**The Effect of Individual and Group Punishment on Individual and Group-Based Dishonesty**

Janis H. Zickfeld1\*, Mathilde H. Tønnesen1, Christian T. Elbæk1, Sebastian Oelrich1, Thais Cardarelli2, Karolina A. Ścigała3, Stefan Pfattheicher3, Panagiotis Mitkidis1

1Department of Management, Aarhus University, Denmark

2Department of Economics, Aarhus University Denmark

3Department of Psychology, Aarhus University Denmark

\*Corresponding author: Janis H. Zickfeld, [jz@mgmt.au.dk](mailto:jz@mgmt.au.dk). Department of Management, Aarhus University, Fuglesangs Allé 4, 8210 Aarhus V, Denmark.

**Author Note**

All materials, syntax, and data will be made available via the following project page: <https://osf.io/gtd5u/> The authors declare no conflict of interest.

**CRediT Taxonomy**

Janis H. Zickfeld: Conceptualization, Methodology, Funding acquisition, Project administration, Writing – original draft. Mathilde H. Tønnesen, Christian T. Elbæk, Sebastian Oelrich, Thais Cardarelli, Karolina A. Ścigała, Stefan Pfattheicher, Panagitis Mitkidis: Conceptualization, Methodology, Writing – review & editing.

**Abstract**

Economic dishonesty is a widespread behavior that has substantial implications for organizations and societies. Recent studies suggest that decision making in groups or commitment to other individuals can further increase such dishonesty in contrast to individual decision making. Various interventions have been suggested to curb dishonesty, with classical economic approaches emphasize the use of punishment by focusing among others on the risk (“how likely am I to get caught?”). However, the effects of punishments on dishonesty are mixed and it is currently unclear if punishment can be effective in collaborative settings. Here, we suggest an online study (*N* = 630) collecting UK-based Prolific participants in which we manipulate (1) the type of payoff (whether dishonesty earns money for the individual or for the group), (2) the risk of punishment (no risk or 30% audit chance), and (3) the type of punishment (whether the individual is punished or the whole group in case of getting caught). The study will shed light on the interplay between individual or group decision making and the specifics of punishment as well as advance the understanding under which conditions punishment might be effective in curbing dishonesty.

**The Effect of Individual and Group Punishment on Individual and Group-Based Dishonesty**

Dishonesty, such as tax fraud[[1]](#footnote-2), is a widespread phenomenon imposing high costs on corporations and society at large (Balafoutas et al., 2015). According to a 2024 report by the Association of Certified Fraud Examiners (ACFE) companies lose around 5% of their yearly revenue to fraudulent and dishonest practices (ACFE, 2024). Therefore, scholars across the behavioral sciences have investigated a plethora of different interventions aimed at reducing dishonesty (Bellé & Cantarelli, 2017; Hertwig & Mazar, 2022). A recent review suggests that psychological interventions, including moral reminders and appealing to social norms as well as social commitment might serve as potential tools in reducing dishonesty (Hertwig & Mazar, 2022). However, many of those interventions have provided mixed or null results (Cagala et al., 2024; Dimant et al., 2020; Dimant & Shalvi, 2022; Verschuere et al., 2018; Zickfeld et al., 2024), indicating that their effectiveness in decreasing dishonesty might be limited or highly context-dependent.

As an alternative to the psychological interventions targeting dishonesty, utility-based models of dishonesty (Becker, 1968) suggest that individuals refrain from dishonest actions if the potential (financial) costs outweigh the potential benefits. For this reason, scholars have proposed more stringent economic interventions to reduce dishonesty (Hertwig & Mazar, 2022) – most prominently the risk and severity of punishment for behaving dishonestly. Punishment, the imposition of a negative outcome in response to unethical behavior, can be differentiated into internal or self-punishment, the punishment being enacted by the transgressing party on themselves, and external punishment (or third-party punishment), the punishment being enacted by a third party on the transgressor (Raihani et al., 2012). Similarly, different types including physical (e.g., getting spanked), social (e.g., being excluded from a group), and economic punishment (e.g., paying a fine) have been discussed in the literature (Mann et al., 2016; Pratt et al., 2017). Here, we focus on external economic punishments, the imposition of monetary fines by a third party, which is the most common type used by institutions in response to organizational dishonesty.

External punishment has been widely discussed and evaluated within criminology with its main goal focusing on deterrence or retribution to restore fairness (Miceli & Segerson, 2007). However, the mainly correlational evidence of institutional and legal punishment in response to crime shows mixed evidence of its effectiveness (Dölling et al., 2009; Pratt et al., 2017). In economics, a large body of literature has investigated the impact of punishment, referred to as audits, in so-called tax evasion games, tasks that model real world tax reporting (Alm & Malézieux, 2021). Across two meta-analyses (Alm & Malézieux, 2021; Blackwell, 2010), risk of punishment (i.e., audit probability) was associated with increased tax compliance (i.e., honesty) and positively interacted with severity of punishment (i.e., fine/penalty rate). However, when considering dishonest reporters only, increased risk of punishment (and its interaction with fine size) predicted honesty negatively. Importantly, many studies employing tax evasion games do not compare risk of punishment to a control condition (i.e., no risk of punishment) and therefore evidence is mostly correlational.

In behavioral ethics, few studies have experimentally tested the effects of punishment on dishonesty. Of these, some studies have provided evidence that introducing a risk of punishment can decrease dishonesty across different experimental and field settings (Bennett, 1998; Bonfim & Silva, 2019; Brink et al., 2019; Chirikov et al., 2020; Peer & Feldman, 2021; Thielmann & Hilbig, 2018). For instance, in an incentivized die-roll task, Thielmann and Hilbig (2018) found that both the risk of punishment and its severity reduced dishonesty. At the same time, several studies providing mixed results have failed to conclusively determine the effect of punishment on dishonesty (Cagala et al., 2021; Gamliel & Peer, 2013; Mahasuweerachai & SrungBoonmee, 2021), with some studies even observing increases in dishonesty with increasing risk of punishment (Bonner et al., 2016; Cagala et al., 2021; Siniver et al., 2022).

Other studies provide a more nuanced, albeit conflicting, picture of the phenomenon at hand. For instance, some studies show that the risk of punishment more effectively reduces dishonesty than its severity (Mahasuweerachai & SrungBoonmee, 2021; Teodorescu et al., 2021). On the contrary, others show that the severity is more important than risk (Engel & Nagin, 2015; Friesen, 2012), or the contingency of the punishment (Bonner et al., 2016) - the application of punitive behavior directly linked to a specific undesirable behavior (Podsakoff et al., 2006). These findings suggest that punishment might be more effective under certain circumstances than others. Due to the complex nature of the effects of punishment, our study primarily focuses on varying punishment risks while keeping punishment severity constant.

Punishment has been mostly tested for individual decisions focusing on retribution of the individual, though recent theories have highlighted the importance of social context and influence (Pratt et al., 2017; Weisel & Shalvi, 2022). Indeed, dishonesty, here defined as *distorting the true state to acquire profits* (Zickfeld et al., 2025), seems to increase when individuals collaborate with others (Leib et al., 2021) or feel commitment towards “partners in crime”(Zickfeld et al., 2024). Such *collaborative dishonesty* (Kocher et al., 2018; Weisel & Shalvi, 2015) has been theorized based on differences in risk-taking between individuals and groups (Fochmann et al., 2021; Jiang & Villeval, 2024). Under increasing risks of punishment, dishonest behavior becomes riskier in that decision makers might face potential penalties and reduced profits. However, there is mixed evidence in whether groups perform riskier decisions than individuals with some observing a *risky shift* (Isenberg, 1986)and others observing a *cautious shift* in groups (Shupp & Williams, 2008). There is accumulating evidence that groups perform more rational decisions (Charness et al., 2007) and based on utility theory, dishonesty can be considered the more rational choice in situations in which potential benefits outweigh their costs (i.e., punishment). Responsibility for and conformity with group members might be additional factors influencing risk taking in groups (Charness & Jackson, 2009).

Indeed, empirical evidence suggests that group decision making increases dishonesty because loyalty and prosocial concerns and norms (i.e., dishonesty benefiting the group or other group members) can overwrite the motivation to tell the truth (Leib et al., 2021; Zickfeld et al., 2024). Based on these findings, in settings in which dishonesty cannot be detected, such loyalty norms should always appeal to dishonesty to benefit the group (as it presents the more rational economic and social choice). Once the risk of punishment increases, group decisions might become more nuanced as dishonesty could incur negative outcomes on the group when detected. Individual punishment becomes more costly in group settings since transgressors are more difficult to detect (Miceli & Segerson, 2007). In such settings, group punishment —i.e., the collective retribution of a community based on few transgressors—can become a viable tool because it punishes the dishonest person with definite probability (Miceli & Segerson, 2007).

