



**Discussion Papers
in Economics**

No. 09/2016

**Whistleblowing: Incentives and situational
determinants**

**Klaus Ulrich Schmolke
University of Erlangen-Nürnberg**

**Verena Utikal
University of Erlangen-Nürnberg**

ISSN 1867-6707

Whistleblowing: Incentives and situational determinants

Klaus Ulrich Schmolke* Verena Utikal**

Abstract

Law makers increasingly try to capitalize on individuals having acquired knowledge of corporate crimes or other misconduct by inducing them to blow the whistle. In a laboratory experiment we measure the effectiveness of incentives on the willingness to report such misconduct to a sanctioning authority. We find that fines for non-reporting insiders, rewards and even simple commands increase the probability of whistleblowing. We find the strongest effect for fines. Situational determinants also influence the willingness to blow the whistle: Insiders who are negatively affected by the misconduct are more likely to blow the whistle than non-affected or profiting insiders. Those (negatively affected) victims are also sensitive to the misconduct's impact on the authority sanctioning the misconduct (public authority or employer): Whistleblowing is more likely if the enforcement authority is negatively affected compared to positively or not affected.

Keywords: Whistleblowing; incentives; situational determinants; experiment.

JEL Classification:C91, D82, K42, M59

* Department of Law, University of Erlangen-Nuremberg, Schillerstraße 1, D-91054 Erlangen, Germany, +49-9131-85-22819, klaus.ulrich.schmolke@fau.de

**Corresponding author, Department of Economics, University of Erlangen-Nuremberg, Lange Gasse 20, D-80403 Nuremberg, Germany, +49-911-5302-229, verena.utikal@fau.de.

1. Introduction

Every once in a while a big scandal shakes the confidence in the integrity of corporate conduct and leads to public outcry. The routine response of the political caste is to come forward with new regulation that imposes ever stricter rules of conduct which come with ever harsher sanctions in case of infringement. However, it is a well-established insight that coming up with new prohibitions and raising sanctions are relatively ineffective regulatory strategies as long as the actual enforcement of the new rules does not keep pace with these measures. Effective enforcement by supervisory authorities and public prosecution, though, is often hampered by the lack of sufficient resources to gather the information necessary to reveal corporate misconduct. Facing this predicament law makers increasingly try to capitalize on individuals having acquired knowledge of such misconduct by inducing them to blow the whistle.¹ This approach is based on the conviction that – had such individuals, typically employees of the delinquent entity, informed public authorities at an early stage – the damage done by corporate scandals like Worldcom, Parmalat (both accounting fraud), Siemens (large scale bribery) or – very recently – Toshiba (accounting fraud) and Volkswagen (fraud on diesel emissions) could have been hugely reduced if not prevented. The cases of Siemens and Volkswagen also very clearly show that corporate entities themselves have strong incentives to integrate whistleblower schemes into their internal compliance systems.

However, potential whistleblowers often hesitate to come to the fore and report to the compliance office of their corporation (internal whistleblowing) or to a public authority (external whistleblowing). This is due to the fact that whistleblowing has some properties of a public good: Whereas the whistleblower often gains little more than the satisfaction from “doing the right thing”, she envisages potentially staggering costs: Beyond direct retaliatory sanctions imposed by the employer² that may culminate in job loss, whistleblowers purportedly face blacklisting from other potential employers in the industry, ostracism from colleagues in addition to the psychological and physical costs caused by stress, public exposure and the like.³ In short, the whistleblower is highly undercompensated for her socially beneficial efforts. This public good problem may be aggravated by deeply rooted social attitudes perceiving whistle-blowing as some kind of deleterious denunciation.⁴

¹ Heyes and Kapur (2008); for a global overview see Lewis (2010); for an example from European Capital Markets Law cf. Fleischer and Schmolke (2012).

² Spedding (2004) lists criticism, poor performance evaluations, punitive transfers and job loss as examples.

³ Alford (2001); Rapp (2007): “Potential whistleblowers face tremendous obstacles beyond direct employer retaliation. [...] Moreover, whistleblowers may fear blacklisting from future employers who suspect disloyalty, as well as social ostracism from their co-workers. Additionally, the psychological burdens associated with whistleblowing, including the effects of public criticism and a lengthy stay in litigation’s limelight, cannot be ignored.”

⁴ As to such attitudes cultivated in Central Europe, which can be explained historically see the literature and the quotes cited by Dowling Jr. (2008).

Against this background, the legislatures around the world come up with new regulatory initiatives in order to tilt the cost-benefit-calculus of the potential informant in favor of blowing the whistle. Furthermore, these initiatives may ultimately change social attitudes towards whistleblowers.⁵ Firstly, legislation focused on the protection of employees in the public service.⁶ Later on law-makers included employees of private entities into their whistleblower protection regimes. Notably, the United States enacted the Sarbanes-Oxley Act 2002 (SOX) in response to some major accounting scandals, including the abovementioned cases of Worldcom and Enron. This act contains protective measures for whistleblowers of private corporations being listed at a US stock exchange (Sec. 806 SOX).⁷ The Dodd-Frank Wall Street Reform and Consumer Protection Act 2010 (Dodd-Frank Act) being a reaction to the financial crisis 2008 further expands the whistleblower protection introduced by SOX.

However, this was only a side issue of the Dodd-Frank whistleblower legislation. Its main feature was the introduction of a whistleblower reward program with regard to the violation of US securities laws (Sec. 922 Dodd-Frank Act)⁸ Whistleblowers thereby became entitled to 10 to 30 percent of monetary sanctions exceeding US\$ 1,000,000 that are imposed by the Securities Exchange Commission (SEC) in consequence of the whistleblower's information. A similar award scheme has very recently been introduced by the Motor Vehicle Safety Whistleblower Act 2015 which aims at fighting infractions in the auto industry. Such bounty programs have a long tradition in US law dating back to 1863 when the False Claims Act came into force. In Europe on the other hand, the idea of incentivizing potential informants by monetary rewards is a rather new approach to bolster up law enforcement. The Dodd-Frank Act inspired EU legislature which adopted whistleblower rules in the new Market Abuse Regulation 2014 (MAR) which came into force in mid-2016.⁹ These rules include the authorization of the EU Member States to "provide for financial incentives to persons who offer relevant information about potential infringements of this Regulation" (Art. 32(4) MAR). Furthermore, with regard to antitrust law enforcement whistleblower bounty schemes make inroads into

⁵ Cf. with regard to the UK Public Interest Disclosure Act 1998 (PIDA) the prediction of Gobert and Punch (2000).

⁶ In the aftermath of the Watergate Scandal the United States enacted the Civil Service Reform Act 1978, which provided the first protective rules for whistleblowers in the federal public service. The Federal Whistleblower Act 1989 aimed at strengthening the protection of federal whistleblowers against employer retaliation further. For further details on the Whistleblower Protection Act 1989, see Ebersole (2011). Other countries followed this example. For an example, take the South Australia Whistleblowers Protection Act 1993.

⁷ Furthermore, such corporations were obliged to install audit committees being responsible for the oversight of the work of accounting firms employed by the corporation, which in turn had to "establish procedures for (...) the confidential, anonymous submission by employees of concerns regarding questionable accounting or auditing matters" (Sec 301 SOX). For further details on Sec 301 SOX see Dworkin (2007).

⁸ Prior to the Dodd-Frank Act the Insider Trading and Securities Enforcement Act 1988 already established a – much more limited – whistleblower bounty program for informing on insider trading. For further details, see Ebersole (2011).

⁹ Regulation (EU) No 596/2014, OJ (2014) L 173/1. In Mid-2016 the Ontario Securities Commission introduced a similar reward scheme for blowing the whistle on securities fraud, while the Quebec Securities Commission refrained from doing so because it perceived that there is "not enough evidence to show that money generated more or better tips" (The Economist, July 16. 2016:60).

Europe as well, but from a different direction.¹⁰ The use of such reward schemes in antitrust law enforcement originates from South Korea. The US Congress, however, refrained from introducing a similar regime in the US.¹¹ All in all, it is safe to say that with regard to whistleblower legislation financial reward programs are currently in the center of reform and policy debates.

While legislatures work on setting the stage for whistleblowers to become effective assistants in enforcing the law, corporations themselves strive to implement efficacious internal whistleblower routines as part of their compliance systems. Thereby, corporations get early access to information about internal misconduct as well as the opportunity to control the way such information is disclosed to the competent supervisory authorities (and the public at large).

All these initiatives and activities circle around the one crucial question: How to make potential informants, notably employees, blow the whistle? Feldman and Lobel (2010) categorized the current regulatory strategies on the table as “Protect – Command – Fine – Pay”. The “Protect” strategy refers to anti-retaliation protections, while the “Pay” strategy designates momentary incentives. In the “Fine” strategy a monetary sanction is imposed for not reporting the misconduct in question. The “Command” strategy confines itself to obliging potential informants to report the misconduct.

At first glance, one might be inclined to discard the latter two strategies without further ado, since the “Fine” strategy is seemingly only a “Pay” strategy in disguise, and the “Command” strategy apparently lacks any financial incentive to change behavior. Taking a closer look, however, reveals that the different framing¹² of the “Fine” strategy may be the pivotal feature that makes its impact on potential informants differ from the “Pay” strategy. While the different framing should have no effect on rational actors at least when being risk neutral (cf. Prediction 4 below), potential informers being loss averse should *ceteris paribus* be more inclined to yield to the threat of being fined than to the expectation of receiving a reward (cf. De Geest and Dari-Mattiacci, 2013). On the other hand, there are significant drawbacks of the “Fine” strategy. If capitalizing on whistleblowers is a reaction to the lack of sufficient resources of the public authorities (or the corporate compliance department) to gather the information necessary to reveal misconduct, the imposition of a fine may be (perceived as) an empty threat. That is, while the “Pay” strategy works as a screening device to reveal private actors possessing relevant information, the “Fine” strategy may not work this way. As to the “Command” strategy relaxing the strict assumptions of the standard model

¹⁰ Cf. UK Competition & Markets Authority (CMA) Informant Reward Policy (2014), and Hungarian Competition Authority (GVH), Regular questions about the cartel informant reward (2010); for further detail, see Polański (2014).

¹¹ Cf. Sullivan at al. (2011): The Korea Fair Trade Commission (KFTC) introduced its Cartel Informant Reward Program in 2002. The adoption of similar measures was discussed in the US in 2011. But no such laws have been enacted so far.

¹² The term framing does not apply in its technical sense insofar, as the end-state/size of payoff of „Fine“ obviously differs from „Pay“. The decision problem is, however, identical: To abstain from whistleblowing comes with real costs („Fine“) or opportunity costs („Pay“), respectively.

might again do the trick. There is a strand of literature suggesting that legal norms as such can shape the preferences of private actors (“preference-shaping effects of law”), thereby influencing behavior without providing further incentives via law enforcement (Sunstein, 1986). Furthermore and apart from this effect, an “expressive function” is attributed to law, by which legal norms appeal to social norms. These social norms, in turn, influence behavior by means of reputational mechanisms and informal sanctions. Again, no direct legal consequences are needed to produce this effect (Sunstein, 1996). Hence by relying on these effects of law and, thus, dispensing with the whole apparatus of law enforcement, the behavior of the addressees of the legal norm (here: potential informants) might be influenced in the intended way at significantly lower costs.¹³ Finally, with regard to internal whistleblowing similar effects may be achieved by corporations/employers via implementing a code of conduct or contractual duties that oblige potential informants to report.

