported each die number within the session. The analysis focuses on the participants' belief about the proportion of individuals reporting '5' as it indicates their belief about the level of dishonesty in the session.

Belief about Group (Dis)honesty

Overall, participants' belief about the level of group dishonesty is consistent with actual behavior. Participants in BaseNoPC and BasePC anticipated the proportion of people who reported '5' to be around 79.6% and 85.0% respectively (p = 0.113, ranksum test). However, when participants were under oath, the beliefs about the level of group dishonesty were 63.51% in OathNoPC and 73.2% in OathPC, which are significantly lower compared to their respective baselines.[34] In addition, the belief about group dishonesty under oath with payoff commonality is significantly higher than that without payoff commonality (p = 0.014, ranksum test), which is consistent with the actual pattern in the data.

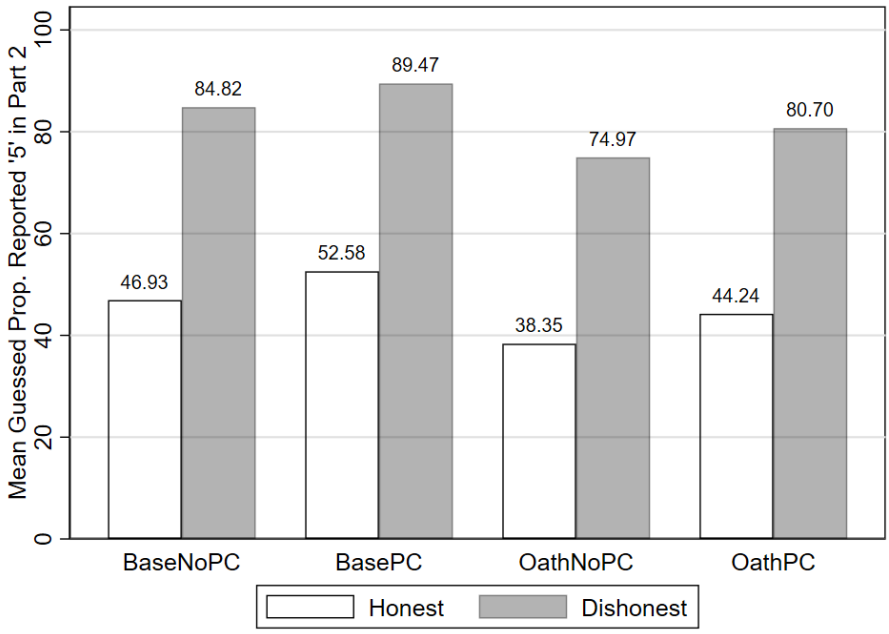

Figure C.1: Belief about Group (Dis)honesty by Type and Treatment

Notes: This figure displays the belief about the proportion of participants reporting '5' in a group setting by treatment, conditioned on their decision in part 2 (honest or dishonest).

Complementing the above analysis, Figure C.1 displays the average guessed proportion of participants who reported '5' in part 2, conditioned on the fact that they lied or not in part 2. The choice of conditioning is because participants could infer about group dishonesty after the chat (which allowed for inference about what others might do) and what they decided for themselves. The pattern of belief of both honest and dishonest participants is consistent with the behavioral data: individuals under oath anticipated a lower level of group dishonesty compared to those in the baseline, and they believed that the impact of oath is weaker when group members earn a common payoff. However, it is interesting to note that in all treatments, while

[34] Using ranksum tests: BaseNoPC vs. OathNoPC, $p < 0.001$; BasePC vs. OathPC, $p = 0.001$.

honest individuals were optimistic about the impact of oath, dishonest individuals were rather pessimistic (honest vs. dishonest $p < 0.001$ in all treatments, ranksum tests). The mean differences between the true proportion and the guess of liars are about 4-7 percentage points, which are significantly smaller than those reported by honest people (between 27 - 39 percentage points, both $p < 0.001$ in BaseNoPC and BasePC; $p = 0.002$ and $p = 0.003$ in OathNoPC and OathPC, respectively, ranksum tests).[35]

Change in Beliefs about Individual (Dis)honesty

In part 1 and 3 of the experiment, participants played the observed cheating game at an individual setting, but differed in that in the latter, they had interacted with two group members via computer chat. How did that change individual belief about others' behavior before and after the group interaction?

