

Motivating Whistleblowers*

Jeff Butler⁺

Danila Serra[♦]

Giancarlo Spagnolo[§]

January 2017

Abstract

We experimentally investigate employees' decisions to blow the whistle against their manager, if they acquire information on the manager's law-breaking activities that benefited the firm but harmed society. We test the effect of financial rewards and public scrutiny on whistleblowing, and we ask whether employees' responsiveness to both incentives depends on whether the social damages caused by the manager's illegal behavior are visible to the general public. Our results suggest that 1) financial rewards ubiquitously increase the likelihood of whistleblowing, 2) public scrutiny increases (decreases) whistleblowing when the negative externalities generated by fraud are visible (invisible) to the public, and 3) political orientation significantly affects responsiveness to public scrutiny.

JEL Codes: K42, C92, D04.

Key words: Whistleblowing, fraud, financial rewards, public scrutiny.

* We thank seminar participants at Luiss and Southern Methodist University, and conference participants at the Bay Area Behavioral and Experimental Workshop, the Texas Experimental Symposium, the North-American ESA conference for useful comments and suggestions. We are grateful to the Wallander and Hedelius foundation for financial support.

§ SITE, Stockholm School of Economics, CEPR and EIEF. Email: spagnologianca@gmail.com.

♦ Southern Methodist University, Department of Economics. Email: dserra@smu.edu.

+ Louisiana State University, Department of Economics. Email: jeffreybutler@lsu.edu.

1. Introduction

Corporate fraud is widespread around the world. A recent survey of over 6000 organizations across 115 countries (2016 Global Crime Survey)¹ shows that one in three organizations, both worldwide and in the US, experienced fraud in the past 24 months, prevalently in the form of asset misappropriation, cybercrime, corruption, as well as procurement and accounting fraud. About 35% of the surveyed firms reported fraud-related losses exceeding 100,000 USD, and 14% of firms reported losses above 1 million dollars.² Dyck, Morse and Zingales (2014) estimated that between 1996 and 2004 about 15% of large publicly-traded US corporations, with assets exceeding 750 million dollars, engaged in fraud. The estimated expected annual cost of fraud for these firms amounts to a staggering \$380 billion dollars.

Due to their informational advantage, employees play a crucial role in uncovering illegal behavior and initiating internal or external investigations through whistleblowing. However, although whistleblowers have been filling the covers of popular journals in recent years, starting with the Enron scandal and ending with the Snowden and Wikileaks-related cases, whistleblowing by employees is quite uncommon. Dyck, Morse and Zingales (2010) analyze 216 securities class action lawsuits filed against large US corporations and find that only about 18% of them were brought forward by an employee. This is unsurprising, given the high costs associated with blowing the whistle. Such costs range from coworkers' disapproval and ostracism, to job loss and difficulties in getting hired within the same industry. The psychological cost caused by conflicting moral norms – loyalty toward the firm on the one hand and fairness concerns on the other – may also contribute to employees' reluctance to report any wrongdoing taking place within their organization, as shown by Waytz, Dungan and Young (2013). Finally, fear of media scrutiny and public disapproval may further reduce employees' willingness to blow the whistle on corporate fraud.

Despite the increasing attention on existing whistleblowers and the widespread awareness of the benefits of whistleblowing, we know very little on how to motivate individuals to report illegal actions or practices taking place within an organization. In this paper, we employ an economics experiment to assess the effectiveness of different policies aimed at incentivizing potential whistleblowers to report unlawful and socially harmful behavior on the part of their superiors. We focus on both monetary and non-monetary incentives. In particular, we assess the effect that: 1) granting financial rewards to whistleblowers and 2) subjecting whistleblowers to public scrutiny, may have on employees' willingness to blow the whistle on corporate fraud. Moreover, we ask whether different sectors or different kinds of fraud require different

¹ <https://www.pwc.com/gx/en/economic-crime-survey/pdf/GlobalEconomicCrimeSurvey2016.pdf>

² Taking into account that most cases of fraud go undetected and that firms self-selecting into a global crime survey are likely to be “cleaner” than those selecting out, the above numbers undoubtedly underestimate the current state of the corporate world.

policies, depending on the extent to which the social costs generated by fraud are visible to and personally felt by the public – think for instance of insider trading versus Medicare fraud.

The question of whether financial rewards/bounties should be used to incentivize whistleblowers is not new. It is, in fact, central to the regulatory debate that followed the 2007-2009 great financial crisis. On the one hand, the US enacted the Dodd-Frank Act that, among other things, allowed whistleblowers to receive financial bounties for bringing information to the Securities and Exchange Commission (SEC) or the Commodity Futures Trading Commission (CFTC).³ On the other, across the Atlantic, while the new EU Directive on financial fraud does introduce the possibility to offer financial rewards for whistleblowers, regulatory agencies remain strongly opposed to rewards for whistleblowers. One concern expressed from many critics is that whistleblower rewards, which are costly to administer, would simply not be effective in eliciting additional valuable information.⁴ However, the US agencies administering these rewards schemes consider them a great success.⁵ Moreover, the existing empirical evidence suggests that whistleblower reward schemes like the ones implemented in the US are effective motivators of whistleblowing. Dyck, Morse, and Zingales (2010) calculated that in sectors where the False Claim Act allows employees to obtain a financial reward corporate fraud is unveiled by employees in 38% of the cases, while this percentage falls to only 18% when the False Claim Act cannot be applied, a highly significant difference.⁶

A more important concern is that financial rewards would “crowd out” ethically-based whistleblowing. This is due to a judgment on the morality of whistleblowing on the basis of the personal returns attached to

³ The US is a pioneer in the enactment on laws and provisions that protect and reward whistleblowers. In 1986 the US strengthened provision of the False Claims Act (FCA), originally passed by Congress in 1863 and signed by President Abraham Lincoln to fight government fraud, allowing among other things for the qui tam, or whistleblower, provisions. It allows any individual or non-governmental organization to file a False Claims Act lawsuit on behalf of the United States Government and, if successful, to obtain up to 30% of recoveries plus fines. Another whistleblower reward scheme against tax evasion was introduced by the Internal Revenue Service in 2006.

⁴ In the UK for example, the two main financial market watchdogs (Bank of England’s Prudential Supervision Authority and Financial Conduct Authorities) in 2014 gave a joint, strongly negative response to a request of opinion from the financial market committee of the UK parliament on rewarding whistleblowers, arguing among other things that: “There is as yet no empirical evidence of incentives leading to an increase in the number or quality of disclosures received by the regulators.”

⁵ The Securities and Exchange Commission (SEC) reported in 2015 that they received 4000 tips from whistleblowers, an increase of 30% from 2012, with the number of tips growing steadily since 2011 probably as a result of increased awareness of the law. According to the IRS, their whistleblower program has helped to bring back 3\$ billion dollars since 2007, with 343\$ million brought back in 2013 and 310\$ million in 2014 (IRS Report 2015).

⁶ In a series of articles published in high level law journals David F. Engstrom showed empirically that several other concerns about distortions linked to the False Claim Act were not justified in the light of the available evidence (Engstrom 2012, 2013, 2014). Findings from a recent experimental study by Schmolke and Utikal (2016) confirm that financial rewards significantly increase the probability of whistleblowing, especially by insiders that are negatively affected by wrongdoing. See Section 2 for more details on this study.

the reporting act.⁷ While both theoretical and experimental research has shown that, in some circumstances, traditionally studied factors, such as financial incentives, may crowd out other types of motivations and lead to perverse outcomes (e.g., Fehr et al., 2001; Fehr and Gächter, 2001; Fehr and List, 2004; Frey, 1997; Gneezy and Rustichini, 2000a, 2000b), there is no study, to the best of our knowledge, assessing how financial rewards may interact with non-monetary incentives in the decision to blow the whistle. We contribute to the existing literature and policy debate by investigating whether crowding out is a reason for concern. We do so by assessing the effectiveness of financial rewards under different experimental conditions where we manipulate the non-monetary incentives associated with blowing the whistle.

The second objective of this paper is to investigate the role that public scrutiny and expected social judgment plays on the decision to report wrongdoing, with the ultimate goal to assess whether, in order to incentivize whistleblowing, the identity of the whistleblower should be kept anonymous or made public. A vast theoretical and experimental literature has shown that individuals' behavior responds positively to the possibility of social observability and judgment (e.g., Andreoni and Bernheim, 2009; Andreoni and Petrie, 2004; Ariely et al., 2009; Benabou and Tirole, 2006; Gerber et al., 2008; Linardi and McConnell, 2011; Xiao and Houser, 2011.) While existing studies suggest that public scrutiny is likely to have a significant effect on whistleblowing, whether the effect will be positive or negative depends on whistleblowers' expectations of how they will be perceived by the public: will they be seen as snitches or as heroes? The answer may lie in the extent to which the public is aware of the social damage caused by corporate fraud and feels directly affected by it. For instance, in 1971 economist Daniel Ellsberg leaked the Pentagon papers concerning American involvement in Vietnam; he is widely viewed as a hero. Much more divided is the public opinion on Edward Snowden, seen by some as a hero and by many as a traitor. Here we ask whether expectations of public praise or condemnation affect individuals' decision to blow the whistle and, consequently, whether whistleblowers should be exposed to or protected from public scrutiny. This is a policy-relevant question, as there do not seem to be clear and unanimous directives on whether the identity of whistleblowers should be safeguarded from the media and, more generally, the public. For instance, in the US, the investigations conducted by the Security and Exchange Commission (SEC) protect the identity of whistleblowers, whereas investigations conducted under the False Claim Act expose whistleblowers by definition, since they require them to file a court case.