There is some empirical analysis from observational data on the effects of incentivizing whistleblowers (see literature section). Unfortunately, observational data often suffers from the informational and the counterfactual problem which makes it impossible to exactly measure the effects of the whistleblower reward program and compare them to alternative incentive designs.¹⁴ By conducting laboratory experiments, our study circumvents this problem. A typical corporate misconduct involves an increase of the wrongdoer’s profit at the expense of others (e.g. employees, shareholders, or the public). Misconduct therefore imposes inequality. In our design we capture this aspect of anti-social behavior, while we dispense with framed scenarios involving fraud for the sake of a cool (= non-suggestive) and parsimonious design. Our experiments reflect all elements in the definition of whistleblowing by Jubb (1999). A player decides whether to increase his profit through misconduct. Others are affected by and learn about the misconduct and can inform the authority. The authority has the option – in case she learns about the misconduct – to punish the wrongdoer. Our focus is to measure the willingness to blow the whistle across different scenarios.

In the first of our two experiments (1 and 2), we focus on incentives to induce whistleblowing. We compare the willingness to blow the whistle in the three of the four strategies described above; that is Reward, Fine and Command. We also compare the impact of different sizes of rewards and fines on the willingness to blow the whistle. Furthermore, our experiments aim at shedding light on whether and, if so, how the willingness to report depends on the way potential informants are affected by the misconduct to be reported. This may reveal the desirability, if not necessity, for regulators and employers implementing whistleblower

¹³ Funk (2007): “First of all, behavior might be changed at lower costs than suggested by the classical ‘Law & Economics’ literature. Merely by passing a law (even if unenforced), citizens might adhere to it.”

¹⁴ As to these problems in general, cf. Apesteguia et al. (2007).

programs to differentiate among potential whistleblowers according to the way they are affected by the misconduct in question.¹⁵

In experiment 2 we study whether and to what extent the willingness to blow the whistle depends on how and whether the authority sanctioning the misconduct (public authority or employer/enterprise) is affected by the reported misconduct. Concededly, as far as the willingness to blow the whistle depends on the expected sanction handed out by the authority (cf. Predictions 7a and 7b), this may be a “non-issue” with regard to external whistleblowing to public authorities since such authorities as well as the individual agents acting on the authorities’ behalf are typically unaffected by the misconduct they are commissioned to punish. However, the results of experiment 2 may shed light on the comparative effectiveness of internal whistleblowing. To illustrate this point, it suffices to recall the examples on corporate scandals mentioned in the introduction. There we see at least in some of the cases that the misconduct in question carried some utility for the corporation in the short run (while being deleterious in the long run). Finally, both experiments provide the opportunity to check on the impact of certain character traits on the willingness to blow the whistle. Our main findings are: Financial incentives and even simple commands increase the probability of whistleblowing. Situational determinants also influence the willingness to blow the whistle: Insiders who are negatively affected by the misconduct are more likely to blow the whistle than non-affected or profiting insiders. Those (negatively affected) victims are also sensitive to the misconduct’s impact on the sanctioning authority: Whistleblowing is more likely if the authority is negatively affected compared to positively or not affected.

2. Related Literature

There is a huge and constantly growing body of literature on whistleblowing. This is true for economics, and even more so for law.¹⁶ First of all, there is some empirical analysis from observational data on the effects of incentivizing whistleblowers. Most notably, the SEC¹⁷ Office of the Whistleblower publishes annual reports on the Dodd-Frank Whistleblower Program. The latest report published at the end of 2015 shows that the number of tips received by the SEC is constantly rising (2015: 3,923). Almost half of those whistleblowers who were granted a reward were current or former company employees. Dworkin and Baucus (2007) and Dyck at al. (2010) analyzed actual law suits where whistleblowing was involved. Dworkin and Baucus (2007) examined 33 cases of employees whose employment was wrongfully terminated for whistleblowing. They compared the characteristics of internal and external whistleblowers

¹⁵ The way people are affected by a certain misconduct may, for example, have a significant impact on their internal motivation to blow the whistle, cf. Feldman and Lobel (2010). Further details on this study *infra* 2.

¹⁶ Cf. the survey of Dasgupta and Kesharwani (2010) and the references given in the comparative study of Thüsing and Forst (2016).

¹⁷ The Securities Exchange Commission (SEC) is the Federal US authority regulation capital markets and enforcing securities law.

and the consequences that ensue after their reporting. Dyck et al. (2010) studied 216 large scale corporate fraud cases. With regard to our experiment, their main finding is that the portion of employee whistleblowers rose significantly where US law provides high monetary rewards under the False Claims Act.

Unfortunately, observational data often suffers from the informational and the counterfactual problem which makes it impossible to exactly measure the effects of the whistleblower reward program and compare them to alternative incentive designs. Nevertheless, it is rather difficult to nearly impossible to carry out investigations into unethical conduct in the workplace by observation. Therefore, many research studies of whistleblowing use self-reported willingness to blow the whistle as explanatory variable (Victor et al., 1993). However, since the correlation between whistleblowing intent and actual whistleblowing seems to be rather low (cf. Mesmer-Magnus and Viswesvaran, 2005), it is worthwhile to focus on revealed rather than stated willingness by means of a more objective experimental design.

Despite that, up to now the number of experimental studies on whistleblowing is comparatively limited.¹⁸ There are different strands of studies, each dealing with whistleblowing in a certain legal context. Breuer (2013) distinguishes between “traitorous whistleblowing” and “watchdog whistleblowing”. While the first type describes reporting misbehavior by cartel members, bribers or corrupt officials who break up conspiracies from the inside, the second one refers to whistleblowing by a third person who is not responsible for the illegal conduct, e.g. an observing bystander. Although our paper contributes to the “watchdog whistleblowing” literature, we start with a short overview on experiments on “traitorous whistleblowing”.

Traitorous whistleblowing

The most prominent studies analyze the effects of leniency programs, i.e. the reduction of fines in case of reporting a cartel, in antitrust law. Two well-known studies, Apesteguia et al. (2007) and Bigoni et al. (2012), investigate whether such leniency programs thwart cartels and promote competition as intended or, to the contrary, stabilize cartels because whistleblowing can be used to punish defecting cartel members. The results are mixed: While in the one-shot game conducted by Apesteguia et al. (2007) leniency reduced cartel formation and lowered market prices in comparison to an antitrust regime only operating with fines¹⁹, adding a reward for the whistleblower comprising of the fines paid by the other cartel members led to high numbers of cartel formation as well as high market prices. In the repeated game conducted by Bigoni et al. (2012), however, leniency resulted in fewer cartels, while the prices of surviving cartels grew compared to traditional antitrust law. Adding a reward for the reporting cartel member lowered the deterrence of cartel

¹⁸ For an overview, see also Breuer (2013).

¹⁹ Cf. also Hinloopen and Soetevent (2008) for similar results in a repeated game.

formation in comparison to leniency, but led to very high reporting rates, thus destabilizing cartels and leading to lower prices.

With regard to bribery and corruption Schikora (2011) has shown in a repeated game setting that leniency for the whistleblower deters bribery, but stabilized the corrupt transactions taking place. Asymmetric leniency in favor of the official partly offset this stabilization effect. Engel et al. (2013) compare symmetric punishment of the briber and the receiver of the bribe with asymmetric punishment favoring the briber. They found that in the latter case more corruption occurs and attributed this to the stabilizing effect of the briber's credible threat to inform the law authorities in case of defection.

Watchdog whistleblowing

Incentives

Our experiment contributes to the literature on behavior of potential whistleblowers who are bystanders, and have therefore not taken part in the illegal conduct. There are several experimental laboratory studies and experimental surveys that analyze the contextual framework that affects whistleblowing behavior.

Feldman and Lobel (2010) contributed significantly to the research on incentivizing whistleblowers by conducting a scenario-based experimental survey. The questionnaires build on the same factual scenario putting the participant in the position of an employee who discovers fraudulent behavior of her employer. What varies is the legal environment, such as protection, high and low reward, duty, and fine. The main findings of the survey are: Participants predict higher reporting action by themselves than by others ("holier than thou effect"). The more severe the participants rate the misconduct, the more likely is their willingness to report. For participants with a high internal motivation to report the willingness to blow the whistle across the different legal environments did not vary much. However, for participants with a low internal motivation to report the willingness to blow the whistle was highest in the environments with high rewards or fines. While Feldman and Lobel (2010), therefore, have unearthed some remarkable results on whistleblowing behavior, the difference in the willingness to report among the various legal environments is rather small and inconclusive with regard to the policy implications. A laboratory experiment that analyzes the effect of reward on whistleblowing behavior is by Breuer (2013). Results show that in the context of tax evasion monetary rewards increase the willingness of bystanders to blow the whistle. The larger the reward, the stronger the increase in reporting turns out. Stikeleather (2016) in his laboratory experiment dealing with a corporate theft scenario also shows an increase in the whistleblowing rate if rewards are granted. He furthermore examines the correlation between the wage level and the whistleblowing rate as well as the conviction of the employee to be morally obliged to blow the whistle. He found a significantly positive correlation in both regards. The higher these latter effects (boosting the internal motivation for blowing the

whistle) are, the lower the net expected economic benefit the employer gains by granting additional financial rewards for whistleblowers.

The theoretical work of Heyes and Kapur (2008) model the question, what an optimal enforcement policy on whistleblowing in terms of responsiveness and the size of penalties should look like given different theories on the motivation of whistleblowers. This work complements our study insofar as it examines a follow-up question on what we are mainly interested in, namely whether and to what extent a potential whistleblower is motivated to report by different regulatory designs as well as by how she is affected by the misconduct.

To summarize, conclusive, experimental evidence on the effect of and differences between Reward, Fine or Command on whistleblowing is still missing. This is our contribution.

Situational and individual determinants

Several studies have analyzed the effect of different situational and individual determinants on whistleblowing. A couple of studies focus on the situation of the potential whistleblower. In particular, they deal with the bystander effect that might affect willingness to blow the whistle. Evidence is mixed. While Choo et al. (2015) in a laboratory experiment and Gao et al. (2015) in an experimental survey find that the number of bystanders increases the willingness to blow the whistle, Robinson et al. (2014), Miceli and Near (1988) and Miceli et al. (1991) find no effect or even a positive effect. Importantly, in all these bystander studies the potential informants are either just observers and not affected by the misconduct or the effect on the potential informant is not explicitly controlled for. In our experiment we measure the willingness to blow the whistle of potential informants who are positively, negatively or not affected by misconduct.

The evidence on the relationship between disposition or personality and the likelihood to blow the whistle (as surveyed by Miceli and Near, 2010) is mixed. Bjørkelo et al. (2010) tested the “Big Five” personality traits and their association with whistleblowing. They found that extraversion is positively related to whistleblowing, while agreeableness is negatively correlated to whistleblowing. There were no significant relations to neuroticism, openness to experience, and conscientiousness. The findings on extraversion and agreeableness go along with anterior empirical evidence (LePine and van Dyne, 2001), although the theory on the relation of agreeableness and whistleblowing is not clear-cut (cf. Miceli and Near, 2010).

There is no clear evidence that locus of control has an effect on the willingness to blow the whistle. Whereas Chiu (2003) finds that agents with internal locus of control will be more likely to blow the whistle compared to agents with external locus of control, Miceli et al. (1991) do not find this effect.

In an experimental study Waytz et al. (2013) tested how the valuation of fairness on the one hand and loyalty on the other hand influence the willingness to report unethical behavior. They find that valuing

fairness/justice over loyalty promotes the willingness to report. Manipulating participants to endorse fairness/justice led to a higher whistleblowing rate compared to manipulating them to endorse loyalty. Waytz et al. (2013) infer from these results that the willingness to report is the result of a fairness-loyalty trade-off.

In their survey described above Feldman and Lobel (2010) find that women are more likely to blow the whistle than men. However, the overall evidence on such gender differences is mixed (cf. the literature surveyed in Miceli and Near, 1992 and Miceli et al., 2008). Also, there is some theory (summarized by Miceli et al., 2008), suggesting that men are more likely to blow the whistle than women: This theory points to the difference in positions and professions occupied by men and women as well as to gender differences in the willingness to conform with a majority opinion.

