The Hidden Costs of Whistleblower Protection∗

Niklas Wallmeier†

January 26, 2018

Abstract

We conduct a laboratory experiment to analyze cooperative behavior between a manager and an employee in the presence of misbehavior and protected whistleblowing. Before taking part in a trust game with her employee, a manager has the opportunity to embezzle money at the expense of a third party. Her behavior is observed by the unaffected employee who may trigger an investigation by a report. We vary the framework with respect to monetary incentives and anonymity in case of a report and compare misbehavior, reporting and cooperative behavior across treatments. Our results suggest that a whistleblower law could deter wrongdoing, but could also have a detrimental effect on cooperation in organizations when it increases the likelihood for fraudulent whistleblowing.

JEL-Codes: C91, D73, K42, M51

Keywords: corporate fraud, corruption, laboratory experiment, business ethics, whistleblowing.

∗ The author would like to thank Ralph Bayer, Leonie Gerhards, Christos Litsios, Gerd Mühlheußer, Andreas Roider, Karl Schlag, Simeon Schudy and Kathrin Thiemann as well as participants of the LEOH conference (Cologne) and the RWI research seminar (Essen) for their valuable comments and suggestions. Pamela Mertens and Olaf Bock provided excellent research assistance. Financial support by the Fritz Thyssen Foundation (Grant 10.13.2.097) is gratefully acknowledged.

† Department of Economics, University of Hamburg, Von-Melle-Park 5, 20146 Hamburg, Germany; niklas.wallmeier@wiso.uni-hamburg.de.
1 Introduction

1.1 Motivation

In an era of corporate fraud causing severe damages, whistleblowing is found to be a major source of fraud detection. Consequently, whenever a large corporate scandal is unveiled by insiders, public discussions emerge how to support employee whistleblowers in coming forward by providing legal protection. This paper investigates experimentally the behavioral effects of protection in the form of incentivized and anonymous whistleblowing in two dimensions. We are interested how the reporting behavior of employees and the compliance of managers change after whistleblower protection is introduced. To our best knowledge, this is the first paper that analyzes how the cooperative climate between employer and employee within an organization is affected by changes in the legal framework. The results suggest that an institutional change increasing expected whistleblowing drives down managerial wrongdoing, but also leads to a decline in productive cooperation.

The extensive and widespread economic damage of corporate fraud is well documented. In a survey by the Economist Intelligence Unit (2015), 75% of surveyed companies reported they had become a fraud victim in the previous year, which is an increase of 14 percentage points from 2012 to 2015, while the Association of Certified Fraud Examiners (2014) find that the average loss caused by fraud amounts to 5% of annual revenues. In a long-term study, Dyck et al. (2017) estimate the average yearly damage of the U.S. economy due to detected and undetected fraud in the range from $180 to $360 billion. Accordingly, detection and deterrence of corporate fraud as has become a major target for international policy makers.

This study analyzes the cost and benefits of legal protection for employee whistleblowers which is one relevant instrument to fight corporate fraud. There are intuitive arguments for the use of insider knowledge for law enforcement. First of all, a share of fraud cases cannot be detected by external actors due to their lag of necessary insider knowledge. Therefore, whistleblowers increase the share of fraud cases that can be detected. Second, whistleblowing might not only facilitate law enforcement, but the mere thread of insiders reporting could deter wrongdoing ex-ante.

Correspondingly, evidence in favor of whistleblowing as an instrument for crime deterrence is prominent in economic literature. For example, Dyck et al. (2010) provide

---

1 We focus on whistleblowing as organization members’ disclosure of illegitimate practices under the control of their employers, to organizations may be able to effect action as defined by Near and Miceli (1985).
evidence on the general importance of non-traditional governance actors for fraud detection. Investigating fraud in the U.S. economy between 1996 and 2004, they find employee whistleblowers involved in detection in 17% of the cases having a larger share than the SEC, auditors, or the media. Furthermore, the fraction of cases detected with the help of whistleblowers has increased over the past years. The Annual Global Fraud Survey finds for 2015 that 41% of the detected fraud cases were exposed by whistleblowers (Economist Intelligence Unit 2015). According to the Association of Certified Fraud Examiners (2014), employees were the source in 49% of tips leading to the detection of fraud. These numbers strongly suggest that whistleblowing has already become a major resource for crime detection.

Yet, becoming a whistleblowers comprises a veritable negotiation for an organization member, since she potentially faces costs from a breach of loyalty and career risks. Academic research in business ethics particularly identifies the fear of retaliation, e.g., a dismissal or a denied promotion, as a major obstacle for whistleblowers that has to be overcome, or eventually thwarts whistleblowing (see, e.g. Near and Miceli 1986; Alford 2001; Rehg et al. 2008; Cassematis and Wortley 2013). As a consequence, whistleblowers might be encouraged to come forward by legally protecting them from retaliation.

To this end, international organizations as the G20 group, or the OECD requested protection for whistleblowers (OECD 2016) and legislators made an effort to increase the legal certainty. Prominent examples can be found in the United States with the Sarbanes-Oxley Act - passed in reaction to the whistleblowing-induced collapses of Enron and WorldCom (Healy and Palepu 2003), the Dodd-Frank Act, and the Public Interest Disclosure Act in the UK.

The most-frequent features of whistleblower protection are protection of employment, i.e. guaranteeing income (see e.g., Kohn et al. 2004, pp. 97) and allowing for anonymous reporting. These schemes should increase the willingness to report misbehavior, thereby help to uncover a larger share of fraud, but also - the organization anticipating this increase - deter the misbehavior in the first place.

However, these legal approaches are discussed controversially, since these benefits might come at a cost. To increase the legal certainty for the whistleblower reporting may not only be protected (or provided with incentives), but this protection must also be obtainable at a sufficiently low cost. Consequently, legislation often do not condition the protection grant on a successful investigation, instead a wide spread content of whistleblowing laws is the low barrier of a ‘reasonable belief’ to obtain the protection (Kohn 2015 out of 23 surveyed countries have implemented a specific whistleblowing law (see Thüising and Forst 2016).

\[^2\] 15 out of 23 surveyed countries have implemented a specific whistleblowing law (see Thüising and Forst 2016).

\[^3\] Nine of fifteen countries allow for anonymous reporting (Thüising and Forst 2016).
et al., 2004, pp. 92). While obviously unfounded complaints are deterred with this standard, one resulting adverse effect may be nevertheless an increase in fraudulent claims. That means blowing the whistle although no underlying fraud has happened to reap the benefits from protection and thereby inducing a damage for the organization and the regulatory agency (Callahan and Dworkin, 1992; Howse and Daniels, 1995; Givati, 2016). On the one hand, these fraudulent claims would cause damage for the organization due to loss in reputation form being investigated. On the other hand, the effort for screening claims for their adequacy would drive down the efficiency of the authorities.

