




Internal whistleblowing systems without proper sanctions may backfire

Sebastian Krügel^{1,2} · Matthias Uhl² 

Accepted: 17 February 2023
© The Author(s) 2023

Abstract

Internal whistleblowing systems are supposed to fight misconduct within organizations. Because it is difficult to study their efficacy in the field, scientific evidence on their performance is rare. This is problematic, because these systems bind substantial resources and might generate the erroneous impression of compliance in a company in which misconduct is prevalent. We therefore suggest a versatily extendable experimental workhorse that allows the systematic study of internal whistleblowing systems in the lab. As a first step, we tested the efficacy of whistleblowing systems if internal punishment for misconduct is mild and hesitant which is usually the case in practice, as several fraud surveys confirm. Our results show that under these conditions almost nobody blew the whistle, and misconduct occurred even more frequently with than without a whistleblowing system. The institutionalization of whistleblowing seemed to crowd out the intrinsic motivation to act compliantly. Moreover, when a whistleblowing system was either unavailable or not used, misconduct was highly contagious and spread quickly. Yet, when we implemented severe and ensured punishment for misconduct, whistleblowing systems could deter wrongdoing. In such a setting, people were willing to blow the whistle and the prevalence of misconduct dropped substantially. Altogether, our results highlight the interaction between institutions and preferences and can support the design of compliance measures within organizations. For compliance managers a key takeaway is that if companies preach a zero-tolerance policy, they should practice it as well. Otherwise, they might even worsen the situation.

Keywords Misconduct · Whistleblowing · Punishment · Crowding out

✉ Matthias Uhl
matthias.uhl@thi.de

¹ School of Governance, Technical University of Munich, Munich, Germany

² Faculty of Informatics, Technische Hochschule Ingolstadt, Esplanade 10,
D- 85049 Ingolstadt, Germany

1 Introduction

Company executives use internal whistleblowing systems to exploit the information advantage of their employees concerning observed misconduct in their workplace environments. Since the Sarbanes-Oxley Act of 2002, many companies have even been obliged to implement such a system and to ensure whistleblowers' anonymity. Following the capital market crisis of 2008, regulations were tightened by the Dodd-Frank Act. Company executives prefer internal instead of external whistleblowing because the latter usually is much more detrimental to the organization (see, e.g., Near & Miceli, 1995, 1996, 2016; Kaptein, 2011; Lee & Xiao, 2018). Internal whistleblowing systems are intended to detect wrongdoing within an organization early and to keep cases of misconduct firmly under control before they are possibly reported to an external party. Most importantly, internal whistleblowing systems are expected to have a deterrence effect on potential wrongdoers because misconduct is less likely to go undetected (see, e.g., Miceli, Near, & Dworkin, 2009; Kaptein, 2011; Johannesen & Stolper, 2017; Wilde, 2017; Amir, Lazar, & Levi, 2018). In the present study, we are mainly concerned with this deterrence effect of internal whistleblowing systems.

Before implementing a whistleblowing system, a company should have a clear idea about how to respond to internal reports and how to deal with identified wrongdoers. Penalties for misconduct not only help discipline the wrongdoer (Miceli et al. 2009), they may also be important to the potential whistleblower (Kaptein 2011). Clear internal penalty scales are therefore particularly important when companies are reluctant to refer cases of misconduct to external parties such as law enforcement agencies due to high costs, bad publicity, possible brand damage or even consumer boycotts. The allure to dealing with wrongdoers exclusively internally is apparent if the misconduct concerns only self-imposed company guidelines to comply with certain good practice rules regarding, for instance, environmentally friendly production, animal-friendly farming or fair trade. But even in cases in which breaches of law are involved and disclosure to the authorities is legally mandatory, companies have an incentive to keep cases of misconduct initially under control to try to downplay wrongful acts. On a less grim perspective, companies may also benefit from an internal managing of the situation by having the chance to first verify potential allegations, avoid future misconduct that may result in even higher future costs or because it gives them the opportunity to develop a communication strategy.

However, if company executives first and foremost try to avoid public disclosure of the misconduct, then dismissing of employees or cutting contractual payments might not be the most favored choice because doing so could lead to subsequent lawsuits, attracting public attention. Instead, executives might be inclined to use milder forms of punishment, such as sending warning letters or cutting optional bonus payments, if they are willing to punish misconduct at all. It may therefore be of little surprise that many employees do not even report observed wrongdoing because they believe that nothing can or will be done by the organization (see, e.g., Near, Rehg, Van Scotter, & Miceli, 2004; Miceli et al., 2009).

How widespread employees' perception of mild and hesitant punishment of misconduct is, becomes evident in several fraud surveys conducted by Ernst & Young

among employees of large companies. For instance, in Europe, the Middle East, India and Africa, 51–52% of the respondents were not aware of clear internal penalties for breaking their employer's anticorruption policies (Ernst&Young, 2013b, 2015). Similarly, worldwide, only 35–45% of the respondents believed that people in their companies have been penalized in the past for non-compliance with standards of good conduct (Ernst&Young, 2014, 2013a). In the Asia-Pacific region, 49% of the respondents even thought that the senior management in their companies would ignore unethical behavior to achieve revenue targets (Ernst&Young, 2017). While these figures do not necessarily mean that misconduct in companies is indeed punished only hesitantly and mildly, they do show that employees often believe this to be the case. A large fraction of employees is either unaware of any clear internal penalties for misconduct, does not believe that people have been punished for misconduct or even expects their executives to ignore unethical behavior, as long as this behavior is supposedly beneficial for the company.

Moreover, even if companies claim to operate whistleblowing systems, this does not mean that these systems actually work very well. Frequently, internal compliance systems appear to be more like window dressing than effective tools of fraud prevention or detection (Krawiec 2003). In a field study among almost 250 firms, for example, Soltes (2020) found that in 20% of the cases, at least one obstacle occurred during the mere attempt to report alleged misconduct. Obstacles included, for instance, web redirects to incorrect pages, email bounce backs, or disconnected phone lines. Soltes (2020) even reports one case where the called person did not even know that this was the whistleblowing hotline.

To the extent that prevailing practices are perceived such that misconduct is investigated only hesitantly and punished at most only mildly within organizations, the likely impact of a whistleblowing system is an open question. Conventional wisdom about whistleblowing is often flawed and research findings are frequently counterintuitive (Miceli et al. 2009). It is particularly questionable to what extent an internal whistleblowing system helps to deter misconduct when the wrongdoer believes that he or she has little to fear. Because it is difficult to study the efficacy and deterrence effect of whistleblowing systems in the field, we ran a controlled lab experiment to address these questions. Of course, misconduct and whistleblowing are complex problems, and many factors can play a role. Past studies examined a number of personal, situational, and organizational factors in this regard, but mostly in terms of the reporting of misconduct and the whistleblowers themselves (see, e.g., Dozier & Miceli, 1985; Gundlach, Douglas, & Martinko, 2003; Reuben & Stephenson, 2013; Bartuli, Djawadi, & Fahr, 2016; Carpenter, Robbett, & Akbar, 2018; Liu et al., 2018; Choo et al., 2019; Butler, Serra, & Spagnolo, 2020).

In the present study, instead, we are interested in the impact of the institutional existence of a whistleblowing system on the occurrence of misconduct within an organization. Our starting point here is the presumption that reported misconduct is often either not investigated or leads to only mild internal consequences for the wrongdoer. In this respect, we wish to emphasize that our experiment addresses internal, not external whistleblowing. While external whistleblowing may often be based on moral outrage (Edward Snowden is arguably a case in point), internal whistleblowing usually is more about the company suffering damage if the misconduct leaks

out. Internal whistleblowers often are the long-term and loyal employees—those who care most about the company (Near and Miceli 2016) and have a greater identification with the organization (Liu et al. 2018). Internal whistleblowing can be viewed as a form of prosocial behavior that may benefit the organization as well as the whistleblower (Dozier and Miceli 1985). Whistleblowers report misconduct internally because they fear the risk to the company and to their own jobs rather than because of general moral concerns of the stigmatized act. This is also why managers and executives prefer employees to report through internal rather than external channels. They want to treat misconduct internally because of its risk and potential damage to the company, not so much because of its moral aspects.

As essential characteristics in the context of internal whistleblowing, we therefore considered that first (undetected) misconduct is beneficial for the wrongdoer to whom, for instance, increased bonus pay, or career improvements may accrue. Second, misconduct has a small but positive risk of discovery by an external party, such as the authorities or the media. And third, if the misconduct is publicly revealed, the company's interests will be severely damaged through, for instance, substantial monetary penalties, exemption from certain markets, loss of image or consumer boycotts.

A whistleblowing system in such a situation can be viewed as an institutionalized punishment mechanism that can be initiated by other employees of the company who observe the misconduct. The potential whistleblowers can report the wrongdoer, but they cannot inflict the penalty. The responsibility for the latter rests on the organization or its executives. The crucial questions are how often employees make use of such a system and how much it deters wrongdoing if the attached penalty for misconduct is either rather mild or possibly not even inflicted on the wrongdoer for the reasons described above. Especially when considering that many whistleblowers risk a lot themselves, such as retaliation or dismissal (see, e.g., Kaptein, 2011; Near & Miceli, 2016; Lee & Xiao, 2018). The European Barometer on corruption, for instance, found that 81% of Europeans who witnessed corruption did not report it, and a third of those said the main reason was fear of retaliation (European Commission, 2017). In our experiment, whistleblowing was completely anonymous, and the whistleblower neither had to fear retaliation nor dismissal. The whistleblower in our experiment had to bear a financial cost for reporting someone else because whistleblowing in companies is typically viewed as a challenging decision-making process (Nicholls et al. 2021) that comes with various psychological costs (see, e.g., Schultz et al., 1993; Kaplan & Whitecotton, 2001).