Few studies deal with the character of the sanctioning institution, i.e. the authority. An exception is the survey by Gao et al. (2015). They show that the willingness to blow the whistle is higher when the authority is external compared to internal. Evidence on how whistleblowing is affected by whether and how the authority is affected by the misconduct in question is absent in the literature, although there is evidence that punishment differs when the authority is not affected compared to being negatively affected. Fehr and Fischbacher (2004) show that unaffected agents will still punish misbehavior, but significantly less than negatively affected agents. If willingness to blow the whistle depends on expected punishment by the authority for the wrongdoer it might therefore be crucial whether authority is affected by the misconduct or not.

Beyond the aspects already mentioned, our experimental study furthermore adds to this literature and data in the following respects: Since in our experiments the perceived misconduct is very unspecific, because it is based on the presumed preference of the participants in favor of an equal distribution of resources, we are able to derive more general conclusions on the relative effect of incentives on potential whistleblowers that are detached from a specific kind of illegal conduct. Furthermore, there are subgroups of potential whistleblowers that benefit or suffer from misconduct. Therefore, we can compare the willingness of (potential) whistleblowers who are affected by the misconduct and of whistleblowers who are not. Also, our study analyzes potential whistleblowers reactions to (un)affected authorities. We can compare willingness to blow the whistle to an authority who is affected by the misbehavior and to an authority which is not.

3. Experimental Design

We conducted two stand-alone laboratory experiments. Experiment 1 focuses on the effect of financial incentives on whistleblowing. Experiment 2 deals with the status of the punishing authority.

A typical corporate misconduct involves increasing the delinquent’s profit at the expense of others (e.g. employees, shareholders, or the public). Misconduct therefore imposes inequality. We capture this aspect of anti-social behavior by letting a participant decide whether to keep a status quo allocation (which imposes the same profit on all participants) or to switch to an unequal allocation that increases his profit. All other participants except one learn about the decision and can inform the ignorant participant who can reduce the wrongdoer’s profit. In order to minimize experimenter demand effects, the experiments were neutrally framed and we avoided all suggestive expressions. However, we interpret the selection of the unequal allocation as *Misconduct*, the transfer of information to the ignorant participant as *Whistleblowing*, and the reduction of participant A’s profit as *Punishment*. We will also label the three stages of the game accordingly throughout the paper.

Stage 1: Misconduct There are five participants A-E in experiment 1 and seven participants A-G in experiment 2. All participants receive an endowment of 100 points. Participant A can decide whether to keep this status quo X or to misconduct and switch to allocation Y which affects participants differently. Y yields a higher payoff for participants A and B (and F in experiment 2), a lower payoff for participant C (and G in experiment 2), and the same payoff for participants D and E. See Table 1 for a detailed description of payoffs. Note that while allocation Y is efficiency-stable in experiment 1, it is efficiency-decreasing in experiment 2.

| Participant | Allocation X | Allocation Y |
|-------------|--------------|--------------|
| A | 100 | 120 |
| B | 100 | 120 |
| C | 100 | 60 |
| D | 100 | 100 |
| E | 100 | 100 |
| F* | 100 | 120 |
| G* | 100 | 60 |

Table 1: Allocation X (status quo) and Y (misconduct), *: Participants F-G only in Experiment 2.

Stage 2: Whistleblowing All Participants except one learn about the decision of participant A. In experiment 1 participant E remains ignorant, in Experiment 2 (depending on the treatment) either participant E, F, or G remains ignorant. However, the knowing participants can inform him. To model psychological or other costs imposed on the whistleblower, reporting is costly. It costs 6 points. To avoid

the bystander effect (e.g. to save reporting costs), only one knowing participant is randomly selected. If he decided to report, he pays the reporting costs and the ignorant participant is informed about participant A's decision. If he decided not to report, he does not have to pay the reporting costs, and the ignorant participant remains uninformed.

Stage 3: Punishment If the ignorant participant learns about participant A's decision, he assumes the role of authority, i.e. he can punish participant A by deducting points. Punishment is costless.

Experiment 1: Incentives for Whistleblowing

To analyze the effect of incentives on whistleblowing we conducted six treatments in a between-subject design. The *No Incentives* treatment was conducted as described. In *High Reward* the randomly selected participant in stage 2 paid the reporting costs of six points and additionally received a reward of 50 points in case he reported the decision. In *Low Reward* he additionally received ten points. In *High Fine* the randomly selected participant in stage 2 paid the reporting costs of six points in case he reported the decision and paid a fee of 50 points in case he did not report the decision. In *Low Fine* he paid ten points. In *Command* there were no additional financial incentives, but all participants were informed in the instructions that "participants B-D are obliged to inform participant E if participant A chose allocation Y". Table 2 supplies an overview on the treatments.

Experiment 2: Status of Sanctioning Authority

In experiment 2 we study the comparative effectiveness of internal whistleblowing. More concretely, we examine whether and to what extent the willingness to blow the whistle depends on whether and how the authority sanctioning the misconduct is affected by the reported misconduct. Experiment 2 consists of three treatments, again in a between-subject design. . In the *Not Affected* treatment the authority is not affected and remains with her endowment of 100 points. In *Positive* the authority is positively affected and receives 120 points in allocation Y. In *Negative* the authority is negatively affected and receives 60 points in allocation Y. To keep total payoffs after misconduct constant across treatments, we added two further players F and G to the experimental design of experiment 1 and varied which participant was in the role of authority. Participant E, F, and G were in the role of authority in *Not Affected*, *Positive*, and *Negative*, respectively. There were no additional incentives for whistleblowing, just as in the *No Incentives* treatment above. Table 2 supplies an overview on the treatments.

Procedures

To obtain all relevant decisions, we applied the strategy method by Selten (1965). Participants decided whether to report participant A's behavior for both allocation choices X and Y. Punishment was elicited for both allocation choices, too. We also elicited participants' beliefs on others' behavior. Beliefs were

incentivized with two additional points for a correct answer. For experiment 1 we conducted 19 sessions between April 2015 and July 2016. All sessions were conducted at the LERN (University of Erlangen-Nuremberg) with a total number of 505 participants (64% female). For experiment 2 we conducted 17 sessions in June 2015 with a total number of 364 participants (49% female). 5 sessions were conducted at the LERN (University of Erlangen-Nuremberg) with a total number of 140 participants (64% female). 13 sessions were conducted at the experimentUM (TU Munich) with a total number of 224 participants (45% female). For both experiments one point translated into 0.05 euros. Experiments took about 35 minutes, average income of participants was 9.02 euros (10.01 \$ at that point in time) including a show-up fee of 4 euros (4.45\$). The experiments were programmed with z-Tree (Fischbacher, 2007). We recruited participants using the online recruiting system ORSEE (Greiner, 2004). Each subject sat at a randomly assigned PC terminal and was given a copy of instructions.²⁰ A set of control questions was provided to ensure the understanding of the game. The experiment did not start until all subjects had answered all questions correctly. The whole procedure as well as the structure of the game was common knowledge. At the end of the experiment participants filled out a questionnaire including questions on gender, sociodemographics, personality traits and locus of control.

| | Treatment | Description | # participants | # groups |
|-------------------------------------|------------------|---|-----------------------|-----------------|
| Experiment 1 (Incentives) | No Incentives | | 85 | 17 |
| | Low Reward | Reporting the misconduct yields a reward of 10 points | 80 | 16 |
| | High Reward | Reporting the misconduct yields a reward of 50 points | 90 | 18 |
| | Low Fine | Not reporting the misconduct results in a fine of 10 points. | 75 | 15 |
| | High Fine | Not reporting the misconduct results in a fine of 50 points. | 90 | 18 |
| | Command | 'If participant A chose Y participants B-D are required to inform participant E.' | 85 | 17 |
| Experiment 2 (Authority) | Not Affected | The authority is not affected by the misconduct. | 119 | 17 |
| | Negative | The authority is negatively affected by the misconduct. | 126 | 18 |
| | Positive | The authority is positively affected by the misconduct. | 119 | 17 |

Table 2: Treatments of the experimental design

²⁰ A translation of the instructions can be found in the appendix.

4. Predictions

Prediction 1: *There will be whistleblowing in No Incentives.*

Prediction 2a: *Whistleblowing is more likely for disadvantaged players compared to neutral and advantaged players.*

Prediction 2b: *Whistleblowing is more likely for neutral players compared to advantaged players.*

Misconduct results in inequality between participants. Participant A's profit after misconduct is still equal to the profit of participants who are positively affected by the misconduct (advantaged), but larger than the profit of participants who are not affected (neutral) or negatively affected (disadvantaged). Assume people have standard preferences as in Fehr and Schmidt (1999). This means they dislike inequality, in particular they dislike others having more than they have (negative inequality aversion α) or less than they have (positive inequality aversion β). Importantly, negative inequality aversion is larger than positive inequality aversion ($\alpha > \beta$).

For our experiment, this means that neutral or disadvantaged participants would like to see the inequality through misconduct removed or (because $\alpha > \beta$) at least reversed. From a potential whistleblower's perspective, the only way to let this happen, is to blow the whistle such that authority can deduct points from participant A. Although the potential whistleblower's profit is not affected by punishment directly, the difference between his payoff and participant A's payoff decreases. Hence, the potential whistleblower's utility increases.

The participants' gain in expected utility through expected punishment and, therefore, the willingness to blow the whistle depends on the individual Fehr-Schmidt parameters and the whistle-blowers type, i.e. the difference between payoffs of participant A and the potential whistleblower. Thus, we expect disadvantaged players to blow the whistle more often than neutral players. Punishment cannot increase the utility of potential whistleblowers who profit from the misconduct (advantaged). We therefore do not expect them to blow the whistle without any further incentives.

Reciprocal behavior leads to similar predictions. Negative reciprocity will make whistleblowing more likely for disadvantage than for neutral players, whereas positive reciprocity will make whistleblowing less likely for advantaged players compared to neutral players.

Prediction 3: *Compared to No Incentives whistleblowing is as likely or more likely in Low and High Fine as well as in Low and High Reward.*

In comparison to the *No Incentives* treatment abstaining from whistleblowing in the *Reward* and *Fine* treatments comes with (opportunity) costs largely exceeding reporting costs. According to the standard homo oeconomicus theory we therefore expect the whistleblowing rates to not decrease in these two treatments.

Prediction 4: Potential informants will always blow the whistle in the *Reward* and *Fine* treatments.

In all four treatments whistle blowing is the payoff-maximizing strategy. Assuming rational profit-maximizing actors (*homines oeconomici*) we expect the whistleblowing rate to be 100% in the *Fine* treatments as well as in the *Reward* treatments. While De Geest and Dari-Mattiacci (2013) in their analysis on the comparative effects of carrots (*Reward*) and sticks (*Fine*) on incentivizing agents rightly point out that risk aversion may result in a comparatively stronger incentive effect when using sticks, we do not expect this factor to have any influence on the outcomes in our treatments. This expectation is based on the fact that the expected utility of reporting is always larger than the expected utility of not reporting. Even if we relax the assumption of actors exclusively interested in profit maximization and acknowledge a dampening effect of external rewards on the intrinsic motivation to blow the whistle (crowding-out effect; cf. Feldman and Lobel 2010), we do not expect the whistleblowing rate to fall below 100% in the *High Reward* treatment. This expectation is based on the assumption that the reward size in *High Reward* is sufficiently large to (over)compensate the crowding-out effect (cf. Walton 2012). There might be crowding out in *Low Reward*, though.

Prediction 5: Whistleblowing will not increase with the size of the *Fine* or *Reward*.