Furthermore, in a cooperate context, efficiency does not solely rely on lawful behavior, but also on productive cooperation. For such cooperation to flourish, employers and employees need to share resources and confidential information, which requires a sufficient level of trust. This trust between co-workers might be negatively affected if employees use sensitive information to file a complaint. For example, the employer’s motivation for dismissal may not be punishment, but reputational concerns which make it unbearable to retain a whistleblowing - fraudulent or honest - in the organization and continue the collaboration. These concerns may not only occur for actual observed behavior, but also for expected reactions to the legislation. Given anonymity, the manager may now expect a larger share of her employees to blow the whistle on her, leading her to decrease cooperation. Therefore, if a whistleblower protection law encourages more active reporting behavior, or even if it increases the expected frequency of whistleblowing, it may cause an atmosphere of distrust within an organization, which has detrimental effects on beneficial cooperation (Dworkin and Near, 1997).

This paper focuses on these twofold hidden costs of whistleblowing and evaluates them against the benefits of detected and deterred misbehavior due to protection laws in an experimental setting.

1.2 Research Question, Framework and Results

Our main goal is to investigate the influence of whistleblower protection on reporting behavior, compliance and cooperation. Therefore, we create a workplace setup in the lab in which a manager and an employee share information and cooperate productively. At the beginning of a period, the manager has the opportunity to embezzle money and increase her payoff at the expense of a real third party. Her choice is always observed by...
the employee. While the employee’s payoff is neither negatively nor positively affected by the embezzlement, she can become a whistleblower and trigger an investigation by reporting misbehavior to an authority, irregardless of the actual decision of the manager. In contrast to other studies, we model the authority as an automaton that responds perfectly to a report reflecting the standard reasonable belief. In consequence, the manager can tell from an investigation that the whistle was blown. If a report is filed and an investigation happens, reputational cost from the investigation for the manager arise and - if embezzlement is detected - she has also to pay a fine that partly reinstates the third party.

At the end of a period, the manager and the employee interact in a trust game [Berg et al., 1995]. As the sender, the manager decides first which share of her endowment to trust to her employee or to take from the employee’s endowment. If this amount is positive, i.e. productive cooperation takes place, it is tripled, and sent to the employee who can in turn decide which fraction of the amount received she wants to return. If the manager choose to take money, that means beneficial cooperation takes not place, the amount is simply transferred and the period ends.

We alter the framework in two ways: Compared to a baseline treatment, representing the status quo legislation without any protection, we consider the two most frequent instruments of whistleblower protection, namely incentives and the provision of anonymity. Incentives are based on the most common feature of employment protection and take the rather mild form of guaranteed income, i.e. guaranteeing the employee that the manager cannot take any of her endowment if she files a report. Anonymity allows the employee to report without revealing her action to the manager. Since an investigation would only happen after the trust game, she can be assured of the manager not condition her cooperation on the decision to blow the whistle.

By introducing incentivized and anonymous reporting one by one, we change the environment stepwise towards a stronger protection for the whistleblower. This setup allows to track the influence of the protection mechanisms on the employee’s willingness to blow the whistle truthfully and fraudulently, as well as the compliance behavior and the manager’s willingness to cooperate. In the context of whistleblowing, a laboratory approach has two major advantages compared to the field where only detected fraud is observable, such that the true amount of misbehavior remains unknown. Furthermore,

\footnote{see Mechtenberg et al. (2017) on the informativeness of a whistleblower report if the authority has to evaluate a complaint and Chassang and Padró i Miquel (2016) on the informational content of an investigation for the employer.}

\footnote{For studies on bounties as whistleblower rewards see e.g. Schmolke and Utikal (2016); Buccirossi et al. (2017); Butler et al. (2017).}
we only observe reporting behavior given the state of compliance. That means, we can account for truthful reporting when fraud was conducted and for fraudulent reporting in the case of compliance, but not for the hypothetical behavior in the state that has not been realized. Choosing a laboratory approach solves both of these informational and counterfactual issues. In addition, a number of studies show a high out-of-lab correlation in unethical behavior (see for discussion Abeler et al., 2016).

Our results show that both incentivized and anonymous reporting increase honest reporting and, in turn, increase compliance. That means whistleblower protection affects the behavior of all parties in the intended way. At the same time, both instruments induce adverse incentives for the employees and lead to an increase in fraudulent whistleblowing. For the managers’ willingness to cooperate, we find an inverse relation to the frequency of investigations they experience. This phenomenon can be explained best by the perception of (fraudulent) whistleblowing as an unkind behavior, which negatively affects the manager’s trust in her employee. The joint use of incentives and the provision of anonymity for reporting leads to a peak of investigations and drives down cooperation significantly.

The following section will review related work from business, sociology and economics literature. Sections 3 and 4 present the experimental design and the behavioral predictions, which will be analyzed in section 5 before section 6 discusses the results.

2 Related Literature

This study is the first to investigate the relation of whistleblowing and cooperation considering i) an unaffected whistleblower and ii) an affected real third party. Furthermore, we are not aware of studies that apply the concept of anonymous reporting to the whistleblowing context.

The study closest to our paper is by Mechtenberg et al. (2017) on the protection of whistleblowers. In a theory-guided lab experiment, they investigate the influence of different whistleblower protection laws on compliance, reporting behavior and retaliation against the whistleblower. In addition, they analyze the investigation decision of the regulatory agencies given the different legal frameworks. They find the desired increasing effect of whistleblower protection on reporting. However, when the legal protection also fosters fraudulent reporting, whistleblowing becomes a less informative signal to the regulatory agency such that a higher number of reports does not necessarily materialize in a higher number of investigations. In a framework where the employees are heterogeneous with respect to their productivity the dismissal decision of the manager could
be either driven by externally given efficiency concerns or by preferences for retaliation. We complement this study by internalizing the productivity of the collaboration. Still the manager could retaliate against the whistleblowing as a pure punishment. But if the reputational damage is not too high, it depends on whether she trusts her employee enough for the collaboration to be profitable, rather than on externally given productivity. Therefore, this framework applies to broader range of employee whistleblowers who are endangered by a dismissal and are inclined to conduct a fraudulent claim.