On the downside, group punishment also incurs costs on innocent individuals (Pereira & Prooijen, 2018). Indirect evidence suggests that loyalty concerns in group settings can undermine potential effects of punishment or whistleblowing behaviors (Batolas et al., 2023; Jiang & Villeval, 2024; Rullo et al., 2024; Solaz et al., 2019). Loyalty concerns with the risk of potential group punishment can appeal to maximizing the group’s benefit either through dishonesty or highlighting the responsibility and the negative costs that dishonesty can cause on the group outcome. Social responsibility indeed seems to reduce risk-taking behavior (Charness & Jackson, 2009). Importantly, there is limited direct evidence on whether interventions tested in individual settings, such as introducing the risk of punishment, are also applicable and effective in group settings and whether group punishment can be more effective than individual punishment.

As an exception, three studies have tested the effect of punishment on dishonesty in group settings (Bonfim & Silva, 2019; Fochmann et al., 2021; Jiang & Villeval, 2024).[[2]](#footnote-3) Using an incentivized die-roll task, Bonfim and Silva (2019) tested the effect of the risk of punishment (i.e., the risk of an audit) and a moral reminder on dishonesty in groups of three participants either facing a payoff based on their individual reports or based on the groups’ reports. Introducing the risk of punishment reduced dishonesty, but more strongly in the group setting in which payoffs were determined by all group member’s reports and punishment affected all group members. However, the study did not test situations with (1) individual punishment in a group payoff setting or (2) group punishment in an individual payoff setting.

Employing a tax evasion task, participants in Fochmann et al. (2021) were asked to indicate the earnings of a fictious company that were subject to 25% corporate tax either individually or in groups of three. Both individuals and groups, respectively, could generate higher payoffs by evading taxes. In addition, in each round there was a risk of punishment with an audit occurring with a probability of 30%. Findings show that participants were more dishonest in group settings compared to an individual setting. However, the study did not *manipulate* punishment but applied it across all participants consequently making it difficult to ascertain whether punishment or the group context influenced dishonesty.

Employing an incentivized die-roll task, Jiang and Villeval (2024) studied the social dilemma of profit maximization versus group punishment in groups of varying sizes (3 vs. 6). The results showed that dishonesty was increased (compared to chance level) across all types of groups, but that larger groups exhibited higher reports. Based on these findings, it was concluded that group punishment might be less effective in larger groups, possibly due to diffusion of responsibility. However, the study did not include a no punishment control and did not test for individual dishonesty.

These studies allow two preliminary predictions. First, replicating previous findings (Leib et al., 2021), it seems that people are more dishonest in group (vs. individual) settings regardless of the potential risk of punishment. Second, punishment appears more effective when the entire group is punished compared to when an individual group-member is punished. However, there is no systematic evidence on how individual (vs. group) punishments interact with decisions benefitting the individual (vs. group).

**The Current Research**

In the current work, and following recent calls (Vainapel et al., 2019), we aim to expand the current state-of-the-art in behavioral ethics and collaborative corruption, by systematically testing the effect of individual- and group-based punishment on dishonesty in individual and group-settings. Thereby, we provide three central contributions. First, we aim to replicate the finding that group-based (vs. individual) decision-making increases dishonesty (Leib et al., 2021; for studies using tax evasion see Lohse & Simon, 2021; Matthaei & Kiesewetter, 2020). Second, we aim to replicate the effectiveness of punishment in group (vs. individual) settings (Bonfim & Silva, 2019). Third, we investigate the effectiveness of individual or group punishment for individual- or group-based dishonesty by crossing these factors. To the best of our knowledge, the full interaction between individual- and group-based decision making and individual or group punishment has not yet been tested systematically.

Based on a recent study (Fochmann et al., 2021), we test dishonesty using a corporate tax report game, a common and validated task to assess dishonesty (Alm & Malézieux, 2021). In this game, each three participants represent an accounting department of a fictious company and are asked to either individually (individual reports) or collaboratively (average of individual reports in the group) report the corporate income of a fictious company across five rounds, which is subject to a corporate tax of 25%. To maximize possible incentives, participants can evade taxes and underreport such tax-liable income. Reporting honest results yields an individual incentive of £0.5, while moderate dishonesty increases this to £1 and full dishonesty to £2. Payoffs are chosen based on one randomly selected round. For punishment, we manipulate the individual risk of punishment (no punishment vs. 30% audit chance[[3]](#footnote-4)) and vary the type of punishment (individual vs. group punishment). To test the effects of punishment in individual and group settings we vary the payoff structure (individual vs. group pay-off). An overview of the main manipulations and experimental cells is provided in *Table 1* and *Figure 1b*.

**Predictions**

Based on rational choice theory and the deterrence model of crime (Becker, 1968) we can derive the expected income function for each experimental treatment based on common equations employed in the tax evasion literature (Allingham & Sandmo, 1972; Alm & Malézieux, 2021). To align with the literature on tax evasion, we refer to the fixed amount of company income as the *true income* (*I*), the reported amount of company income by the individual or group as the *declared company income* (*R*), the share of the declared company income paid as taxes as the *tax rate* (*t*), the tax rate times the declared income as the applicable tax (*T*), the risk of punishment as the *audit rate* (*p*), and the severity of punishment as the *penalty rate* (*f*). Across the study, the true company income (*I*) is fixed at 1000, the tax rate (*t*) at 25%, and the penalty rate (*f*) at 2 (referring to the lost amount of taxes plus a fine of the same size) based on Fochmann et al. (2021; see Method or Figure 1 for more details). The applicable tax (*T*) equals the declared company income (*R*) times the tax rate (*t*).

Equation 1 depicts the *company success* (i.e., company income after taxes) when not being caught underreporting the company income (or when reporting the true income). The company success when not being caught (*IN*) is determined by the difference between the true company income (*I* = 1000) and the declared company income(*R*) times the tax rate of 25% (*t*). For example, for a declared company income of 0 the company success is 1000 (1000 – 0 \* 0.25).

Equation 1

*IN* = *I* – *t \* R*

Equation 2 depicts the *company success* when being caught (*IC*). In this case, the penalty equals twice the unpaid taxes. If the declared company income equals the true company income (i.e., honest reporting, *R* = *I*) the penalty equals 0 and thus Equation 1 applies. However, if the declared company income differs from the true company income (i.e., dishonest reporting, *R* < *I*) the penalty equals two times the difference in true and declared company income times the tax rate (i.e., the unpaid tax). For instance, declaring a company income of 0 and being audited a penalty of two times 25% of 1000 (i.e., 500) is applied and the company successwhen being caught (*IC*) is 500.

Equation 2

*IC* = *I* – *t* \* *R* – *f \** [*t \** (*I* – *R*)]

To calculate the *expected company income* *E*(*I*)we use Equation 3, which sums the probability of an audit (*p*) times the company success when being caught (*IC*) and the probability of not being audited (1-*p*) times the company success when not being caught (*IN*). In case of audit rate being 0% or honest reporting, Equation 3 is identical to Equation 1. However, in case of underreporting the true company income, expected company income needs to consider the probability of an audit and the company success when being caught. In this case, reporting a declared company income of 0 results in an expected company income *E*(*I*) of 850 (0.3 \* 500 + 0.7 \* 1000).

Equation 3

Expected company income *E*(*I*) = *p* \* *IC* + (1- *p*) \* *IN*

In the current study, we implement a factorial design manipulating punishment risk (audit rate 0% vs. 30%), punishment type (individual vs. group), and payoff structure (individual vs. group; see Method and Figure 2 for more details). Below and in Table 1, we provide an example for each combination of treatments and for reporting the company income as honest (*R* = 1000), moderately dishonest (*R* = 500), and fully dishonest (*R* = 0). [[4]](#footnote-5)

As the control treatment, we implement a situation in which the audit rate is 0% and there is no risk of punishment (see Table 1, A). In this situation, the expected company income is based on Equation 1. We further manipulate the payoff structure across treatments, with incomes either based on individual reports (*RI*) or group reports (*RGr*) - with the group report being the *average* of the three individual reports in a group (Figure 1). Importantly, in the group payoff structure reports are first made independently and individually without the possibility for communication and only later averaged for the group report. We implement five rounds of the task to model temporal effects over time. However, to control for possible learning or expectations effects participants receive no feedback after the specific round, but only after completing all five rounds. Therefore, they will not be able to know the other participant’s reports, the group average report, or whether an audit occurred before the end of the game. This is to ensure control of potential learning effects and models real-world situations in which there is often a temporal delay of possibly implemented audits.