We expect the willingness to blow the whistle to be as high in the *High Fine/High Reward* as in the *Low Fine/Low Reward* treatment. This prediction follows from our explanation of prediction 4: Assuming rational profit-maximizing participants (*homines oeconomici*) the comparatively higher expected payoff should lead to a whistleblowing rate of 100 %. This reasoning does not depend on the size of the gain in the expected payoff as long as there is any. Or to put it differently: Since there is no gain to be had from not blowing the whistle in the *Fine* treatments a higher fine does not have a more deterrent effect on not blowing the whistle than a lower fine as long as the fine amounts at least to 7 points.²¹ The same is true for the incentivizing effect of a reward as long as the reward size amounts at least to 7 points.

If we relax the homo oeconomicus assumption and acknowledge the existence of a crowding-out effect there may be a size effect with regard to the *Reward* treatments. This is because the reward size in the *Low Reward* treatment may be too low to compensate the crowding-out effect. This may, in turn, even decrease

²¹ Cf. Becker (1968) who introduced the idea to standard economic theory that a fine reduces the expected utility of the sanctioned behavior for the rational actor (*homo oeconomicus*) the more the higher the fine is.

the whistleblowing rate in *Low Reward* compared to *No Incentives* which may, thus, be below 100% (cf. Feldman and Lobel 2010; Walton 2012). Since we expect a 100% whistleblowing rate in the *High Reward* treatment under the same assumptions (cf. explanation to Prediction 4), there may be an increase in the whistleblowing rate in *High Reward* compared to *Low Reward*.

Prediction 6: *Whistleblowing is more likely in Command than in No Incentives.*

According to the standard homo oeconomicus model, that is, assuming a participant only interested in the size of her payoff, there should be no significant differences in the *Command* treatment from the *No Incentives* treatment with regard to the reporting rate. However, as mentioned in the introduction, if we relax the assumptions of the standard model *Command* may incentivize participants due to the preference-shaping property of stating a norm in favor of whistleblowing or its “expressive function”, respectively.

Prediction 7a: *When the authority is positively affected by the misconduct compared to when it is negatively or not affected whistleblowing is less likely for disadvantaged players and neutral players.*

Prediction 7b: *When the authority is negatively affected by the misconduct compared to when it is not affected whistleblowing is more likely for disadvantaged players and less likely for neutral players.*

Players will blow the whistle, i.e. enable authority to punish participant A if their gain in expected utility through punishment exceeds their reporting costs. From the point of view of the authority punishment will be higher if authority is negatively affected by the misconduct than when it is not affected or even positively affected (again assuming Fehr-Schmidt preferences). More specifically, a negatively affected authority has to deduct 60 points to restore equality between player A and itself. An authority that is not affected by misconduct has to deduct 20 points to restore equality between player A and itself. A positively affected authority will not punish at all, because although it might suffer from inequality (with respect to other players) it cannot restore equality by punishing player A since payoffs are equal between player A and the authority.

As a consequence, potential informants who face a positively affected authority and, furthermore, anticipate the aforementioned punishment behavior will not spend the reporting costs to blow the whistle. Whistleblowing is more likely if potential informants face a negatively affected authority or an authority that is not affected: On the one hand, advantaged players will also refrain from blowing the whistle to a negatively or not affected authority because their utility cannot increase by punishment. On the other hand, disadvantaged players have the highest expected utility through punishment if the authority is negatively affected because punishment will not only remove inequality for the authority but also for the disadvantaged players. An authority that is not affected will at least mitigate inequality. Thus, for disadvantaged players

whistleblowing will be more likely in *Negative* than in *Not Affected*, more likely in *Negative* than in *Positive*, and more likely in *Not Affected* than in *Positive*.

Neutral players have the highest expected utility through punishment when the authority is not affected because punishment will not only remove inequality for the non-affected authority but also for all neutral players. Punishment of a negatively affected authority will reverse inequality between player A and all neutral players. For neutral players punishment of a negatively affected authority still has a positive effect on expected utility but punishment might be too high for the taste of some neutral players which might make them refrain from blowing the whistle. To sum up: For neutral players whistleblowing will be less likely in *Negative* than in *Not Affected*, but more likely in *Negative* than in *Positive* and more likely in *Not Affected* compared to *Positive*.

5. Results

Unless noted otherwise, the reported test statistics refer to a two-sided Fisher's exact test.

Experiment 1 - Incentives for Whistleblowing

We turn towards the results of the incentives treatments in experiment 1 first. Table 3 gives descriptive statistics of participants' behavior, in particular participants' B-D willingness to blow the whistle. For completeness we also report the frequencies of misconduct by participant A and the amount of punishment assigned to participant A for choosing X and Y by participant E across treatments.

Table 3 gives two pieces of evidence that choosing allocation Y is indeed perceived as inappropriate behavior in all treatments. First, the vast majority of participants A refrains from choosing Y. Second, punishment is significantly higher for allocation Y than for allocation X (Wilcoxon signed-rank test, $p < .01$).

Table 4 shows that although few participants A chose the misconduct (Y), it is perceived as a realistic event by most potential whistleblowers. This perception secures that behavior of potential whistleblowers is crucial.

Result 1: *There is whistleblowing in No Incentives.*

Evidence for Result 1 Even in the *No Incentives* treatment with no additional incentives, 33% of potential whistleblowers blow the whistle on average (See Table 3). The frequency is significantly different from zero (binomial test, $p < .01$).

Result 2: *Whistleblowing is more likely for disadvantaged players compared to neutral and advantaged players in No Incentives.*

Evidence for Result 2 Table 5 show that the majority of players who blow the whistle in *No Incentives* are negatively affected by the misconduct. 71% of disadvantaged players blow the whistle, whereas only 18 % of neutral players and 12 % of advantaged players do so. Disadvantaged players are significantly more likely to blow the whistle than neutral players and advantaged players. We do not find that the willingness to blow the whistle differs between neutral and disadvantaged players. See Table 6.

Result 3 *Compared to No Incentives, whistleblowing is more likely in Low Fine, High Fine, Low Reward, and High Reward.*

Result 4 *Whistleblowing is more likely in Low and High Fine than in Low and High Reward.*

Evidence for Result 3-4 All four treatments increase the willingness to blow the whistle on average (see Table 3). Paying whistleblowers or fining participants who refrained from blowing the whistle significantly increased whistleblowing compared to *No Incentives* ($p < .01$ for all treatments). Furthermore, fines work better than rewards, on average.²²

Result 5a: *Advantaged players' willingness to blow the whistle increases with the size of the reward.*

Result 5b *Financial incentives mitigate the difference in willingness to blow the whistle between player types.*

Evidence for Result 5a & 5b The size of the reward or fine does not have a significant effect on the willingness to blow the whistle on average²³. However, treatment effect sizes strongly differ across types. Table 5 shows that compared to the *No Incentives* treatment, all types are significantly more likely to blow the whistle when offered a *High* or *Low Reward* or threatened by a *High* or a *Low Fine*.²⁴ Whereas neutral and disadvantaged players' willingness to blow the whistle does not differ across the financial incentives treatments, advantaged players' willingness is affected. For advantaged players fines work better than rewards ($p < .01$) and high rewards are more effective to induce whistleblowing than low rewards ($p < .1$). We do not find evidence that players' willingness to blow the whistle is affected by the size of the fine, though.

²²*High and Low Fine vs. High and Low Reward:* $p < .05$.

²³*High Reward vs. Low Reward:* $p = .121$; *High Fine vs. Low Fine:* $p = .699$.

²⁴*Low Reward vs. No Incentives:*

advantaged: $p < .1$; neutral: $p < .01$; disadvantaged: $p < .05$;

High Reward vs. No Incentives:

advantaged: $p < .01$; neutral: $p < .01$; disadvantaged: $p < .1$;

Low Fine vs. No Incentives:

advantaged: $p < .01$; neutral: $p < .01$; disadvantaged: $p = .178$;

High Fine vs. No Incentives:

advantaged: $p < .01$; neutral: $p < .01$; disadvantaged: $p < .05$;

Whereas willingness to blow the whistle for disadvantaged players is already very high without incentives (71%) and is not strongly affected by additional incentives, willingness of neutral and advantaged players increases if there is a reward or fine (see result 4) and adjusts almost to the level of disadvantaged players (see Table 5 and Table 6): We do not find differences across types' willingness to blow the whistle when the reward is high or when there are low or high fines; in these treatments whistle rates are very high for all types of potential informants (between 0.78 and 1.00). However, since the low reward has only a weak effect on advantaged players' willingness to blow the whistle, disadvantaged players and even neutral players are more likely to blow the whistle than advantaged players when the reward is low.

Result 6: *Whistleblowing is more likely in Command compared to No Incentives.*

Evidence for Result 6: A – compared to financial incentives – weaker, but still significant positive impact on whistleblowing is caused by *Command*, where we find 55% of participants willing to blow the whistle ($p < .05$). Besides, advantaged players are significantly more likely to blow the whistle in *Command* ($p < .1$). For neutral and disadvantaged players the positive treatment effect of *Command* fails to be significant. In *Command*, disadvantaged players are more likely to blow the whistle than neutral players ($p < .1$).

Table 7 gives the results of a probit regression analysis that adds several controls to extend and control for the robustness of the non-parametric test results above. Column 1 confirms the general treatment effects. Compared to the *No Incentives* treatment, *High Fine* and *High Reward*, *Low Fine* and *Low Reward*, and *Command* increase the willingness to blow the whistle – with decreasing effect size. Column 2 controls for the situational determinants of potential whistleblowers. It confirms that disadvantaged players are more likely and advantaged players are less likely to blow the whistle than neutral players. Column 3 adds further controls.²⁵ Results are robust. Furthermore, we find that men are significantly more likely to blow the whistle than women. Whistleblowers are significantly more likely to believe others will blow the whistle, too ($p < .01$). Other additional controls do not have a significant effect on the willingness to blow the whistle.

²⁵ Controls are age and gender of the potential whistleblower, measures of the Big 5 personality traits, internal and external loci of control measures, and beliefs on other players' behavior.

| | Frequency of misconduct | Frequency of whistleblowing | Punishment for misconduct (Y) | Punishment for X |
|---------------|-------------------------|-----------------------------|-------------------------------|------------------|
| No Incentives | .06 | .33 | 66.41 (46.15) | 15.23 (33.41) |
| Low Reward | .19 | .77 | 34.38 (32.24) ²⁶ | 6.69 (24.92) |
| High Reward | .33 ²⁷ | .89 | 42.33 (43.16) | 7.22 (24.21) |
| Low Fine | .13 | .91 | 58.13 (42.80) | 15.10 (31.43) |
| High Fine | .11 | .94 | 55.56 (40.00) | 9.17 (25.68) |
| Command | .12 | .55 | 53.53 (37.74) | 6.23 (24.21) |

Table 3: Descriptive statistics on average behavior for experiment 1 (incentives)

| | Frequency of misconduct | Frequency of others' average willingness to blow the whistle | Punishment for misconduct (Y) |
|---------------|-------------------------|--|-------------------------------|
| No Incentives | .51 | .39 | 53.57 (41.54) |
| Low Reward | .48 | .77 | 52.92 (38.08) |
| High Reward | .50 | .75 | 54.43 (43.78) |
| Low Fine | .47 | .84 | 42.67 (31.36) |
| High Fine | .29 | .94 | 60.56 (42.62) |
| Command | .41 | .58 | 62.86 (43.33) |

Table 4: Descriptive statistics on average beliefs of potential whistleblowers in experiment 1

| | Advantaged | neutral | disadvantaged |
|---------------|------------|---------|---------------|
| No incentives | 0.12 | 0.18 | 0.71 |
| Low reward | 0.44 | 0.88 | 1.00 |
| High Reward | 0.78 | 0.94 | 0.94 |
| Low fine | 0.87 | 0.93 | 0.93 |
| High Fine | 0.94 | 0.89 | 1.00 |
| Command | 0.47 | 0.41 | 0.76 |

Table 5: Whistle blow frequencies across types

²⁶ In *Low Reward* punishment is significantly lower compared to *No Incentives* (Wilcoxon rank-sum, $p < .05$).