Given no difference in the die numbers observed in part 1 and 3 in each treatment (the smallest p-value is 0.767, signrank tests), if group interaction change the belief participants have about others' (dis)honesty at the individual level, it can be captured by the difference between the beliefs of others' (dis)honesty reported for part 1 and 3.[36]

Overall, participants in the BaseNoPC and BasePC became more pessimistic about the level of individual dishonesty. The average change in their belief about the proportion of participants reporting '5' is 12.22 in BaseNoPC and 14.48 in BasePC ($p = 0.661$, ranksum test). That is, after group interaction, participants in the baseline conditions anticipated an upward shift in individual dishonesty by about 12-14 percentage points. The shifts in belief are far smaller in the two oath treatments. On average, the change in their belief about the level of individual dishonesty is 2.65 in OathNoPC and and 6.14 in OathPC, which are significantly smaller than their respective baselines.[37] This means that while participants in the baseline conditions anticipated group interaction to make people become more dishonest in part 3, those in OathNoPC and OathPC seemed to believe that people would be, more or less, as (dis)honest as before the group interaction and oath-taking. This may be the case because the effects of group interaction and oath on dishonesty are opposing each other.

To complement the above analysis, Figure C.2 displays the average change in belief about the proportion of individuals reporting '5' in part 1 and 3, conditioned on the fact that they lied or not in part 2.[38] This explains why the changes in belief are smaller in OathNoPC and OathPC. Those who were honest and dishonest in the group setting in BaseNoPC and BasePC tended to update their belief about individual dishonesty upward (that is, expecting higher level of dishonesty after group interaction). The patterns look different in OathNoPC and OathPC – honest people revised their belief about individual dishonesty downward (that is, expecting higher level of honesty after group interaction) due to the introduction of oath. However, liars in the group setting in OathNoPC and OathPC updated their belief about individual dishonesty upward in a similar fashion to those in the baseline conditions.[39]

Finally, comparing the pooled baseline conditions to pooled oath treatments, the difference in the change in beliefs about individual dishonesty is significant both for honest and dishonest people ($p < 0.001$ and $p = 0.007$, respectively, ranksum tests).[40] This means that without any oath, people anticipate group interaction to drive individual dishonesty upwards. On the other hand, oath leads to a lower anticipation of individual

[35] Excluding those who viewed '⚄' in part 2 yield the same results.

[36] Change in beliefs is calculated by subtracting the guess about the proportion of participants reporting '5' for part 1 from the guess for part 3. Thus, a positive value means that participants expect the others to become more dishonest after group interaction, and more honest for a negative value.

[37] Using ranksum tests: BaseNoPC vs. OathNoPC, $p < 0.001$; BasePC vs. OathPC, $p = 0.002$. There is no difference between OathNoPC and OathPC, $p = 0.501$.

[38] As done for the analysis of the belief about group dishonesty, this choice of conditioning is because participants might adjust their beliefs of individual dishonesty after having interacted with others in their group and what they decided themselves.

[39] Using ranksum tests: Honest vs. Dishonest; OathNoPC, $p < 0.001$; OathPC, $p = 0.004$. Excluding those who observed '⚄' in part 2 yield the same results.

[40] This claim is robust to excluding those who saw '⚄' in part 2.