Table 1. Overview of predictions of expected company income *E*(*I*) and expected gain from dishonesty for levels of fully dishonest (*Ri* = 0), moderately dishonest (*Ri* = 500), and honest (*Ri* = 1000) reporting across the manipulations of payoff structure, punishment risk, and punishment type.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **Punishment Risk (Audit Rate, *p*)** | | | | | | Declared Income (*Ri*) |
|  |  | 0% |  | 30% |  |  |  |  |
|  |  |  |  | **Punishment Type** | | | |  |
|  |  |  |  | *Individual* (*RI*) |  | *Group* (*RGr*) |  |  |
|  |  | Expected Company Income *E*(*I*) | Expected Gain from Dishonesty | *IE* | Exp. Gain | *IE* | Exp. Gain |  |
| **Payoff Structure** | *Individual* (*RI*) | (A) *Ind-No*  *E(I)* = *I* - *tRI* |  | (C) *Ind-Ind*  *E(I)* = *p* (*I - tRI*) + (1-*p*) {*I* - *tRI* - *f*[*t*(*I* – *RI*)]} | | (E) *Ind-Gr*  *E(I)* = *p* (*I - tRI*) + (1-*p*) {*I* - *tRI* - *f*[*t*(*I* – *RGr*)]} | |  |
| 1000 | 1.33 | 850 | 1.13 | 900 | 1.29 | 0 (FD) |
| 875 | 1.17 | 800 | 1.07 | 800 | 1.14 | 500 (MD) |
| 750 |  | 750 |  | 700 |  | 1000 (H) |
| *Group* (*RGr*) | (B) *Gr-No*  *E(I)* = *I* - *tRGr* |  | (D) *Gr-Ind*  *E(I)* = *p* (*I - tRGr*) + (1-*p*) {*I* - *tRGr* - *f*[*t*(*I* – *RI*)]} | | (F) *Gr-Gr*  *E(I)* = *p* (*I - tRGr*) + (1-*p*) {*I* - *tRGr* - *f*[*t*(*I* –*RGr*)]} | |  |
| 916.66 | 1.50 (1.10 \* *l*) | 766.66 | 0.92 | 816.66 | 1.04 | 0 (FD) |
| 875 | 1.43 (1.05 \* *l*) | 800 | 0.96 | 800 | 1.02 | 500 (MD) |
| 833.33 |  | 833.33 |  | 783.33 |  | 1000 (H) |

Note. FD = fully dishonest (*Ri* = 0), MD = moderately dishonest (*Ri* = 500), H = honest (*Ri* = 1000); *l* (loyalty parameter) = 1.36.

A graph of different colored lines

Description automatically generated

Figure 1. Simulated expected company income (*IE*) for levels of fully dishonest (*Ri* = 0), moderately dishonest (*Ri* = 500), and honest ((*Ri* = 1000) reporting and combinations of group members reporting (FD = fully dishonest, MD = moderately dishonest, H = honest) across the manipulations of payoff structure, punishment risk, and punishment type. Blue box and lines refer to bonus payoff > £0, while red box refers to bonus payoff < £0. Thick blue line refers to bonus payoff of £2, blue box refers to bonus payoff of £1, dashed blue line refer to bonus payoff of £0.5 and red box refers to bonus payoff of £0.

For the individual payoff structure with no punishment risk (*Ind-No;* Table 1, A; Figure 1, A), an individual being honest and reporting the declared company income at 1000 (*Ri*)would generate an expected company incomeof 750 (1000 – 0.25 \* 1000). In comparison, an individual or group being fully dishonest and declaring an income of 0 (*Ri*) would generate an expected company income of 1000 when not being caught. Therefore, the expected gain from full dishonesty is 1.33 times as high as the expected gain from honesty.

For the group payoff structure with no punishment risk (*Gr-No;* Table 1, B; Figure 1, B), the expected incomes are identical for group reports (*RGr*), but as they depend on the individual reports of the three group members (*R1*, *R2*, *R3*), expected incomes at the individual level differ based on what the other two participants report. For example, for the group payoff structure the expected company income (at the group level) for participant *1* being fully dishonest and reporting 0 as the company income (*R1*) depends on the reported company income of participant *2* (*R2*) and *3* (*R3*). If both other participants are fully dishonest (*R1* = *R2* = *R3* = 0) the average group report is *RGr* = 0 and the expected company outcome is 1000. However, if the other participants are honest (*R1* =0; *R2* = *R3* = 1000), the average group report is, *RGr* = 666.67 and the expected company income is 833.33. We simulated the group report (*RGr*) and the expected company income based on combinations of honesty, moderate dishonesty, and full dishonesty for the three participants (Figure 1, B) and calculated the average expected company income per type of report. For a declared company income of participant, *1*, *2*, or *3* of 0 (fully dishonest) the average expected company income (considering possible combinations of reports of the other participants) would be 916.66. For a moderately dishonest company income by one of the participants (*R1* = 500) the average expected company income is 875 and 833.33 for honest reporting (*R1* = 0; Table 1). In this context, the expected gain from full dishonesty is only 1.09 times higher than the gain from honesty. Therefore, based on this simulation, dishonesty in the group payoff structure would be on average less profitable and honesty more profitable compared to the individual payoff structure. However, the highest expected company income of 1000 is still expected for all participants reporting fully dishonest (*R1* = *R2* = *R3* = *RGr* = 0). Therefore, based on previous empirical findings we expect loyalty to group members to increase dishonest reporting for group payoff structure compared to the individual payoff structure when risk of punishment is absent (Leib et al., 2021; Weisel & Shalvi, 2022; Zickfeld et al., 2024). We account for this by multiplying the expected payoffs by a *loyalty* parameter (*l*). The loyalty parameter should model the fact that in group settings social norms might play a role in increasing the average payoff (Pratt et al., 2017). A recent meta-analysis observed an overall effect of *Hedges’ g* = 0.17 of commitment on dishonesty (Zickfeld et al., 2024; identical to *Cohen's d* for large sample sizes as in this case). Converting this to an Odds Ratio reveals 1.36. Therefore, we multiply the expected gain from full dishonesty compared to honesty (1.10) with 1.36 in case of *Gr-No* (which is 1 in *Ind-No* as no loyalty applies), resulting in a larger gain compared to *Ind-No* (1.50). This adjustment reflects the increased propensity for dishonesty due to the presence of commitment or loyalty within the group and increases the expected payoff compared to the *Ind*-*No* treatment (Table 1, B). Based on this prediction and previous findings highlighting that group reports increase dishonesty we predict that:

H1. Group payoff structure increases dishonesty (i.e., non-compliance in the tax evasion game) compared to the individual payoff structure (baseline) for the no punishment treatment (*Gr-No* vs. *Ind-No*).

Further, we manipulate risk of punishment by introducing an audit rate of 30%. The expected gain from dishonesty is thereby reduced as the probability of not being punished decreases to 70%. Taking into account the expected company payoff for the punishment treatments (see Table 1) and previous findings that the risk of punishment reduces dishonesty (e.g., Thielmann & Hilbig, 2018), we predict:

H2. Risk of punishment (vs. no punishment) reduces dishonesty (i.e., non-compliance).

Finally, we manipulate the type of punishment by either basing punishment on individual reports (*Ri*) or group reports (*RGr*; Table 1) and crossing it with payoff structure. This results in four possible treatment combinations: individual payoff structure with individual punishment (*Ind-Ind*;Table 1, C; Figure 1, C), group payoff structure with individual punishment (*Gr-Ind*;Table 1, D; Figure 1, D), individual payoff structure with group punishment (*Ind-Gr*; Table 1, E; Figure 1, E), and group payoff structure with group punishment (*Gr-Gr*, Table 1, F; Figure 1, F). Thereby, we systematically vary how Equation 3 is derived depending on the type of payoff and the type of punishment (see Table 1 for an overview).

For individual payoff structure with individual punishment (*Ind-Ind*;Table 1, C; Figure 1, C), expected company payoffs are derived based on Equation 3. For full dishonesty (*Ri* = 0), expected company income when not being caught is 1000, but is reduced to 500 when being caught. As the expected probability of an audit is 30%, this results in an expected company income of 850 (Table 1, C). This is reduced to 800 for moderate dishonesty (*Ri*= 500) and 750 for honesty (*Ri* = 1000; Figure 1, C). Expected gain by full dishonesty over honesty thereby reduces to 1.13 compared to the control treatments with no punishment, reflecting that the gain from dishonesty is reduced when the risk of punishment increases. This is a standard tax statement constellation for a group of companies where the individual firms are affiliates or associates of the parent company but not subsidiaries (IRC Section 1504; Treasury Regulation §1.1502-1). While individual and group reports are prepared, the individual tax statements are used by tax authorities and the associates or affiliates are fined based on their individual misreporting.