²⁷ In *High Reward* the likelihood of misconduct is significantly higher compared to *No Incentives* ($p < .1$).

| | advantaged vs. neutral | advantaged vs. disadvantaged | neutral vs. disadvantaged |
|---------------|---------------------------|---------------------------------|------------------------------|
| No Incentives | 1.000 | .001*** | .005** |
| Low Reward | .023** | .001*** | .482 |
| High Reward | 1.000 | 1.000 | 1.000 |
| Low Fine | .338 | .338 | 1.000 |
| High Fine | 1.000 | 1.000 | .486 |
| Command | 1.000 | .157 | .080* |

Table 6: p-values of Fisher's exact test for differences in whistle blow frequency across types, * p<0.01, ** p<0.05, * p<0.1.**

| | (1) | (2) | (3) |
|------------------|---------------------|---------------------|---------------------|
| Low Reward | 0.241*** (0.040) | 0.238*** (0.038) | 0.192*** (0.040) |
| High Reward | 0.301*** (0.038) | 0.289*** (0.038) | 0.240*** (0.040) |
| Low Fine | 0.296*** (0.035) | 0.279*** (0.036) | 0.241*** (0.038) |
| High Fine | 0.335*** (0.036) | 0.321*** (0.037) | 0.283*** (0.039) |
| Command | 0.140*** (0.054) | 0.145*** (0.050) | 0.099* (0.053) |
| Advantaged | | -0.105* (0.063) | -0.104* (0.063) |
| Disadvantaged | | 0.219*** (0.050) | 0.242*** (0.048) |
| Male | | | 0.090* (0.053) |
| Further controls | NO | NO | YES |
| Observations | 303 | 303 | 303 |
| Pseudo R2 | .2194 | .3078 | .3685 |

Table 7: Probit regression, reporting marginal effects, dependent variable: whistle blow after misconduct, omitted variable in column 1: no incentives, in column 2-3: no incentives and neutral, * p<0.01, ** p<0.05, * p<0.1.**

Experiment 2 - Status of the sanctioning authority

Experiment 2 analyzes whether potential whistleblowers react to how the sanctioning authority is affected by the misconduct. To begin with, we see similar behavioral patterns as in experiment 1 (Table 8). First, the majority of participants A refrains from choosing allocation Y. Second, the average punishment for allocation Y is higher than for allocation X (Wilcoxon signed-rank test, $p < .01$). Again, these two findings support that choosing allocation Y is perceived as misbehavior. Third, potential whistleblowers once more see allocation Y as a realistic event. About 50% expect participant A to choose allocation Y (see Table 9). Whistleblowing is different from zero in all treatments (binominal test, $p < .01$). Besides, there is a higher willingness to blow the whistle for disadvantaged players compared to neutral ($p < .01$) or advantaged players ($p < .01$) in all treatments (see Table 10). This confirms result 1 and result 2 for whistleblowing after an efficiency-decreasing misconduct.

Result 7 *Disadvantaged players are more likely to blow the whistle when the authority is negatively affected by misconduct compared to not affected or positively affected.*

Evidence for Result 7 As predicted, average punishment for participant A by the sanctioning authority (see Table 8) is lower when the authority is positively affected by the misconduct (distribution Y) than when it is negatively or not affected (Wilcoxon rank-sum, $p < .01$ each). Average punishment is higher when the authority is negatively affected compared to not affected by the misconduct, although this difference is not significant (Wilcoxon rank-sum, $p = .38$).

As actual punishment, average punishment beliefs by the potential whistleblowers (see Table 9)²⁸ are lower when the authority is positively affected compared to negatively affected or not affected (Wilcoxon rank-sum test, $p < .01$, each). Besides, it is higher when the authority is negatively affected compared to not affected (Wilcoxon rank-sum test, $p < .05$).

Table 8 shows that whistleblowing on average is in fact more likely when the authority is negatively affected compared to positively or not affected. Treatment effect sizes and significance varies across player types (see Table 10 and 11): Disadvantaged players are significantly more likely to blow the whistle when the punishment authority is negatively affected by misconduct compared to not affected or positively affected. Neutral and advantaged players also blow the whistle slightly more often when the authority is negatively affected by misconduct, but the treatment difference fails to be significant.

²⁸ Expected punishment of potential whistleblowers is quite accurate for negatively or not affected authorities, but clearly overstating for positively affected authorities.

Results are robust to adding controls²⁹ as confirmed by a probit regression analysis (see Table 12). Here, a negatively affected authority not only affects disadvantaged players, but all types on average (as in column 1). Again, we find that men are more willing to blow the whistle than women. This is true especially for disadvantaged players. As in experiment 1 whistleblowers are significantly more likely to believe others will blow the whistle, too ($p < .01$). Other additional controls do not have a significant effect on the willingness to blow the whistle.

| | Frequency of misconduct | Frequency of whistleblowing | Punishment for misconduct (Y) | Punishment for X |
|--------------|-------------------------|-----------------------------|-------------------------------|------------------|
| Negative | .35 | .34 | 73.29 (36.80) | 11.76 (27.39) |
| Not Affected | .33 | .30 | 61.94 (37.23) | 2.78 (11.50) |
| Positive | .33 | .26 | 7.71 (17.37) | 9.47 (13.51) |

Table 8: Descriptive statistics on behavior in experiment 2 (authority)

| | Frequency of misconduct | Frequency of others' average willingness to blow the whistle | Punishment for misconduct (Y) |
|--------------|-------------------------|--|-------------------------------|
| Negative | .49 | .30 | 74.59 (43.22) |
| Not Affected | .40 | .47 | 57.34 (42.72) |
| Positive | .53 | .35 | 45.42 (43.29) |

Table 9: Descriptive statistics on beliefs of potential whistleblowers in experiment 2 (authority)

| | Advantaged | Neutral | Disadvantaged |
|--------------|------------|---------|---------------|
| Negative | .18 | .29 | .76 |
| Not Affected | .19 | .22 | .44 |
| Positive | .12 | .18 | .41 |

Table 10: Whistleblowing frequencies across types in experiment 2 (authority)

| | Advantaged | Neutral | Disadvantaged |
|---------------------------|------------|---------|---------------|
| Negative vs. Not Affected | 1.000 | .746 | .040** |
| Negative vs. Positive | .703 | .392 | 0.021** |
| Positive vs Not Affected | .701 | .723 | .813 |

Table 11: p-values of Fisher's exact test for differences in frequency of whistleblowing across treatments in experiment 2 (authority), * $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.**

²⁹ Same controls as in experiment 1, see Table 7. We also controlled for subject pool effects since experiment 2 was conducted in two distinct laboratories.

| | (1) all types | (2) advantaged | (3) neutral | (4) disadvantaged |
|-------------------------|-------------------------|--------------------------|-----------------------|-----------------------------|
| Positive | 0.012 (0.077) | -0.068 (0.088) | 0.046 (0.117) | 0.156 (0.158) |
| Negative | 0.213** (0.084) | 0.063 (0.099) | 0.060 (0.117) | 0.650*** (0.118) |
| Advantage | -0.098 (0.070) | | | |
| Disadvantage | 0.346*** (0.080) | | | |
| Male | 0.144** (0.062) | 0.045 (0.091) | 0.062 (0.088) | 0.380*** (0.134) |
| Further controls | Yes | Yes | Yes | Yes |
| Observations | 260 | 87 | 86 | 87 |
| Pseudo R2 | .2112 | .1906 | .2752 | .3377 |

Table 12: Probit regression, reporting marginal effects, dependent variable: whistle blow after misconduct, omitted variable in column 1: not affected and neutral, in column 2-4 not affected, * p<0.01, ** p<0.05, * p<0.1.**

6. Discussion

Incentives for whistleblowing

Our evidence confirms Prediction 1 stating willingness to blow the whistle in *No Incentives*. *Reward*, *Fine*, and *Command* manipulations significantly increase the likelihood of potential informants to blow the whistle in comparison to the *No Incentives* treatment. This result accords with our Predictions 3 and 6. While higher willingness to blow the whistle in the *Reward* and *Fine* treatments conforms theoretically with a behavioral model suggesting that agents are only interested in final monetary payoffs, the significant increase of whistleblowing in *Command* compared to *No incentives* does not. This result supports the theory of a preference-shaping function of law. With regard to an expressive function of law, the results are, however, not clear-cut, since the subjects of our experiment acted anonymously in a one-shot game. Thus, the appeal to social norms cannot have influenced the subjects' behavior in a direct way, that is by tying a reputational loss or other social sanction to the concrete action of non-reporting. That does not exclude an indirect influence of such social norms, inducing subjects to comply with common rules in general, because they otherwise feel shame or a loss in self-esteem (cf. Sunstein, 1996). This, however, presumes that agents have already formed a preference to comply with given rules, when they enter the experiment.

Furthermore, the results show clearly, that *Command* is not as effective in inducing participants to blow the whistle as *Fine* or *Reward*, both of which offer monetary incentives. For neutral and disadvantaged players this weaker effect even fails to be significant, although with regard to disadvantaged players this is probably due to the already very strong willingness to blow the whistle in the *No Incentives* treatment.

At first glance, one might therefore be inclined to discard the *Command* strategy from a policy perspective altogether. This would, however, be premature, since a regulator or corporation adopting a *Command* strategy to induce whistleblowing may at least raise the whistleblowing rate modestly at nearly no cost. This tilts the cost-benefit calculus of regulatory strategies to induce whistleblowing in favor of *Command*. Furthermore, in the long run *Command* strategies might have the potential to change the cultural attitudes of people towards whistleblowing including those who are not in the position of potential whistleblowers. This, in turn, may reduce the costs of whistleblowing with regard to the threat of retaliation, ostracism and the like significantly.

In our experiment, whistleblowing is more likely in the *Fine* treatments than in the *Reward* treatments. The whistleblowing rate falls short of 100% in both, the *Fine* and the *Reward* treatments, even though it is close to 100% except for *Low Reward*. Most notably, the whistleblowing rate in *Low Reward* is considerably lower than in the other three treatments. These results go against our prediction 4. They are, however, in line with the results of Sutter et al. (2010) with regard to their findings on the size of the participants' contributions in the treatments where punishment and reward institutions are imposed exogenously. At first sight, this result suggests favoring a regulatory strategy that fines/sanctions potential informants who know of the misconduct in question but nevertheless stay silent. The lack of enforcement and administration costs ex post in case the potential informant complied and blew the whistle equally argues in favor of a stick (*Fine*) over a carrot (*Reward*) strategy (assuming a high compliance rate). However, at a second glance this conclusion may be premature for the following reason: In our experiment, the probability of being "detected" is the same for participants B-D who refrain from blowing the whistle in the *Fine* and *Reward* treatments. In reality, however, regulation on incentivising whistleblowers is typically considered where the enforcement authority (public authority or employer) lacks the information to sanction misconduct. Thus, the probability of detecting the "silent bystander" is very low. In such scenarios, therefore, reward schemes appear to be the superior regulatory option to induce an insider to share her private information about misconduct with enforcement authorities.

We do not see a strong crowding-out effect in the *Reward* treatments as described in Feldman and Lobel (2010). To the contrary, we observe an increase in the likelihood of potential informants to blow the whistle for high and low rewards. This may be due to the fact that even in *Low Reward* the loss in intrinsic motivation is overcompensated by the utility gained through the reward. To put it differently, our results do not rule out a crowding-out effect in case the reward were set (even) lower.