In order to address our research questions in a controlled setting and carefully measure individuals' willingness to report corporate wrongdoing, we employ a framed laboratory experiment that simulates the relationships between employees and manager within a firm. In our basic set-up, managers have the chance

⁷ See e.g. Carson, Verdu and Wokutch (2008) and references therein.

to engage in law-breaking behavior to benefit themselves and their employees at the expense of other subjects, playing in the role of members of the public. The employees, which, importantly, are not victims but beneficiaries of the manager's illicit behavior, are given the option of blowing the whistle on their manager. Whistleblowing leads to the imposition of a monetary penalty on the manager. We conduct different treatments where we manipulate the presence of both financial rewards and public scrutiny on whistleblowers. In particular, in some treatments whistleblowing entails a net cost to the employee, while in other treatments whistleblowing engenders a net financial gain. To incorporate non-pecuniary social image motives, in some treatments participants assigned the role of member of the public are allowed to send costless judgmental messages to the employees who choose to blow the whistle. To induce variation in employees' expectations of positive or negative public judgment, we also manipulate across treatments whether members of the public are aware of the costs imposed on them by manager malfeasance. This variation also allows us to investigate whether financial rewards and public scrutiny have a different impact on whistleblowing, and therefore are differently desirable or undesirable, when applied to different kinds of fraud or different industries.

Our results provide strong support for the use of financial rewards as motivators for whistleblowing: employees are significantly more likely to blow the whistle when doing so entails a personal financial gain. Our findings on the effect of public scrutiny on whistleblowing confirms our expectations. When the public is made aware of the costs imposed on them by manager misbehavior, public scrutiny increases the likelihood of whistleblowing. The opposite is true when the public does not know about the extent to which they have been personally harmed by corporate fraud. This suggests that the visibility of the social costs of fraud affects whistleblowers' expectations of how they will be perceived by the public, i.e., as heroes when the social costs are visible and as snitches when they are invisible. Contrary to the crowding-out hypothesis, we do not find that financial rewards are less effective when non-monetary motivations to blow the whistle are likely to be stronger, i.e., when the public is aware of the cost imposed on them by managers' lawbreaking and can signal approval or disapproval to the whistleblower. As an interesting ancillary result, we find that political orientation significantly affects employees' responsiveness to incentives. In particular, while both right-leaning and left-leaning subjects respond to financial incentives, only left-leaning employees are concerned about public scrutiny and the possibility of social approval or disapproval.

Overall, our investigation provides novel and important insights on the design and implementation of policies aimed at increasing the occurrence of employees' whistleblowing on manager wrongdoing. Our findings suggest that 1) rewarding whistleblowers is broadly effective and therefore highly desirable; 2) crowding out of non-pecuniary motivations for blowing the whistle should not be a reason for concern, and 3) the identity of whistleblowers should be concealed and protected from the public in cases of fraud and

in industries where it is difficult for the public to feel personally affected by corporate misbehavior – think of insider trading or accounting fraud – while it should be exposed to the public when the opposite holds – think of fraud that affects access to and/or compromises the quality of public services.

The paper is organized as follows. Section 2 reviews the literature on whistleblowing. Section 3 describes the experiment and presents our hypotheses. Section 5 reports our empirical findings and Section 6 concludes.

2. Literature Review

In recent years, an increasing number of researchers from a variety of fields have turned their attention to the study of individuals' willingness to blow the whistle on wrongdoing. As a result, the whistleblowing literature is fast-growing. This is partly due to the definition of whistleblowing being quite broad and therefore encompassing a large set of illegal behaviors and reporting mechanisms. Any individual that exposes unlawful, dishonest or simply unethical actions or practices taking place within a private or a public organization is in fact a whistleblower. Depending on the situational and institutional context, we can then distinguish between “watchdog” and “traitorous” whistleblowers (Spagnolo, 2008; Breuer, 2013). The former are bystanders that did not take part in the crime and may either benefit or suffer from it, while the latter are accomplices that played an active role in the wrongdoing. Whistleblowers can be further classified as “internal” or “external”, based on the nature and identity of the agency to which they report the illegal activity. Finally, we can distinguish between “peer-to-peer” whistleblowing – think of employees reporting each other's activities – and whistleblowing on somebody of a different rank, such as an employee whistleblowing on a manager, or a citizen whistleblowing on a public official. In this paper, we focus on employees' whistleblowing on the wrongdoing of individuals of superior rank within the same firm, i.e., a team leader, a manager or a CEO. Moreover, we are interested in cases where the potential whistleblowers did not take part in the crime – even though they benefited from it – and therefore are “watchdog” whistleblowers.

The first formal economic analyses of rewards programs for whistleblowers focus traitorous whistleblowers (accomplice-witnesses). Spagnolo (2004) first analyzes these programs within a dynamic model of collusion that captures the strategic features of any illegal relationship with hold up problems within the criminal team. It is shown that offering a reward to the first self-reporting party financed by the fines paid by the remaining parties generate additional deterrence effects through “distrust” in partners in crime, and - with finite fines - can lead to the first best of full deterrence (with zero probability of inspection from law enforcers). Aubert et al. (2006) extends the study of rewards to whistleblowers in collusion cases focusing

on organizations and on employees blowing the whistle on their firm's misbehavior, accounting for the additional deterrence but also several inefficiencies this may generate, such as forcing firms to reduce employees' rate of turnover or to adopt a more "innocent" internal image. Friebel and Guriev (2012) study rewards for watchdog (innocent) whistleblowers focusing on accounting management (e.g. overstatement of financial results), and also shows that besides deterring such unlawful behavior by making it more costly for the management, they may also have negative effects on firms' productive efficiency.⁸ Felli et al. (2016) show how rewarding whistle-blowing can be used as a designer tool to prevent opportunistic behavior, that takes the form of collusion or blackmail, on the part of members of a hierarchical structure. The only theoretical investigation of whistleblowing that focuses on behavioral motivations, to the best of our knowledge, is that of Heyes and Kapur (2008). They manipulate and derive optimal penalties associated to wrongdoing and the optimal government's responsiveness to whistleblowers tip-offs under different assumptions on the intrinsic motivations of potential whistleblowers.

Empirical studies of whistleblowing are more common. However, they present fundamental measurement and identification challenges caused by the inability to observe the overall populations of infringements and of individuals willing to blow the whistle if made of aware of such infringements. As a result, the existing studies focus either on the infringements that have been discovered (e.g., Dyck et al., 2010) or on scenario-based survey data where respondents self-report their willingness to blow the whistle under different conditions (e.g., Feldman and Lobel, 2010). Measurement and identification issues have led to the recent surge of experimental studies of whistleblowing. Laboratory experiments are particularly valuable to study deterrence of crimes, as they allow to directly observe both wrongdoing and whistleblowing, and measure responsiveness to changes in incentives in a controlled environment.

In the following subsections we review the existing experimental studies on whistleblowing. While there are very few other studies that investigate whistleblowing in an organizational environment similar to ours, there exist numerous studies of whistleblowing under different institutional and legal contexts, including whistleblowing on corrupt public officials and whistleblowing on illegal cartel formation. We first review studies of whistleblowing of wrongdoing within a firm (Section 2.1) and then move on to studies of whistleblowing in other contexts (Section 2.2).

⁸ When top managers over-report earnings, lower level managers are more likely to realize it when their division's performance is poor. Top management may then offer a bribe to prevent the manager from blowing the whistle, and this must be higher the higher the rewards for whistleblowers. This provides lower level managers with an additional payoff when his division performs poorly, thereby reducing her incentives to exert effort.

2.1 Whistleblowing on wrongdoing within a firm

There is only a handful of experimental investigations of whistleblowing on wrongdoing within a firm. The study most closely related to ours is that of Schmolke and Utikal (2016). They employ a laboratory experiment to investigate whistleblowing in a neutrally framed environment where one subject may increase his payoff at the cost of increasing inequality among other players who can then report this behavior to a third subject, the potential whistleblower. Blowing the whistle leads to punishment and redistribution of payoffs to recreate the initial conditions. The authors study the effects of rewards versus fines for not blowing the whistle. They also manipulate whether and how the reporting subject and the enforcing authority are positively or negatively affected by the first subject's decision and investigate the corresponding effects on whistleblowing. The most relevant finding for the purpose of our study is that, controlling for other factors, monetary rewards are very effective in increasing the probability of whistleblowing, even if they are of rather limited size.⁹

Reuben and Stephenson (2013) conduct an experiment where individuals observe cheating behavior by other members of their team and can blow the whistle on them, causing the whole group to be penalized. The results show that whistleblowing is less common when groups can choose their members, that whistleblowers tend to be subsequently shunned from groups and that endogenous group formation produces groups where lying is more common and less likely to be unsanctioned than when groups are randomly assigned. Contrary to Reuben and Stephenson (2013), we are interested in employees' whistleblowing on the wrongdoing of a manager rather than a peer, and in cases of wrongdoing that generate costs on third parties (the public) while benefiting the member of the firm/team.