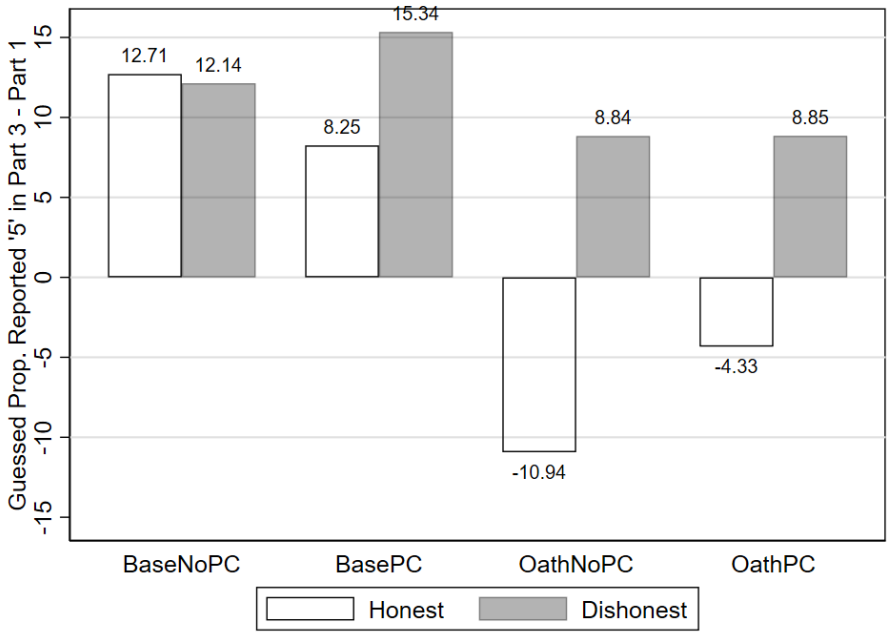

Figure C.2: Change in Beliefs about Individuals' Honesty After group interaction

Notes: This figure displays the change in beliefs about the proportion of participants reporting to '5' in an individual setting after group interaction in part 2 by treatment, conditioned on the decision in part 2 (honest or dishonest).

dishonesty, which comes from honest people being optimistic, while dishonest people being *less* pessimistic about others' dishonesty.

These analyses lead to the following observation:

Observation 2. Overall, beliefs are consistent with behavior. People anticipate oath to reduce lying in group, and that the impact of oath is larger when group members do not earn common payoffs. In addition, without an oath, people expect group interaction to push individual dishonesty upwards. When oath is introduced, however, people expect a lower shift in individual dishonesty after group interaction, resulting from honest people being more optimistic and dishonest people being less pessimistic.

C.3 Response Times Under Oath

This section explores the time participants spent before reaching the decision in part 2.[41] Participants interacted with group members via chat before making the decision, which exposed them to the arguments in favor of honesty and/or dishonesty. Further, those under oath might find it hard to justify their decision to lie. How do these moral dilemma reflect in the response time?

There are two opposing psychological theories regarding the decision time and lying behavior (see Köbis et al. (2019) for a meta-analysis). The first posits *intuitive honesty* on the grounds that lying is psychologically costly. Given that one already knows the truth, truth-telling should be quick and intuitive, while lying is deliberate and requires time. In contrast, the other theory asserts *intuitive dishonesty*, by which truth-telling requires self-control and resistance to temptations. This means that lying is an intuitive response, while truth-telling is a slow reaction.

[41] After the group chat, participants had 180 seconds to enter the die number. When hitting the button 'Submit', z-Tree records the time remaining displayed on the decision screen. Subtracting this value from 180 seconds can used as a proxy of response time the participant spent in the decision-making. In the event that participants exceeded the allotted time, z-Tree records the value as '99999'. This is replaced by '-1', such that the response time is proxied at 181 seconds.

Table C.2 displays the average response time (in seconds) in part 2 by treatment. Groups that observed '⚄' were excluded so the value reflects the time spent prior to deciding whether to lie or not. Overall, participants spent on average 8-10 seconds in the baseline conditions (p = 0.345, ranksum test). Response time increased under oath, but only when group members did not earn a common payoff. Participants in OathNoPC spent about 13.5 seconds while those in the OathPC spent about 8 seconds (p < 0.001, ranksum test). This could mean that participants in OathNoPC needed more time possibly due to the conflict between the honesty norm and the arguments made by their group members. In contrast, group members in OathPC must have already agreed to coordinate on their reports. As such, the response time in OathPC remains unaffected because of group agreement and possibly due to the shared justification to lie.

Table C.2: Response Time (in seconds) in Part 2

	(1) BaseNoPC	(2) BasePC	(3) OathNoPC	(4) OathPC	(1-2)	(1-3)	(2-4)	(3-4)
					p-values			
All	10.49 (12.29)	8.18 (4.45)	13.48 (16.07)	8.02 (5.97)	0.345	0.134	0.301	<0.001***
N	96	93	93	96				
Low States	11.26 (15.04)	8.67 (4.88)	15.31 (19.35)	8.15 (6.22)	0.752	0.076*	0.264	0.001***
N	57	57	54	60				
High States	9.36 (6.48)	7.42 (3.58)	10.95 (9.55)	7.81 (5.61)	0.247	0.801	0.772	0.161
N	39	36	39	36				
Liars	10.06 (10.67)	8.25 (4.48)	11.56 (10.54)	8.06 (5.99)	0.320	0.450	0.322	0.010***
N	88	87	68	81				
Non-liars	15.25 (24.62)	7.17 (4.12)	18.72 (25.31)	7.80 (6.05)	0.944	0.151	0.985	0.028**
N	8	6	25	15				