For the group payoff structure with individual punishment (*Gr-Ind;* Table 1, D; Figure 1, D) the prediction reverses. The highest expected company income (916.66) is predicted with the participant being honest (*R1* = 1000), while both other participants are fully dishonest (*R2* = *R3* = 0). Similarly, the lowest expected company income (683.33) is predicted with the participant being fully dishonest (*R1* = 0), while both other participants are honest (*R2* = *R3* = 1000). Individuals in the group payoff gain from a high group report but lose for the individual punishment when their individual report is audited and dishonest. In this scenario, individuals gain from being honest in a dishonest group - they gain from the dishonesty of the group, but do not face any consequences if their individual behavior is honest. As for all group payoff structure treatments, we expect loyalty pressure influencing reporting that profits the group (i.e., full dishonesty) in this treatment. However, this would go against the individual motivation to avoid individual punishment, and we expect the motivation to reduce individual costs to be as least as strong as the loyalty towards the group (i.e., loyalty parameter is 1). Average expected company income is predicted to be highest for honesty (833.33), followed by moderate dishonesty (800) and full dishonesty (766.66; Table 1, D). This results in an expected gain of full dishonesty over honesty of .92, which is why we expect the lowest degree of dishonesty in this treatment. This setting is similar to when a parent company has subsidiaries and prepares consolidated financial statements. The group benefits from the retained earnings that are higher if a subsidiary evades taxes and can redistribute these within the group. The tax evasion of the subsidiary is punished individually, as it retains its status as a separate legal entity. Such issues pertain especially in international tax settings and are often achieved through transfer pricing, where profits can be shifted to lower tax countries (Diller et al., 2024; see also IRC Section 482, 1504).

For individual payoff structure with group punishment (*Ind-Gr*;Table 1, E; Figure 1, E), the prediction is exactly opposite to the *Gr-Ind* treatment. The highest expected company income (950) is now for individuals being fully dishonest (*Ri* = 0), while the group report should be as accurate as possible based on the other participants reporting honest (i.e., *R2* = *R3* = 1000; *RGr* = 666.67). Similarly, the lowest expected company income (650) is for the individual being honest (*Ri* = 1000) with the remainder of the group being fully dishonest (i.e., *R2* = *R3* = 0; *RGr* = 333.33). This is since individual gain is based on the individual report which increases with higher dishonesty, while the potential punishment is based on the group report which decreases in size for groups tending to behave honestly. In this scenario individuals gain from being dishonest in an honest group resulting in a *wolf in sheep herd* effect. The average expected company income is 900 for full dishonesty, 800 for moderate dishonesty, and 700 for honesty (Table 1, E), resulting in an expected gain of full dishonesty over honesty of 1.29 and we would expect the highest degree of dishonesty for this treatment among the punishment treatments. In a tax context, recent studies show that tax evasion of one firm (individual payoff) can have spillover effects on peer firms (group punishment), reducing their market values when the cheating firm is caught without being directly involved (Bauckloh et al., 2021). In a group setting, parent companies and thus the entire group, are regularly fined for misconduct of subsidiaries, piercing the corporate veil. For instance, in 2024, parent company META was fined $840 mil (€798 mil) by the European Commission for violations by one of her group members, subsidiary Facebook, for abusive market practices (European Commission, 2024). Similarly, in 2017, Volkswagen group was fined $4.3 bn in the US for emission fraud (“Dieselgate”). In this case, mainly two subsidiaries were cheating (Audi and Porsche) who benefitted individually, as their sports cars relied on high emission engines (US Department of Justice, 2017).

For group payoff structure with group punishment (*Gr-Gr*;Table 1, F; Figure 1, F), the prediction is similar to the *Ind-Ind* treatment. The highest expected company income (850) is expected for all participants in the group being fully dishonest (*R1* = *R2* = *R3* = *RGr* = 0), while this is lowest (750) for all participants in the group being honest (*R1* = *R2* = *R3* = *RGr* = 1000). Average expected company income is predicted at 816.66 for full dishonesty, 800 for moderate dishonesty, and 783.33 for honesty (Table 1, F), resulting in an expected gain of full dishonesty over honesty of 1.04. Since the potential gain is based on the group report there is an incentive to misreport to keep the group report as low as possible, while the potential loss is also based on the group report. This creates a loyalty conflict with loyalty appealing to increase group income but also minimize the chances of others being punished at the same time. We expect this potential conflict to reduce the loyalty parameter to 1 and therefore predict this treatment to show a slightly lower level of dishonesty as *Ind-Ind* but a slightly higher level of dishonesty as *Gr-Ind*. The Gr-Gr constellation is common within companies, where individual units, branches, or facilities misreport their earnings individually. This is consolidated in one tax statement within the firm. Since this involves only one legal entity, the authorities punish the entire firm for misreporting and thus all units equally based on the firm’s overall report (group report). On the other hand, all units benefit equally as well, because tax savings can be distributed among the units. Cartels can be seen as an example for a setting across firms, where firms collude on price setting, practically through reporting higher costs to justify higher prices. In cartels, the entire group is punished, regardless of how much individual firms engaged in the cartel or if they really did inflate costs (EU: Articles 101, 102 Treaty on the Functioning of the European Union; US: Bourveau et al., 2020).

Considering the individual predictions (Table 1), we would expect an interaction between payoff structure and punishment risk. While, we expect dishonesty to be higher for the group payoff structure (*Gr*-*No*) than the individual payoff structure (*Ind-No*) when no punishment is implemented, we predict this to reverse when introducing punishment (regardless of the type of punishment) in that group payoff structure with punishment (*Gr*-*Ind* and *Gr*-*Gr*) should result in lower dishonesty than individual payoff structure with punishment (*Ind-Ind* and *Ind-Gr*). There is limited evidence for such an interaction effect in the published literature though, which needs to be interpreted with caution as there exist few studies that manipulate both payoff structure and punishment risk (Zickfeld et al., 2024). Based on the expected company income across dishonesty levels, we predict:

H3. Dishonesty will be higher for group payoff structure compared to the individual payoff structure when no punishment is possible (i.e., H1), but this will reverse for situations in which punishment is possible with the individual payoff structure showing higher dishonesty compared to the group payoff structure.

When looking at the predictions for each type of punishment across payoff structure, the expected gain from (full) dishonesty is lower for the individual punishment treatments compared to the group punishment treatments (Table 1). Given that there is limited empirical evidence on the effectiveness of group punishment (Bonfim & Silva, 2019) we predict a main effect of punishment with individual punishment being on average (across individual and group payoff structure) more effective in reducing dishonesty compared to group punishment:

H4. Individual punishment (vs. group punishment) will show stronger effects in reducing dishonesty across payoff structure.

When considering the interaction of type of punishment and payoff structure we would expect a similar effect of individual punishment for both types of payoff structure. Based on the predictions (Table 1), individual punishment should reduce dishonesty more strongly compared to group punishment and to a similar degree for both individual payoff structure and group payoff structure. Therefore, we expect no interaction effect between type of punishment and payoff structure:

H5. Individual punishment is expected to reduce dishonesty more strongly compared to group punishment across both types of payoff structure.

At the same time, we expect group punishment to be more effective in reducing dishonesty for group payoff structure (*Gr*-*Gr*) compared to individual payoff structure (*Ind-Gr*).

H6. We expect group punishment to reduce dishonesty more strongly for the group (vs. individual) payoff structure treatment.



Figure 2. Schematic overview of a) procedure of the study, b) experimental manipulations of the study, c) procedure of the tax report game, and d) calculation of indices in tax report game.

**Method**

An overview of the proposed study design is provided in Table 2. All participants will provide informed consent before starting the study. The study’s procedure and all materials have been approved by the research ethics committee at Aarhus University (BSS-2023-064).

**Participants**

**Planned Sample.** We will recruit UK-based participants via Prolific, equally balancing men and women and screening for English language fluency and 100% approval rate. Participants will be paid £1.5 as a base payment and might earn an additional bonus payment between 0 and £2 based on their or their group’s reporting. We will stop data collection once we have recruited 630 participants after exclusions. Participants will be automatically screened out during recruitment when a) failing at least one comprehension check twice, b) failing two attention checks, or c) not being matched with a partner. Based on previous studies and two pilots on comprehension (Supplementary Note 4) we expect around 30% of participants to be screened out or excluded based on these criteria. Recruitment is stopped once 630 participants are retained after exclusions.

**Sample Size Justification.** Given the limited availability of previous studies testing group punishment, we estimated the expected effects in each experimental cell based on our predictions and previous data (Supplementary Note 1) and simulated power using the main models. Focusing on the interaction effect in H4, using an ordered beta regression, the estimated effects, and an alpha level of 0.05, a sample size of around 400 (5 decisions for each participant; 2000 observations) is necessary to achieve 95% power (Supplementary Figure 2). In addition, we simulated a model accounting for our smallest effect size of interest when testing main effects in H1 and observed between 86% and 90% power at around 600 total participants and around *n* = 100 participants per treatment (Supplementary Figure 2). Based on these simulations, we set the final sample size at 630 participants (3150 observations, 105 participants per cell, 35 groups per cell). A detailed justification and further details on the simulation for each hypothesis are provided in Supplementary Note 1.

**Design & Procedure**

We will employ a 2 (Punishment Risk: 30% vs. 0%) x 2 (Punishment Type: Individual vs. Group; nested in punishment risk) x 2 (Payoff Structure: Individual vs. Group) between-subjects design (see Table 1, Figure 1b).