We find that with mild and hesitant punishment of the reported, whistleblowing was almost absent in our experiment. The institutionalization of such a “toothless” whistleblowing system even induced a level of misconduct that was significantly higher than if no whistleblowing system was present. When the whistleblowing system came with harsh and certain punishment, however, a substantial level of whistleblowing could be observed and the system unfolded its desired deterrence effect on wrongdoers.

2 Experimental design and procedure

Subsequently, we first describe the design of our experiment. Some important design features are discussed in a separate subsection thereafter. In the last subsection, we provide additional procedural details about the experiment.

2.1 Design

The experiment was set up to compare the participants' behavior across three treatments in a between-subjects design: one treatment with a whistleblowing opportunity (the *Whistle* treatment) and two treatments without such an opportunity (the *No-Whistle* treatments). Because *Whistle* was the main treatment of interest, we describe the design of this treatment first and detail the differences in the other two treatments thereafter.

As already mentioned, we reduced the problem to a situation in which (undetected) misconduct is individually profitable but imposes a risk of severe damage on the entire organization. A company in our experiment was represented by a group of participants. Group members did not work together on a certain project. Rather, they just happened to be colleagues in the same company, each of them receiving his or her own salary. Because employees typically stay in the same company over a longer period of time, the participants in our experiment played a repeated game in a partner's design. We chose a group size of three and a total number of 10 periods. Thus, prior to the first period, participants were matched at random in groups of three, and the group composition remained the same over all 10 periods. In addition, each participant received a unique membership number in his or her own group ("1", "2" or "3") and kept this number until the end of the game. Participants could earn a reputation because the other group members could match observed actions with the membership numbers across periods—mimicking reputation building in a company.

In each period, each participant could choose between two alternatives, labeled "alternative A" and "alternative B." A participant choosing alternative A received period earnings of 100 experimental currency units (ECU). A participant choosing alternative B received period earnings of either 100 or 150 ECU (more on this below). At the end of the game, a participant's earnings over all 10 periods were summed up and converted to euro at a rate of 1 ECU=0.01 euro. However, alternative B, which was individually more profitable than alternative A, involved a risk that the entire group incurred a loss. Each time a group member chose alternative B, the computer played out a lottery that erased all total earnings of all three group members, with a probability of 0.003. This feature and its parameterization were borrowed from Abbink, Irlenbusch and Renner's (2002) bribery game, with the difference being that in their design, players engaging in misconduct put only their own earnings at risk. Moreover, the lottery in our experiment was played out only after all 10 periods were over.¹ For each group, the computer counted the number of alternative B choices over all 10 periods and performed the lottery the respective number of times. As soon as

¹ In Abbink et al. (2002), the lottery was always played out immediately after corruption took place. Participants who were hit by the lottery lost all of their previous earnings and were excluded from further play.

one of these draws indicated a lethal result, all of the entire group's earnings were lost. Losing all earnings from play is presumably the harshest form of punishment feasible in the lab. In our experiment, it presents the discovery of misconduct by an external party as a total collective loss.

After all of the participants chose between alternatives A and B in a given period, they were informed about the other group members' choices through the unique membership numbers. In *Whistle*, each participant now had the option to report each group member who had chosen alternative B in the current period by simply ticking a box: "report member X." Participants were informed that only group members choosing alternative B but not group members choosing alternative A could be reported. This asymmetry in reporting possibilities implied a stigmatization of the risky alternative B and thus declared the punishable behavior. The fact that a group member who chose alternative A could not be reported also avoided possible "punishment wars" (see, e.g., Nikiforakis, 2008). Reporting a group member came at a cost of 10 ECU per reported member. This cost was supposed to capture the trouble and discomfort of blowing the whistle (Schultz et al. 1993; Kaplan and Whitecotton 2001) and was subtracted from a participant's period earnings. Other than that, a whistleblower had nothing to fear for reporting a group member.

A participant who chose alternative B but was not reported by any group member earned 150 ECU in the respective period. This resembles a situation in which company executives have no information about the misconduct and the respective employee gets bonus pay (=50 ECU) for improving a company's revenue or the like when the means used for that end are not disclosed to the employer. On the other hand, a participant who was reported by at least one group member for choosing alternative B earned either 100 or 150 ECU with equal probability (regardless of whether he or she was reported by one or two group members).² Thus, reporting a group member initialized a mechanism that inflicted a punishment on the wrongdoer in about half of the cases. The punishment for choosing alternative B was rather mild, however, as it referred only to the suspension of the bonus pay. Altogether, this is supposed to resemble the situation portrayed by Ernst & Young's Fraud Surveys, where—at least, in the eyes of the employees—misconduct was not reliably punished and sometimes probably even ignored within a company.

After all of the participants decided on whether or not to report group members who chose alternative B, the interaction was regarded as completed for the current period. Participants received full information about their own current period earnings, including possible costs for filing reports concerning their group members. A participant who chose alternative B was additionally informed about whether he or she had been reported by at least one other group member. In this case, a participant learned that he or she was reported but not by whom or by how many group members. Blowing the whistle in this sense was an anonymous act.

² If whistleblowing is truly anonymous, then compliance managers or other company executives cannot distinguish whether several reports came from the same person or from different people. Thus, an internal investigation should be initialized as soon as one person blows the whistle, and the probability of punishment should not increase with the number of reports.

Both *No-Whistle* treatments followed the same basic procedure as *Whistle*, including in particular the risk of a collective loss that went along with alternative B. After all of the participants chose between alternatives A and B in a given period, they were informed about their group members' choices. Yet, they had no option to blow the whistle on participants who chose alternative B. As in *Whistle*, full information about their own current period earnings was disclosed to all at the end of each period. Because participants of the *No-Whistle* treatments could not report a group member for choosing alternative B, we adapted the earnings for this alternative. In one treatment, participants earned 150 ECU in each period during which they chose alternative B (*No-Whistle (Never Detected)*). In the other treatment, participants received period earnings of either 100 or 150 ECU, with equal probability, for choosing alternative B (*No-Whistle (Always Detected)*). Together, both *No-Whistle* treatments constituted the lower and upper bounds regarding the profitability of alternative B in *Whistle*. Both *No-Whistle* treatments mainly served as controls in our effort to assess the efficacy and deterrence effect of a whistleblowing institution. However, as the name suggests, *No-Whistle (Never Detected)* may be viewed as a situation in which a company does not have a whistleblowing system and misconduct therefore always goes unnoticed by company executives. Likewise, *No-Whistle (Always Detected)* may be considered a situation in which company executives always observe the appearance of misconduct themselves, but the internal punishment for wrongdoing is again mild and only hesitantly inflicted on the wrongdoer.

2.2 Discussion of the design

Clearly, there are many factors—personal, situational and organizational—that surround misconduct and whistleblowing in companies (for a review and discussion of some of these factors, see, e.g., Miceli et al., 2009; Lee & Xiao, 2018). Previous experimental studies on whistleblowing typically focused on the characteristics of whistleblowers and how whistleblowing can be encouraged. For example, previous studies examined personal traits of potential whistleblowers (Bartuli et al. 2016), the effect of financial incentives on whistleblowing (Schmolke & Utikal, 2018; Stikeleather, 2016; Butler et al., 2020), the effect of profit sharing (Carpenter et al. 2018), and the role of diffusion of responsibility when it comes to whistleblowing (Choo et al. 2019). In the present study, we were mainly interested in the effect of a whistleblowing institution on the occurrence of misconduct.

Which behavior qualifies as misconduct in companies is usually prescribed by the law, society or the market, regardless of whether employees or executives share this view (in fact, wrongdoers obviously do not). In our experiment, the dividing line between good and bad conduct was the risk of severe damage to the company and its employees that the latter entailed, but not the former. In our view, this is the main reason why company executives implement an internal whistleblowing system and why they prefer internal over external reporting. The same applies to most whistleblowers. They, too, prefer internal reporting channels and only rarely go external with their information (Near and Miceli 2016). This is the approach of the present study on corporate misconduct on which alternative B is based.

Therefore, alternative B in our experiment did not only impose a risk on the person choosing this alternative, but also on her or his group members or “colleagues” (where the magnitude of this risk was empirically calibrated; see below). Wrongdoers are willing to take this risk for themselves and, crucially, are also willing to impose it on others. It is this feature of stochastically harming others which makes alternative B morally relevant in our experiment. We intentionally abstained from labeling alternative B “immoral” or “wrong” because we did not want to confound our findings with the effect of the specific semantics that we would be choosing. Recall, however, that choosing alternative B was in fact stigmatized by being the only of the two options that could be reported. In this sense, in the whistleblowing treatment, participants were sensitized for (or educated on) the kind of behavior that was undesired from an institutional perspective. It should be acknowledged, however, that this was a more subtle hint at the undesirability of alternative B than straight out labeling it “misconduct.”

The situation of misconduct and internal whistleblowing that we had in my mind in our experiment is exemplarily expressed in a famous letter by Sherron Watkins, a whistleblower of Enron, who was anonymously expressing some severe accounting discrepancies to the company’s CEO, Kenneth Lay:

“Has Enron become a risky place to work? For those of us who didn’t get rich over the last few years, can we afford to stay? [. . .] I am incredibly nervous that we will implode in a wave of accounting scandals. My eight years of Enron work history will be worth nothing on my resume, the business world will consider the past successes as nothing but an elaborate accounting hoax.” (The New York Times, 2002).

In those lines, all three elements that we consider essential become evident: (1) some people became rich by misconduct, but (2) there is a risk that the company gets busted by it, and (3) in that case, all employees more or less bear damage. Moreover, Sherron Watkins was not concerned with the moral aspect of the reported misconduct. She was only worried about the possibility that the company, including herself, could suffer severe damage if the misconduct were discovered.