Bartuli et al. (2016) examine whistleblowing in an experimental context very similar to ours: the potential whistleblower is an employee that benefits from the wrongdoing of the manager, such wrongdoing generates losses to a third party and blowing the whistle is costly. The authors are interested in the relationship between personality traits and the likelihood to blow the whistle rather than in testing policies aimed at incentivizing whistleblowing.¹⁰ Similarly, Waytz et al. (2013) investigate the relationship between propensity to blow the whistle and a specific individual trait: the subjective valuation of fairness/justice

⁹ Schmolke and Utikal (2016) also find that, in line with Fehr and Fischbacher's (2004) results for generic punishment behavior, individuals negatively affected by the first decision are more likely to blow the whistle than non-affected or profiting individuals.

¹⁰ They find that employees who are more altruistic and more concerned about ethical issues, are more likely to blow the whistle. The Honesty-Humility personality scale is also significantly associated with whistleblowing. For survey-based study of personality and whistleblowing, see also Miceli and Near (2001) and Feldman and Lobel (2010).

over loyalty. Using survey questions, the authors find evidence of a significant loyalty/fairness tradeoff in the decision to blow the whistle.

2.2 Whistleblowing in other contexts

The most well developed literature on whistleblowing has to do with illegal cartel formation among firms. The existing studies look at situations where one of the parties of a potential or actual illegal transaction/relationship is offered amnesty (if the illegal action was consumed) and a monetary reward for blowing the whistle about the illegal agreement to law enforcers. Note that this literature focuses on “traitorous” rather than “watchdog” whistleblowers (Spagnolo, 2008). Apesteguia, Dufwenberg and Selten (2007) were the first to study rewards to whistleblowers in an experiment on illegal cartel formation in the context of static Bertrand competition. Any member of a cartel may blow the whistle on its existence, in which case cartel members must pay a fine. Incentives for whistleblowing firms are varied across treatments: i) no incentives; ii) a reduced fine for whistleblowers (leniency); and iii) a monetary bonus for whistleblowers. Their results suggest that whistleblower leniency significantly reduces market prices and increases the incidence of whistleblowing but does not reduce the likelihood of cartel formation. Whistleblower bonuses instead do not reduce market prices relative to the no incentives treatment, but do, however, produce the highest likelihood of whistleblowing.

In a repeated game version of an analogous leniency experiment, where subjects could experiment and learn the subtleties of a rather complex strategic environment, Bigoni et al. (2012) found that amnesty for the first whistleblower result in fewer collusive agreements but higher prices in the surviving ones. Adding a monetary reward for the first whistleblower leads to very high reporting rates, thus destabilizing collusion at an increasing pace, and resulting in strong cartel deterrence and very low prices as theory predicted (see Spagnolo 2004, 2008). A number of other experimental studies focus on the effectiveness of leniency policies providing amnesty or asymmetric legal treatment to accomplice-witnesses that blow the whistle against collusion without the use of monetary rewards, including Hamaguchi and Kawagoe (2009), Hinlopen and Soetevent (2008), Bigoni et al. (2015), Cotten and Santore (2016) among others.

Another strand of the literature has investigated whistleblowing in the context of corrupt transactions between public officials and citizens/firms. In a recent laboratory experiment, Abbink and Wu (2013) simulate both one-shot and repeated transactions between firms and public officials where firms can get illegal services through the payment of a bribe. The authors find that the possibility for one party to obtain amnesty from prosecution and a monetary reward when blowing the whistle has a strong deterrence effect on one-shot illegal transactions and a limited effect on repeated corrupt relationships. A number of other

experimental studies (e.g., Abbink et al., 2014; Serra, 2012; Schikora, 2011) have investigated the effect of guaranteeing amnesty (without monetary rewards) to whistleblowers and have generated similar results in terms of reduction in corruption. Finally, Breuer (2013) studies the effects of financial rewards for whistleblowers in a laboratory experiment on tax evasion. He finds a strong positive effect of rewards on subjects' willingness to blow the whistle on the tax declaration of another subject and little evidence of crowding out of non-monetary motivations. Another interesting finding is that individuals correctly anticipate the effect of financial rewards on whistleblowing, resulting in overall lower tax evasion in the presence of rewards.

In sum, the existing literature, whether it investigates “whatchdog” or “traitorous” whistleblowing, and whether it simulates a firm environment, illegal cartel formation or corrupt transactions, has mainly focused on the effect of financial rewards and/or amnesty on the propensity to report wrongdoing. Our contribution to this literature is threefold. First, we expand our understanding of the effectiveness of financial rewards on whistleblowing by asking whether they may crowd-out employees' non-monetary motivations to blow the whistle on corporate fraud. This is a first-order question, since the assumption of crowding-out of intrinsic motives¹¹ is one of the primary reasons why financial rewards are opposed in the international policy debate on whistleblowing. Second, we examine how social image concerns and expectations of public approval or disapproval affect the propensity to blow the whistle. This is a largely unexplored question. In fact, while there is a growing literature on the effect of image motivations on behavior (e.g., Andreoni and Bernheim, 2009; Andreoni and Petrie, 2004; Ariely et al., 2009; Benabou and Tirole, 2008; Gerber et al., 2008; Linardi and McConnell, 2011; Xiao and Houser, 2011), there are no studies, to the best of our knowledge, investigating the relationship between whistleblowing and public scrutiny. Nevertheless, this is an important relationship as the results of our analysis would guide the design of policies either guaranteeing the anonymity of whistleblowers or exposing their identity to the public. Third, we ask whether different kinds of wrongdoing, possibly taking place in different industries, require different kinds of policies. In particular, we differentiate between cases of fraud generating negative externalities to society that are easily visible to the public, and cases of fraud leading to social costs that the public is less likely to recognize and disapprove of.

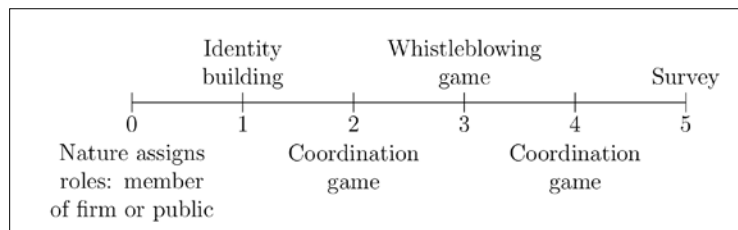
¹¹ Starting from Titmuss (1970), an increasing number studies have shown that economic, or extrinsic incentives, may lower non-monetary incentives and end up having undesired effects on behavior. See, for instance, Fehr et al. (2001), Fehr and Gächter (2001), Fehr and List (2004), Frey (1997), Gneezy and Rustichini (2000a, 2000b). For a survey of the literature, see Gneezy, Meier and Rey-Biel (2011).

3. The Experiment

3.1 Design

Participants are randomly assigned either the role of “member of a firm” or the role of “member of the public”. Each firm is made of three subjects, while the public is made of six participants. The experiment consists of four Stages and only one Stage is randomly chosen for payment at the end of the experimental session. Figure 1 displays the experimental stages.

Figure 1
Stages of the Experiment



Stage One has the purpose to lower the social distance between the members of a firm, generate a shared firm identity and, ideally, create a sense of loyalty between team members. In this stage, the three members of a firm engage in a series of team-building tasks. The first task is the Kandinsky and Klee painting elicitation module first developed in Tajfel, Billig, Bundy, and Flament (1971). Each member of the firm gets credit if at least one member of the firm associates a painting to its artist correctly. The second task consists in adding a set of two two-digit numbers. As before, each member of the firm earns money if at least a member of the firm solves a problem correctly. The third task is a multiplication task; each subject has to multiply a set of two-digit numbers. Individual payoffs are determined as in the previous team-building tasks. The members of the public engage in the same three tasks but their payoffs are determined exclusively by their own performance. At the end of each task, firm members are informed of their own performance and the overall firm performance, which generates their earnings. Members of the public are informed only of their own performance.

Stage Two consists in a one-shot minimum-effort coordination game aimed at measuring within firm cohesion. Each member of a firm plays the game with the other two members. Each member of the public plays the game with two other members of the public. Participants are not informed of the outcome of the game and their earnings until the end of the experimental session.