A schematic overview of the procedure is provided in Figure 1a. Participants will provide informed consent before being matched with two other participants from the experiment using the SMARTRIQS interface in Qualtrics (Molnar, 2019). All three participants will make up a group (referred to as a *company*) and will be randomly assigned to one of the six treatments. All participants in a group will be assigned to the same treatment (*Figure 1a*). Participants are not able to communicate with the other group members and all participants will be anonymous. However, the decisions of the group might impact their outcome, dependent on the treatment. Participants in the Ind-No and Ind-Ind treatments will not interact further. Participants in the Ind-Gr might be impacted by the other group members if a punishment is enacted (and the whole group is punished). Participants in the Gr-No, Gr-Ind, and Gr-Gr will be impacted by the other group members by the average group report. Importantly, participants will not be able to see the final group reports or individual reports before the end of the game to control for possible learning and expectation effects.

Participants will receive adapted instructions about the tax report game (Fochmann et al., 2021; Supplementary Note 2). They will complete five comprehension items regarding the instructions and two attention checks (Supplementary Note 3). If they fail at least one of the comprehension items they will be shown instructions and the comprehension items once more. If they fail at least one of the items again, they will be asked to return the submission and screened out. Similarly, participants failing both attention checks will be screened out. After instructions, they will complete an item on commitment to the other group members (before the game). Then they will start with a practice round of the tax report game. After the practice round, they will complete five rounds of the tax report game. They are only shown the results of the practice and five rounds after completing all rounds. In the end, one round of the tax report game is randomly picked to determine individual bonus payoff. Before being shown the final outcomes, participants will complete an item measuring commitment to the other group members. Finally, they will complete items on moral anger and guilt, an item on risk aversion, and trait Honesty-Humility, as well as demographic information.

**Measures & Materials**

**Tax Report Game.** The main measure of dishonesty is a tax report game based on Fochmann et al. (2021). The game is illustrated in *Figure 1c*. In this game each triad represents a *company,* and their task is to report the true company income across five rounds. In each round, the true company income is fixed at 1000 points and participants are asked to report this income (before tax). Each report is subject to a 25% corporate tax. Assuming full honesty, the final company success (net income) would be 750 points. As participants can report any income from 0 to 1000 points, dishonest participants can report an income below 1000 points and pay lower taxes (*Figure 1d*).

Individual bonus payoffs are determined by a conversion rule based on the company success in each round and the final bonus is determined by selecting one of the five rounds randomly. Company success is transformed into British pounds for individual bonus payoffs using the following rule: Company success below 750 results in a bonus payoff of £0 (*dishonesty caught*), company success equal to 750 results in a bonus payoff of £0.50 (*honesty*), company success larger than 750 but smaller than 1000 result in a bonus payoff of £1.00 (*moderate dishonesty*), and company success of 1000 result in a bonus payoff of £2.00 (*full dishonesty*). Individual payoff is always reduced to £0 when dishonest income reporting is audited.

Experimental factors are implemented in the tax report game as follows: For individual payoffs, participants will report the company income for all five rounds and their bonus is determined by the individually reported net income (after taxes). For example, if a participant correctly reports a true company income (before taxes) of 1000 points, they will receive an individual bonus payoff of £0.5 (750 = 1000 – 25% x 1000). For group payoffs, participants will also report the company income individually for all five rounds, but their bonus is determined by reports of the other two group members. Each group receives the same income based on the average income of all three individual reports. For instance, if participant A reports a company income of 1000, participant B a company income of 500, and participant C a company income of 750, then a mean of 750 will be selected as the group-based reported company income. Each group member will therefore receive an individual bonus payoff of £1 (812.5 = 1000 – 25% x 750).

The risk of punishment is implemented by varying a 0% or 30% audit rate. Audits are constant across rounds. Thus, for each round participants face the exact same risk of an audit. Further, we vary the type of punishment by implementing an individual punishment or group punishment. In the individual punishment treatments, each individual report is audited, and the penalty is applied at the individual level. In the Ind-Ind treatment this means that each group member independently faces an audit rate of 30% for their individual report and penalties are applied to the individual payoff. In the Gr-Ind treatment this means that each group member independently faces an audit rate of 30% for their individual report and penalties are also applied to the individual payoff. However, individual punishment does not affect the group payoff which is still calculated on the individual reports (even if an individual audit occurs). In the group punishment treatments, each group is audited with an audit rate of 30% and the penalty is applied at the group level. In the group payoff – group punishment treatment this means that the group average is audited at a 30% rate and a penalty is applied to each individual payoff. In the Ind-Gr treatment this means that the group is audited at a 30% rate and the penalty is applied to each individual payoff.

If the audit reveals dishonesty (i.e., non-compliance by reporting a lower company income than 1000) a penalty will be added to the company income two times the evaded tax. This penalty consists of the company repaying the evaded tax, plus a fine that is equivalent to the evaded tax. For instance, if an individual is audited and reported a company income of 0, the company first must repay the evaded tax (0.25 \* [1000 – 0] = 250) plus the same amount as a fine (250). The individual share will then be calculated based on 1000 – penalty (i.e., 500 points, bonus payoff of £0). In the individual punishment treatment this penalty is applied to the individual income in case of an audit. In the group punishment treatment this penalty is applied to the group-based income in case of an audit. Thus, if one group member is caught cheating, this will impose a penalty on the whole group even though the other group members might be honest. This mirrors real life situations in which employees are punished (e.g., receiving less bonuses or lower wages) due to penalties imposed on the company due to the dishonesty of one individual. Whether an audit occurred in a specific round and its consequences are only revealed at the end of the game to control for possible learning and expectation effects.

**Social Commitment.** Right before and after the tax report game, participants will complete a commitment measure adapted from Zickfeld et al. (2023). Participants will be asked how much they feel committed to their two group members on a scale from 1 to 7 (not committed at all to very much committed).

**Moral Anger & Guilt.** After the tax report game, participants will complete three items each measuring felt anger (*mad, irritated, angry*) and guilt during the game (*remorseful, sorry, guilty*; “How guilty did you feel for your actions during the game?”; “How angry did you feel at the other players during the game?”) on a scale from 1 (not at all) to 7 (very much; Motro et al., 2018).

**Stress.** We will include one item to assess self-reported stress (Ścigała et al., 2025) asking participants to rate “I am stressed right now.” On a 5-point scale from 1 (“strongly disagree”) to 5 (“strongly agree”).

**Risk Aversion.** We will include a one-item measure on risk aversion (“Generally speaking, would you characterize yourself as someone who is willing to take risks, or as someone who is avoiding risks?”) on a scale from 1 (absolute risk aversion) to 9 (absolute risk seeking) as employed in a previous study (Casal et al., 2016). While there exist multiple measures to assess risk taking or risk aversion, research on tax evasion has recommended employing simple single item scales (Alm & Malézieux, 2021).

**Honesty-Humility.** We will use the 4-item Honesty-Humility scale (De Vries, 2013) on a 5-point scale from “strongly disagree” to “strongly agree” (e.g., “I find it difficult to lie”).

**Demographic Questions.** Participants will also complete demographic items including participant’s self-identified gender (male, female, non-binary, prefer not to say), years in age, and nationality (UK vs. other).

**Analysis Plan**

**Exclusion Criteria.** We will exclude participants a) failing at least one of the five comprehension checks twice (these are automatically excluded during recruitment), b) who are not matched with two partners (e.g., due to participant availability), c) participants who show a duplicate in their Prolific ID (we will only keep the first observation), d) participants failing two attention checks (these are automatically excluded during recruitment).

**Calculation of Variables.** We will calculate a compliance score in each round and for each individual based on the reported company income (*R*) / true company income (*I*). Two of the Honesty-Humility items will be reverse-coded, and all four items averaged into an Honesty-Humility score. The three guilt and the three anger items will be averaged into a guilt and anger score. For all scores, reliabilities will be assessed using McDonald’s w.

**Statistical Tests**

For all analyses, we will set our alpha level at 0.05. Given the expected distribution of compliance scores based on previous studies (e.g., Zickfeld et al., 2025), most main analyses will be computed using an ordered beta regression model in the *glmmTMB* package (Bolker, 2016). We will add random intercepts according to participants nested in triads for all models. For all models we will compute average marginal effects using the *marginaleffects* package (Arel-Bundock, 2023). All analyses will be performed in the statistical environment *R* (version 4.2.1; R Core Team, 2022).

**Manipulation check.** We will compute a multilevel model with social commitment as the outcome and payoff structure as the predictor. We will repeat the model by adding punishment (contrast coded: punishment risk -2/3 (no punishment) vs. 1/3 (individual punishment) vs. 1/3 (group punishment); punishment type: 0 (no punishment) vs. –½ (individual punishment) vs. ½ (group punishment)) in a second step. We will compute two models, one with social commitment before the game as the outcome and one with social commitment after the game as the outcome. We follow up with post-hoc comparisons comparing each treatment to the baseline (Ind-No). Using Tukey adjustment to adjust for multiple comparisons. In addition, we will run one model with difference in social commitment (pre vs. post) as the outcome and payoff structure, punishment, and their interactions as predictors.