Furthermore, our work is related and contributes to three strands of the literature on whistleblowing. Recent experimental studies cover the effect of (monetary) incentives on the willingness to blow the whistle. [Bartuli et al. (2016)] analyze whistleblowing in a context where the employee faces a conflict between ethical considerations and monetary interests. They find employees who are more altruistic and more aware of ethical issues are more likely to refrain from supporting fraud and report wrongdoing. [Schmolke and Utikal (2016)] measure the effectiveness of incentives on the willingness to report. Fines for non-reporting, rewards and also commands increase the probability of whistleblowing in their setup. If whistleblowers are affected by the misconduct themselves, reporting is more likely if the enforcement authority is negatively affected as well. [Butler et al. (2017)] investigate the effect of monetary rewards on whistleblowing in the presence of potential crowding out of intrinsic motivation (see [Benabou and Tirole 2003; Gneezy et al., 2011], for theoretical foundation and overview about crowding out). They find an enhancing effect of monetary rewards on the willingness to report and no substantial crowding out of non-monetary motivations.

Another strand considers the effects of whistleblower protection schemes on efficiency. [Heyes and Kapur (2009)] develop a model which allows to operationalize several behavioral motivations for whistleblowing and they show that the optimal whistleblower protection regime depends on which motivation is the driving force. [Friebel and Guriev (2012)] show that the possibility of whistleblower protection might harm a firm’s productive efficiency if wrongdoers within the hierarchy ”bribe” other members of the organization. They show that whistleblower protection might reduce effort incentives. [Felli et al. (2016)] provide a model in which incentivized whistleblowing can prevent opportunistic behavior that takes the form of collusion or blackmail between supervisors and employees within an organization.

Our study investigates also the influence of incentives on reporting and the efficiency of whistleblower protection. It contributes to the existing literature by testing the effect of rather mild financial incentives, that is the guarantee to keep the endowment instead of an additional reward. Furthermore, it extends the relationship with the organization
the employee could blow the whistle on by adding a productive collaboration.

As a third strand, a large number of studies identify the fear of retaliation as an obstacle to reporting (see, e.g., Near and Miceli 1986; Alford 2001; Rehg et al. 2008; Cassematis and Wortley 2013). Highlighting the role of retaliation, Chassang and Padró i Miquel (2016) employ a cheap-talk approach in which an employee can send a report to a monitor, who in turn can then decide whether to intervene. They show that, in environments where anonymous reporting is not feasible, the optimal intervention policy must garble the whistleblower’s message, because a very responsive policy would lead to retaliation and prevent reporting in the first place. They assume that retaliation is costly for the manager such that she has commit to a retaliation strategy conditional on whistleblowing ex-ante. In contrast, we allow the manager to evaluate beneficial retaliation against potentially even more profitable cooperation. Reuben and Stephenson (2013) focus on the willingness to report lies of other members of the organization and highlight the retaliation associated with whistleblowing. They find that former whistleblowers are less likely to be chosen by an organization, even if it complies to the rules and would not have to fear an investigation. This career risk for a whistleblower is also included in our framework. Furthermore, we investigate whether managers continue to ostracize whistleblowing even if this means to pass on an opportunity for profitable cooperation.

In addition, the context of whistleblower protection is related to leniency programs in cartel prosecution. In such programs, cartel members who report their activities to the authorities are rewarded by a fine reduction. The most prominent studies analyze the effects of the reduction of fines or rewards in case of reporting a cartel. They find that a leniency program provided better outcomes than a pure fine regime or the introduction of rewards (Apesteguia et al. 2007; Hinloopen and Soetevent 2008; Feltovich and Hamaguchi 2016), but also a stabilizing effect for the remaining cartels (Bigoni et al. 2012).

As in the case of whistleblower protection, the analysis of such programs is concerned with the reporting of illegal activities from within the respective entity, but there is a crucial difference, since cartel members who report are wrongdoers themselves, while the typical whistleblower is an innocent bystander.

Besides the specific context of collusion, experimental studies have systematically investigated truth-telling in general (see e.g., Gneezy 2005; Mazar et al. 2008; Fischbacher and Föllmi-Heusi 2013; Muehlheusser et al. 2013; Abeler et al. 2016). Findings demonstrate that participants cheat for their own monetary advantage, but less than predicted

\[9\text{see also the surveys by Spagnolo 2008 Marvão and Spagnolo 2014}\]
by standard economic theory, already when there are no negative externalities to a third party. This feature of an affected third party is added in experiments about corruption. The usual setup models an opportunity for two players to collude beneficially at the cost of a not involved third party (see e.g., Abbink et al., 2002; Barr and Serra, 2009). Experimental studies have also investigated the punishment behavior of unaffected third parties. It has been shown that people are willing to punish violation of norms or unethical behavior in the lab, both if punishment is incentivized or not. This behavior can even be found if punishment is costly (see e.g., Fehr and Gächter, 2002; Fehr et al., 2002; Fehr and Fischbacher, 2004). Our study contributes to this literature by combining elements from all three strands in a unified setting.

3 Experimental Design

The game played in each period To investigate the influence of (protected) whistleblowing on misbehavior, reporting and cooperative behavior, we combine a whistleblowing game with a modified trust game. In this experiment, the subjects are assigned with the role of a manager, an employee, or a third party. Those who become a manager maintain their role throughout the experiment, whereas both the other two roles are reshuffled after each of the 16 periods. Before a period starts, groups of three are randomly formed with one subject of each role, such that subjects face a stranger matching and cannot infer any information from previous periods.

While the third party is completely passive, both the other roles have to make two decisions. In the whistleblowing game, the manager does or does not embezzle money in the first stage, which generates an exclusive revenue \( B \) for her and induces a cost \( C \) for the third party. In stage two, the employee needs to decide whether to file a complaint conditional on the compliance decision of the manager.\(^{10}\) Note that she is able to report an illegal action, irregardless whether it has happened or not. If a complaint is filed, cost from the investigation \( K \) arise for the manager in any case. If thereby an illegal activity is detected, a fine \( F \) is added and compensatory fees are paid to the third party \( R \) that partly cover the damage \( C \).

The trust game starts in stage three. The manager decides by choosing the level of the investment \( t \) whether beneficial cooperation takes place, and if so, to which extent. She can choose a negative amount, which would only mean a transfer from the employee’s

\(^{10}\) Using the strategy method (Selten, 1965) allows to keep track of reporting behavior independent of the compliance behavior which solves the counterfactual problem. Since the employee is not directly affected by the embezzlement, Brandts and Charness (2011) suggest that using the strategy method on reporting should not yield different results in this context.
endowment to her own payoff. This represents the opportunity to retaliate against the employee by taking some of her income. If she chooses to trust, i.e. \( t \) is positive, this amount is multiplied by three and transferred to the employee. Note that this multiplier captures the social benefit of cooperation. While taking endowment from the employee leaves the aggregated payoff for the group unchanged, trusting a positive amount increases this payoff. Finally in stage four, depending on the received investment, the employee can return an amount back to her manager. The general timing of a period is presented in Figure 1.