In Stage Three, participants play the Whistleblowing Game. Subjects retain the role of either member of the firm or member of the public. Within a firm, one participant is randomly chosen to be the “manager” and the remaining two participants are assigned the role of “employees”. The employees engage in a real-effort task consisting in adding two digit numbers, as in Stage Two of the experiment. Each correct answer generates private earnings at a piece rate of 2 ECU and also contributes to a firm fund at a piece rate of 1 ECU. The firm fund is later distributed back to the manager (one half of the fund) and the employees (one fourth each). The manager gets a fixed wage of 24 ECU and has the chance to double the firm’s fund by engaging in a more difficult real-effort task (multiplications of two-digit numbers, as in Stage One of the experiment) and getting more than 7 answers correctly. Alternatively, the manager can augment the fund by choosing to break the law. Breaking the law generates money to the firm but causes a monetary loss of 2 ECU to each of the 6 members of the public.¹²

As before, members of the public are only involved in individual decision-making. Like the employees, they engage in a real effort task consisting in adding two digit numbers. The task generates 2 ECU for each correct answer. However, the final earnings of each member of the public also depends on the rule-breaking choice of the managers of the firms in the session.

We measure employees’ willingness to blow the whistle by using the strategy elicitation method. We ask each employee whether they would blow the whistle if they found out that the manager broke the law. Blowing the whistle requires the employee to pay a monetary cost of 5 ECU and imposes a monetary penalty of 14 ECU to the manager if he or she broke the law. We compute final earnings by randomly choosing one of the two employees in each firm and implementing the stated whistleblowing decision conditional on the matched manager’s behavior.

Stage Four concludes the experiment with a minimum-effort coordination game identical to the game subjects played in Stage Two. After participating in the experiment, subjects fill in a questionnaire. As part of the survey, all subjects are presented with four actual whistleblowing cases that differ both in the extent to which the negative externalities caused by the illegal behavior are visible to the public and in the presence of financial rewards for whistleblowers. The four cases are the Snowden case, the Enron case, the UBS case and the Tenet case.¹³ We chose these cases because the negative externalities caused to the public are clearly

¹² In order to keep the decision to break the law comparable across firms, we do not reveal the size of the firm fund to the manager before eliciting his or her decision to break the law.

¹³ For information on the Snowden case, see: https://www.nytimes.com/2014/01/02/opinion/edward-snowden-whistle-blower.html?_r=0 . For the Enron case, see: <http://news.bbc.co.uk/2/hi/business/5335214.stm>. For the UBS case, see: <http://www.wsj.com/articles/SB10000872396390444017504577645412614237708>. For information on the Tenet case, see: <http://www.corporatecrimereporter.com/news/200/tenet-healthcare-to-pay-514-million-gets-non-prosecution-agreement-two-units-with-no-assets-to-plead-guilty>.

visible in the Snowden and the Tenet case, less visible in the UBS case and invisible in the Enron case. Moreover, the whistleblowers in the UBS and Tenet cases received financial compensation while those in the Enron and Snowden cases did not. As part of the survey, we also register subjects' political orientation by asking subjects to place themselves on a 0-10 political spectrum, where higher numbers correspond to more right-leaning preferences.

3.2 Treatments

We employ three treatment variations, by manipulating the presence of financial rewards for whistleblowers, the extent of public scrutiny to which whistleblowers are exposed, and the visibility of the social costs that the manager's illegal actions cause to the public.

1. *Reward vs. No Reward*: In the *Reward* condition, employees that blow the whistle against their manager get a financial reward of 10 ECU to offset the cost (5ECU) of blowing the whistle;
2. *Public Scrutiny vs. Private Whistleblowing*: Under *Public Scrutiny*, the members of the public are given the chance to send messages of approval or disapproval to whistleblowers. The messages take the form of a smiley face, a frowny face or a neutral face. Each member of the public can also choose to send no message to whistleblowers. Sending a message comes at no cost to the member of the public and does not lead to any monetary reward or penalty for the whistleblower.
3. *Visible vs. Invisible Externalities*: Under *Visible Externalities*, the members of the public are made aware of the monetary losses they suffer (or could suffer) due to managers' illegal actions. In contrast, under *Invisible Externalities* the members of the public are informed about manager wrongdoing but do not know that such wrongdoing affects their own earnings negatively.

The interactions between our three treatment manipulations generate eight experimental conditions, as shown in Table 1.

3.3 Implementation

We employed a 2x2x2 design, for a total of eight treatments. We conducted 18 sessions involving 324 participants at the University of California, Santa Barbara's Experimental and Behavioral Economics laboratory (EBEL), as shown in Table 1. Each subject participated only in one session and one treatment. In each session, 6 subjects were randomly assigned the role of other members of the public (MPs) and

between 6 and 18 subjects were randomly assigned the role of members of a firm, for a total of between 2 and 6 firms per session. Members of a firm made decisions independently from all the other firms participating in a session.

Table 1
Sessions and treatments

Treatments	Invisible Externalities		Visible Externalities		Total	
	Sessions	Subjects	Sessions	Subjects	Sessions	Subjects
No Rewards & Private Whistleblowing	2	33	2	36	4	69
No Rewards & Public Scrutiny	2	36	2	39	4	75
Rewards & Private Whistleblowing	3	60	2	36	5	96
Rewards & Public Scrutiny	2	36	3	48	5	84
Total	9	165	9	159	18	324

In referring to the subject roles, the experimental environment and decision sets, we used the same contextual labels we have used in Section 2.1 when describing the game. We did this as we believe that psychological and social factors play a significant role in individuals' decisions to engage in and report on unlawful behavior, and we wanted to make sure that subjects understood the decision-making context.¹⁴

The experiment had four stages plus a questionnaire. Subjects were presented with the instructions for each stage of the experiment on their computer screen right before that stage began. Only one randomly selected stage of the experiment was used for actual payments. Experimental payoffs were converted from ECUs to US\$ at the exchange rate of \$1 for 2 ECU. The experiment was programmed in z-Tree (Fischbacher, 2007) and subjects were recruited among pre-registered UCSB students using ORSEE (Greiner, 2015). In order to guarantee anonymity, at the beginning of each session subjects were randomly assigned an identification number, which they kept for the duration of the experiment. At no point did we ask subjects to reveal their names during the experiment, and although actual names were used during the payment process for accounting purposes, we informed the subjects that we would not register their names and, therefore, we would not be able to link them to the choices made in the experiment. Each session lasted between 60 and 90 minutes, with average earnings of \$29 per subject including a \$10 show-up fee.

¹⁴Framing effects have been found in a large set of pro-social games, including public goods games (Andreoni, 1995; Cookson 2000; Rege and Telle 2004; among the others), and dictator games (Eckel and Grossman, 1996; Brañas-Garza, 2007). For a recent study of how frames significantly affect first- and second-order beliefs see Dufwenberg, Gächter, and Hennig-Schmidt (2011).

3.4 Predictions

In order to generate our predictions, we need to make assumptions on employees' motives for blowing the whistle. Recall that reporting wrongdoing is costly. Therefore, under the assumption that individuals are motivated purely by extrinsic incentives, we should see no whistleblowing in the absence of financial rewards, no matter the other treatment manipulations. Our first prediction follows.

Prediction 1: If individuals are purely money maximizers, financial rewards will increase the likelihood that an employee will blow the whistle. The effectiveness of financial rewards will be the same across the public scrutiny and visibility treatments.

Additional predictions are generated if we allow individuals' behavior to reflect an endogenous mix of extrinsic/monetary incentives, intrinsic motivations, and reputational concerns linked to social (dis)approval, as in the general framework developed by Benabou and Tirole (2006), hereafter **BT**. Unlike a classical conception of reputation, in BT's framework individuals have a direct preference over their "social image," which is others' beliefs about one's type conditional on observed actions. Our experiment introduces an important wrinkle that complicates the straightforward application of BT's framework to our setting. The first wrinkle is that, mirroring real-world heterogeneity in opinions about particular whistleblowers, blowing the whistle may be interpreted differently by different audiences. For instance, it may be seen as anti-social, violating the pervasive moral norm of group loyalty. Alternatively, it may be interpreted as a manifestation of a pro-social proclivity to punish wrong-doing. In light of this, our experimental treatments manipulate factors which we believe may affect how whistleblowing is perceived and use this variation to make predictions. We stop short, however, of attempting to extend BT's theoretical framework to a multiple-audience signaling model which is beyond the scope of the current exercise. Another wrinkle in our design is the possibility of an "expressive role" of the legal system, which may serve to clarify what constitutes desirable or pro-social behavior in complex situations, such as whistleblowing with overlapping constituencies which is absent from private incentive frameworks. In our context, rewarding whistleblowing may lead some individuals to interpret it as pro-social *qua* reward.