**H1.** We will compute an ordered beta regression with compliance as the DV and payoff structure (individual: 0; group: 1) focusing only on participants in the no punishment treatments.

**H2 & H4.** We will compute an ordered beta regression with compliance as the DV. We will add punishment (no vs. individual vs. group) as the predictor with two contrast codings (punishment risk -2/3 vs. 1/3 vs. 1/3; punishment type: 0 vs. –½ vs. ½).

**H3 & H5.** We will compute an ordered beta regression with compliance as the DV. As predictors we will add payoff structure (individual: 0; group: 1) and punishment (punishment risk -2/3 vs. 1/3 vs. 1/3; punishment type: 0 vs. –½ vs. ½) and their interactions. To test for the absence of an interaction effect we set our smallest effect size of interest (SESOI) at h2 = .008 (Cohen’s *f* = .09; Cohen’s *d* = .18) based on previous findings and average effect sizes in the literature (Aguinis et al., 2005; Lovakov & Agadullina, 2021). We will test for equivalence using the *marginaleffects* package.

**H6.** We will compute an ordered beta regression with compliance as the DV and payoff structure (individual: 0; group: 1) as the predictor, focusing only on participants in the group punishment treatments.

**Honesty-Humility.** We repeat the full model in H3/H5 with Honesty-Humility (mean centered) as an additional predictor and its interaction with payoff structure and punishment.

**Time.** We repeat the model in H3/H5 with a variable coding for round number and its interaction with payoff structure and punishment.

**Risk Aversion.** We repeat the model in H3/H5 with the measure on risk aversion (mean centered) as an additional predictor and its interaction with payoff structure and punishment.

**Demographics.** We repeat the model in H3/H5 with gender (male: 0, female: 1), age (mean centered), and nationality (other: 0, UK: 1) as additional predictors and their interaction with payoff structure and punishment. We only include nationality if at least 10% of responses report a nationality different than the UK.

**Moral Emotions & Stress.** We conduct three multilevel regression models with moral anger, guilt, or stress as the dependent variable. As predictors we will add payoff structure (individual: 0; group: 1) and punishment (contrast coded: punishment risk -2/3 vs. 1/3 vs. 1/3; punishment type: 0 vs. –½ vs. ½) and their interactions. Further, we will conduct three multilevel regressions with moral anger, guilt, or stress as the dependent variable and compliance (mean centered) as the predictor.

Table 2. Study Design Template for the Proposed Study.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Research Question** | **Hypothesis** | **Sampling Plan** | **Analysis Plan** | **Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis** | **Interpretation given different outcomes** | **Theory that could be shown wrong by the outcomes** |
| Manipulation Check. Are individuals showing more commitment to their group members in the different treatments? | Manipulation Check. Group payoff and group punishment increases commitment before the game compared to the individual payoff (baseline). | Previous studies have found effect sizes between *d* = 0.20 and 1.01 for commitment ratings in group settings compared to individual baselines (Zickfeld et al., 2023). Given power simulations for the other models (Supplementary Material Note 1) we obtain a power of > 90% with the current sample size. | Multilevel model with social commitment before the game as the DV and payoff structure as the predictor in a first step. In a second step, punishment is added (contrast coded: punishment risk -2/3 vs. 1/3 vs. 1/3; punishment type: 0 vs. –½ vs. ½). We follow up with post-hoc comparisons comparing each treatment to the baseline (Ind-No). Using Tukey to adjust for multiple comparisons. | The smallest effect size of interest (SESOI) was set based on previous findings in tax-compliance studies (Zickfeld et al., 2025). This was set to *d* = 0.15 for main effects and h2 = .008 for interaction effects. Power simulations suggested that given the final sample size (*N* = 630) and the repeated design more than 90% power would be obtained. | The null hypothesis is rejected if the pairwise comparisons with the group treatments (Gr-No, Gr-Ind, Ind-Gr, Gr-Gr) and the baseline are statistically significant, the 95% CI is not entirely included in the equivalence bounds, and social commitment is higher for group treatments compared to the baseline.  The null hypothesis is not rejected if the 90% TOST CI of the pairwise comparison is entirely included in the equivalence bounds. | If the null hypothesis is rejected the manipulation is considered successful.  If the null hypothesis is not rejected the manipulation is considered unsuccessful and we will investigate which treatments failed to increase feelings of commitment. |
| RQ1. Are individuals more dishonest when deciding in a group compared to individually? | H1. Group payoff increases dishonesty (i.e., non-compliance in the tax evasion game) compared to the individual payoff (baseline) for the no punishment treatment. | We simulated a model accounting for our SESOI (*d* = .15) when testing H1and observed 90% power at around 630 participants (Supplementary Note 1, Supplementary Figure 1). | Ordered beta regression with participants nested in triads as random intercepts, payoff structure (individual: 0; group: 1) as the predictor and compliance as DV. We will only focus on the no punishment treatments. | See above. | H10 is rejected if the main effect of payoff structure is statistically significant, the 95% CI is not entirely included in the equivalence bounds, and compliance is higher for individual payoff compared to group payoff.  H10 is not rejected if the 90% TOST CI of the main effect of payoff structure is entirely included within the equivalence bounds. | If the hypothesis is rejected/not rejected, this would provide evidence for/against the theory that dishonesty increases in group contexts (Leib et al., 2021; Weisel & Shalvi, 2022). |
| RQ2. Does the risk of punishment reduce dishonesty? | H2. A main effect of risk of punishment. Risk of punishment (vs. no punishment) reduces dishonesty (i.e., non-compliance) | We simulated a model accounting for our SESOI (*d* = .15) when testing H1and observed 90% power at around 630 participants (Supplementary Note 1). | Ordered beta regression with participants nested in triads as random intercepts and compliance as DV. punishment (no vs. individual vs. group) as the predictor with two contrast codings (punishment risk -2/3 vs. 1/3 vs. 1/3; punishment type: 0 vs. –½ vs. ½). | See above. | H20 is rejected if the main effect of punishment risk is statistically significant, the 95% CI is not entirely included in the equivalence bounds, and compliance is higher for punishment risk compared to no punishment risk.  H20 is not rejected if the 90% TOST CI of the main effect of punishment risk is entirely included within the equivalence bounds. | If the hypothesis is rejected/not-rejected, this would provide evidence confirming/ disconfirming utility-based models of dishonesty (Becker, 1968) that expect a decrease of dishonesty with an increase of risk of punishment. |
| RQ4. Does individual punishment more strongly decrease dishonesty compared to group punishment? | H4. A main effect of type of punishment. Individual punishment will show stronger effects in reducing dishonesty compared to group punishment. | We simulated a model accounting for our SESOI (*d* = .15) when testing H1and observed 90% power at around 630 participants (Supplementary Note 1). | See above. | H40 is rejected if the main effect of punishment type is statistically significant, the 95% CI is not entirely included in the equivalence bounds, and compliance is higher for group punishment compared to individual punishment.  H40 is not rejected if the 90% TOST CI of the main effect of punishment type is entirely included within the equivalence bounds. | There is no explicit theory about more effectiveness of individual or group punishment strategies. |
| RQ3. Does the effectiveness of punishment risk depend on the type of decision making? | H3. An interaction effect of payoff structure and punishment risk. Dishonesty will be higher for group payoff structure compared to the individual payoff structure when no punishment is possible (i.e., H1), but this will reverse for situations in which punishment is possible with the individual payoff structure showing higher dishonesty compared to the group payoff structure. | We simulated a model accounting for the expected interaction effect (h2 = .01/.08) when testing H3 and observed a power of 95% at 400 participants (Supplementary Note 1). | Ordered beta regression with participants nested in triads as random intercepts and compliance as the DV. As predictors we will add payoff structure (individual: 0; group: 1) and punishment (punishment risk -2/3 vs. 1/3 vs. 1/3; punishment type: 0 vs. –½ vs. ½) and their interactions. | See above. | H30 is rejected if the interaction of punishment risk and payoff structure is statistically significant, the 95% CI is not entirely included in the equivalence bounds, and compliance is lower for group-based decision making for now punishment but higher for punishment.  H30 is not rejected if the 90% TOST CI of the interaction effect of punishment type and payoff structure is entirely included within the equivalence bounds. | If the hypothesis is rejected, this would provide evidence against the theory that in general dishonesty increases in group contexts (Leib et al., 2021; Weisel & Shalvi, 2022) but provide more nuanced evidence that this could be depended on whether punishment is likely. |
| RQ5. Does group punishment or individual punishment interact with risk of punishment? | H50. We expect no interaction effect of type of punishment and payoff structure. Group payoff increases dishonesty compared to individual payoff (across all punishment treatments). | We simulated a model accounting for our SESOI (h2 = .008) when testing H50 and observed 90% power at around 600 participants (Supplementary Note 1). | See above. | H50 is rejected if the interaction of punishment type and payoff structure is statistically significant.  H50 is not rejected if the 90% TOST CI of the interaction effect of punishment type and payoff structure is entirely included within the equivalence bounds. | There is limited theorization on the interaction of punishment risk and punishment type. If H50 is rejected this would provide evidence that the effect of dishonesty increasing in groups (Weisel & Shalvi, 2022) is moderated by the enforcement of specific punishment types. |
|  |  |
| RQ6. Is group punishment more effective for group-based decision making compared to individual decision making? | H6. We expect group punishment to reduce dishonesty more strongly for the group payoff structure treatment compared to the individual payoff structure treatment. | We simulated a model accounting for our SESOI (*d* = .15) when testing H1and observed 90% power at around 630 participants (Supplementary Note 1). | Ordered beta regression with participants nested in triads as random intercepts, payoff structure (individual: 0; group: 1) as the predictor and compliance as DV. We will only focus on the group punishment treatments. | See above. | H60 is rejected if the main effect of payoff structure is statistically significant, the 95% CI is not entirely included in the equivalence bounds, and compliance is higher for group payoff compared to individual payoff.  H60 is not rejected if the 90% TOST CI of the main effect of payoff structure is entirely included within the equivalence bounds. | If this hypothesis is rejected this would provide evidence that group punishment can be effective for group-based decision making (Bonfim & Silva, 2019) |