In our experiment, the chosen probability of discovery for each choice of alternative B was very low, with $p = 0.003$. A participant who chose alternative B ten times increased the group’s overall probability of discovery by $\Delta p = (1 - 0.003)^n - (1 - 0.003)^{(n+10)}$, where n is the total number of B-choices of both other group members. Thus, a participant who chose alternative B ten times increased the overall probability of discovery by only 0.0279 to 0.0296, depending on the number of B-choices by his or her group members. If all three group members chose alternative B every period, the group’s overall probability of discovery was only $p = 1 - (1 - 0.003)^{30} = 0.086$.

While the probability of discovery of misconduct is likely to be low in real situations outside the lab as well, it was not our goal to mimic this probability in our experiment. The aim of the present study was to determine the efficacy and deterrence effect of a whistleblowing system given that there is misconduct, not the effect of the probability of discovery on the occurrence of misconduct. Thus, we had to

calibrate the probability of discovery such that some participants would be willing to bear the risk of alternative B (i.e., the potential wrongdoer) while some would not be (i.e., the potential whistleblower). This calibration task was an empirical exercise, not a theoretical one.

In Abbink et al. (2002), the possibility of losing all earnings with a probability of $p = 0.003$ significantly and substantially decreased misconduct without completely wiping out wrongdoing. Many participants in their experiment appeared to have a severe problem with the possibility of a total loss, despite its low probability. Nonetheless, we conducted two pilot sessions of the *No-Whistle (Always Detected)* treatment, in which we set the probability of discovery after each choice of alternative B to $p = 0.01$ because we feared that $p = 0.003$ might be too low in our experiment. Participants in those pilot sessions chose alternative B on average only 17% of the time. As we were afraid of potential floor effects with respect to the *Whistle* treatment, we set the probability of discovery to $p = 0.003$, as in Abbink et al. (2002), and conducted all treatments with this probability. An analysis of post-experimental questionnaire data further confirmed the success of our calibration exercise, as the vast majority of participants stated that the probability of discovery was a serious obstacle to choosing Alternative B, despite its low level (see [Appendix B](#)).

Playing out the lottery of discovery at the end of the game instead of right after each period during which misconduct took place was mainly a pragmatic design choice to collect the behavior of all participants in all 10 periods. But even in real corporations, misconduct is hardly ever discovered right after its occurrence, and the passage of time does not undo wrongful acts. A similar argument applies to whistleblowing. While it may reduce the likelihood of future misconduct, it cannot undo previous misconduct. Therefore, we kept the probability of discovery constant, independent of the occurrence of whistleblowing. Analogously, we did not reduce the collective penalty after discovery even if the wrongdoer was internally punished for the discovered misconduct. The vital question that we addressed and modeled in our experiment is about the impact of a whistleblowing system in the context of mild internal punishment of wrongdoing. It is unlikely that such punishment would pacify the authorities or consumers when they pass their verdict on a company. In fact, mild internal punishment may even lead to liability (see, e.g., Miceli et al., 2009).

Notice that our whistleblowing system enabled people to blow the whistle completely anonymously. Although coworkers in companies can also blow the whistle in absence of an institutionalized whistleblowing system, it is only the formal institution which assures whistleblowers anonymity through its technical implementation. While potential whistleblowers in companies often fear retaliation (see, e.g., European Commission, 2017), they did not have to be afraid of such attacks in our experiment. However, reporting a group member came at a cost for the whistleblower to capture the discomfort of blowing the whistle in real situations (see, e.g., Lee & Xiao, 2018, and the references cited therein). Although there is convincing evidence that financial rewards can encourage whistleblowing (see, e.g., Stikeleather, 2016), many companies have not yet implemented such incentive structures (Association of Certified Fraud Examiners, 2016) and there often seems to be some dispute among practitioners as to whether financial rewards for whistleblowing should be used (PwC,

Table 1 Overview of treatments and sessions

Treatment	Sess.	Location	Particip. (in total)	Period earnings of:	
				alt. A	alt. B
<i>No-Whistle</i> (<i>Always Detected</i>)	2	Jena	60	100	$(0.5 \circ 100$
	2	Munich	60		$\oplus 0.5 \circ$ $150)$
<i>No-Whistle</i> (<i>Never Detected</i>)	2	Jena	57	100	$(1.0 \circ 150)$
	2	Munich	60		
<i>Whistle</i>	2	Jena	60	100	if report-
	2	Munich	57		ed: $(0.5 \circ$ $100 \oplus 0.5$ $\circ 150)$ if not reported: $(1.0 \circ 150)$

^aThe table shows from left to right the treatment, number of sessions, locations of the sessions, total number of participants and the individual period earnings of alternative A and B. The period earnings of alternative B are summarized in the form of lotteries where $(p_1 \circ x_1 \oplus p_2 \circ x_2)$ denotes a probability distribution with probabilities p_1 and p_2 and, and corresponding elementary events x_1 and x_2 .

2013). In our experiment, we therefore decided against incentivizing whistleblowing by means of financial rewards.

2.3 Procedural details

Table 1 gives an overview of the treatments and corresponding sessions that were conducted for the experiment. For each treatment, we ran four sessions with a grand total of 354 participants. Ten sessions consisted of 30 participants each, and two sessions had a participant number of 27. Two sessions of each treatment were conducted at the laboratory of the Max Planck Institute of Economics in Jena, Germany, and the other two sessions took place at the laboratory of the Technical University in Munich, Germany. The participants were students of all majors who were invited to the experiment using ORSEE (Greiner 2015), and each session was run on computers with the software z-tree (Fischbacher 2007). As usual, written instructions were handed out to each participant at the beginning of the experiment and were additionally read aloud by an experimenter.³ Control questions prior to the first period ensured comprehension of the instructions. Each session took about 50 min, including payment of the participants. On average, participants earned 11.80 euro from playing plus a show-up fee of 2.50 euro in Jena and of 4.00 euro in Munich.

3 Hypotheses

As mentioned in Sect. 2, the experiment was designed to compare the behavior of our participants between treatments. To this aim, we empirically calibrated the probability of discovery for each choice of alternative B in our experiment such that the proportion of B-choices in both *No-Whistle* treatments was at approximately medium

³ For the instructions used in the experiment, see supplementary material.

levels, where some participants would choose B and others would not. With a probability of 0.003 this was roughly the case. Certainly, it is surprising that such a small probability has such a large impact on behavior, as was also found in Abbink et al. (2002). However, this was not to be investigated in the present study and we do not attempt to rationalize this effect through a certain behavioral model and some utility or probability weighing functions afterwards. In our study, we wanted to examine how the availability of a whistleblowing system affects the prevalence of B-choices, given that a significant proportion of participants has a problem with alternative B. The latter was ensured through pilot sessions and both control treatments set the benchmark for B-choices (see also [Appendix B](#)).

Therefore, we compared the prevalence of B-choices in the *Whistle* treatment with the prevalence of B-choices in both *No-Whistle* treatments. We first put forth a purely economic hypothesis in which no psychological effect of whistleblowing on the reported are considered. In this sense, it is assumed here that participants' preference for choosing B is only constrained by the monetary implications of the whistleblowing institution and their belief about the probability that it will be used. From this purely monetary perspective, both *No-Whistle* treatments span the range in which the prevalence of B-choices in *Whistle* should fall. Because alternative B is individually more profitable in *No-Whistle (Never Detected)* than in *No-Whistle (Always Detected)*, there should be more B-choices in the former than in the latter. In the *Whistle* treatment, the profitability of alternative B depends on the participants' propensity to report group members for choosing this alternative.

Even though nobody should bear the costs of reporting a group member in the last period and, by backward induction, this holds for all periods, we expected the participants to use the whistleblowing system despite its costs throughout all periods. Blowing the whistle in our experiment basically is a punishment opportunity that inflicts a penalty on a participant for choosing B in 50% of the cases. Someone who chooses alternative B risks the entire group's earnings for his or her own personal gain. It is a well-established finding in the larger economic literature on public-goods experiments that people are willing to make use of costly punishment opportunities against selfish group members, even in (quasi-) one-shot interactions (see, e.g., Fehr & Gächter, 2000, 2002; Anderson & Putterman, 2006; Carpenter, 2007) and in cases in which the impact-to-cost ratio of punishment is not very favorable for the punisher (see, e.g., Egas & Riedl, 2008; Nikiforakis & Normann, 2008). In repeated public-goods experiments with unchanged group compositions (i.e., with a partner's design), as in our experiment, punishment typically occurs even more frequently. Notice that punishment in public-goods experiments does not undo or increase any past contributions. Punishment in public-goods experiments represents an opportunity to change future behavior, just like in our experiment.

While the profitability of alternative B in the *Whistle* treatment varied with the actual extent of whistleblowing in our experiment, both *No-Whistle* treatments constituted the lower and upper bounds of B's profitability. If every participant will always be reported for choosing alternative B, then the expected profits for B-choices are the same in the *Whistle* and *No-Whistle (Always Detected)* treatments. If, on the other hand, no participant will ever be reported for choosing alternative B, then a B-choice in the *Whistle* treatment would lead to a sure bonus payment of 50 ECU, just like in

the *No-Whistle (Never Detected)* treatment. Hence, from a purely monetary perspective, the prevalence of B-choices in the *Whistle* treatment should lie in between those of both *No-Whistle* treatments, with the magnitude of our whistleblowing system's deterrence effect depending on its usage. This constitutes our first hypothesis that we refer to as the Economic Hypothesis:

Economic Hypothesis. The prevalence of B-choices will be lower in the No-Whistle (Always Detected) treatment than in No-Whistle (Never Detected) and the Whistle treatment will be in between.