In accord with Prediction 5, we do not find a significant increase in the whistleblowing rate on average depending on the size of reward and fines. That means, that the expected utility gain from the *High Fine/Reward* in comparison to *Low Fine/Reward* is – on average – not strong enough to raise the

whistleblowing rate. One reason is that whistle blowing rates are already very high in *Low Fine* and in *Low Reward* (especially for disadvantaged players) which does not leave much room for increase. Prima vista, this appears to suggest a regulation strategy that confines itself to rather modest financial incentives, at least as a first step. This conclusion is seemingly supported by the consideration that – in the real world – too high rewards or fines may result in too many low-quality tips by whistleblowers who fear being fined or are (over)eager to get a reward. Such low-quality tips put a strain on the sparse resources of enforcement authorities and distract them from pursuing high-quality tips. However, we refrain from jumping to such a conclusion since the perception of reward and fine sizes may be highly context-dependent and subjective. Furthermore, we actually do observe a significant increase with regard to *High Reward* in comparison to *Low Reward* for the subset of advantaged participants, and it can be assumed that a regulator also aims at effectively incentivizing such bystanders profiting from the misconduct in question. Lastly and, in our view, most importantly, in the real world whistleblowing may often come with staggering costs for the informant due to retaliatory measures (cf. Introduction), which are not captured in our experimental design. To compensate for these costs, a rather high reward seems to be indicated. This appears to be true at least as long as there are no protective measures in place that effectively neutralize or frustrate harmful retaliation.

Situational determinants of whistleblowing

As to the likelihood of blowing the whistle across types, our finding that whistleblowing is more likely for disadvantaged players compared to neutral and advantaged ones (Result 2) in *No Incentives* conforms with Prediction 2a. As a consequence to the very high whistleblowing rate of disadvantaged players in *No Incentives* the increase in the likelihood to blow the whistle for disadvantaged players due to additional incentives is relatively low. With regard to *Command* it even fails to be significant (cf. Results 5a & 5b). From a policy perspective, these results suggest that whistleblower incentive schemes may not be necessary with regard to the victims of the misconduct in a wide range of scenarios. This suggestion conforms with the observation that, in reality, the victim of the misconduct (= the disadvantaged party) is not the primary addressee of incentive schemes to further whistleblowing anyway. From a legal point of view this can be explained as follows: If the misconduct is not concealed from the victim (open misconduct) or the victim otherwise knows about the misconduct he generally has sufficient incentives to (1) inform the enforcement authorities in order to stop the misconduct or make the perpetrator repair the harm done and/or (2) claim damages before the courts. If the misconduct is concealed from the victim it does not have any private information to share with the authorities and is therefore no suitable addressee of incentives to blow the whistle. However, there is an exception to this reasoning where the harm done to the informed party (“victim”) is not severe enough to motivate him to act. In the legal literature, the problem of private law enforcement in case of dispersed low-value damage is well-known and intensely debated. The enforcement problem may be aggravated where there is no individual claim to self-enforce. In the latter case, however,

the „victim“ may alternatively inform the competent public authority or the private party who has a claim in order to stop the misconduct and/or make the perpetrator repair the harm done.

As a consequence, whistleblower incentives primarily address neutral or advantaged insiders. However, insofar our study contradicts Prediction 2b, since we do not observe a significant difference in the willingness to blow the whistle between neutral and advantaged types in *No Incentives*. Furthermore, in this treatment the whistleblowing rates of these types are rather low. For such types disadvantageous inequality is not relevant: They do not suffer enough from the caused inequality to blow the whistle. In contrast, both types are highly responsive to financial incentives. As a consequence, for both neutral and advantaged players as the main targets of whistleblower regulation in real life financial incentive schemes, at least of the *Reward* kind (see explanation above) and coming along with sufficient protection from retaliation, promise to be highly effective in making them blow the whistle.

Apart from its comparatively modest impact on the willingness to blow the whistle across all player types *Command* only raises the likelihood to blow the whistle significantly for advantaged players. Thus, from a policy perspective our results suggest not to focus on a *Command* strategy as an exclusive regulatory measure, but rather to deploy it as a (nearly) costless “add-on”.

No Incentives in experiment 1 and *Not Affected* in experiment 2 differ in only one aspect. While in *Not Affected* the misconduct decreases total welfare, it does not affect total welfare in *No Incentives*. Therefore, one might expect a higher willingness to blow the whistle in *Not Affected*. However, this is not the case. Results are similar in both experiments, no player type is more willing to blow the whistle when total welfare is destroyed. If at all, we observe a tendency for disadvantaged player to be less willing to blow the whistle in this situation. One can conclude that the willingness to blow the whistle does not depend on the misconduct’s impact on total welfare.

Status of sanctioning authority

Experiment 2 provides evidence that the willingness to blow the whistle at least for one type depends on how and whether the sanctioning authority is affected by the reported misconduct. More concretely, we find that unaffected bystanders and insiders for whom the misconduct carries utility are not affected by the status of authority. Their willingness to blow the whistle is rather low no matter how the authority is affected by the misconduct. For victims of the misconduct, however, we find significant differences. From disadvantaged insiders, a negatively affected authority receives information more frequently than an authority who is positively or not affected. Thus, Prediction 7a is confirmed for disadvantaged players. At least insofar companies or employers implementing internal whistleblowing schemes might consider to stress in their codes of conduct or elsewhere that they themselves suffer from the misconduct to be disclosed by the potential whistleblower. However, this suggestion comes with the substantial qualification that the

victims of the respective misconduct are generally not the main target of a whistleblower scheme as explained above.

Individual determinants of whistleblowing

In both experiments we measured the influence of various control variables on the willingness to blow the whistle. In line with Miceli et al. (1991) and Miceli and Near (1988) we find that men are more willing to blow the whistle than women. As to the relation of the “Big Five” personality traits and whistleblowing our results show no significant correlation. In contrast to Bjørkelo et al. (2010) there is neither a significant positive relation between extraversion and whistleblowing nor a negative relation of agreeableness and whistleblowing. This may be due to the difference in experimental design. While Bjørkelo et al. (2010) asked municipality employees whether they actually had blown the whistle in the past, our experiment did not require reporting participants to “come to the fore”. They instead stayed anonymous throughout the whole duration of the experiment. Reporting only required to choose this option on the screen and no further social interaction. Furthermore, the results of Bjørkelo et al. (2010) might suffer from a selection bias, since those employees who actually returned the questionnaire (around 1/3) may be more extravert than those who did not return the questionnaire. Contradicting Dozier and Miceli (1985) and Chiu (2003) we do not find that locus of control is negatively correlated to whistleblowing.

However, we do find that beliefs on others’ behavior are important. There is a positive correlation of the willingness to blow the whistle and the belief that others will do so as well. One interpretation is that people trap into the “false consensus” effect. That means that people overestimate whether their individual behavior (in this case whistleblowing) is normal and typical also for others.

7. Conclusion

How to make insiders report misconduct to a sanctioning authority? In two laboratory experiments we analyzed the influence of (financial) incentives and situational frame conditions on insiders’ willingness to blow the whistle. Contrary to other studies, we do not measure stated intention to blow the whistle. Our contribution is an experimental design that measures actual whistleblowing behavior in a controlled environment, and that is easily adaptable to other incentive strategies and situational determinants on the willingness to report misconduct.

Retaliation and protection against retaliation are apparent candidates for extensions. For the time being, we excluded this aspect for reasons of simplicity. Including a retaliation threat in our experimental design may at least partly offset the incentive effect of the other strategies because of uncertainty of whistle blowing costs. So far, we focused exclusively on punishment as a “remedy” for the misconduct. Another natural extension is the compensation for the losses the victims incurred because of the misconduct. Finally, with regard to the comparative whistleblowing rates of neutral and advantaged players in *Reward* and *Fine*, it would promise insightful results to add extensions with differing, rather low probabilities of detecting non-reporting participants.

In our current experimental design the framing was neutral and the misconduct was designed as anti-social behavior. There was no fraud involved. Although we can show that our participants perceived the unequal allocation as inappropriate behavior anyhow, an extension that strengthens anti-social behavior with fraudulent behavior would be a straightforward amplification. A loaded framing could also be worth investigating. Participants in our experiments were university students. One the one hand, students will become employees eventually and university misconduct is a relevant issue, too. One the other hand, conducting the experiment with potential whistleblowers on real corporate misconduct would be a promising road for future research.

References

- ALFORD, C. FRED “Whistleblowers and the Narrativew of Ethics.” *Journal of Social Philosophy*, Vo. 32 2001, pp. 402-418.
- APESTEGUIA, J., DUFWENBERG, M., AND SELTEN, R. “Blowing the whistle.” *Economic Theory*, Vol. 31 (2007), pp. 143-166.
- BECKER, G. “Crime and Punishment: An Economic Approach.” *Journal of Political Economy*, Vol. 76 (1968), pp. 169-217.
- BIGONI, M., FRIDOLFSSON, S., LE COQ, C., AND SPAGNOLO, G. “Fines, Leniency, and Rewards in antitrust.” *RAND Journal of Economics*, 43(2) (2012), pp. 368-390.
- BJØRKELO, B., EINARSEN, S., AND MATTHIESEN, S. “Predicting proactive behaviour at work: Exploring the role of personality as an antecedent of whistleblowing behaviour.” *Journal of Occupational and Organizational Psychology*, Vol. 83 (2010), pp. 371-394.
- BREUER, L. “Tax Compliance and Whistleblowing – The Role of Incentives.” *The Bonn Journal of Economics*, Vol. II(2) (2013), pp. 7-45.
- CHIU, R. “Ethical judgment and whistleblowing intention: Examining the moderating role of locus of control.” *Journal of Business Ethics*, Vol. 43(1-2) (2003), pp. 65-74.
- CHOO, L., GRIMM, V., HORVATH, G., AND NITTA, K. *Remuneration Structure and Whistleblowing Behaviour: An Experimental Investigation*. Working Paper (2016). University of Erlangen-Nuremberg.
- COMPETITION & MARKETS AUTHORITY (CMA) “Informant Reward Policy”, 2014, online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/299411/Informant_rewards_policy.pdf.
- DASGUPTA, S. AND KESHARMANI, A. “Whistleblowing: A Survey of Literature.”, *IUP Journal of Corporate Governance*, Vol. 9(4) (2010), pp. 57-70.
- DE GEEST, G. AND DARI-MATTIACCI, G. “The Rise of Carrots and the Decline of Sticks.” *University of Chicago Law Review*, Vol. 80 (2013), pp. 341-392.
- DOWLING, JR., D. “Sarbanes-Oxley Whistleblower Hotlines Across Europe: Directions Through the Maze.” *The International Lawyer*, Vol. 42 (2008), pp. 1-57.
- DOZIER, J. AND MICELI, M. “Potential predictors of whistle-blowing: A prosocial behavior perspective.” *Academy of Management Review*, Vol. 10(4) (1985), pp. 823-836.
- DWORKIN, T. “SOX and Whistleblowing.” *Michigan Law Review*, Vol. 105 (2007), pp. 1757-1780.
- DWORKIN, T. AND BAUCUS, M. “Internal vs. External Whistleblowers: A Comparison of Whistleblowing Processes.” *Journal of Business Ethics*, Vol. 17 (1998), pp. 1281-1298.
- DYCK, A., MORSE, A., AND ZINGALES L. “Who Blows the Whistle on Corporate Fraud?” *Journal of Finance*, Vol. 65 (2010), pp. 2213-2253.
- EBERSOLE, D. “Blowing the Whistle on the Dodd-Frank Whistleblower Provisions.” *Ohio State Entrepreneurial Business Law Journal*, Vol. 6(1) (2011), pp. 123-174.

ENGEL, C., GOERG, S., AND YU, G. "Symmetric vs. Asymmetric Punishment Regimes for Bribery." *MPI Collective Goods Preprint*, No. 2012/1, online: <http://ssrn.com/abstract=1983969>.

FEHR, E. AND FISCHBACHER, U. "Third-party punishment and social norms." *Evolution and Human Behavior*, Vol. 25(2) (2004), pp. 63-87.