Note. DV = Dependent variable; SESOI = smallest effect size of interest; CI = confidence interval; TOST = two one-sided t-tests

**References**

ACFE. (2022). *Occupational Fraud 2022: A Report to the Nations by ACFE*. https://legacy.acfe.com/report-to-the-nations/2022/

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology*, *90*(1), 94.

Allingham, M. G., & Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, *1*(3–4), 323–338.

Alm, J., & Malézieux, A. (2021). 40 years of tax evasion games: A meta-analysis. *Experimental Economics*, *24*(3), 699–750. https://doi.org/10.1007/s10683-020-09679-3

Arel-Bundock, V. (2023). *marginaleffects: Predictions, comparisons, slopes, marginal means, and hypothesis tests* [Manual]. https://CRAN.R-project.org/package=marginaleffects

Balafoutas, L., Beck, A., Kerschbamer, R., & Sutter, M. (2015). The hidden costs of tax evasion.: Collaborative tax evasion in markets for expert services. *Journal of Public Economics*, *129*, 14–25.

Batolas, D., Perkovic, S., & Mitkidis, P. (2023). Psychological and Hierarchical Closeness as Opposing Factors in Whistleblowing: A Meta-Analysis. *Journal of Business and Psychology*, *38*(2), 369–383. https://doi.org/10.1007/s10869-022-09849-5

Bauckloh, T., Hardeck, I., Inger, K. K., Wittenstein, P., & Zwergel, B. (2021). Spillover effects of tax avoidance on peers’ firm value. *The Accounting Review*, *96*(4), 51–79.

Becker, G. S. (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy*, *76*(2), 169–217.

Bellé, N., & Cantarelli, P. (2017). What causes unethical behavior? A meta-analysis to set an agenda for public administration research. *Public Administration Review*, *77*(3), 327–339.

Bennett, R. J. (1998). Taking the sting out of the whip: Reactions to consistent punishment for unethical behavior. *Journal of Experimental Psychology: Applied*, *4*(3), 248–262. https://doi.org/10.1037/1076-898X.4.3.248

Blackwell, C. (2010). A meta-analysis of incentive effects in tax compliance experiments. *Developing Alternative Frameworks for Explaining Tax Compliance*, *59*, 97.

Bolker, B. (2016). Getting started with the glmmTMB package. *Vienna, Austria: R Foundation for Statistical Computing. Software*. https://cran.r-hub.io/web/packages/glmmTMB/vignettes/glmmTMB.pdf

Bonfim, M. P., & Silva, C. A. T. (2019). Inhibitory Mechanisms on Dishonesty of Groups and Individuals. *Journal of Accounting, Management and Governance*, *22*(2), 205–226.

Bonner, J. M., Wang, C. S., & Greenbaum, R. L. (2016). Punishment Contingency and Unethical Behavior: The Role of Uncertainty and Justice Perceptions. *Academy of Management Proceedings*, *2016*(1), 16910. https://doi.org/10.5465/ambpp.2016.70

Bourveau, T., She, G., & Žaldokas, A. (2020). Corporate Disclosure as a Tacit Coordination Mechanism: Evidence from Cartel Enforcement Regulations. *Journal of Accounting Research*, *58*(2), 295–332. https://doi.org/10.1111/1475-679X.12301

Brink, W. D., Eaton, T. V., Grenier, J. H., & Reffett, A. (2019). Deterring Unethical Behavior in Online Labor Markets. *Journal of Business Ethics*, *156*(1), 71–88. https://doi.org/10.1007/s10551-017-3570-y

Bundesministerium der Finanzen. (2023). *Ergebnisse der steuerlichen Betriebsprüfungen der Länder 2022—Bundesfinanzministerium—BMF-Monatsbericht Oktober 2023*. Bundesministerium der Finanzen. https://www.bundesfinanzministerium.de/Monatsberichte/2023/10/Inhalte/Kapitel-3-Analysen/3-2-steuerliche-betriebspruefung-2022.html

Cagala, T., Glogowsky, U., & Rincke, J. (2021). Detecting and preventing cheating in exams: Evidence from a field experiment. *Journal of Human Resources*.

Cagala, T., Glogowsky, U., Rincke, J., & Schudy, S. (2024). Commitment requests do not affect truth-telling in laboratory and online experiments. *Games and Economic Behavior*, *143*, 179–190. https://doi.org/10.1016/j.geb.2023.11.014

Casal, S., Kogler, C., Mittone, L., & Kirchler, E. (2016). Tax compliance depends on voice of taxpayers. *Journal of Economic Psychology*, *56*, 141–150. https://doi.org/10.1016/j.joep.2016.06.005

Charness, G., & Jackson, M. O. (2009). The role of responsibility in strategic risk-taking. *Journal of Economic Behavior & Organization*, *69*(3), 241–247.

Charness, G., Karni, E., & Levin, D. (2007). Individual and Group Decision Making under Risk: An Experimental Study of Bayesian Updating and Violations of First-Order Stochastic Dominance”. *Journal of Risk and Uncertainty*.

Chirikov, I., Shmeleva, E., & Loyalka, P. (2020). The role of faculty in reducing academic dishonesty among engineering students. *Studies in Higher Education*, *45*(12), 2464–2480. https://doi.org/10.1080/03075079.2019.1616169

De Vries, R. E. (2013). The 24-item brief HEXACO inventory (BHI). *Journal of Research in Personality*, *47*(6), 871–880.

Diller, M., Lorenz, J., Schneider, G., & Sureth-Sloane, C. (2024). Is tax transfer pricing harmonization a panacea? Real effects of global tax transparency and standards consistency. *The Accounting Review*, 1–32.

Dimant, E., & Shalvi, S. (2022). Meta-nudging honesty: Past, present, and future of the research frontier. *Current Opinion in Psychology*, *47*, 101426. https://doi.org/10.1016/j.copsyc.2022.101426

Dimant, E., Van Kleef, G. A., & Shalvi, S. (2020). Requiem for a nudge: Framing effects in nudging honesty. *Journal of Economic Behavior & Organization*, *172*, 247–266.

Dölling, D., Entorf, H., Hermann, D., & Rupp, T. (2009). Is deterrence effective? Results of a meta-analysis of punishment. *European Journal on Criminal Policy and Research*, *15*, 201–224.

Engel, C., & Nagin, D. (2015). Who is afraid of the stick? Experimentally testing the deterrent effect of sanction certainty. *Review of Behavioral Economics*, *2*(4), 405–434.

European Commission. (2024, November 14). *Commission fines Meta* [Text]. European Commission - European Commission. https://ec.europa.eu/commission/presscorner/detail/en/ip\_24\_5801

Fochmann, M., Fochmann, N., Kocher, M. G., & ... (2021). Dishonesty and risk-taking: Compliance decisions of individuals and groups. In *Journal of Economic …*. Elsevier. https://www.sciencedirect.com/science/article/pii/S0167268121000822

Friesen, L. (2012). Certainty of Punishment versus Severity of Punishment: An Experimental Investigation. *Southern Economic Journal*, *79*(2), 399–421. https://doi.org/10.4284/0038-4038-2011.152

Gamliel, E., & Peer, E. (2013). Explicit risk of getting caught does not affect unethical behavior. *Journal of Applied Social Psychology*, *43*(6), 1281–1288. https://doi.org/10.1111/jasp.12091

Hertwig, R., & Mazar, N. (2022). Toward a taxonomy and review of honesty interventions. *Current Opinion in Psychology*, 101410.

IRS. (2024). *IRS releases Strategic Operating Plan update outlining future priorities; transformation momentum accelerating following long list of successes for taxpayers | Internal Revenue Service*. https://www.irs.gov/newsroom/irs-releases-strategic-operating-plan-update-outlining-future-priorities-transformation-momentum-accelerating-following-long-list-of-successes-for-taxpayers

Isenberg, D. J. (1986). Group polarization: A critical review and meta-analysis. *Journal of Personality and Social Psychology*, *50*(6), 1141.