From a behavioral perspective, reporting a B-choice may not only lead to possible monetary consequences for the reported participant, but also cause psychological costs because it signals social disapproval by his or her peers. This, in turn, may lower a participant's utility for choosing alternative B due to, for instance, feelings of shame. In related social dilemma situations as, for instance, the public-goods game, parts of the effectiveness of punishment institutions might be caused by such emotional mechanisms (Bowles and Gintis 2002). Indeed, in a public-goods experiment, Masclet, Noussair, Tucker and Villeval (2003) found that non-monetary punishment opportunities increased cooperation levels, particularly in a repeated game where the group composition remained the same. The power of purely symbolic incentives in a public-goods context has also been demonstrated in the field by Gallus (2016).

If being reported by peers induces feelings of shame in our experiment as well, the participants may choose alternative A in order to avoid social disapproval. If the disutility of shame is high enough, the availability of a whistleblowing system might be even more effective than expected based on pure monetary grounds. This leads to our second hypothesis regarding the prevalence of B-choices that is based on the staggering and robust effect of punishment opportunities on cooperation levels, found in many previous public-goods experiments. We refer to this hypothesis as the Psychological Cost Hypothesis.

Psychological Cost Hypothesis: The prevalence of B-choices will be lower in the No-Whistle (Always Detected) treatment than in No-Whistle (Never Detected) and will be lowest in the Whistle treatment.

However, there is an alternative behavioral perspective that predicts the opposite effect. If the possibility to report B-choices is institutionalized, there exist an external sanctioning mechanism next to an internal sanctioning mechanism of one's guilty conscience for inducing a risk on peers. Gneezy and Rustichini (2000a), for instance, have shown that imposing a fine for misbehavior may shift the respective conduct from the moral domain into the market domain. This may turn convictions into commodities and thus deprive them of their perceived inherent morality (Frey and Oberholzer-Gee 1997; Mellström and Johannesson 2008). Similarly, a whistleblowing institution together with a sanctioning mechanism for the wrongdoers puts an explicit fine on misconduct. If this fine is sufficiently low, as in the case of a mild and hesitant sanctions, people may feel psychologically legitimized to engage in the respective behavior if they are willing to pay this fine. Thus, the presence of a whistleblowing

Table 2 Proportion of B-choices across treatments

	<i>No-Whistle (Always Detected)</i>		<i>No-Whistle (Never Detected)</i>		<i>Whistle</i>	
	Obs.	Mean (SE)	Obs.	Mean (SE)	Obs.	Mean (SE)
All periods	40	0.30 (0.05)	39	0.42 (0.05)	39	0.56 (0.05)
First period	120	0.21 (0.04)	117	0.33 (0.04)	117	0.42 (0.05)

^aThe table shows the means (standard errors) of the relative frequencies of B-choices over all periods as well as in the first period of each treatment. In the former case, the standard errors of the means are based on the group averages. In the latter case, the standard errors are based on individual data because there was no within-group interaction prior to the first period.

institution with weak economic consequences may imply a psychological gain for the wrongdoer. These behavioral considerations lead to our third and final hypothesis on the prevalence of B-choices that, being a reverse version of the Psychological Cost Hypothesis, we refer to as the Psychological Gain Hypothesis.

Psychological Gain Hypothesis: The prevalence of B-choices will be higher in the No-Whistle (Never Detected) treatment than in the No-Whistle (Always Detected) treatment and will be highest in the Whistle treatment.

4 Results

4.1 Crowding out of compliance

4.1.1 Behavioral effects of a whistleblowing system

In the following, we will first investigate the frequency of B-choices as an indicator of wrongdoing and second the frequency of submitted reports as an indicator of whistleblowing which constitute our two dependent variables. Within each treatment, participants' behavior regarding alternative B and whistleblowing did not differ significantly between the sessions conducted in Jena and those conducted in Munich (for more details, see Table A1 in Appendix A). Therefore, we pooled the data from both locations for each treatment.

Table 2 gives a first overview of the general tendency toward B-choices in each treatment. The first two lines of Table 2 show the means and standard errors of the relative frequencies of B-choices over all 10 periods at the group level. The last two lines contain the relative frequencies and standard errors of B-choices in the first period of each treatment using individual data because there was no within-group interaction prior to the first period. As expected, the prevalence of B-choices was lower in the *No-Whistle (Always Detected)* treatment than in *No-Whistle (Never Detected)*. While approximately 30% of all choices were B-choices when the bonus payment for this alternative was uncertain, there were about 42% B-choices when

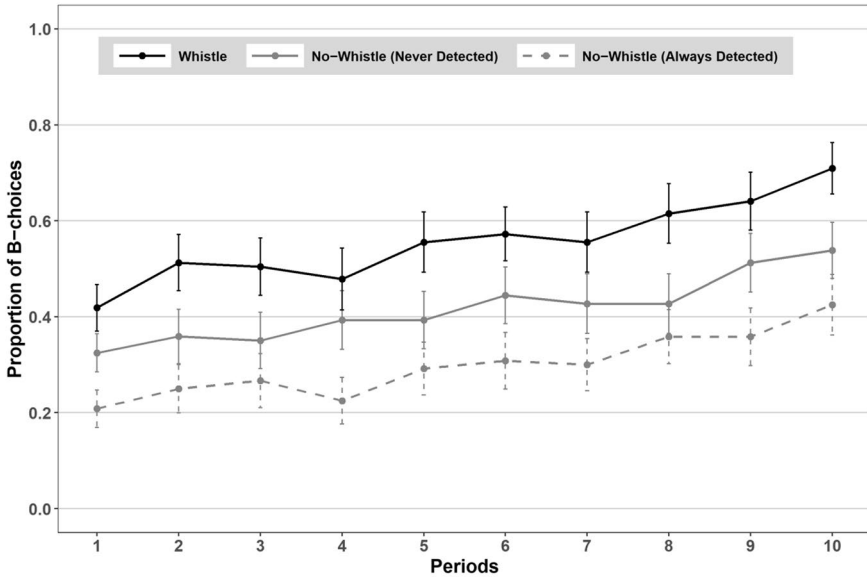


Fig. 1 Proportion of B-choices over periods across treatments

^aThe figure plots the means and standard errors of relative frequencies of B-choices over all 10 periods. As before, the standard errors are based on group-level averages.

the bonus payment for alternative B was certain. In line with the Psychological Gain Hypothesis, however, the prevalence of B-choices was highest in the *Whistle* treatment, with about 56% B-choices overall. Interestingly, this difference was already present in the first period.

In Fig. 1, the relative frequencies of B-choices in each treatment are plotted over all 10 periods. The abovementioned order of treatments regarding the prevalence of B-choices was stable throughout all of the periods. The prevalence of B-choices was always lowest in *No-Whistle (Always Detected)*, followed by *No-Whistle (Never Detected)*. In all of the periods, the prevalence of B-choices was highest in *Whistle*. Common to all treatments was that the prevalence of B-choices rose by period. In *Whistle*, the relative frequency of B-choices increased from approximately 42% in the first period to 71% in the last period. In *No-Whistle (Never Detected)*, the relative frequency rose from 31% in the first period to 54% in the last period, and in *No-Whistle (Always Detected)*, it increased from 21% to about 42%.

The descriptive results regarding B-choices across treatments and periods were largely confirmed in a regression analysis. Table 3 presents the coefficients and standard errors of logit regressions in which choosing alternative B was regressed on treatment dummies and on a variable for the period. Compared to the *No-Whistle (Never Detected)* treatment, which was the base category in the regressions, the overall prevalence of B-choices was lower in *No-Whistle (Always Detected)* and higher in *Whistle* ($p = 0.083$ and $p = 0.054$, respectively, in both Eqs. (1) and (2)). These differences were already present in the first period, but only significant between both *No-Whistle* treatments and not significant at conventional levels when comparing *No-*

Table 3 B-choices across treatments and periods (logit regression)

	Dependent variable: <i>B-choice</i> (= 1 if true)		
	(1) All periods	(2) All periods	(3) 1st period
<i>No-Whistle (Always Detected)</i>	-0.52 (0.30)	-0.53 (0.30)	-0.60 (0.30)
<i>Whistle</i>	0.56 (0.29)	0.57 (0.30)	0.40 (0.27)
<i>(Period - 1)</i>		0.10 (0.02)	
<i>Constant</i>	-0.34 (0.21)	-0.79 (0.22)	-0.73 (0.20)
Obs.	3540	3540	354
Groups	118	118	

^aThe table shows the coefficients of the logit regressions. *No-Whistle (Always Detected)* and *Whistle* are dummy variables indicating the respective treatment. *No-Whistle (Never Detected)* is the base category. *(Period - 1)* indicates the respective period minus 1 and runs from 0 to 9. None of the interactions between each treatment variable and the variable *(Period - 1)* were significant, so they were therefore left out of the regressions. Standard errors are given in parentheses and are clustered at the group level, except for the regression based on first period choices.

Whistle (Never Detected) with *Whistle* ($p = 0.044$ and $p = 0.138$ in Eq. (3), respectively). Moreover, the prevalence of B-choices significantly increased by period ($p < 0.01$ in Eq. (2)). This time trend was the same in all treatments, as none of the interactions between period and treatment were significant, so they were therefore left out of the regressions.