With these caveats in mind, in order to predict how public scrutiny and social judgment may affect whistleblowing we make a handful of plausible assumptions about how social image incentives vary across our treatments. We assume first and foremost that individuals prefer to *appear* pro-social to members of the public. Given the size of the public relative to the firm, we also assume that when the public can express approval or disapproval of whistleblower they are the audience that matters most for social image incentives. Third, we assume that the public is more likely to perceive whistleblowing as a pro-social act when they are aware of the harm imposed on them by manager misbehavior. Intuitively, when the members

of the public are aware that they are being harmed by the firm, they are more likely to want the manager to be punished and, consequently, to socially reward the whistleblower for triggering such punishment. If, instead, the public does not feel directly affected by the manager's wrongdoing, it is possible that it will perceive the whistleblower as somebody that decided to run afoul of a firm-specific moral norm of loyalty,¹⁵ leading to social disapproval. In other words, the visibility of the negative externalities to the public is likely to affect whistleblowers' beliefs on how they will be perceived and judged by the public if they do blow the whistle, i.e., as heroes if the externalities are visible and as snitches if they are not visible. These assumptions lead to our second prediction.

Prediction 2: If individuals are image motivated, allowing for public scrutiny and social judgment will increase whistleblowing more if the negative externalities generated by the firms' illegal behavior are visible to the public than if they are not visible.

Prediction 2 essentially states that for a given set of intrinsic preferences and extrinsic incentives, making an action a stronger signal of pro-sociality induces more of that action.

Next, we consider an interaction between the social image incentives and extrinsic incentives. One key insight of the BT's framework is that extrinsic incentives can make it more difficult for observers to infer an individual's type from actions, reducing the ability to influence social image through behavior. If social image concerns factored heavily into, e.g., behaving pro-socially, adding extrinsic incentives may backfire and "crowd out" such pro-social behavior. In our context, adding financial rewards may affect the public's perception of the motives behind the whistleblowing act and, consequently, the way the whistleblower is socially judged. More or less whistleblowing is possible after offering financial incentives—the net effect depends on the relative weights individuals place on (increased) extrinsic incentives versus (reduced) non-monetary incentives—so we have no prediction overall. However, one might expect the magnitude of the reduction in non-monetary incentives to be larger when whistleblowing was originally a stronger signal of intrinsic pro-sociality. In our context, this corresponds to the negative externalities of fraud being visible to the public. This is the context where whistleblowing is more clearly pro-social. Hence, our last prediction, which formalizes how extrinsic incentives may crowd out image motivations in our setting.

¹⁵ In our discussion, we are abstracting from the social image concerns that individuals may have toward their fellow firm members. A plausible assumption is that employees prefer to appear loyal to fellow firm-members while also wanting to appear pro-social to the members of the public, especially if there exists public scrutiny. When the negative externalities caused by fraud are visible to the public, image motivations toward firm members and those toward members of the public pull employees in different directions. When the negative externalities are invisible to the public, both motivations steer employees away from blowing the whistle.

Prediction 3: If individuals are image motivated, financial incentives may be less effective at eliciting additional whistleblowing when whistleblowers are subject to public scrutiny than when they are shielded from social judgment.

Finally, we expect individuals' political orientation to impact both whistleblowing and social judgment of whistleblowers, although we do not have clear predictions on the sign of the impact. We hypothesize that the left-leaning respondents, by being more concerned about social justice issues, may be more likely to be whistleblower out of concern for the members of the public suffering the social costs of corporate fraud. However, it could also be the case that the right-leaning students, by being more concerned with rule of law and law-breaking, would be more likely to blow the whistle in order to punish such violations.

4. Results

In this section, we start by presenting and discussing the effects of our treatments on employees' willingness to blow the whistle against their manager (Section 4.1). We then present our findings with respect to the members of the public's approval or disapproval of whistleblowers under the different treatments (Section 4.2). We conclude by describing managers' rule-breaking behavior across treatments (Section 4.3).

4.1.1 The decision to blow the whistle

About 33% of employees decided to blow the whistle against their rule-breaking managers. There is considerable variation across treatments, with the percentage of whistleblowers ranging from 6% to 61%, as shown in Figure 2 and Table 2. Since the Visible and the Invisible Externalities treatments simulate different types of illegal actions or different industries where the damages generated by fraud to the public are more or less difficult to identify, we present the results obtained under the two settings separately.

A number of striking results emerge from Figure 2 and Table 2. First, the presence of financial rewards seems to increase the occurrence of whistleblowing among employees. This holds both when whistleblowing takes place in private and when whistleblowers are subject to public scrutiny and judgment. Financial rewards are ineffective only when the externalities caused by fraud are visible to the public and whistleblowers are shielded from social approval or disapproval. In this setting, intrinsic motivations are most likely to play a role in the decision to blow the whistle; therefore, the ineffectiveness of financial rewards may signal crowding out of intrinsic motivations, in line with our Prediction 2. Crowding out vanishes in the presence of public scrutiny.

Figure 2
The effect of rewards and public scrutiny on whistleblowing

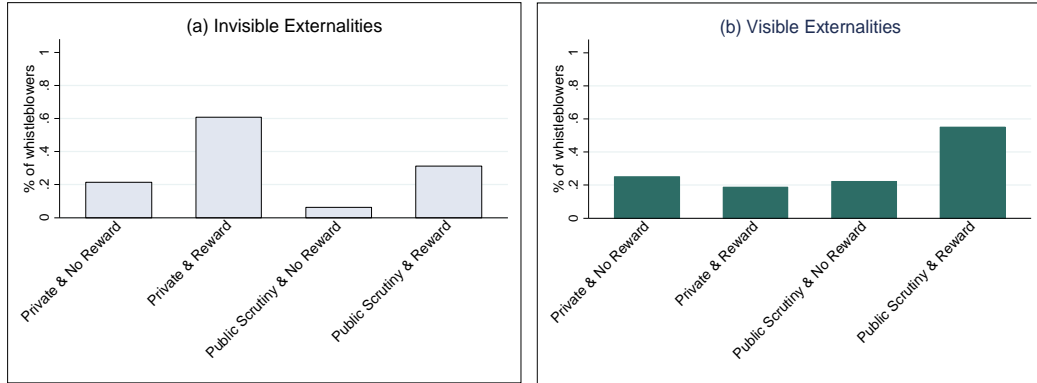


Figure 1 and Table 2 also show that, in accordance with our Prediction 3, public scrutiny has a different effect on whistleblowing depending on how aware the public is of the social damages caused by manager wrongdoing. When the negative externalities are visible to the public, the possibility of public scrutiny seems to increase employees' willingness to blow the whistle, conditional on the presence of financial rewards. The opposite is true when the negative externalities are not visible to the public.

Table 2
Whistleblowing under different treatments

	No Rewards & Private WB	No Rewards & Public Scrutiny	Rewards & Private WB	Rewards & Public Scrutiny
Invisible Externalities	21.43%	6.25%	60.71%	31.25%
H ₀ : Private = Public	p-value = 0.222 (0.249) if rewards=0		p-value = 0.060 (0.058) if rewards=1	
H ₀ : Rewards = No Rewards	p-value = 0.016 (0.018) if private=1		p-value = 0.070 (0.086) if private=0	
Visible Externalities	25.00%	22.22%	18.75%	55.00%
H ₀ : Private = Public	p-value = 0.849 (0.583) if rewards=0		p-value = 0.027 (0.029) if rewards=1	
H ₀ : Rewards = No Rewards	p-value = 0.669 (0.550) if private=1		p-value = 0.039 (0.041) if private=0	

Note: The table reports the percentages of employees that blew the whistle on their manager under different treatment manipulations. P-values are generated by Chi-square tests. P-values from Fisher exact tests in parentheses.

In Table 3, we report estimates from probit regressions where the dependent variable is a dummy equal to 1 if the employee is willing to blow the whistle and 0 otherwise. We start by testing the effectiveness of our Rewards and Public Scrutiny treatments in the setting where the negative externalities of fraud are invisible

(column 1) and where they are visible (column 2) to the public. Next, we extend the analysis to the full sample while interacting our Reward and Public Scrutiny treatment dummies with the Visible Externality dummy (column 3). In column 4, we include a set of control variables that include demographics (age and gender), whether the subject is an economic major and the number of firms in the session. In order to proxy for employees' loyalty to the firm, we include the ratio between firm performance and own performance in Stage 1 of the experiment and the effort level chosen in the minimum effort game of Stage 2. The former variable captures the extent to which each employee may feel "indebted" to the other firm members for the earnings accumulated during the team-building stage. The latter variable is a measure of firm cohesion, as it captures trust and cooperation among firm members. Finally, we include a measure of political orientation generated by our post-experiment survey. We asked subjects where they would place themselves on the left-right spectrum, using a scale from 0 to 10, with higher numbers indicating more right-leaning preferences. The average response among subjects in the role of employees was 3.625 (3.80 in the full UCSB sample), indicating a moderately left-leaning sample. We employ a dummy for left-leaning, which takes the value of 1 if the respondent answered 0, 1, 2 3 or 4; such dummy is equal to 1 for 56% of our participants.