![Figure 1: General Timing](image)

**Cost and reward parameters** We choose the cost and reward parameters such that the aggregated payoffs for the group resulting from the subjects’ decisions correspond to their desirability from a social perspective. That means in terms of social welfare, (i) the manager complying to the law is always better than non-compliance, (ii) detected fraud is better than undetected fraud, and (iii) in case of compliance, reports should not arise.

The intuition is as follows: Clearly, the most preferred outcome would be the absence of misbehavior and reports. In this case neither damage from misconduct nor investigatory costs arise, which leaves all players with just their endowment and the social welfare unaffected \( (0) \). Assuming efficient law enforcement, the least favorable outcome is an undetected embezzlement. Here the manager reaps a benefit \( B (50) \), which is outweighed by the cost for the third party \( C (90) \). This would result in a social net loss of \( B - C (-40) \). Compared to this, a preferable outcome would be a detected embezzlement. The manager would have to pay a fine \( F (60) \) that exceeds her benefit from embezzlement (otherwise embezzlement would be a payoff maximizing strategy) and the costs of the investigation \( K (10) \). On the other side, the third party partially recovers her loss \( R (80) \), such that social welfare loss is reduced to \( C + R - K - F - I (-30) \). The fourth pos-

---

11 We follow a similar argumentation to Hart and Zingales (2017) that a firm does not necessarily
sibility is a fraudulent claim. Reporting although there has been no misconduct means that there is neither a damage for the third party nor a benefit for the manager, but it creates investigation cost $I$ (10) for the manager. This has to be positive to reflect the reputational costs of being investigated for compliance issues. If this would have a cost zero, it would indicate an indifference towards being investigated which is clearly not the case in reality (otherwise investigation should not rely on reports). Also, if the cost of investigation was too large and would outweigh the recovered loss, a social planner would prohibit an investigation. The four possible outcomes of the whistleblowing game are summarized below in Table 1

<table>
<thead>
<tr>
<th>Embezzlement</th>
<th>Whistleblowing</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td>0</td>
<td>-K (-10)</td>
</tr>
<tr>
<td>Yes</td>
<td>B - C (-40)</td>
<td>B + R - C - F - K (-30)</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Change in Welfare After the Whistleblowing Game

For the trust game, we impose a range from -30 to 60 on the amount $t$ that the manager can send to can her employee. We modeled $t$ as a choice from discrete steps of 10 length. Instead of a binary choice for the manager to punish or not to punish, we allow her to gradate the amount she wants to take from the employee. In this way we are able to disentangle the motives for the retaliation. The gradations of the negative values for $t$ give the manager the opportunity to differentiate whether she wants to either recover precisely an experienced loss from an investigation (10), or due to detected cheating (20), or simply guarantee herself a profit in any case if she chooses $t = -30$. For positive values of $t$ the upper bound is set to 60. This guarantees that the employee cannot punish the manager stronger by keeping the entire trusted investment than by filing a report. The endowment is set sufficiently high (100) that neither party could make a loss nor is restricted in her choice options. Figure 2 reports the payoffs for the three roles in a given period conditional on the decisions of the subjects.

maximizes its shareholders welfare by maximizing its market value. If the victim could be fully compensated for the damage, embezzlement reduces to be only a distributional issue and the original state could always be reconstructed.
\[
\pi_{\text{Manager}} = 100 + \text{Embezzlment} \times (50 - (60 \times \text{Report})) - \text{Report} \times 10 - t + s
\]
\[
\pi_{\text{Employee}} = 100 + \begin{cases} 
  t \times 3 - s & \text{if } t > 0 \\
  t & \text{if } t \leq 0 
\end{cases}
\]
\[
\pi_{\text{3rdParty}} = 100 - \text{Embezzlment} \times (90 - (80 \times \text{Report}))
\]

Figure 2: Payoffs

Treatments In this section, we present the design of the treatments used to separate the effects of interest. As mentioned above, we vary the legal environment in two dimensions, namely the status of incentives and of anonymity provision. Altering the protection schemes one by one results in a total of four treatments as depicted in Table 2.

<table>
<thead>
<tr>
<th>Anonymity</th>
<th>Incentives</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Baseline</td>
<td>Only anonymity (AN)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>Only incentives (IN)</td>
<td>Anonymity and incentives (ANIN)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Treatments

The four treatments differ with respect to the date when the manager is informed about the reporting decision (anonymity) and the choices for the manager in the trust game conditional on the reporting decision (incentives).

In the baseline treatment, equivalent to the illustration in Figure 1, the manager knows after stage two about the employee’s reporting decision, i.e. before she chooses \(t\), and is free to choose a negative \(t\) independent of the reporting decision.

In the AN treatment, in which only anonymous reporting is granted, the information about whistleblowing is only disclosed after stage four, i.e. after she chooses \(t\). The choice of \(t\) is again unrestricted for any reporting decision. This change in the timing guarantees that the manager cannot condition her cooperative behavior on the actual behavior of the employee. Nevertheless, the manager has the opportunity to retaliate against her employee, if she suspects an unkind action, although she cannot observe it.

In the IN treatment, in which only incentives are introduced, the manager knows
whether the employee blew the whistle after stage two as in the baseline treatment. In this treatment the feature of employment protection is modeled such that by filing a report the employee can guarantee her status quo payoff. That means, if there has been a report - truthful or fraudulent - \( t \) has to be at least zero.

Finally, in the ANIN treatment, when both incentives and anonymity provision are in place, the manager knows only after her choice on \( t \) whether the employee blew the whistle. In case the manager chose a negative \( t \), it is set ex-post to zero. The specific timing of each treatment is depicted in Figure 3.

These four treatments allow to identify two possible ways in which whistleblower protection might affect trust of managers towards their employees. One possible channel would be a direct influence through changed observed behavior of the employee. If incentives leads to an increase in whistleblowing - truthful or fraudulent - and this is perceived as unkind behavior, trust, and therefore cooperation, might go down. This is captured by the non-anonymous settings. A second possible way would be an indirect effect caused by expected behavior. In the anonymous settings, the manager cannot observe the actual behavior of the employee, but may form a belief about the whistleblowing likelihood. Again, if she expects an unkind behavior, the willingness to cooperate would go down. In this case, the distrust would be caused by institutional framework itself.

With respect to statistical power a within-subject design is preferable. However, confronting the subjects with four different treatments might bear disadvantages. For example, an increased number of treatments could pronounce the issue of order effects. Secondly, presenting several changes of the environment to one subject could provoke an experimenter demand effect and thereby bias the behavior. Introducing the anonymity environment means a larger modification, since it changes the timing of a period, while incentives only changes the allocatable endowment for the trust game. Therefore, we argue that the most natural design is to model the introduction of incentivized whistleblowing as a within-subject variation, and to capture the anonymity provision in a between-
subject design.