Figure 2 gives a first hint as to why the possibility of whistleblowing did not decrease the prevalence of B-choices. In this figure, the proportion of submitted reports over all possible reports is plotted over all 10 periods. With an overall proportion of about 5.5%, whistleblowing barely existed in our experiment throughout all periods. From a total of 651 B-choices in *Whistle*, only 71 were reported.⁴ Thus, in roughly 90% of the cases, participants got away with their B-choice without being reported. While (costly) punishment of selfish group members is used to a considerable extent in almost all previous public-goods experiments, whistleblowing in our experiment did not appear to be seriously considered by the participants. Recall that whistleblowing in our experiment represented nothing but a punishment opportunity in the hands of each group member who observed a B-choice. The main difference with most previous public-goods experiments is that punishment in our experiment did not lead to sure monetary consequences for the punished, as was usually the case in all previous studies. If employees in real organizations shy away from using the whistleblowing system if consequences for the wrongdoer are unclear or uncertain, this can be a huge problem because it might leave executives with the erroneous con-

⁴ In one of the 71 cases in which misconduct was reported, both group members filed a report. In all other cases, misconduct was reported by only one group member

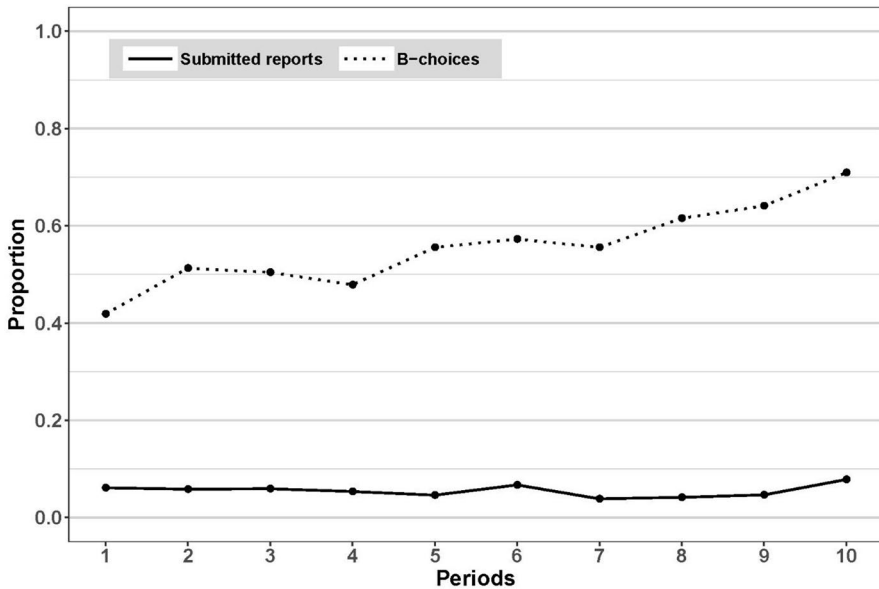


Fig. 2 Proportion of participants' submitted reports over periods in the *Whistle* treatment

clusion that misconduct is not present in their company. Even worse, an ineffective whistleblowing system may increase the prevalence of misconduct, as our results suggest.

4.1.2 Lack of whistleblowing as implicit consent?

The Psychological Gain Hypothesis was based on the behavioral perspective that the mere institutionalization of whistleblowing might increase the prevalence of B-choices. Because a B-choice can be reported and punished, it may no longer induce feelings of shame among the participants. According to this explanation, the institutionalization of whistleblowing crowds out the intrinsic motivation to act in the other group members' interest.

However, given the low level of whistleblowing that occurred in our experiment, an alternative explanation for the observed increase in B-choices may come to mind. It might be the behavioral reluctance to use the whistleblowing system that increases the prevalence of B-choices rather than the presence of an explicit whistleblowing system. According to this explanation, the availability of an internal whistleblowing system affords people the opportunity to anonymously object to observed behavior, and seemingly wrongful acts are regarded as socially accepted until peers raise their voices. A lack of whistleblowing is therefore interpreted as a sign of implicit consent. This mechanism might be interpreted in terms of a recent concept of "shared guilt" (Inderst et al. 2019). If peers do not raise their voices, the occurrence of B-choices is partly their own fault. The forgone option of blowing the whistle leads to a shift in the attribution of guilt from the B-chooser to the potential whistleblower.

Table 4 Behavioral response to being reported in the *Whistle* treatment

	Proportion (B-choice in t B-choice in $t - 1$)					
	Reported in $t - 1$		Not reported in $t - 1$		Not reported in $t - i$	
	Obs. (Subj.)	Prop.	Obs. (Subj.)	Prop.	Obs. (Subj.)	Prop.
Overall	58 (29)	0.81	510 (93)	0.84	429 (84)	0.83
The first time	29 (29)	0.83				
A second time	17 (17)	0.76				
More than twice	12 (9)	0.83				
$t = 2$	6 (6)	0.83	43 (43)	0.86		

^aThe table shows the relative frequencies of B-choices in periods $t > 1$ when the respective participants chose B in the previous period (i.e., $t - 1$). Those situations are divided into cases where participants were reported in the previous period, were not reported in the previous period, and were not reported in any previous period. The situations where participants were reported in the previous period are further subdivided into cases where they were reported the first time, a second time and more than twice. The last two lines of the table contrast the relative frequencies of choosing B in the second period when participants were or were not reported in the first period for choosing B. The respective columns to the left of the relative frequencies contain the number of observations and, in parentheses, the respective numbers of subjects for each situation.

Table 4 presents the behavioral responses to being reported. The table shows the proportions of B-choices in period $t > 1$ among all participants in the *Whistle* treatment who chose alternative B in the respective previous period. The proportions of choosing B again were calculated for the cases in which participants were reported in the previous period, were not reported in the previous period and were never reported in any previous period (including those periods in which they chose A).

The fact that participants repeated alternative B in 81% of the cases when they were reported in the previous period alone speaks against implicit consent as an explanation for the increased prevalence of B-choices. Compared to the proportions when participants were not reported in the previous period or were never reported in any previous period (84% and 83%, respectively), it becomes apparent that being reported appears to have virtually no effect on the propensity to choose alternative B again in the subsequent period. Moreover, it does not seem to matter whether participants were reported the first, the second or more than a second time (the respective proportions of choosing B again were 83%, 76% and 83%).⁵

In the *No-Whistle (Always Detected)* treatment, the overall proportion of choosing alternative B again in the subsequent period was 73%; in the *No-Whistle (Never Detected)* treatment, it was 85%. Particularly with respect to the latter proportion, it turns out that being reported in *Whistle* barely affected subsequent choices. Once participants chose alternative B, they had a strong inclination to choose B again,

⁵ One participant was reported five times during the first nine periods. This participant never switched to alternative A in the subsequent period.

Table 5 Contagion effects (logit regression)

	Dependent variable: <i>B-choice</i> (= 1 if true)		
	(1)	(2)	(3)
<i>No-Whistle (Always Detected)</i>	-0.25 (0.19)	-0.28 (0.32)	
<i>Whistle</i>	0.32 (0.18)	0.53 (0.33)	
<i>(Period - 2)</i>	0.09 (0.02)		0.16 (0.02)
<i>Proportion B (others)</i>	2.99 (0.34)		
<i>B-type</i>		2.23 (0.31)	2.89 (0.22)
<i>B-type × [No-Whistle (Always Detected)]</i>		-0.19 (0.44)	
<i>B-type × Whistle</i>		-0.07 (0.44)	
<i>B-type × (Period - 2)</i>			-0.15 (0.03)
<i>Constant</i>	-1.77 (0.20)	-1.03 (0.23)	-1.66 (0.17)
Obs.	3186	3186	3186
Groups	118	118	118

^aThe table shows the coefficients of logit regressions based on all data except from the first period (i.e., periods 2 to 10). *No-Whistle (Always Detected)* and *Whistle* are dummy variables indicating the respective treatment. *No-Whistle (Never Detected)* is the base category. *(Period - 2)* indicates the respective period minus 2 and runs from 0 to 8. *B-type* is a dummy variable indicating whether a participant chose alternative B in the first period. Standard errors are given in parentheses and are clustered at the group level.

whether or not someone blew the whistle on them. The main difference between the three treatments appears to be the inhibition threshold for choosing alternative B the first time. Indeed, in *No-Whistle (Always Detected)*, 42% of the participants never chose alternative B; in *No-Whistle (Never Detected)*, 30% never did; and in *Whistle*, only 18% never chose B.

Although this does not provide direct evidence for the existence of a crowding-out of participants' intrinsic motivation to act in the other group members' interest, it empirically rules out a second explanation that could have been plausible in light of the very low level of whistleblowing.

4.2 Contagion effects of misconduct

While the results so far have shown that the prevalence of B-choices increased by period, it is not clear yet whether this was a pure time effect or whether choosing alternative B was indeed contagious. With the help of some logit regressions reported in Table 5, we tried to separate both effects. In the first model, we again regressed choosing alternative B on treatment dummies and on a variable for the period. In addition, we calculated the overall proportion of B-choices of both other group members in all previous periods for each participant and used this as another explanatory variable in the regression. Because this variable could only be calculated after the

first period, the regression was run on data from periods 2 to 10. As can be seen in the table, we find evidence of both a pure time effect as well as a separate contagion effect regarding B-choices. The tendency to choose alternative B significantly increased by period, even when we controlled for the behavior of group members (*Period - 2*: $\beta = 0.09$, $p < 0.01$). At the same time, the tendency to choose alternative B significantly increased with the observed propensity of B-choices of both other group members (*Proportion B (others)*: $\beta = 2.99$, $p < 0.01$), which clearly points to a contagion effect of choosing B on top of a pure time effect.

A second way to look at the contagion effects of B-choices is presented in Eqs. (2) and (3) of Table 5. Here, we separated the participants into two types and investigated how their behavior changed over periods. Because there was no interaction within groups prior to the first period, we used the participants' choices between alternatives A and B in period 1 as a proxy of their type. We call participants who chose alternative A in period 1 "A-types" and those who chose alternative B in period 1 "B-types."