Notes: This table presents mean response time (in seconds) in part 2 of the experiment. Standard deviations in parentheses. Observations that observed '⚄' were excluded. Figures are reported for all participants and separately for the following subsamples: those who observed low states (0, 1 or 2 tokens), those who observed high states (3 or 4 tokens), those who decided to lie and those who did not lie. p-values reported using ranksum tests. * p < 0.10, ** p < 0.05, *** p < 0.01

Looking at the response time conditional on the die numbers observed reveals a clear effect of oath when group members do not earn a common payoff. When states are low (i.e., 0, 1 or 2 tokens), participants in OathNoPC spent about 15 seconds before reaching the decision and this is longer than those in BaseNoPC (11.26 seconds, p = 0.076) and OathPC (8.15 seconds, p < 0.001). This means that when there is an economic reason to cheat, oath increases the response times but only when group members do not earn a common payoff. Under payoff commonality, the decision time conditioned on low states is not affected by the oath (p = 0.264, ranksum tests). The response times under high states are similar across treatments. Finally, restricting to only those who eventually lied, those in OathNoPC spent about 12 seconds and this is significant longer than those in OathPC (8 seconds, p = 0.010). Non-liars in OathNoPC spent about 19 seconds to decide and this is significantly longer than their counterparts in OathPC who spent about 8 seconds (p = 0.028, ranksum tests).

In sum, the analysis of response times suggests that oath increases the time required to reach a decision whether or not to lie when group members do not earn a common payoff. The increase is pronounced when there is an economic motive to lie. In addition, the increased response time is reflected in both dishonest and honest participants. Longer response time in OathNoPC could result from a deliberation process imposed by the conflict between honesty norm and group communication. Nonetheless, oath does not affect the response time when group members earn a common payoff, possibly because the rule requires group members to coordinate on their reporting. Besides, having already agreed on the number to report could create a shared sense of justification in misreporting. Finally, in comparison to the psychological theories of decision time (Köbis et al., 2019), the data in this experiment seem to be more in line with dishonesty being intuitive: in BaseNoPC and OathNoPC, truth-tellers seem to take longer time to decide than liars.

These analyses lead to the following observation:

Observation 3. Oath increases the time needed to reach a decision when group members do not earn a common payoff compared to when they do. This increased deliberation is more pronounced among people who have an incentive to lie. Response time remains unaffected under the payoff commonality rule.

Appendix D Appendix Tables

Table D1: Summary Statistics

	(1) BaseNoPC		(2) BasePC		(3) OathNoPC		(4) OathPC		(1-2)	(1-3)	(1-4)	(2-3)	(2-4)	(3-4)
									p-value					
	Mean	SD	Mean	SD	Mean	SD	Mean	SD						
Age (Years)	21.64	4.09	21.23	1.96	22.02	3.09	20.85	1.73	0.796	0.021**	0.342	0.037**	0.219	0.001***
Male (%)	0.44	0.50	0.45	0.50	0.45	0.50	0.46	0.50	0.849	0.849	0.778	1.000	0.929	0.929
SOEP (0-10)	6.26	2.39	6.26	2.12	6.13	2.45	5.57	2.30	0.842	0.875	0.027**	0.973	0.037**	0.050**
Business and Economics (%)	0.54	0.50	0.62	0.49	0.59	0.50	0.46	0.50	0.270	0.505	0.263	0.663	0.027**	0.076*
Observations	102		99		99		102							

Notes: The p-values reported are from chi-square tests for gender and school and ranksum tests for age and SOEP. Regression analyses control for these individual characteristics. *$p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table D2: Number of Groups by Types of Arguments

Number of Groups		Dishonest Arguments		
		Yes	No	Total
Honest Arguments	Yes	37	5	42
	No	46	37	83
	Total	83	42	125

Notes: This table presents the number of groups in which honest and/or dishonest argument is mentioned at least once. Of 134 groups, one group in OathNoPC did not chat, and eight groups (two per treatment) viewed the video where '⚄' was shown, leaving 125 groups. For ex. in 37 out of 125 groups, there were at least one honest and one dishonest arguments mentioned.