**Framing** We framed the experiment in a corporate context and spoke of employers and employees to support the subjects in understanding the hierarchical relation between the players. Furthermore, we chose to phrase the choice about embezzlement in a neutral way and spoke of alternatives to not induce an experimenter demand effect. However, using payoff tables (see Appendix A) control questions, and stating that the third party is affected, we made clear that the precise consequences for the manager as well as for the third party are understood. We gave a legal reminder that the alternative corresponding to embezzlement means a violation of law. Herewith we model an important feature of unethical decision making in the real world, since organizations are clearly aware of illegality of such decisions. The employee’s decision about a report was phrased as ‘filing a complaint’ to make them aware of the social undesirability of embezzlement. Drawing attention to unethical behavior may influence the subjects’ decisions, which would be appropriate for our specific fraud-related research question, though.

**Summary information** The experiment was programmed with the software zTree (Fischbacher, 2007) and conducted in the laboratory of the University of Hamburg, June 2016, using hroot for recruitment (Bock et al., 2014). Before a session, we conducted a number of control questions to ensure that the subjects understood the stranger matching as well as the decisions’ payoff consequences for every group member. To keep the incentives identical for every period over the entire experiment, in the end of a session, one period was randomly drawn for payout. We ran five sessions with a total number of 147 student subjects. The majority of the subject were enrolled in economics or business programs. The subjects received payments between 5.50 and 18.50 Euro (including a show-up fee of 5 Euro) with an average of 10.07 Euro.

4 Behavioral Predictions

In this section, we establish a set of behavioral predictions concerning compliance, truthful and fraudulent whistleblowing as well as cooperation conditional on the environment. Since the protection schemes, that are varied across the treatments, are intended to influence the employee’s behavior in the first place, we start with the predictions on truthful

---

1. Pruckner and Sausgruber (2013) provide evidence that legal reminders influence unethical behavior.
2. Alekseev et al. (2017) survey a wide range of experimental literature with respect to the instructions and find that meaningful language could be useful for understanding the environment. For the context of unethical behavior see Abbink and Hennig-Schmidt (2006); Barr and Serra (2009).
and fraudulent whistleblowing.

**Predictions on truthful whistleblowing (TW)** We first turn our attention to the baseline treatment. According to standard economic theory - in the absence of other-regarding preferences, the employee would be expected not to blow the whistle on embezzlement, since she is not directly affected by the misbehavior and has no further incentives to report. However, a large body of empirical field and experimental studies have accumulated evidence that norms of equity and fairness play an important role (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Andreoni and Miller, 2002). Also there is evidence for punishment from unaffected third parties, both if punishment is incentivized or not. (Fehr and Gächter, 2002; Fehr et al., 2002; Fehr and Fischbacher, 2004). On the other side, the manager may consider a report as disloyal behavior driving down her willingness to cooperate in the trust game. Therefore, we can assume that each employee has to weigh the benefits from a report, i.e. reduced social damage from an unfair behavior, against the costs of possible retaliation in the trust game. Based on the experimental evidence for fairness preferences, we expect that already in the baseline treatment a positive fraction of employees blow the whistle truthfully.

In the AN treatment, this trade-off changes slightly. While the possibility to recover the damage remains the same, the manager now cannot observe reporting and has to make her trust choice independent of the employee’s actual decision. Put differently, the benefit side for the employee stays constant, while the costs become independent of her decision. We conclude from this, that the share of truthful reporting should rise with the introduction of anonymity.

In the IN treatment, in which only incentives is added in contrast to the baseline treatment, the trade-off is affected in a different way. Still, the benefit from reporting the misconduct remains the same. On the other side, the employee still might suffer from lower trust, but she is now insured against losses from retaliation by the manager in the trust game if she reports. Therefore, also in the IN treatment truthful reporting should rise compared to the baseline treatment.

When both protection schemes are in place in the ANIN treatment, the employee again can not be punished in response to her behavior. In addition, reporting offers an insurance against managers who do not trust in either case. As a result, we expect the reporting frequency to increase compared to both the AN and the IN treatment.

Taken all together, the predictions on truthful reporting are summarized as follows:

**TW1:** *In the baseline treatment, truthful reporting will occur with a positive fre-
Predictions on fraudulent whistleblowing (FW) In a similar vein, we can predict the frequency of fraudulent claims. Starting with the baseline treatment, no incentives are provided for the employee to report alleged misconduct and nobody would benefit from this report. Only the manager would suffer a cost. Consequently, fraudulent whistleblowing should not occur. In the AN treatment, the picture remains the same. While the reporting behavior cannot be observed, there are still no incentives to report fraudulently, such that there should be no whistleblowing either. There are also no reasons to punish a manager by a fraudulent claim for experienced behavior, since the subjects interact in a repeated one-shot games with stranger matching.

If instead employment protection is introduced, the incentives compared to the baseline treatment are changed. In the IN treatment, the employee has the opportunity to insure herself against a not trusting manager by guaranteeing herself $t \geq 0$. This means in the decision upon fraudulent whistleblowing she faces a trade-off between imposing the investigation cost on the manager to avoid a potential loss and receiving a lower level of trust in response to this unkind behavior. Based on experimental evidence for subjects’ risk and loss-aversion (Harrison and Rutstrom, 2008), we expect an increase in fraudulent whistleblowing in this treatment.

The trade-off changes further in the ANIN treatment, when reporting is anonymous in addition. The employee can still insure herself against the potential loss, but the manager cannot retaliate against her actual behavior. Thus, the share of fraudulent whistleblowing should increase compared to the IN treatment. In contrast to truthful reporting, a fraudulent claim is an illegal action, similar to non-compliance by the manager. Since evidence from experiments on unethical behavior (Gneezy 2005; Mazar et al. 2008; Fischbacher and Föllmi-Heusi 2013; Abeler et al. 2016) suggests that individuals attach importance to moral concerns, even if unethical actions are unobserved and the expected willingness to cooperate by the manager is low, fraudulent reporting should not be maximal even in this treatment.

**FW1:** Fraudulent reporting will not occur in the baseline treatment

**FW2a:** Fraudulent reporting will be more frequent with anonymous reporting than with non-anonymous reporting when also incentives are given, but not without in-
centives

FW2b: Fraudulent reporting will be more frequent with incentives than without for a given status of anonymity

Predictions on embezzlement (EM) Predictions about the compliance behavior can be derived directly from the expected truthful whistleblowing behavior. Choosing to misbehave only pays off, if there is no whistleblowing. That means the higher the probability that the whistle will be blown, the lower the expected payoff from cheating. Thus, we can formulate the predictions on embezzlement inversely to those on truthful reporting. In addition, again with respect to evidence for costs of unethical behavior and fairness preferences, we expect the managers to embezzle with less than maximal frequency, even if it was maximizing expected profit.