The results of Eq. (2) in Table 5 show that this simple classification into types distinguishes the participants' behavior regarding overall levels of B-choices to a considerable extent. Participants who chose alternative B in the first period chose alternative B in all subsequent periods more than twice as often as participants who chose alternative A in period 1 (*B-type*: $\beta = 2.23$, $p < 0.01$). Moreover, we did not find significant differences regarding the overall level of B-choices of each type between the three treatments. Therefore, we pooled the data from all treatments to look at the behavioral changes among each type over periods.

Equation (3) in Table 5 presents the estimation results of a logit model in which we regressed choosing alternative B on a variable for periods, a dummy variable for being a B-type and the interaction between both variables. The main effect of B-type again indicates that participants who chose alternative B in the first period chose alternative B in all subsequent periods significantly more frequently than those who chose alternative A in period 1 did (*B-type*: $\beta = 2.89$, $p < 0.01$). The main effect of the period variable shows that A-types significantly increased their proclivity to choose alternative B over periods 2 to 10 (*Period - 2*: $\beta = 0.16$, $p < 0.01$). The negative and significant interaction effect between period and B-type suggests that this dynamic was systematically different for B-types (*B-type* \times (*Period - 2*): $\beta = -0.15$, $p < 0.01$). In fact, the regression indicates no significant linear time trend for B-types over periods 2 to 10 (*(Period - 2) + B-type* \times (*Period - 2*): $\beta = 0.01$, $p = 0.76$).

Figure 3 visualizes the different dynamics between A- and B-types in each treatment. B-types had a high propensity to choose alternative B during the first periods of the game and more or less remained at this level throughout all subsequent periods. On the other hand, A-types started with a low propensity to choose alternative B at the beginning of the game and gradually increased their willingness for B-choices by period. This shows how dangerous misconduct in a company can become if it is not harshly punished. Wrongdoing might spread and infect even the "good" people.

4.3 The effects of severe and certain punishment

As explained in the introduction, employees' perception of mild and hesitant punishment for wrongdoing might be widespread in many companies, as the results of Ernst

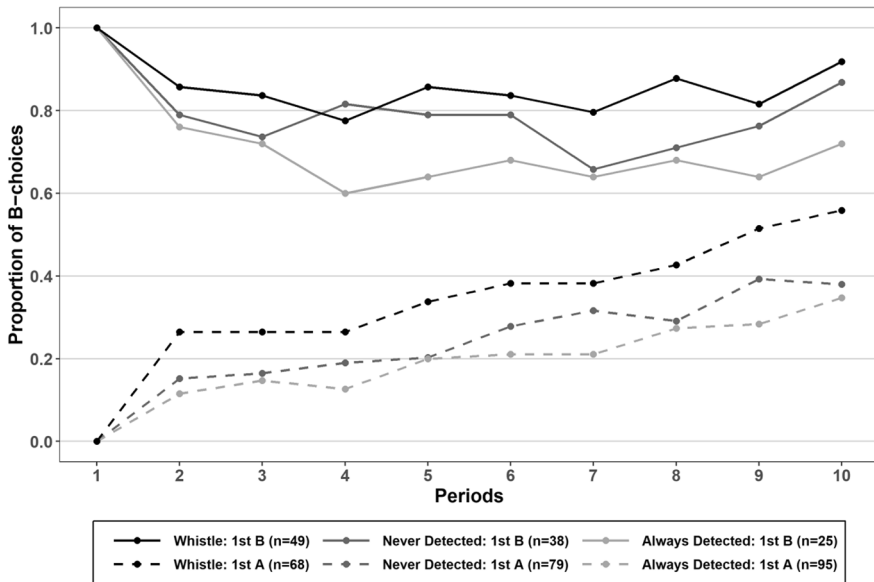


Fig. 3 Proportion of B-choices over periods across the different types

& Young's Fraud Surveys appear to show. However, mild and hesitant punishment of misconduct is usually not what executives and compliance managers publicly report as being their attitudes toward internal wrongdoing. A typical statement of an anonymous chief financial officer in one of Ernst & Young's Fraud Surveys was: "Strong rules. Zero tolerance. You pay a bribe; you're fired" (Ernst&Young, 2013a). Thus, executives and compliance managers claim to take each instance of misconduct seriously and to punish it harshly. They probably even believe this situation exists in their companies and that every employee knows about it.

Therefore, we conducted another treatment in which whistleblowing was associated with stronger and certain punishment to investigate whether this would have the desired effects. Again, we were focusing on the frequency of B-choices as an indicator of wrongdoing as well as the frequency of reporting as an indicator of whistleblowing as our dependent variables. The purpose of this treatment was not to disentangle whether the severity or certainty of punishment is more effective. The purpose of this treatment was to investigate whether the much-vaunted zero-tolerance policy is as promising as is often assumed. And zero tolerance usually means both: severe and certain punishment.

We were therefore interested whether people are willing to report other people who are misbehaving from the group's or company's point of view in a zero-tolerance setting. And, likewise, whether a whistleblowing institution has a deterrence effect in such an environment. Neither is necessarily clear a priori. Since misconduct occurs in secret, companies depend on employees being willing to report wrongdoers. The implementation of a zero-tolerance policy in a top-down manner is completely ineffective if employees are reluctant to report misconduct. But this reluctance might

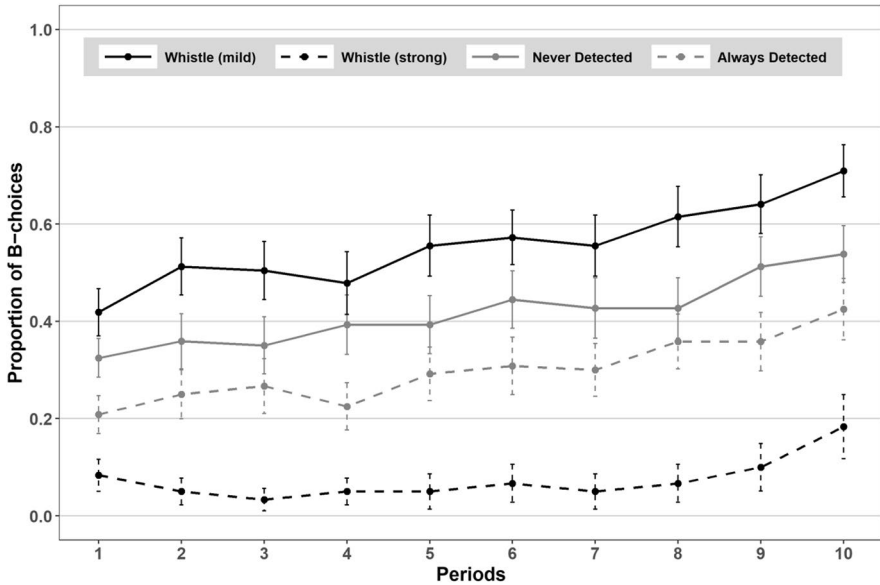


Fig. 4 Proportion of B-choices over periods across treatments, including severe and certain punishment

even be reinforced rather than diminished if the wrongdoer has to expect severe internal penalties. Or the potential whistleblower does not react at all to the level and likelihood of punishment of the wrongdoer.

While we could not fire a participant in our experiment, we increased the monetary consequences for the punished. This treatment was exactly the same as the whistleblowing treatment described so far. The only difference in this new treatment was that if a participant was reported by a group member for choosing alternative B, the participant's earnings were reduced to 25 ECU with certainty in the respective period. Thus, punishment for choosing alternative B was certain and more substantial. We ran two sessions of this treatment in the laboratory of the Max Planck Institute of Economics in Jena, with a total of 60 participants (i.e., 20 groups). The results regarding B-choices are presented in Fig. 4.

Clearly, if whistleblowing is associated with more severe and certain punishment for choosing alternative B, it will have the desired deterrence effect. Under such conditions, the prevalence of B-choices was substantially reduced in all periods ($p < 0.01$ in each period and against any other treatment based on two-sided t -tests). Moreover, we did not find any evidence regarding contagion effects of choosing alternative B because the prevalence of B-choices remained at a low level throughout all periods (except for some small endgame effects in the last period).

Maybe most interestingly, the overall proportion of whistleblowing increased from 5.5% in the mild and uncertain punishment treatment to 27% in the severe and certain punishment treatment. Thus, with stronger punishment for choosing alternative B, the participants considered reporting a B-choice to a non-negligible degree. Notice that this was the case although the monetary cost of blowing a whistle was kept constant at 10 ECU as compared to our treatment with mild and hesitant punish-

ment. From a total of 44 B-choices in the severe and certain punishment treatment, 19 were reported. It therefore seems that a non-negligible proportion of people was willing to blow the whistle even though they knew that they could not undo past misbehavior. This means that participants got away with their B-choice without being reported in only 57% of the cases, compared to 90% in the mild and uncertain punishment treatment.

5 Discussion and conclusion

Fighting misconduct is a pressing issue for companies and is anything but simple. The first and probably greatest problem is that companies usually do not possess information about current misconduct because wrongdoing takes place secretly. The implementation of a whistleblowing system is propagated as a countermeasure. The institutionalization of whistleblowing shall help to uncover misconduct within organizations and, as a consequence, impede wrongdoing.