Table 3
Treatment Effects

Dep. Variable: Dummy equal to 1 if employee blew the whistle, 0 otherwise					
	Invisible Ext.	Visible Ext.	All	All	All
	(1)	(2)	(3)	(4)	(5)
Rewards	1.06*** (0.000)	0.42 (0.111)	1.06*** (0.000)	1.03*** (0.000)	0.89*** (0.001)
Public Scrutiny	-0.76*** (0.000)	0.49* (0.070)	-0.76*** (0.000)	-0.66** (0.014)	-0.85* (0.078)
Visible Externalities			-0.20 (0.575)	-0.05 (0.890)	0.24 (0.306)
Visible x Reward			-0.64* (0.079)	-0.62* (0.089)	-1.22*** (0.000)
Visible x Public Scrutiny			1.25*** (0.000)	1.12*** (0.003)	0.60 (0.224)
Public x Reward					0.31 (0.541)
Visible x Reward x Public					1.04* (0.071)
Constant	-0.79*** (0.002)	-0.99*** (0.000)	-0.79*** (0.002)	-1.15 (0.562)	-1.03 (0.619)
Controls	No	No	No	Yes	Yes
Observations	74	70	144	144	144

Note: We report marginal effects. Robust standard errors are clustered at the session level. Controls are: age, gender, economics major, left-leaning political preferences, number of firms in the session, ratio between firm performance and own performance in team building task, and effort chosen in minimum effort task. None of the controls is statistically significant. P-values in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The first clear finding concerns the effectiveness of financial rewards. The estimates reported in Table 3 show that rewarding whistleblowers significantly increases employees' willingness to blow the whistle. Rewards are less effective in industries or cases of fraud where the public feels directly affected by the firm's law breaking behavior. On the other hand, we do not find support to our Prediction 3. Indeed, financial rewards do not seem to crowd out image motivations, i.e. they are not less effective when the whistleblower is subject to public scrutiny. Our first result follows:

Result 1: (a) Financial rewards increase whistleblowing and do not crowd-out non-monetary image-driven motivations;

(b) Financial rewards are less effective when the negative externalities caused by fraud are visible to the public, in the absence of social judgment.

Part (a) of Result 1 goes against the predictions of the BT model, suggesting that in the organizational context simulated in our experiment, extrinsic incentives do not lead potential whistleblowers to fear a more negative social judgment of their actions. Part (b) of Result 1 is also not consistent with the standard crowding-out prediction of the BT model, but could be explained by self-signaling, i.e. the notion that if an individual is his/her own audience, then even absent public judgment we might expect financial incentives to crowd out pro-social behavior.

The second clear result emerging from Table 3 is that employees' responsiveness to the possibility of public scrutiny and judgment depends on the visibility of the negative externalities to the public. In particular, public scrutiny reduces the likelihood of whistleblowing when the damage done by fraud is visible to the public, and increases it when the opposite holds. This suggests that, in line with our expectations, the visibility of the negative externalities affects whistleblowers' expectations of how they will be perceived by the public. In particular, the observed patterns are in line with employees caring about social judgment directly and holding the plausible belief that the public disapproves of whistleblowing unless it is directed against managers that clearly generated harm to them. Hence, our second result:

Result 2: The possibility of social judgment increases (decreases) whistleblowing when the negative externalities are visible (invisible) to the public.

The estimates concerning the visibility of the negative externalities caused by fraud provide further insights into individual willingness to blow the whistle against a rule-breaking manager. In the absence of financial rewards and public scrutiny, employees' propensity to blow the whistle does not seem to differ across sectors and/or kinds of fraud that are characterized by different degrees of public awareness of the social

costs of the illegal behavior the manager engaged in. However, under the threat of social judgment of the whistleblower – i.e., when the identity of the whistleblower is not kept anonymous – whistleblowing is more likely to occur in sectors where the social costs of fraud are visible to public, and more so if rewards are offered to whistleblowers. In the absence of public scrutiny, the visibility of the externalities has no effect on whistleblowing. Our third result follows:

Result 3: The visibility of the negative externalities caused by fraud increases whistleblowing only when the whistleblower is subject to public scrutiny and judgment.

In the next subsection we explore the relationship between propensity to blow the whistle, public scrutiny and political orientation.

4.1.2 The Interaction between Political Orientation and Public Scrutiny

As discussed in Section 2.2, in our post-experiment survey, we collected data about our participants' political orientation. In columns 4 and 5 of Table 3 we controlled for political orientation and found no significant effects on the propensity to blow the whistle. Since the mechanisms through which political views determine whistleblowing may vary with our experimental treatments, in Table 4 we report marginal effects from multiple separate probit models in which we look at right-leaning subjects and left-leaning subjects separately. In columns 2 and 4 we add the same set of controls employed in Table 3.

The estimated marginal effects suggest that motivations to blow the whistle vary with political orientation conditional on public scrutiny. In particular, right-leaning subjects seem to respond only to monetary incentives. On the other hand, the behavior of left-leaning individuals reflects the results highlighted in Section 4.1.1, i.e., the fact that public scrutiny affects the propensity to blow the whistle positively if the negative externalities of fraud are visible to the public and negatively if they are not. This suggests that left-leaning individuals are more concerned about social approval and, at the same time, expect the public, at least the part of it whose opinion they care about, to generally disapprove of whistleblowing when it is unaware of the negative externalities suffered because of law-breaking behavior, and approve of it when such externalities are known. While we cannot directly test for this with our existing data since we did not elicit employees' beliefs about social disapproval of their actions, the analysis of the social messages sent by the public (Section 4.2) partly confirms this interpretation of our results.

Table 4
Political orientation and response to treatments

	Dep. Variable:			
	Dummy equal to 1 if employee blew the whistle, 0 otherwise			
	Right-leaning	Right-leaning	Left-leaning	Left-leaning
	(1)	(2)	(3)	(4)
Rewards	0.76 (0.244)	0.85* (0.087)	1.34*** (0.000)	1.24** (0.021)
Public Scrutiny	-0.66 (0.331)	-0.63 (0.290)	-4.43*** (0.000)	-4.76*** (0.000)
Visible Externalities	0.24 (0.737)	1.10 (0.279)	0.11 (0.625)	0.41 (0.251)
Visible*Reward	-1.39 (0.174)	-2.01* (0.052)	-1.37*** (0.001)	-1.44** (0.038)
Visible*Public	1.09 (0.208)	-0.02 (0.988)	3.78*** (0.000)	3.73*** (0.000)
Public*Reward	0.44 (0.562)	0.61 (0.330)	2.86*** (0.000)	3.66*** (0.001)
Visible*Public*Reward	0.20 (0.863)	1.04 (0.425)	-1.24 (0.127)	-0.82 (0.486)
Constant	-0.67 (0.239)	4.69 (0.187)	-0.84*** (0.000)	-11.42*** (0.001)
Controls	No	Yes	No	Yes
Observations	63	63	81	81

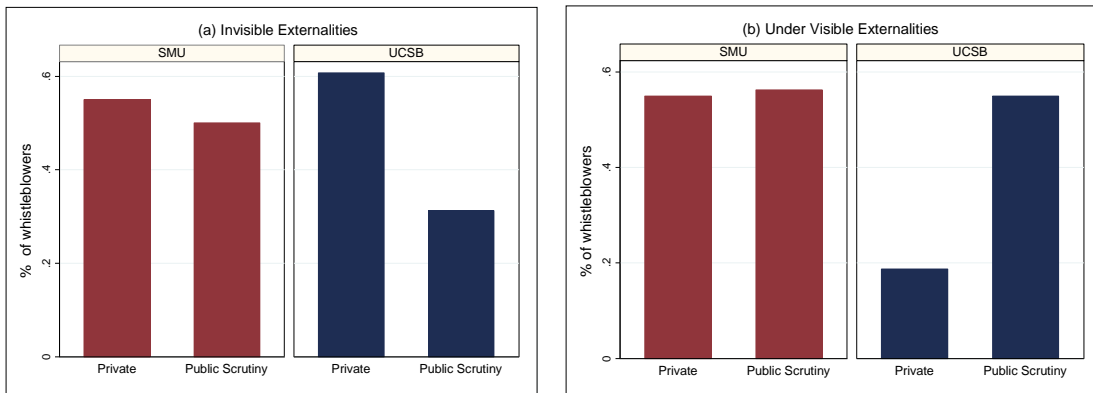
Note: The table reports marginal effects (dprobit). Robust standard errors are clustered at the session level. P-values in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Our finding concerning the differential impact of public scrutiny on left-leaning and right-leaning populations was unexpected, and suggests that our general results might be influenced by the relatively high number of left-leaning students at UCSB. We therefore searched for validation using a different sample of students characterized by predominantly right-leaning political views. We conducted a subset of our treatments at Southern Methodist University in Dallas, Texas. Since financial rewards had the same effect on the behavior of right- and left-leaning individuals in the UCSB sample, we only conducted treatments where financial rewards were present. As a result, at SMU we implemented a 2 by 2 design, varying public scrutiny and the visibility of externalities only, as shown in Table 5. We conducted 2 sessions per treatment, with between 3 and 5 firms per session, involving a total of 153 SMU students.