FEHR, E. AND SCHMIDT, K. "A theory of fairness, competition, and cooperation." *Quarterly Journal of Economics*, Vol. 114(3) (1999), pp. 817-868.

FELDMAN, Y. AND LOBEL, O. "The Incentives Matrix: The Comparative Effectiveness of Rewards, Liabilities, Duties, and Protections for Reporting Illegality." *Texas Law Review*, Vol. 88 (2010), pp. 1151-1212.

FISCHBACHER, U. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental economics* 10.2 (2007), pp. 171-178.

FLEISCHER, H. AND SCHMOLKE, K.U. "Financial Incentives for Whistleblowers in European Capital Markets Law." *European Company Law*, Vol. 9 (2012), pp. 250-259.

FUNK, P. "Is There an Expressive Function of Law?" *American Law and Economics Review*, Vol. 9 (2007), pp. 135-159.

GAO, J., GREENBERG, R., AND WONG-ON-WING, B. "Whistleblowing intentions of lower-level employees: The effect of reporting channel, bystanders, and wrongdoer power status." *Journal of Business Ethics* 126.1 (2015), pp. 85-99.

GREINER, B. "The Online Recruitment System Orsee 2.0 - a Guide for the Organization of Experiments in Economics." (2004) University of Cologne, Department of Economics.

GOBERT, J. AND PUNCH, M. "Whistleblowers, the Public Interest, and the Public Interest Disclosure Act 1998." *Modern Law Review*, Vol. 63 (2000), pp. 25-54.

HEYES, A. AND KAPUR, S. "An Economic Model of Whistle-Blower Policy." *Journal of Law, Economics, & Organization*, Vol. 25 (2008), pp. 157-182.

HINLOOPEN J. AND SOETEVENT, A. "Laboratory evidence on the effectiveness of corporate leniency programs" *RAND Journal of Economics*, Vol. 39 (2008), pp. 607-616.

HUNGARIAN COMPETITION AUTHORITY (GVH) "Regular questions about the cartel informant reward." (March 2010), http://www.gvh.hu/en/other/6429_en_regular_questions_about_the_cartel_informant_reward.html?query=cartel%20informant%20reward.

JUBB, P. "Whistleblowing: A restrictive definition and interpretation." *Journal of Business Ethics*, Vol. 21(1) (1999), pp. 77-94.

LEPINE, J.. AND VAN DYNE, L. "Voice and Cooperative Behavior as Contrasting Forms of Contextual Performance: Evidence of Differential Relationships With Big Five Personality Characteristics and Cognitive Ability." *Journal of Applied Psychology*, Vol. 86 (2001), pp. 326-336.

LEWIS, D. (ed.) "A Global Approach to Public Interest Disclosure. What Can We Learn from Existing Whistleblowing Legislation and Research?" (Edward Elgar, Cheltenham) (2010).

MESMER-MAGNUS, J.. AND VISWESVARAN, C. "Whistleblowing in Organizations: An Examination of Correlates of Whistleblowing Intentions, Actions, and Retaliation." *Journal of Business Ethics*, Vol. 62 (2005), pp. 277-297.

MICELI, M. AND NEAR, J. "Individual and Situational Correlates of Whistle-Blowing." *Personnel Psychology*, Vol. 41 (1988), pp. 267-281.

MICELI, M. AND NEAR, J. "Blowing the Whistle." (Lexington Books, New York) (1992).

MICELI, M. AND NEAR, J. "When do observers of organizational wrongdoing step up? Recent US research on the factors associated with whistleblowing." in David B. Lewis (ed.), *A Global Approach to Public Interest Disclosure* (2010), pp. 74-90.

MICELI, M., DOZIER, J., AND NEAR, J. "Blowing the Whistle on Data Fudging: A Controlled Field Experiment." *Journal of Applied Social Psychology*, Vol. 21(4) (1991), pp. 271-295.

MICELI, M., NEAR, J., AND DWORKIN, T. "Whistleblowing in Organizations." (Routledge, New York) (2008).

PATTON, J., STANFORD, M., AND BRARRATT, E. "Factor structure of the Barratt impulsiveness scale." *Journal of clinical psychology*, Vol. 51(6) (1995), pp. 768-774.

POLAŃSKI, J. "Informant reward schemes in competition enforcement" Working Paper (October 2014), online: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2457450.

RAPP, G. "Beond Protection: Invigorating Incentives for Sarbanes-Oxley Corporate and Securities Fraud Whistleblowers." *Boston University Law Review*, Vol. 87 (2007), pp. 91-156.

ROBINSON, S. N., ROBERTSON, J.C., AND CURTIS, M.B. "The effects of contextual and wrongdoing attributes on organizational employees' whistleblowing intentions following fraud." *Journal of business ethics*, Vol. 106.2 (2012), pp. 213-227.

SEC, OFFICE OF INSPECTOR GENERAL "Evaluation of the SEC's Whistleblower Program." (2013), online: <https://www.sec.gov/about/offices/oig/reports/audits/2013/511.pdf>.

SEC, OFFICE OF THE WHISTLEBLOWER "2015 Annual Report to Congress on the Dodd-Frank Whistleblower Program." (2015), online: <https://www.sec.gov/about/offices/owb/annual-report-2015.pdf>.

SELTEN, R. "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes." *Seminar für Mathemat. Wirtschaftsforschung u. Ökonometrie* (1965).

SHIKORA, J. "Bringing good and bad Whistle-blowers to the Lab", Munich Discussion Paper No. 2011-4, online: <http://epub.ub.uni-muenchen.de/12161/>.

STIKELEATHER, B. "When do employers benefit from offering workers a financial reward for reporting internal misconduct?" *Accounting, Organizations and Society*, Vol. 52 (2016), pp. 1-14.

SULLIVAN, K., BALL, K., AND KLEBOLT, S. "The Potential Impact of Adding a Whistleblower Rewards Provision to ACPERA." *Antitrust Source* (October 2011), pp. 1-6, online: http://www.americanbar.org/content/dam/aba/publishing/antitrust_source/oct11_sullivan_10_24f.authcheckdam.pdf.

SUNSTEIN, C. "Legal Interference with Private Preferences." *University of Chicago Law Review*, Vol. 53 (1986), pp. 1129-1174.

SUNSTEIN, C. "On the Expressive Function of Law." *University of Pennsylvania Law Review*, Vol. 144 (1996), pp. 2021-2053.

SUTTER, M., HAIGNER, S., AND KOCHER, M. “Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations.” *Review of Economic Studies*, Vol. 77 (2010), pp. 1540-1566.

THÜSING, G. AND FORST, G. (eds.) “Whistleblowing – A Comparative Study.” (Springer, Cham, Heidelberg, New York, Dordrecht, London) (2016).

VICTOR, B., TREVINO, L., AND SHAPIRO, D. “Whistleblowing on Unethical Behavior: The Influence of Justice Evaluations and Social Context Factors.” *Journal of Business Ethics*, Vol. 12 (1993), pp. 253-263.

WALTON, N. “Crowding Theory and Executive Compensation” *Theoretical Inquiries in Law*, Vol. 13 (2012), pp. 429-456.

WAYTZ, A., DUNGAN, J., AND YOUNG, L. “The whistleblower’s dilemma and the fairness-loyalty tradeoff.” *Journal of Experimental Social Psychology*, Vol. 49 (2013), pp. 1027-1033.

Appendix A: Instructions to Experiment 1

General Instructions

We would like to welcome you to this economic experiment. Thank you for participating in our study.

If you read the following instructions carefully, you may – depending on your decisions made throughout the experiment – earn money. Therefore, it is very important, that you read the following text thoroughly. If you have any questions, please raise your hand above the upper rim of your cubicle. We will come to your place and answer your question in private.

During the experiment you are not allowed to speak to the other participants, to use your cell phone, or to start other programs on the computer. Ignoring these rules leads to exclusion from the experiment and from any payment.

During the experiment we do not speak of Euro, but of points. Your whole income will be computed in points, at first. In the end, the points you have earned during the experiment will be converted at the following rate:

1 point = € 0.05.

For showing up and participating in the experiment you earn four additional euros. At the end of today's experiment your whole income will be paid in cash.

On the following pages we explain the course of the experiment in detail.

The Experiment

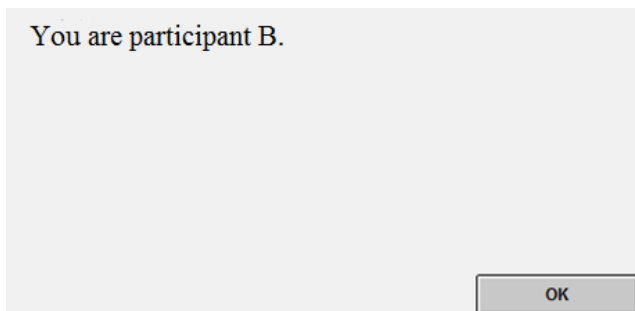
Summary

- In this experiment there are five types of participants: A, B, C, D and E.
- Participants are divided into groups of 5 each, of whom each Type A-E is represented by one participant.
- Each participant gains 100 points (distribution X).
- Participant A decides whether to stick to this distribution (X) or to change it to distribution Y.
 - Distribution Y is as follows: Participants A and B receive 120 points, participant C receives 60 points, and participants D and E receive 100 points.
- Participants B, C, and D learn of participant A's choice.
- Participant E, for the time being, stays ignorant to A's choice.
- Participants B, C and D can choose to inform participant E on the choice A has made.
- If participant E learns of A's choice, she may delete as many points of A as she chooses. The deleted points are forfeit and not transferred to any other participant.

Course of the Experiment in Detail

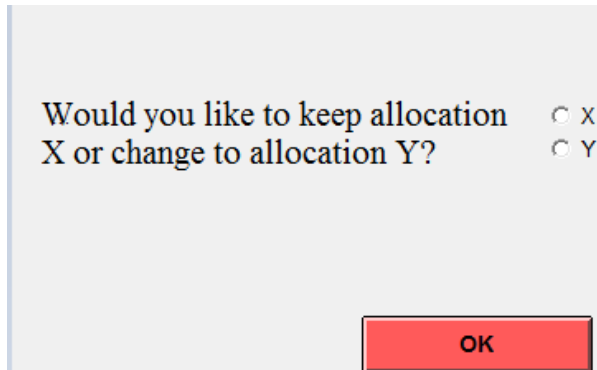
Screen 1 (displayed to all participants)

All participants come to know their respective type. For example:



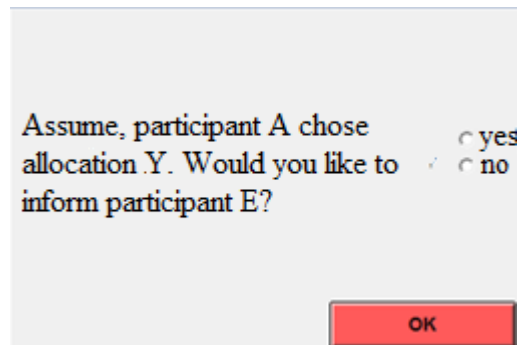
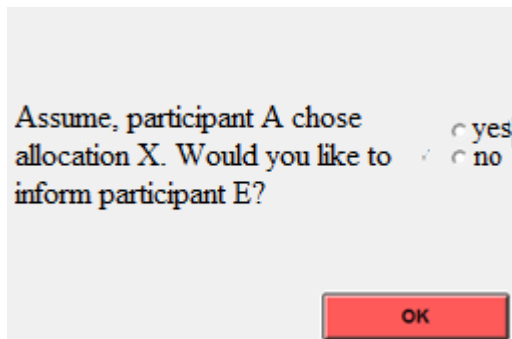
Screen 2 (displayed to participant A)

Participant A chooses between distribution X and Y.