Because it is hardly possible to study the effects and efficacy of whistleblowing systems in the field, we ran a controlled lab experiment. Our experiment was based on the view that internal whistleblowing mainly addresses the risk of damage to the company rather than the moral component of the reported behavior. In the experiment, we therefore modeled only the economic consequences of misconduct and reduced this situation to its essential elements. Nevertheless, the proposed game can be extended in many directions to investigate other factors of misconduct and whistleblowing. For instance, one could add a pre-play communication stage in which players can either agree on certain “codes of conduct” or are simply informed about them. Alternatively, one could implement a stronger education of players regarding undesirable behavior and clearly label alternative B “misconduct.” It would be interesting to see how these ways of sensitization and education affect (i) the prevalence of B-choices and (ii) the participants’ willingness to blow the whistle. One could also examine how different degrees of misconduct affect whistleblowing by varying the level of possible damage to the group or how financial rewards help to induce whistleblowing among employees. Or one could introduce possibilities for retaliation to make whistleblowing riskier. These are just a few directions in which the experiment could be easily adapted.

The experimental setting of the present study contained some features that should have been conducive to whistleblowing. Notwithstanding, we did not see a strong urge to report peers among the participants. In the mild and uncertain punishment treatment, whistleblowing was virtually absent. While weak and uncertain consequences for the wrongdoer might have been a reason for the sparse reporting in this treatment, the penalty for wrongdoing was much more severe and surely inflicted in the severe and certain punishment treatment. Indeed, in such a setting, the participants were willing to report their group members and the chances of getting away with misconduct were considerably reduced. Future research may examine whether it is necessary to increase both the severity and the likelihood of the punishment, or whether changing one of the two factors is sufficient in this context.

Another reason for the lack of whistleblowing was brought to our attention by the participants themselves. A few of them remarked after the experiment that they did not report their group members because reports did not decrease the probability of suffering a loss and that we should possibly change this in future experiments. Obviously, they did not understand that although their report could not undo past misconduct, it could well prevent future wrongdoing, at least in the severe and certain punishment treatment. In this sense, participants were unaware of the shadow of the future that their reporting might cast. For the deterrence effect of an internal whistleblowing system, however, this is a crucial mechanism. Employees must understand that one important aspect of whistleblowing is that it might prevent future misconduct. Thus, companies should possibly emphasize the importance of whistleblowing with respect to its influence on future behavior, to increase reporting among their employees. Along the way, companies can manage employees' perceptions concerning the pursuit and punishment of misconduct, which seems to be a frequent problem as several fraud surveys suggest. This highlights the importance of educating employees about misconduct, its consequences, and the function of internal whistleblowing systems.

However, internal whistleblowing systems may not only fail to fulfill their minimal requirement of disclosing misconduct. They may even induce wrongdoing, especially in companies with lenient punishments. The institutionalization of whistleblowing appears to crowd out people's conscience to act in a compliant way. The outer institution seems to substitute for the inner inhibition threshold to abstain from misconduct. If the punishment for wrongdoing is at best only mild, then people might be willing to bear the risk of being punished and become wrongdoers more often. This result is in stark contrast to our hypotheses derived from the extensive public-goods literature, which documents a profound deterrence effect of punishment on selfish behavior, even if the punishment is merely symbolic.

The adverse effects of the whistleblowing system, however, are in line with evidence from behavioral research indicating that there indeed exists an interaction between institutions and preferences (see, e.g., Frey & Oberholzer-Gee, 1997; Gneezy & Rustichini, 2000a; Mellström & Johannesson, 2008; Belot & Schröder, 2015). This literature suggests that material incentives may sometimes erode people's social preferences and that it would be a cardinal fault to consider preferences to be purely exogenous (Bowles 2016). Our results underline this concern and highlight the importance of accounting for possible interactions between institutions and preferences when designing incentive schemes to foster ethical behavior.

The endogeneity of participants' preferences in our context becomes clear when considering our analysis on contagion effects in Sect. 4.2. Not only were our participants' preferences on choosing the risky alternative B influenced by the institutionalization of whistleblowing. Their choices were also influenced by observing their peers. Although many participants initially started with choosing the safe option, they ended up in playing the risky game when they saw that others did so as well. It were the wrongdoers that incited the compliant ones to do wrong and not the compliant ones that inspired the wrongdoers to comply (see also Kandul & Uhl, 2016). This infectivity of observable wrongdoing provides yet another reason to strive for a working environment in which wrongdoing is contained through effective whistleblowing.

In sum, companies should be aware that whistleblowing systems may not only waste resources but could even crowd out compliance. This implication gains relevance in light of the pending political obligation to implement these measures. The upside of our results, however, is that whistleblowing systems can become a powerful tool with which to fight misconduct. At the very least, our findings suggest that a whistleblowing institution in combination with severe and assured penalties for the wrongdoer can be effective. This may mean that companies have to commit credibly to referring possible cases of misconduct to law enforcement agencies or to dismissing convicted wrongdoers themselves. If companies preach a zero-tolerance policy, they should practice it as well. Otherwise, they might even worsen the situation. Loosely following Gneezy and Rustichini (2000b), misconduct might be a case of “punishing enough or not punishing at all.”

6 Appendix A

Table A1 B-choices and whistleblowing in Jena and Munich

	<i>No-Whistle (Always Detected)</i>		<i>No-Whistle (Never Detected)</i>		<i>Whistle</i>		Whis.
	Groups	Mean (SD)	Groups	Mean (SD)	Groups	Mean (SD)	
Jena	20	0.28 (0.27)	19	0.41 (0.33)	20	0.53 (0.33)	0.05
Munich	20	0.32 (0.31)	20	0.42 (0.31)	19	0.59 (0.32)	0.06
M.W.U.		0.66		0.86		0.58	
t-test		0.60		0.90		0.54	

“The table shows the means (standard deviations) of the relative frequencies of B-choices for both locations across all treatments. The proportions of submitted reports over all possible reports for both locations are reported in the last column (“Whis.”). The last two lines display the p-values of two-sided Mann-Whitney *U*-tests and unpaired *t*-tests regarding possible differences in the overall tendency toward B-choices between sessions conducted in Jena and Munich. The standard deviations as well as statistical testing are based on group-level averages. The total number of groups in each treatment and each location are contained in the “Groups” columns.

7 Appendix B

7.1 Participants’ views on alternative B

Even though we calibrated the probability of discovery through pilot sessions of the control treatments such that some participants were willing to choose alternative B while others were not, it may be worthwhile to highlight the problem of the (supposedly negligible) risk of alternative B from the participants’ point of view. In a questionnaire at the end of the experiment we asked all the participants about their rationale regarding their choices in a free-text question. Almost all participants

answered this question and we classified each of their answers into one of the following categories:

- (i) The participant does not find the risk of alternative B problematic.
- (ii) The participant considers the risk of alternative B to be problematic for him/herself.
- (iii) The participant finds it problematic that risk is transferred to others through his/her choice of alternative B.
- (iv) The participant finds the risk of alternative B problematic, but chose B (from time to time) after other group members chose B.

If a participant stressed the problem of the risk to him- or herself as well as to the other group members, we classified the answer under category (ii). Participants who did not answer the question or whose answers could not be classified in one of the above categories, were listed in the category "Miscellaneous."

Table B1 shows the results of this categorization for both control treatments. The table contains representative statements from two to three participants in each category as well as the proportion of participants in each category for both control treatments. Despite the very low risk attached to alternative B, the vast majority of participants indicated unease with the probability of discovery. By far the largest proportion expressed a considerable concern regarding the risk to themselves (categories (ii) + (iv): 72% and 53% in *No-Whistle (Always Detected)* and *No-Whistle (Never Detected)*, respectively); 8% and 16% of the participants in the respective control treatment expressed their concerns mainly with respect to the group. Only 14% and 24% of the participants in *No-Whistle (Always Detected)* and *No-Whistle (Never Detected)*, respectively, stated that they considered the risk to be so small that it hardly played a role or that they found the higher earnings of alternative B attractive enough to take the risk.

The reasoning expressed by the participants in the post-experimental questionnaire thus confirms the successful calibration of the probability of discovery: for some participants the risk associated with alternative B was not a problem, while for others (in fact the majority) it was.

Table B1 Participants' views on the inherent risk of Alternative B

Risk of B was	Representative Statements	Proportion of Participants	
		<i>No-Whistle (Always Det.)</i>	<i>No-Whistle (Never Det.)</i>
... not a problem for the participant.	<p>"I always chose option B because I am a very risk-loving person and the probability was low enough in my view."</p> <p>"If all participants had chosen answer B, the probability of losing would still be very low."</p>	14.2%	23.9%

Table B1 Participants' views on the inherent risk of Alternative B

Risk of B was	Representative Statements	Proportion of Participants	
		No-Whistle (Always Det.)	No-Whistle (Never Det.)
... a problem for the participant him/herself.	<p><i>"Better safe than sorry."</i></p> <p><i>"I didn't think it made any sense to take the risk of earning nothing just for another 5 Euros."</i></p> <p><i>"Even though the risk of losing the entire earnings is extremely low, I didn't want to take the risk of losing everything. For this, the appeal of getting just 50% more money was not high enough. In order to take the risk of the loss, the possible earnings should have doubled in my opinion."</i></p>	64.2%	36.8%
... a social/moral problem for the participant.	<p><i>"Since I am socially committed, the welfare of all is in my interests. This means that I always took option A to avoid risking a loss for all."</i></p> <p><i>"Small risk for higher earnings, but teammates were concerned about safety - consequently considerateness."</i></p> <p><i>"I didn't want to be the fool to rob the entire group of its profits. I didn't do anything wrong and I'll have a peace of conscience that it wasn't my greed that might have made the group lose money."</i></p>	8.3%	16.2%
... a problem for the participant, but B was contagious.	<p><i>"At first solidarity, but then it escalated because the others only chose B."</i></p> <p><i>"Actually, I always wanted to choose option A. But since the other two members of my group always chose option B, I didn't accept that they were putting me at risk and earning more than I did. So, I chose option B as well."</i></p>	7.5%	16.2%
Miscellaneous Participants		5.8%	6.8%
		120	117

^aThe table shows a classification of the participants in the control treatments with respect to their views on alternative B. The classification is based on free-text answers in a post-experimental questionnaire, in which the participants were able to state their reasoning about their behavior during the experiment. The second column contains representative statements from two to three participants of the respective category. Columns three and four indicate the proportions of the participants in the respective category for the two control treatments. The category "Miscellaneous" contains participants who did not answer the corresponding question or whose answers could not be assigned to one of the other categories.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11573-023-01144-w>.