EM1: Embezzlement will occur with less than maximal frequency in the baseline treatment
EM2a: Embezzlement will be less frequent with anonymous reporting than with non-anonymous reporting for a given status of incentives
EM2b: Embezzlement will be less frequent with incentives than without for a given status of anonymity

Predictions on cooperative behavior (CO) Concerning the willingness to cooperate, the behavioral predictions are ambiguous. To start with, standard economic theory would predict the manager not to cooperate at all. Since she would not expect the employee to return anything of the trusted amount, she maximizes her payoff by choosing the smallest possible $t$. However, evidence from experimental studies strongly suggests that subjects show reciprocal behavior and trust their counterparts (Fehr and Schmidt, 2006).14 If the trusting behavior is unaffected by the whistleblowing environment, no differences in cooperation should arise across the treatments. Alternatively, if managers do perceive reporting as an unkind behavior, according to the previous predictions, cooperation may vary with the institutional framework as well as with the number of whistleblowing cases. That means the higher the reporting frequency of the employee the lower the willingness to trust of the manager. If this behavioral response can be found for truthful claims, the effect could be even more pronounced for fraudulent whistleblowing, since a false report is a less reasonable cause for breaching loyalty than a truthful allegation.

14 Across studies, subjects usually invest half of their endowment in trust games and receive approximately the invested amount in return.
Furthermore, cooperation may depend on observed as well as on expected whistle-
blowing, such that the impact of increased reporting can also play out in the treatments
that feature anonymity. While the manager can respond directly to truthful and fraud-
ulent claims in the baseline and the IN treatment, she can also form expectations about
reporting behavior when whistleblowing is anonymous. As for the compliance decision
discussed above, the manager expects a certain probability for a truthful or fraudulent
report, and takes her cooperation decision in the anonymity treatments also with re-
spect to anticipated behavior. The resulting hypotheses for cooperative behavior are
summarized below.

**CO1**: Cooperation will occur with a positive frequency in the baseline treatment

**CO2**: Cooperation will decrease with expected and observed investigations

**CO3**: The decreasing effect of investigations on cooperation will be stronger for fraudulent claims

5 Results

This section analyzes how the treatments representing different legal frameworks affect
the subjects’ behavior. We compare the treatment differences to identify the effects of
incentives and anonymity provision on aggregated compliance, reporting and cooperative
behavior. To test for statistical significance, we use non-parametric tests with subject-
role-level averages as observational units. For between-subject differences we apply a
Mann-Whitney-U Test, while we account for within-subject differences with a Wilcoxon
Signed-Rank Test. The figures below report the fraction of subjects which opted for a
particular decision. There is one bar for each of the four treatments. As for the behav-
ioral predictions, we start with the reporting behavior of the employees and investigate
whether the whistleblower protection induces the desired change in the willingness to
blow the whistle.

**Truthful whistleblowing** Since we use the strategy method for the employee’s deci-
sion on reporting, we do not only observe actual whistleblowing, but are able to track
separately how the willingness to report truthfully as well as fraudulently evolves across
the treatments, independent of the compliance decision of the manager.

Figure 4 displays the fractions of employees choosing truthful whistleblowing in the
respective treatments. In the baseline treatment, 71.7% decide to report cheating con-
ditionally on this misconduct actually happens. This supports our prediction TW1 and
is in line with experimental findings for fairness preferences.

To evaluate the effect of the protection schemes on the willingness to report, we compare the outcome of the AN and the IN treatment relative to the baseline treatment. For the AN treatment, when employees can report anonymously, the fraction rises to 83.6% although this increase is not statistically significant ($p < 0.200$). On the other hand in the IN treatment, we find a significant rise of 16 percentage points to 86.7% ($p < 0.003$). The highest fraction of whistleblowing is found in the ANIN treatment with 89.5%. Introducing incentives in addition to anonymity leads to a significant increase of 5.9 percentage points compared to the AN treatment ($p < 0.003$).

These results provide evidence that both protection schemes affect the employee’s trade-off in the desired direction to drive up truthful whistleblowing, thereby lending support to the predictions TW2a and TW2b.

**Result 1** *Truthful reporting occurs in the absence of protection.* *(Support for TW1)*

**Result 2** *Truthful reporting increase with both incentives and anonymity.* *(Support for TW2a, TW2b)*

![Bar chart showing truthful claims across treatments](image)

**Fraudulent whistleblowing** Analogously to the paragraph above, Figure 5 displays the willingness to conduct a fraudulent report. Surprisingly, we find overall 12.9% of the employees would blow the whistle although there was no misbehavior already in the baseline treatment. This result does not support the prediction, since in the absence of incentives fraudulent whistleblowing was not expected to occur. A possible explanation could be negative reciprocal behavior induced by undesired decisions by managers in previous rounds. Although the employees cannot target the managers whom behavior
they disliked, some employees might still want to punish managers in general.

In line with prediction FW2a, we find that the share of fraudulent reporting does not increase significantly in the AN treatment (21.7%, \( p < 0.129 \)). This suggests that anonymity alone does not add incentives to conduct fraudulent behavior.

However, as expected, introducing employment protection - and thereby incentivizing whistleblowing - leads to a significant jump in fraudulent reports to 30.8% \( (p < 0.001) \). When anonymous reporting becomes an additional feature in the ANIN treatment, the share of employees willing to file a fraudulent claims peaks with 54.0% \( (p < 0.001) \). Both these findings support the predictions FW2a and FW2b and suggest that subjects react also to the adverse incentives of whistleblower protection.

**Result 3** Employees are willing to report fraudulently already in the absence of incentives. (Rejects FW1)

**Result 4** Employees’ willingness to report fraudulently increase with the introduction of incentives and with anonymity only when incentives are in place. (Support for FW2a, FW2b)

![Figure 5: Fraudulent Claims Across Treatments](image)

Considering the results for truthful and fraudulent whistleblowing, the experiment already provides evidence for costs as well as for benefits of whistleblower laws. Protection does favor desired behavior and increase righteous reports, but produces adverse effects for fraudulent claims at the same time.

**Embezzlement** Previous results indicate that under whistleblower protection cheating would be reported more often. Further it is of interest whether this changed behavior induced by the legal environment is anticipated by the managers and drives down illegal
behavior in advance. Therefore, the focus turns to the compliance decisions of the managers, depicted in Figure 6.

In the baseline treatment, a fraction of 41.3% opting for embezzlement, although there are no incentives for the employee to report non-compliance. This supports the prediction EM1 that managers anticipate altruism of the employees or, additionally exhibit their own fairness preferences.