Table 5
SMU Sessions and Treatments

Treatments	Invisible Externalities		Visible Externalities		Total	
	Sessions	Subjects	Sessions	Subjects	Sessions	Subjects
Rewards & Private Whistleblowing	2	42	2	42	4	84
Rewards & Public Scrutiny	2	33	2	36	4	69
Total	4	75	4	78	8	153

Figure 3
The effect of public scrutiny on whistleblowers: SMU vs. UCSB



In line with our expectations, the SMU sample significantly differ from the UCSB population in terms of their political orientation. Only 29% of the SMU subjects in the role of employee (33% in the full sample) are left-leaning, versus 56% of the UCSB employees ($p=0.000$). If the responsiveness to public scrutiny depends on political orientation, we should observe public scrutiny to have less of an impact on whistleblowing in the SMU sample. This is clearly shown in Figure 3, which compares the responsiveness of SMU and UCSB students to public scrutiny under visible and invisible externalities.

Table 6 reports estimates from probit regressions on the probability of whistleblowing in the SMU sample. The small sample size prevents us from conducting the analysis separately for left-leaning and right-leaning subjects. Instead, in columns 2 and 3, we interact the public scrutiny and visible externality treatment dummies with our measure of political orientation. The estimates in column 1 show that in the aggregate, both public scrutiny and the visibility of the externalities, and their interaction, have no impact on whistleblowing. However, the estimates in columns 2 and 3 reveal that these null results are caused by the behavior of the right-leaning subjects that constitute the majority of the sample. When interacting treatments

and political orientation, we find that, as expected, public scrutiny does not affect right-leaning subjects' decision to blow the whistle both under visible and invisible externalities. Left-leaning people are not more likely than right-leaning people to blow the whistle when whistleblowing is private and the negative externalities are invisible to the public. They are however less likely to blow the whistle under public scrutiny when the negative externalities of fraud are invisible to the public and more likely to blow the whistle when the externalities are visible to the public. These findings confirm the results obtained in the UCSB sample. We can therefore state our fourth result:

Result 4: Political orientation significantly impacts the effect of public scrutiny on whistleblowing: right-leaning subjects respond only to monetary incentives, while left-leaning subjects respond to public image concerns

Table 6
Political orientation and response to treatments – SMU sample

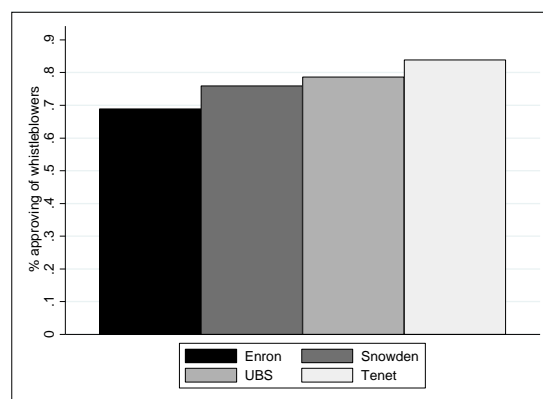
	Dep. Variable: Dummy equal to 1 if employee blew the whistle, 0 otherwise		
	(1)	(2)	(3)
Public Scrutiny	-0.13 (0.775)	-0.22 (0.676)	-0.69 (0.371)
Visible Externalities	0.00 (1.000)	-0.36 (0.463)	-0.54 (0.327)
Public x Visible	0.16	0.01 (0.993)	0.43 (0.639)
Left-leaning		-0.75 (0.205)	-0.69 (0.279)
Left x Public Scrutiny		-4.54*** (0.000)	-5.19*** (0.000)
Left x Visible Externalities		1.11 (0.271)	1.65* (0.091)
Left x Public x Visible		4.89*** (0.000)	4.75*** (0.001)
Constant	0.13 (0.657)	0.43 (0.253)	-2.47 (0.419)
Controls	No	No	Yes
Observations	70	70	69

Note: The table reports marginal effects for continuous variables and the effect of a change from 0 to 1 for dichotomous variables. Controls are: age, gender, economics major, left-leaning political preferences, number of firms in the session, ratio between firm performance and own performance in team building task, and effort chosen in minimum effort task. P-values in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

4.2 Social Judgment of Whistleblowers

A central hypothesis of our study is that individuals' expectations of social approval or disapproval from the general public may have a significant impact on their decision to blow the whistle against manager wrongdoing that advanced the firm at the expense of the general public. Our finding with respect to the differential responsiveness to public scrutiny conditional on the visibility of the negative externalities to the public suggests that expectations of positive or negative social judgment are indeed important. In this section we investigate the social judgment of whistleblowers under different conditions. We start by analyzing individual answers to post-experiment survey questions eliciting opinions on the social appropriateness or inappropriateness of actual whistleblowing cases. We then analyze the messages sent to whistleblowers by the members of the public in our social scrutiny treatments.

Figure 4
Social judgment of four whistleblowing cases (survey)



As part of our post-experiment survey, all study participants were presented with four actual whistleblowing cases – the Snowden case, the Enron case, the UBS case and the Tenet case – and asked to evaluate the social appropriateness of each case. As discussed in section 2.2, we chose these cases because they vary in the visibility of the negative externalities that illegal behavior caused to the public and in the presence of financial rewards for the whistleblower. The social costs of the unlawful actions unmasked by the whistleblower are clearly visible in the Snowden and the Tenet case, less visible in the UBS case and almost invisible in the Enron case. Moreover, financial rewards were present in the UBS and Tenet cases and not in the Enron and Snowden cases. In order to minimize ordering effects, the four cases were presented in the above order but not one after the other. Subjects first were presented with the Snowden case, then answered a number of unrelated questions collecting demographics and attitudinal preferences, then saw the Enron

case, followed by more unrelated questions. The UBS case came afterwards, followed by more questions before the appearance of the Tenet case. For each whistleblowing scenario, we provided a summary of the case and we asked subjects to rank the appropriateness of the whistleblower’s decision.

Figure 4 reports the percentages of survey participants stating that the decision made by the whistleblower is socially acceptable. The social acceptability of whistleblowing is lowest in the Enron case and highest in the Tenet case. Pairwise comparisons between cases suggest that both the visibility of the externalities and the presence of financial rewards increase the social acceptability of the whistleblowing act. Naturally, this is only suggestive evidence. In order to more scientifically evaluate attitudes toward whistleblowers under different conditions we analyze the messages that the members of the public sent to whistleblowers in our public scrutiny treatments.

Table 7
Percentage of members of the public sending smiley faces to whistleblowers (UCSB)

	No Rewards	Rewards
Invisible Externalities	58.33%	83.33%
Visible Externalities	33.33%	72.22%
H ₀ : Rewards = No Rewards if Visible==0	p-value = 0.178 (0.185)	
H ₀ : Rewards = No Rewards if Visible==1	p-value = 0.035 (0.042)	
H ₀ : Visible = Invisible Ext. if Rewards=0	p-value = 0.219 (0.207)	
H ₀ : Visible = Invisible Ext. if Rewards=1	p-value = 0.481 (0.403)	
Note: p-values generated by Chi-square tests. P-values from Fisher exact tests in parentheses.		

Overall, across all treatments, 15% of members of the public decided to send no message to the whistleblowers, 63% sent a message of approval, 6% sent a message of disapproval and the remaining 17% sent a neutral message. Table 7 reports the percentages of members of the public that sent a message of approval under the different treatment manipulations. The presence of rewards leads to social approval, especially when the negative externalities caused by the managers’ illegal activities are visible to the public. In contrast, the visibility of the externalities per se does not seem to affect approval of whistleblowers.

In Table 8, we conduct probit regressions on the probability to send a message of approval as opposed to a neutral message or a message of disapproval. In the first column we only include our treatment dummies, whereas in columns 2 and 3 we employ the same set of controls as in Tables 3, 4 and 6. In column 3, we also interact our political orientation variable with the visible externalities treatment dummy. The estimates

confirm that the presence of financial rewards to the whistleblower increases the probability that the public will approve of the whistleblower, while the visibility of the social cost of fraud does not. We also find that left-leaning subjects are less likely to send a message of approval to a whistleblower when they are unaware of the negative externalities generated by the manager’s illegal behavior. The coefficient of the interaction between the left dummy and the visibility treatment suggests that the propensity of left-leaning subjects to send a smiley face to the whistleblowers increases when the negative externalities are visible to the public, although we do not have enough power for the estimated coefficient to be statistically significant. Our final result follows:

- Result 5:** (a) Financial rewards increase the social approval of the whistleblower.
 (b) Political orientation impacts the social approval of the whistleblower, conditional on the visibility of the externalities caused by fraud to the public.