Acknowledgements We thank Max Albert, Marcus Giamattei, Hartmut Kliemt, Johann Graf Lambsdorff and the seminar participants at the Max Planck Institute of Economics, the Technical University of Munich, the University of Giessen, the University of Kassel, the University of Passau and the ESA World Meeting 2018 for their valuable comments as well as Friedrich Gehring for programming the experiment. We gratefully acknowledge the funding by the Max Planck Institute of Economics and the Technical University of Munich.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data Availability The data that support the findings of this study are available from the corresponding author upon request.

Declarations

Compliance with ethical standards The experiment was approved by the Institutional Review Board of the Max Planck Institute of Economics and experimentUM (social science lab of Technische Universität München). The investigation was conducted according to the principles expressed in the Declaration of Helsinki. Written consent was obtained from the participants.

Competing interests The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethics approval Approval was obtained from the ethics committee of the Max Planck Institute of Economics, Jena, Germany, and the Technical University of Munich, Germany. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abbink K, Irlenbusch B, Renner E (2002) An experimental bribery game. *J Law Econ Organ* 18:428–454
- Amir E, Lazar A, Levi S (2018) The Deterrent Effect of Whistleblowing on Tax Collections. *Eur Acc Rev* 75:939–954
- Anderson CM, Putterman L (2006) Do non-strategic sanctions obey the law of demand? The demand for punishment in the voluntary contribution mechanism. *Games Econ Behav* 54:1–24
- Association of Certified Fraud Examiners (2016) 2016 *Global fraud survey*. <https://www.acfe.com/-/media/files/acfe/pdfs/2016-report-to-the-nations.ashx>
- Bartuli J, Djawadi BM, Fahr R (2016) Business ethics in organizations: an experimental examination of whistleblowing and personality. IZA Discussion Paper No.10190
- Belot M, Schröder M (2015) The spillover effects of monitoring: a field experiment. *Manage Sci* 62:37–45
- Bowles S (2016) The moral economy: why good incentives are no substitute for good citizens. Yale University Press
- Bowles S, Gintis H (2002) Social capital and community governance. *Econ J* 112:419–436
- Butler JV, Serra D, Spagnolo G (2020) Motivating whistleblowers. *Manage Sci* 66:605–621
- Carpenter JP (2007) The demand for punishment. *J Econ Behav Organ* 62:522–542
- Carpenter J, Robbett A, Akbar PA (2018) Profit sharing and peer reporting. *Manage Sci* 64:4261–4276
- Choo L, Grimm V, Horváth G, Nitta K (2019) Whistleblowing and diffusion of responsibility: an experiment. *Eur Econ Rev* 119:287–301
- Dozier JB, Miceli MP (1985) Potential predictors of whistle-blowing: a prosocial behavior perspective. *Acad Manage Rev* 10:823–836
- Egas M, Riedl A (2008) The economics of altruistic punishment and the maintenance of cooperation. *Proceedings of the Royal Society of London B: Biological Sciences*, 275: 871– 878
- Ernst&Young (2013a) *Growing Beyond: a place for integrity—12th Global Fraud Survey Report*
- Ernst&Young (2013b) Navigating today's complex business risks: Europe, Middle East, India and Africa—Fraud Survey 2013. Report
- Ernst&Young (2014) *Overcoming compliance fatigue: Reinforcing the commitment to ethical growth—13th Global Fraud Survey Report*
- Ernst&Young (2015) Fraud and corruption: the easy option for growth? Europe, Middle East, India and Africa—Fraud Survey 2015. Report

- Ernst&Young (2017) Economic uncertainty, unethical conduct: how should overburdened compliance functions respond? —Asia-Pacific Fraud Survey 2017. Report
- European Commission (2017) Special Eurobarometer 470: Corruption - Summary. <https://europa.eu/eurobarometer/api/deliverable/download/file?deliverableId=63944>
- Fehr E, Gächter S (2000) Cooperation and punishment in public goods experiments. *Am Econ Rev* 90:980–994
- Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415:137
- Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Exp Econ* 10:171–178
- Frey BS, Oberholzer-Gee F (1997) The cost of price incentives: an empirical analysis of motivation crowding-out. *Am Econ Rev* 87:746–755
- Gallus J (2016) Fostering public good contributions with symbolic awards: a large-scale natural field experiment at Wikipedia. *Manage Sci* 63:3999–4015
- Gneezy U, Rustichini A (2000a) A fine is a price. *J Legal Stud* 29:1–17
- Gneezy U, Rustichini A (2000b) Pay enough or don't pay at all. *Q J Econ* 115:791–810
- Greiner B (2015) Subject pool recruitment procedures: organizing experiments with ORSEE. *J Economic Sci Association* 1:114–125
- Gundlach MJ, Douglas SC, Martinko MJ (2003) The decision to blow the whistle: a social information processing framework. *Acad Manage Rev* 28:107–123
- Inderst R, Khalmetski K, Ockenfels A (2019) Sharing guilt: how Better Access to Information May Backfire. *Manage Sci* 65:3322–3336
- Johannesen N, Stolper T (2017) The Deterrence Effect of Whistleblowing: An Event Study of Leaked Customer Information from Banks in Tax Havens (December 21, 2017). *CEifo Working Paper Series*, No. 6784
- Kandul S, Uhl M (2016) Inspirations or Incitements? Ethical mind-sets and the Effect of Moral examples. *J Behav Experimental Econ* 65:146–153
- Kaplan SE, Whitecotton SM (2001) An examination of auditors' reporting intentions when another auditor is offered client employment. *AUDITING: A Journal of Practice & Theory* 20:45–63
- Kaptein M (2011) From inaction to external whistleblowing: the influence of the ethical culture of organizations on employee responses to observed wrongdoing. *J Bus Ethics* 98:513–530
- Krawiec KD (2003) Cosmetic compliance and the failure of negotiated governance. *Wash Univ Law Q* 81:487–544
- Lee G, Xiao X (2018) Whistleblowing on accounting-related misconduct: a synthesis of the literature. *J Acc Literature* 41:22–46
- Liu Y, Zhao S, Li R, Zhou L, Tian F (2018) The relationship between organizational identification and internal whistle-blowing: the joint moderating effects of perceived ethical climate and proactive personality. *RMS* 12:113–134
- Masclet D, Noussair C, Tucker S, Villeval M-C (2003) Monetary and non-monetary punishment in the voluntary contributions mechanism. *Am Econ Rev* 93:366–380
- Mellström C, Johannesson M (2008) Crowding out in blood donation: was Titmuss right? *J Eur Econ Assoc* 6:845–863
- Miceli MP, Near JP, Dworkin TM (2009) A word to the wise: how managers and policy-makers can encourage employees to report wrongdoing. *J Bus Ethics* 86:379–396
- Near JP, Miceli MP (1995) Effective-whistle blowing. *Acad Manage Rev* 20:679–708
- Near JP, Miceli MP (1996) Whistle-blowing: myth and reality. *J Manag* 22:507–526
- Near JP, Miceli MP (2016) After the wrongdoing: what managers should know about whistleblowing. *Bus Horiz* 59:105–114
- Near JP, Rehg MT, Van Scotter JR, Miceli MP (2004) Does type of wrongdoing affect the whistle-blowing process? *Bus Ethics Q* 14:219–242
- Nicholls AR, Fairs LRW, Toner J, Jones L, Mantis C, Barkoukis V, Perry JL, Micle AV, Theodorou NC, Shakhverdiev S, Stoicescu M, Vesic MV, Dikic N, Andjelkovic M, Grimau EG, Amigo JA, Schomöller A (2021) Snitches get Stitches and End up in ditches: a systematic review of the factors Associated with Whistleblowing Intentions. *Front Psychol* 12:1–20
- Nikiforakis N (2008) Punishment and counter-punishment in public good games: can we really govern ourselves? *J Public Econ* 92:91–112
- Nikiforakis N, Normann H-T (2008) A comparative statics analysis of punishment in public-good experiments. *Exp Econ* 11:358–369
- PwC (2013) *Striking a balance: Whistleblowing arrangements as part of a speak up strategy – revised paper*. <https://pwc.blogs.com/files/130813-striking-a-balance.pdf>

- Reuben E, Stephenson M (2013) Nobody likes a rat: on the willingness to report lies and the consequences thereof. *J Econ Behav Organ* 93:384–391
- Schmolke KU, Utikal V (2018) Whistleblowing: Incentives and situational determinants. Available at SSRN 3198104
- Schultz JJ, Johnson DA, Morris D, Dyrnes S (1993) An investigation of the reporting of questionable acts in an international setting. *J Accounting Res* 31:75–103
- Stikeleather B (2016) When do employers benefit from offering workers a financial reward for reporting internal misconduct? *Acc Organ Soc* 52:1–14
- Soltes E (2020) Paper Versus Practice: a Field Investigation of Integrity Hotlines. *J Accounting Res* 58:429–472
- The New York Times (2002) Text of Letter to Enron’s Chairman After Departure of Chief Executive. <https://www.nytimes.com/2002/01/16/business/text-of-letter-to-enron-s-chairman-after-departure-of-chief-executive.html>, January 16.
- Wilde J (2017) The Deterrent Effect of Employee Whistleblowing on Firms’ Financial Misreporting and Tax Aggressiveness. *Acc Rev* 92:247–280

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.