Comparing this to incentivized reporting in the IN treatment, we find a significant drop of 17.1 percentage points to 24.2% which corresponds to a decrease of 41% in illegal behavior ($p < 0.003$). When instead the treatment AN is contrasted, in which anonymity is granted to the employee, also a lower share of 31.6% decides to embezzle money from the third party. However, this decline is not statistically significant ($p < 0.578$). In the ANIN treatment, when both protection schemes are in place, only 7.9% of the managers decide to behave illegally. This means significant declines in contrast to the AN treatment ($p < 0.001$) as well as to the IN treatment ($p < 0.03$).

**Result 5** Managers choose to cheat in the absence of whistleblower protection, but not to the maximal extend. (Support for EM1)

**Result 6** The frequency of cheating decreases with incentives and anonymity. (Support for EM2a, EM2b)

![Bar Chart](image)

Figure 6: Embezzlement Across Treatments

These results suggest that the managers anticipate the reporting behavior of the employees correctly and adjust their cheating frequency downwards with increased probability for whistleblowing. Therefore, the results are in line with the predictions CH2a and CH2b and provide evidence for a beneficial deterrence effect of both whistleblower protection schemes on managerial misbehavior.
Cooperation Having analyzed the whistleblowing and compliance behavior, we now evaluate the willingness to cooperate and the level of cooperation. Figure 7a depicts the share of managers that chose to send a positive amount to their employees in the respective treatment. With respect to cooperation, the share of managers making a positive decision is the most relevant variable as treatment differences can be interpreted independent of the parameters of this experiment. Since the predictions about cooperative behavior depend on the observed and expected whistleblowing frequency, Figure 7b reports the combined truthful and fraudulent whistleblowing cases across the treatments.

We find a fraction of 30.4% of the managers choosing to cooperate in the baseline treatment (whistleblowing cases: 39.6%), which supports our prediction CO1.

The willingness to cooperate slightly increases in the AN treatment (33.5%, \( p < 0.851 \)), while the number of whistleblowing cases goes down to 38.2% (\( p < 0.892 \)). Although we cannot provide statistical significance, the change in cooperation shows the predicted upward adjustment to the frequency of investigations.

In the IN treatment, a similar picture emerges. We find an increase for the overall number of investigations compared to the baseline treatment (45.4%, \( p < 0.345 \)), and especially for investigations from fraudulent reports (24.6%, \( p < 0.001 \)). Correspondingly, the share of the managers choose to cooperate decreases to 25.8% (\( p < 0.258 \)) in response to higher number of whistleblowing cases. These results are in line with the hypotheses CO2 and CO3, which predict an inverse relation between investigations and cooperation. Nevertheless, they cannot already confirm the predictions, since the treatment effects between the baseline treatment and the AN and IN treatments for whistleblowing and cooperation are not found to be statistically significant.

Considering the ANIN treatment on the other hand, only 17.1% of the managers decide to cooperate, which is a significant drop compared to the AN treatment (\( p < 0.019 \)). This coincides with the highest number of whistleblowing cases (57.9%)\(^{15}\) and even more strikingly, with the highest number of investigations caused by fraudulent claims (50.0%)\(^{16}\).

Result 7 Managers choose to cooperate in the absence of whistleblower protection. (Support for CO1)

Result 8 The willingness to cooperate declines with an increased expected number of

\(^{15}\)Number of whistleblowing cases in the ANIN treatment is significantly higher than in the AN treatment (\( p < 0.003 \)) and in the IN treatment (\( p < 0.030 \))

\(^{16}\)Number of fraudulent whistleblowing cases in the ANIN treatment is significantly higher than in the AN treatment (\( p < 0.001 \)) and in the IN treatment (\( p < 0.002 \))
investigations. (Support for CO2, CO3)

Since the ANIN treatment produces significant increases in both overall reporting cases and fraudulent claims, we cannot assign the decrease in cooperation based on these results to either of these factors in particular. However, the data suggests that fraudulent whistleblowing plays the dominant role, since it accounts for over 85% of the whistleblowing cases, which is 52 percentage points larger than in the AN treatment. In contrast, the total whistleblowing cases increase by only 20 percentage points. By disentangling cooperation in the ANIN treatment with respect to the compliance decision, we find only 15.7% of the managers investing if they did not embezzle, while those few who opted for misbehavior choose to cooperate in every third case.

Additionally, contrasting the two anonymity treatments provides an intuition that expected fraudulent claims cause the trust to decrease. Note that the only difference between the two treatments is given by the incentivized whistleblowing in the ANIN treatment. This means the manager cannot observe any change in actual behavior, but only form an expectation about the employee’s choice when she decides about cooperation. Thus, the employee can already report truthfully in the AN treatment without being retaliated, while there are no incentives for fraudulent claims. The managers anticipate an even higher willingness to report truthfully in ANIN treatment by cutting down
illegal behavior substantially. This means, truthful reports cannot arise more often, since embezzlement is almost completely deterred. Only opportunistic whistleblowing could increase and cause a damage for the manager. From this we conclude that the treatment effect is strong evidence for arising distrust from whistleblower protection - especially from the increased likelihood of fraudulent claims.

In addition, it is not only of interest whether cooperation takes place, but also to which degree, we turn our focus from the overall willingness to cooperate to the level of cooperation. Since the design allows to vary the level of investment, the arising distrust may lead some managers to adjust the amount that is trusted to the employee instead of shutting down cooperation in general. To account for this, we consider only those managers who chose to cooperate and show the actual investment relative to the maximal amount possible (Figure 8).

The results provide a clear picture that there are no treatment differences present for the size of cooperation. Independent of the protection scheme, the trusted share lies within a range of 40 to 44 % of the endowment, which roughly corresponds to the average investment level across experimental studies (Fehr and Schmidt, 2006).

**Result 9** The trust level of those managers who cooperate remains constant independent of whistleblower protection.

![Figure 8: Cooperative Level Across Treatments](image)

The results on cooperative behavior provide evidence that the willingness for cooperation depends inversely on the frequency of observed and expected whistleblowing and in particular of fraudulent claims. This supports the hypothesis that whistleblower protection can have a detrimental effect on welfare beyond the idiosyncratic costs of fraudulent claims. While we find treatment differences for the share of managers that are willing
to cooperate, neither incentives nor anonymous reporting seems to affect the amount subjects are willing to invest.

6 Discussion

With this paper, we shed light on the potential hidden costs of whistleblower protection. In a workplace setup, a manager could opportunistically embezzle money at the expense of a third party while being observed and potentially reported by her employee before they enter a cooperative game. We varied the framework in two dimensions to capture two prominent features of whistleblower protection laws: First, changing the timing of the information about the reporting decision for the manager allows to provide anonymous reporting for the employee. Second, restricting the manager’s options at choice in the cooperation game conditionally on a report, enables the employee to insure herself against retaliation from the manager.