Table D3: Shares of Dishonest and Honest Arguments

	(1) BaseNoPC	(2) BasePC	(3) OathNoPC	(4) OathPC	(1-2)	(1-3)	(2-4)	(3-4)
					p-values			
Dishonest	61.33%	54.81%	56.82%	43.50%	0.699	0.571	0.403	0.252
Honest	13.67%	9.70%	19.84%	22.12%	0.301	0.306	0.069*	0.973
N	32	31	30	32				

Notes: This table presents the average shares of dishonest and honest arguments circulated within each group by treatment. Of 134 groups, one group in OathNoPC did not chat, and eight groups (two per treatment) viewed the video where '⚄' was shown, leaving 125 groups. * $p < 0.10$

Table D4: Summary Statistics of Videos Shown

	(1) BaseNoPC		(2) BasePC		(3) OathNoPC		(4) OathPC		(1-2)	(1-3)	(1-4)	(2-3)	(2-4)	(3-4)
									p-value					
	Mean	SD	Mean	SD	Mean	SD	Mean	SD						
Part 1	3.60	1.75	3.58	1.77	3.62	1.80	3.69	1.78	0.927	0.945	0.723	0.872	0.658	0.787
Part 2	3.00	1.67	2.94	1.68	3.09	1.63	3.00	1.65	0.764	0.656	0.965	0.455	0.738	0.672
Part 3	3.69	1.63	3.65	1.66	3.74	1.63	3.70	1.60	0.866	0.822	0.967	0.698	0.832	0.849
Observations	102		99		99		102							

Notes: The table summarizes the mean and standard deviations of the die outcomes shown on the video (1 to 6, inclusive). The p-values reported are from ranksum tests.

Table D5: Number of Sessions and Participants, by Treatment and Mode of Payment

Treatments	Bank Transfer	Cash
BaseNoPC	4 (72)	2 (30)
BasePC	4 (90)	1 (9)
OathNoPC	5 (87)	1 (12)
OathPC	5 (81)	1 (21)
Number of sessions (participants)	18 (330)	5 (72)

Notes: This table summarizes the number of sessions (number of participants in parenthesis) by treatment and mode of payment (bank transfer vs. cash).

Appendix E Appendix Figures

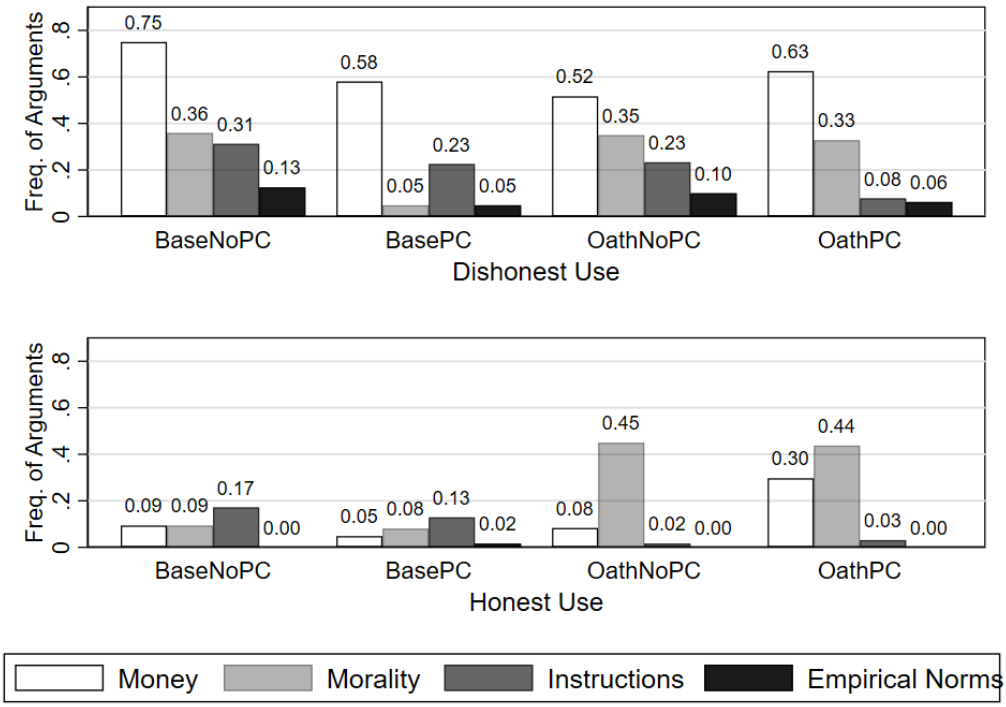

Figure E1: Frequencies of Each Type of Arguments

Note: This figure displays the frequencies of each type of arguments used in the group chat during part 2 of the experiment. Dishonest use (Upper Panel) Honest Use (Lower Panel). The category 'Morality' includes messages related to 'Honesty' and 'Oath'. The category 'Instructions' includes messages related to 'Insecurity' and 'Rules'.