Screen 3 (displayed to participants B-D)

Whilst participant A chooses, participants B, C, and D choose, whether they want to inform participant E on participant A's choice. They choose for both possible distributions (X and Y). According to B-D's choice E is or is not informed (only) about the distribution actually chosen by A. [Command: If participant A chose Y, participants B-D are obliged to inform participant E.]



Whether participant E is informed about A's choice, is determined in the following way:

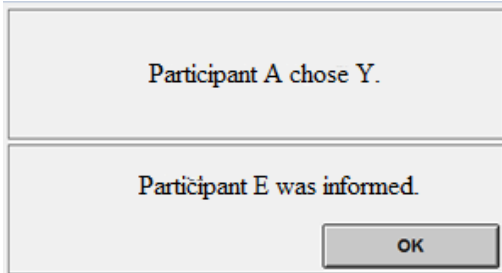
One of the participants B-D is selected randomly.

- Participant E is informed about A's choice, if the selected participant (B, C, or D) beforehand opted for informing participant E. Furthermore, 6 points are deducted from the selected participant's account. [Low Reward: The selected participant receives 10 points. High Reward: The selected participant receives 50 points.]
- Participant E is not informed about A's choice, if the selected participant beforehand opted for not informing participant E. [No Incentives & Command: In this case no points are deducted from de selected participant's account. Low Fine: In this case 10 points are deducted from de selected participant's account. High Fine: In this case 50 points are deducted from de selected participant's account.]

Screen 5 (displayed to participants A-D)

Participants A-D learn of A's choice and whether participant E has been informed about this choice.

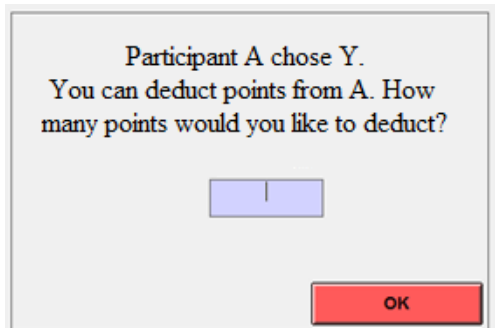
Exemplary screen:



Screen 6 (displayed to participant E)

Participant E is or is not informed about participant A's choice. If participant E is informed, she can delete as many of participant A's points as she likes. The maximum amount of points participant E can delete equals the amount of points assigned to participant A, that is participant E can delete maximally 100 (distribution X) or 120 (distribution Y) points, respectively.

Exemplary screen:



Participant E decides about a deduction from participant A's stock of points for both possible distributions (X and Y). Only the choice of participant E with regard to the distribution actually chosen by participant A is subsequently implemented provided that participant E is informed about this choice of participant A by the randomly chosen participant (B, C, or D).

While participant E makes his choice, participants A-D answer several questions at the screen.

Screen 7 (displayed to all participants)

All participants are informed about the choices made by participants A and E as well as about their final number of points earned.

Control Questions

Please raise your hand after you have answered the following control questions. The experiment starts as soon as all participants have answered the questions correctly.

1. How many points do the five (types of) participants receive according to distribution X?
A _____ B _____ C _____ D _____ E _____
2. How many points do the five participants receive according to distribution Y?
A _____ B _____ C _____ D _____ E _____
3. How is the participant determined who chooses whether E is or is not informed about A's choice?

4. If participant E is informed about A's choice, what can she do?

5. What can participant E do, if she is not informed about A's choice?

6. Assume, participant B who chose to inform participant E about A's choice is picked randomly. What will happen next automatically?

7. Assume, participant C who chose NOT to inform participant E about A's choice is picked by random. What will happen now automatically?

8. How many points can participant E delete maximally?

Appendix B: Instructions to Experiment 2

Note that the instructions for Experiment 2 were slightly different than described in the design section. Instead of assigning the role of authority to E, F, or G (across treatments), it was always participant G who was in the role of authority. However, participant G's payoff through misconduct was manipulated in the three treatments.

General Instructions

We would like to welcome you to this economic experiment. Thank you for participating in our study.

If you read the following instructions carefully, you may – depending on your decisions made throughout the experiment – earn money. Therefore, it is very important, that you read the following text thoroughly. If you have any questions, please raise your hand above the upper rim of your cubicle. We will come to your place and answer your question in private.

During the experiment you are not allowed to speak to the other participants, to use your cell phone, or to start other programs on the computer. Ignoring these rules leads to exclusion from the experiment and from any payment.

During the experiment we do not speak of Euro, but of points. Your whole income will be computed in points, at first. In the end, the points you have earned during the experiment will be converted at the following rate:

1 point = € 0.05.

For showing up and participating in the experiment you earn four additional euros. At the end of today's experiment your whole income will be paid in cash.

On the following pages we explain the course of the experiment in detail.

The Experiment

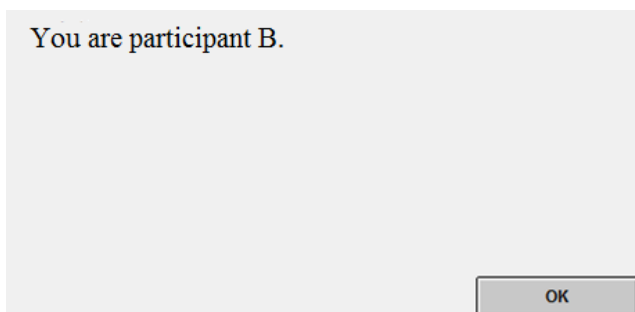
Summary

- In this experiment there are seven types of participants: A, B, C, D, E, F and G.
- Participants are divided into groups of 7 each, of whom each Type A-G is represented by one participant.
- Each participant gains 100 points (distribution X).
- Participant A decides whether to stick to this distribution (X) or to change it to distribution Y.
 - Distribution Y is as follows:
 - [Not Affected:* Participants A and B and C receive 120 points, participants D and E receive 60 points, and participants F and G receive 100 points.
 - Negative:* Participants A and B and C receive 120 points, participants D and E receive 100 points, and participants F and G receive 60 points.
 - Positive:* Participants A receives 120 points, participants B and C receive 100 points, participants D and E receive 60 points, and participants F and G receive 120 points.]
- Participants B, C, D, E, F and G learn of participant A's choice.
- Participant G, for the time being, stays ignorant to A's choice.
- Participants B, C, D, E and F can choose to inform participant G on the choice A has made.
- If participant G learns of A's choice, she may delete as many points of A as she chooses. The deleted points are forfeit and not transferred to any other participant.

Course of the Experiment in Detail

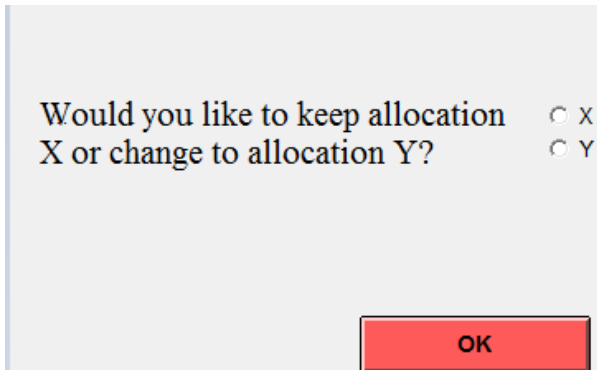
Screen 1 (displayed to all participants)

All participants come to know their respective type. For example:



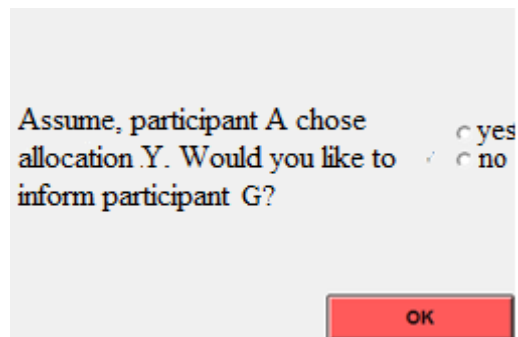
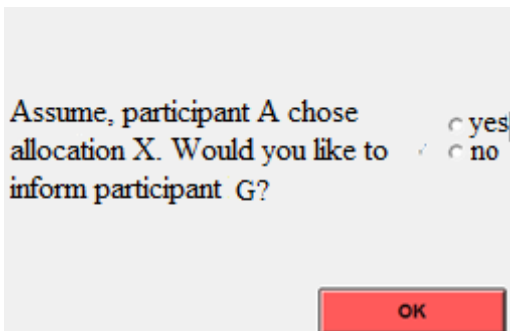
Screen 2 (displayed to participant A)

Participant A chooses between distribution X and Y.



Screen 3 (displayed to participants B-F)

Whilst participant A chooses, participants B, C, D, E and F choose, whether they want to inform participant G on participant A's choice. They choose for both possible distributions (X and Y). According to B-F's choice participant G is or is not informed (only) about the distribution actually chosen by A.



Whether participant E is informed about A's choice, is determined in the following way:

One of the participants B-F is selected randomly.

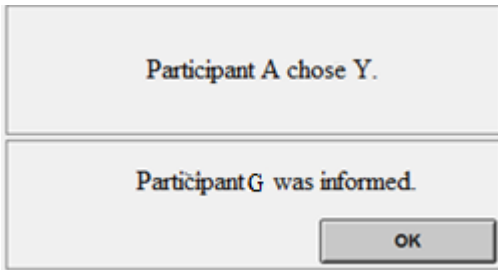
- Participant G is informed about A's choice, if the selected participant (B, C, D, E or F) beforehand opted for informing participant G. Furthermore, 6 points are deducted from the selected participant's account.

Participant E is not informed about A's choice, if the selected participant beforehand opted for not informing participant E. In this case no points are deducted from the selected participant's account.

Screen 5 (displayed to participants A-F)

Participants A-F learn of A's choice and whether participant G has been informed about this choice.

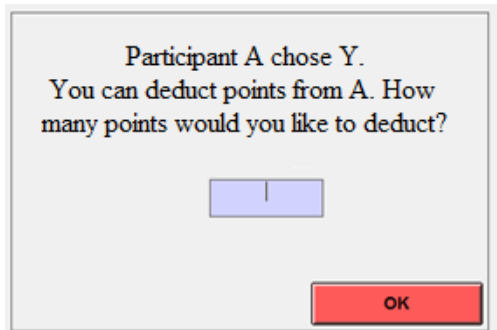
Exemplary screen:



Screen 6 (displayed to participant G)

Participant G is or is not informed about participant A's choice. If participant G is informed, she can delete as many of participant A's points as she likes. The maximum amount of points participant G can delete equals the amount of points assigned to participant A, that is participant G can delete maximally 100 (distribution X) or 120 (distribution Y) points, respectively.

Exemplary screen:



Participant G decides about a deduction from participant A's stock of points for both possible distributions (X and Y). Only the choice of participant G with regard to the distribution actually chosen by participant A is subsequently implemented provided that participant G is informed about this choice of participant A by the randomly chosen participant (B, C, D, E or F).

While participant G makes his choice, participants A-F answer several questions at the screen.

Screen 7 (displayed to all participants)

All participants are informed about the choices made by participants A and G as well as about their final number of points earned.

Control Questions

Please raise your hand after you have answered the following control questions. The experiment starts as soon as all participants have answered the questions correctly.

1. How many points do the five (types of) participants receive according to distribution X?

A _____ B _____ C _____ D _____ E _____ F _____ G _____

2. How many points do the five participants receive according to distribution Y?

A _____ B _____ C _____ D _____ E _____ F _____ G _____

3. How is the participant determined who chooses whether G is or is not informed about A's choice?

4. If participant G is informed about A's choice, what can she do?

5. What can participant G do, if she is not informed about A's choice?

6. Assume, participant B who chose to inform participant G about A's choice is picked randomly. What will happen next automatically?

7. Assume, participant C who chose NOT to inform participant G about A's choice is picked by random. What will happen now automatically?

8. How many points can participant G delete maximally?