Our results confirm that both instruments of whistleblower protection raise the desired effects with an increased willingness to report truthfully illegal behavior by the employees which is anticipated by the managers inducing a reduction in illegitimate practices. This suggests that whistleblower laws offer a rich potential for fighting the damage of corporate fraud through both increased deterrence and detection. Furthermore, the findings demonstrate that also adverse effects of whistleblower protection arise. Since employment protection provides an incentive for reporting in general, it does not only increase truthful reporting, but also triggers fraudulent whistleblowing by the employees.

A novel finding of this paper relates to further costs associated with opportunistic reporting. We point out the importance of observed and expected whistleblowing as unkind behavior for the cooperative climate in an organization - especially fraudulent whistleblowing. From this, we conclude that whistleblower protection can have a detrimental effect on welfare beyond the idiosyncratic costs of fraudulent claims.

We chose a simple design for the whistleblowing environment, where the employee has precise knowledge about the state of illegal behavior of her superior. Also the employee does not face the risk of leaks under anonymity and investigation as well as incentives are guaranteed consequences of a report. This captures the intended increase in legal certainty for the whistleblower. In reality, when for some laws not all of these assumptions are met, uncertainty may also influence the behavior under the different protection regimes and cause a lower responsiveness of employees (see e.g., Chassang and Padró i Miquel 2016, Mechtenberg et al. 2017). However, our approach has the advantage that the results are not driven by ambiguity or risk aversion and serve as a
benchmark for future studies that relax these assumptions.

The results of our study provide some policy implications. First of all, when a whistleblower law is designed, the benefits of reduced fraud may not only be evaluated against inspection and reputational costs arising from fraudulent claims, but hidden costs from forgone cooperation have to be taken into account as well. Further, a legislator could pass different tailor-made laws for different sectors, since the importance of cooperation may well vary between the industries of an economy. For example, laws that apply for organizations where efficiency is rather driven by compliance than by cooperation, could be designed similar to the one presented in this paper. In contrast, if a company’s success heavily depends on productive cooperation, the policy should acknowledge this by avoiding an excessive amount of fraudulent claims at the cost of non-maximal deterrence. This could be achieved by providing only anonymity and no further incentives, or alternatively, conditioning further incentives on a successful investigation (compare to Mechтенбerg et al. 2017).

Taken together, our results suggest that a carefully designed whistleblowing policy allows to manage the trade-off between reduced damage from fraud and costs arising from opportunistic reporting and thereby to enhance overall welfare.

References


OECD (2016): “Committing to Effective Whistleblower Protection.”


A Translated Instructions

Welcome to today’s experiment! If you read the following instructions carefully, you can earn a significant payment - depending on your decisions.

Please note, that from now on and during the whole experiment no communication is allowed. If you have any questions, please direct these at one of the experimenters. Neglecting these rules result in exclusion from this experiment and all payments.

All your decisions during this experiment will remain anonymous and cannot be related to you by either the experimenters nor the fellow subjects. Your earnings will be accounted in points. The points you acquire during this experiment will be exchanged for Euro at the end. The exchange rate is: 10 points = 50 eurocent.

**General procedure:**

There are **three roles** in this experiment: Manager, employee and victim. These roles are assigned randomly. If you are drawn into the role manager, you’ll maintain this role throughout the entire experiment. If you start with one of the other two roles, your role will be drawn randomly before each period. In each period you are part of a group consisting of exactly one manager, one employee and one victim. Also the group composition will result from a random draw in every period.

The experiment is divided into two parts consisting of multiple periods. Beneath you find the procedure of a period in part 1. For the second part, you’ll receive instructions on your screen immediately before it starts.

**Procedure of period in part 1:**

Every subject is endowed with 100 points. After the roles are assigned, the manager chooses between two alternatives (CIRCLE or TRIANGLE). CIRCLE as no payoff consequences for any member of the group. TRIANGLE represents violating the law, resulting in a gain (50 points) for the manager, and a loss (90 points) for the victim. Again, there are no consequences for the employee.
After the manager has made her choice about CIRCLE and TRIANGLE, the employee has to decide whether she wants to file a complaint. This decision is taken separately for both alternatives (complaint if CIRCLE was chosen; complaint if TRIANGLE was chosen). Filing a complaint causes costs for the manager in any case (10 points). If CIRCLE has been chosen and complaint has been filed, the manager has to pay an additional fine (60 points). The victim receives partial compensation for her damage (80 points).

The table below displays all possible combinations of the decisions made by the manager and the employee as well as its respective payoffs for all group members.

<table>
<thead>
<tr>
<th>Arbeitgeber wählt Alternative</th>
<th>Arbeitnehmer entscheidet über Anzeige</th>
<th>Auszahlungen</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Arbeitgeber</td>
</tr>
<tr>
<td>Kreis</td>
<td>Nein</td>
<td>0</td>
</tr>
<tr>
<td>Kreis</td>
<td>Ja</td>
<td>-10</td>
</tr>
<tr>
<td>Dreieck</td>
<td>Nein</td>
<td>50</td>
</tr>
<tr>
<td>Dreieck</td>
<td>Ja</td>
<td>-20</td>
</tr>
</tbody>
</table>

Subsequently, all group members are informed about the chosen alternative[ and whether there has been a complaint].

To conclude a period the manager and the employee play an investment game. First, the manager chooses an amount x between -30 and 60 points. Negative figures mean that points are taken from the employee. Positive mean that points are sent to the employee. If the manager deduces points from the employee these points are transferred and the investment game ends. If the manager sends a positive amount to the employee, it will be multiplied by three. In this case, the employee chooses an amount y between 0 and 3 · x which she would like to return to the manager. There are no consequences for the victim in the investment game.

Payoffs in the investment game:
Manager = - x + y points,
Employee = max(x, 3 · x) – y points,
Victim = 0.

At the end of a period all of the group members are informed whether there was a complaint and your surplus adds up from your endowment (100 points), your revenue from the decisions made (see table) and your revenue from the investment game.

Summary of a period in part 1
1. Manager chooses alternative CIRLCE or TRIANGLE (violation of law)
2. Employee decides upon reporting
3. Every member of a group learns about the chosen alternative [and the reporting decision]
4. Manager and employee engage in an investment game
5. Every member of a group learns about the reporting decision

5./6. The surplus is computed

After you have completed the second part and a questionnaire, one period is drawn for payout. You’ll receive the points you earned in that period converted according to the exchange rate plus 5 Euro as show up fee.

Thank you for participating and good luck!