Jiang, S., & Villeval, M. C. (2024). Dishonesty as a collective-risk social dilemma. *Economic Inquiry*, *62*(1), 223–241. https://doi.org/10.1111/ecin.13171

Kocher, M. G., Schudy, S., & Spantig, L. (2018). I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. *Management Science*, *64*(9), 3995–4008.

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

Lohse, T., & Simon, S. A. (2021). Compliance in teams–implications of joint decisions and shared consequences. *Journal of Behavioral and Experimental Economics*, *94*, 101745.

Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, *51*(3), 485–504.

Mahasuweerachai, P., & SrungBoonmee, T. (2021). The relative effectiveness of punishment severity and risk of apprehension in deterring dishonest behavior: Experimental results from a developing country. *Kasetsart Journal of Social Sciences*, *42*(2), Article 2.

Mann, H., Garcia-Rada, X., Hornuf, L., & Tafurt, J. (2016). What deters crime? Comparing the effectiveness of legal, social, and internal sanctions across countries. *Frontiers in Psychology*, *7*, 85.

Matthaei, E., & Kiesewetter, D. (2020). *A problem shared is a problem halved? Risky tax avoidance decisions and intra-group payoff conflict*. https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3626982

Miceli, T. J., & Segerson, K. (2007). Punishing the Innocent along with the Guilty: The Economics of Individual versus Group Punishment. *The Journal of Legal Studies*, *36*(1), 81–106. https://doi.org/10.1086/509274

Molnar, A. (2019). SMARTRIQS: A Simple Method Allowing Real-Time Respondent Interaction in Qualtrics Surveys. *Journal of Behavioral and Experimental Finance*, *22*, 161–169. https://doi.org/10.1016/j.jbef.2019.03.005

Motro, D., Ordóñez, L. D., Pittarello, A., & Welsh, D. T. (2018). Investigating the Effects of Anger and Guilt on Unethical Behavior: A Dual-Process Approach. *Journal of Business Ethics*, *152*(1), 133–148. https://doi.org/10.1007/s10551-016-3337-x

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

Pereira, A., & Prooijen, J.-W. van. (2018). Why we sometimes punish the innocent: The role of group entitativity in collective punishment. *PLOS ONE*, *13*(5), e0196852. https://doi.org/10.1371/journal.pone.0196852

Podsakoff, P. M., Bommer, W. H., Podsakoff, N. P., & MacKenzie, S. B. (2006). Relationships between leader reward and punishment behavior and subordinate attitudes, perceptions, and behaviors: A meta-analytic review of existing and new research. *Organizational Behavior and Human Decision Processes*, *99*(2), 113-142.

Pratt, T. C., Cullen, F. T., Blevins, K. R., Daigle, L. E., & Madensen, T. D. (2017). The empirical status of deterrence theory: A meta-analysis. In *Taking stock* (pp. 367–395). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315130620-14/empirical-status-deterrence-theory-meta-analysis-travis-pratt-francis-cullen-kristie-blevins-leah-daigle-tamara-madensen

R Core Team. (2022). *R: A language and environment for statistical computing.* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Raihani, N. J., Thornton, A., & Bshary, R. (2012). Punishment and cooperation in nature. Trends in ecology & evolution, 27(5), 288-295.

Rullo, M., Presaghi, F., Baldner, C., Livi, S., & Butera, F. (2024). Omertà in intragroup cheating: The role of ingroup identity in dishonesty and whistleblowing. *Group Processes & Intergroup Relations*, *27*(1), 41–61. https://doi.org/10.1177/13684302231164722

Ścigała, K. A., Zickfeld, J. H., Bahník, S., Bjørsting, C. K., Capraro, V. … Pfattheicher, S. (2025). Megastudy on workplace deviance: Rule breaking at work. *Manuscript in preparation.*

Shupp, R. S., & Williams, A. W. (2008). Risk preference differentials of small groups and individuals. *The Economic Journal*, *118*(525), 258–283.

Siniver, E., Tobol, Y., & Yaniv, G. (2022). Collective Punishment and Cheating in the Die-Under-the-Cup Task. *Experimental Psychology*. https://econtent.hogrefe.com/doi/10.1027/1618-3169/a000543

Solaz, H., De Vries, C. E., & de Geus, R. A. (2019). In-group loyalty and the punishment of corruption. *Comparative Political Studies*, *52*(6), 896–926.

Teodorescu, K., Plonsky, O., Ayal, S., & Barkan, R. (2021). Frequency of enforcement is more important than the severity of punishment in reducing violation behaviors. *Proceedings of the National Academy of Sciences*, *118*(42), e2108507118. https://doi.org/10.1073/pnas.2108507118

Thielmann, I., & Hilbig, B. E. (2018). Daring dishonesty: On the role of sanctions for (un)ethical behavior. *Journal of Experimental Social Psychology*, *79*, 71–77. https://doi.org/10.1016/j.jesp.2018.06.009

US Department of Justice. (2017, January 11). *Office of Public Affairs | Volkswagen AG Agrees to Plead Guilty and Pay $4.3 Billion in Criminal and Civil Penalties; Six Volkswagen Executives and Employees are Indicted in Connection with Conspiracy to Cheat U.S. Emissions Tests | United States Department of Justice*. https://www.justice.gov/archives/opa/pr/volkswagen-ag-agrees-plead-guilty-and-pay-43-billion-criminal-and-civil-penalties-six

Vainapel, S., Weisel, O., Zultan, R., & Shalvi, S. (2019). Group moral discount: Diffusing blame when judging group members. *Journal of Behavioral Decision Making*, *32*(2), 212–228.

Verschuere, B., Meijer, E. H., Jim, A., Hoogesteyn, K., Orthey, R., McCarthy, R. J., Skowronski, J. J., Acar, O. A., Aczel, B., Bakos, B. E., Barbosa, F., Baskin, E., Bègue, L., Ben-Shakhar, G., Birt, A. R., Blatz, L., Charman, S. D., Claesen, A., Clay, S. L., … Yıldız, E. (2018). Registered Replication Report on Mazar, Amir, and Ariely (2008). *Advances in Methods and Practices in Psychological Science*, *1*(3), 299–317. https://doi.org/10.1177/2515245918781032

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

Weisel, O., & Shalvi, S. (2022). Moral currencies: Explaining corrupt collaboration. *Current Opinion in Psychology*, *44*, 270–274. https://doi.org/10.1016/j.copsyc.2021.08.034

Zickfeld, J. H., Karg, S. T. S., Engen, S. S., Gonzalez, A. S. R., Michael, J., & Mitkidis, P. (2024). Committed (dis)honesty: A systematic meta-analytic review of the divergent effects of social commitment to individuals or honesty oaths on dishonest behavior. *Psychological Bulletin*, *150*(5), 586–620. https://doi.org/10.1037/bul0000429

Zickfeld, J. H., Ścigała, K. A., Weiss, A., Michael, J., & Mitkidis, P. (2023). Commitment to honesty oaths decreases dishonesty, but commitment to another individual does not affect dishonesty. *Communications Psychology*, *1*(1), 27.

Zickfeld, J. H., Ścigała, K. A., Elbæk, C. T., Michael, J., Tønnesen, M. H., Levy, G., ... & Mitkidis, P. (2025). Effectiveness of ex ante honesty oaths in reducing dishonesty depends on content. Nature Human Behaviour, 9(1), 169-187.

1. The current manuscript focuses on organizational tax fraud. There are different potentially dishonest practices that might negatively impact organizations or society such as paying a low percentage of taxes through so-called *tax havens*, which is not the focus of the current investigation. [↑](#footnote-ref-2)
2. Siniver et al. (2022) tested group punishment in a small group of students (*n* = 49) in an incentivized die-roll task. After a first round of a die-roll, the experimenters announced that on average cheating occurred and that if such behavior persisted in the next round the whole group would be punished. Against predictions, announcing possible group-punishment increased dishonesty in the second round. Importantly, the sample size of the study is considerably low, and it did not include a treatment without punishment or individual punishment, strongly limiting any conclusions that can be drawn from these findings, which is why we do not discuss these further. [↑](#footnote-ref-3)
3. Audit rate is based on previous study by Fochmann et al. (2021). We acknowledge that in real-life settings audit rates can differ considerably based on company size and region. Some regions apply or plan to apply audit rates that are similar to the current one (e.g., Bundesministerium der Finanzen, 2023; IRS, 2024). [↑](#footnote-ref-4)
4. Notably, the expected company income applies to a single decision as one round is randomly chosen for payoff, while the design of the current study features five repetitions. An updated equation would consider possible learning effects due to experiences in previous rounds. To eliminate possible learning effects, participants only receive information about their final income and possible audits at the end of the study after the five rounds. [↑](#footnote-ref-5)