Table 8
The decision to approve of a whistleblower

	Dep. Variable: Dummy equal to 1 if the MP sent a message of approval, 0 otherwise		
Rewards	0.30** (0.029)	0.39** (0.026)	0.51** (0.013)
Visible Externalities	-0.17 (0.196)	-0.12 (0.403)	-0.28 (0.183)
Left-leaning			-0.38** (0.043)
Left x Visible Externalities			0.25 (0.247)
Controls	No	Yes	Yes
Observations	46	46	46

Note: We report marginal effects for continuous variables and the effect of a change from 0 to 1 for dichotomous variables. Controls are: age, gender, economics major, number of firms in the session, earnings in real effort task. None of the controls are significant. P-values in parentheses. *** p<0.01, ** p<0.05, * p<0.1

4.3 Manager’s Law-Breaking Behavior

Our experiment was meant primarily to investigate employees’ decision to blow the whistle against their manager. As a consequence, our sample of manager is quite small, with a total of 72 observations. A total of 11% of managers decided to break the law to double the firm fund at the expenses of the members of the public. The occurrence of cheating varies across treatments, as shown in Figure 5 and Table 9. A clear pattern we see in the data is the reduction in manager illegal behavior when there exist financial rewards

for whistleblower, suggesting that the manager correctly predicts the effect of rewards on employees' willingness to report wrongdoing. Managers seem also less willing to break the law when the public is made aware of the negative externalities generated by fraud. However, the small sample size prevents us from finding statistically significant differences in manager behavior across treatments. Regression analysis¹⁶ provides evidence of the impact of the manager's skills on the probability of breaking the law. In particular, the better the manager performance in the multiplication task subjects engaged in during Stage 1 of the experiment, the lower the probability that the manager will decide to cheat to augment the firm fund.

Figure 5
Manager's rule-breaking behavior

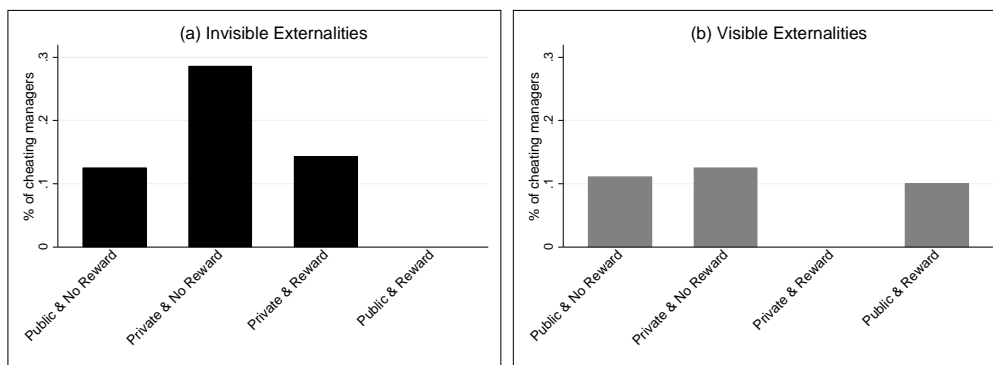


Table 9
Percentages of managers breaking the law

	No Rewards and Private Whistleblowing	No Rewards and Public Scrutiny	Rewards and Private Whistleblowing	Rewards and Public Scrutiny
Invisible Externalities	28.57%	12.50%	8.33%	6.67%
H ₀ : Private = Public	p-value = 0.438 (0.446) if rewards=0		p-value = 0.849 (0.674) if rewards=1	
H ₀ : Rewards = No Rewards	p-value = 0.16 (0.212) if private=1		p-value = 0.635 (0.585) if private=0	
Visible Externalities	12.50%	11.11%	0.00%	5.56%
H ₀ : Private = Public	p-value = 0.929 (0.735) if rewards=0		p-value = 0.310 (0.500) if rewards=1	
H ₀ : Rewards = No Rewards	p-value = 0.126 (0.308) if private=1		p-value = 0.603 (0.564) if private=0	

Note: p-values generated by Chi-square tests. Fisher tests in parentheses. The decline observed when the externalities are visible is also not statistically significant.

¹⁶ The corresponding table is not reported here. It is available from the authors upon request.

5. Conclusion

Our study contributes to the policy debate and growing literature on the motivations and incentives of employees blowing the whistle on corporate fraud. Despite whistleblowing cases have been filling the covers of popular journals in recent years, their occurrence is rare and the vast majority of white-collar crimes remains undetected and unpunished. In this paper, we examined two policies that may motivate employees to blow the whistle on white-collar crimes: the use of financial rewards and the protection(exposure) of whistleblowers' identity from(to) public scrutiny. We also examined the interaction between these two sources of whistleblowing incentives and tested whether financial rewards may crowd-out non-monetary motivations. Finally, we asked whether different policies should be used for different cases of fraud or different industries, depending on whether the public feels directly affected by the negative externalities generated by the illegal activities undertaken within the organization. We employed a specially designed economics experiment, which allowed us to observe willingness to break the law, willingness to blow the whistle on rule breaking, and public reaction to whistleblowing. Crucially, in our setting, manager wrongdoing caused financial losses to 'real' third parties, potential whistleblowers did not take part in the illegal activities but benefited from them, and whistleblowing was costly.

We found strong evidence of the effectiveness of financial rewards on whistleblowing. We did not find any evidence of crowding out of non-monetary image-driven motivations to blow the whistle, as financial rewards are effective both when the whistleblower is shielded from public scrutiny and when he or she is not. Our findings with respect to the relationship between whistleblowing and public scrutiny show that the possibility of social judgment may either act as an incentive or a disincentive to blow the whistle. It acts as an incentive in cases of fraud where the public feels directly affected by the negative externalities caused by corporate fraud and as a disincentive when the opposite holds. This suggests that, in order to maximize whistleblowing, industries and corresponding cases of fraud should be classified based on the negative effects they have on the public, and different policies should be adopted, either protecting or exposing the identity of whistleblowers.

Overall, our results confirm previous research on the effectiveness of financial rewards on whistleblowing and provides novel insights on the interaction between extrinsic incentives and whistleblowers' image motivations. Even more novel is our finding of the importance of social approval or disapproval for the decision to report corporate wrongdoing. Future research could extend our analysis in multiple interesting directions. For instance, it could test whether our results apply also to "traitorous" whistleblowing, i.e., cases of fraud where the potential whistleblower took active part in the illegal activities, and whether making the punishment of the manager probabilistic rather than deterministic significantly alters employees' reporting rates and responsiveness to treatments. Even more interesting would be to incorporate

collective action problems in the decision to blow the whistle, as beliefs about other employees' reporting decisions may significantly affect individuals' willingness to blow the whistle.

References

- Abbink, Klaus & Danila Serra (2012). "Anticorruption Policies: Lessons from the Lab," in Danila Serra & Leonard Wantchekon, eds., *New Advances in Experimental Research on Corruption* (Emerald Books), pp. 77-115.
- Abbink, Klaus & K. Wu. (2013). Reward Self-Reporting to Deter Corruption: An Experiment on Mitigating Collusive Bribery. Monash University, Discussion Paper 42/13.
- Abbink, Klaus, Utteeyo Dasgupta, Lata Ghanghadaran & Tarun Jain. (2014). Letting the Briber Go Free: An Experiment on Mitigating Harassment Bribes. *Journal of Public Economics* 111: 17-28.
- Andreoni, J., & Bernheim, B. D., (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5), 1607-1636.
- Andreoni, J., & Petrie, R.. (2004). "Public goods experiments without confidentiality: a glimpse into fundraising." *Journal of public Economics*, 88(7), 1605-1623.
- Andreoni, Jame (1995). "Warm-Glow Versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments." *The Quarterly Journal of Economics*, vol. 110, no. 1, 1995, pp. 1–21.
- Apestequia J., M. Dufwenberg and R. Selten, (2007). "Blowing the Whistle," *Economic Theory*, vol. 31(1),143-166.
- Ariely, D., Bracha, A., & Meier, S., (2009). "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially." *The American Economic Review*, 99(1), 544-555.
- Aubert, C., Kovacic, W., and Rey P. (2006). "The Impact of Leniency and Whistleblowing Programs on Cartels," *International Journal of Industrial Organization*, 24: 1241-1266.
- Baloria, V. Marquardt, C. Wiedman, C. (2015). "A Lobbying Approach to Evaluating the Whistleblower Provisions of the Dodd-Frank Reform Act of 2010", unpublished working paper. Available at: <http://papers.ssrn.com/sol3/papers.cfm?abstract id=1923310>
- Bank of England Prudential Regulations Authority & Financial Conduct Authority. (2014). "Financial Incentives for Whistleblowers", available at: <https://www.fca.org.uk/news/financial-incentives-for-whistleblowers>.

- Bartuli, J., Djawadi, B., Fahr, R. (2016). "Business Ethics in Organizations: An Experimental Examination of Whistleblowing and Personality", IZA Discussion Paper No.10190.
- Benabou, Roland & Jean Tirole (2004). "Willpower and Personal Rules," *Journal of Political Economy* 112(4): 848-886.
- Bénabou, R., & J. Tirole (2006). "Incentives and prosocial behavior," *The American Economic Review*, 96(5), 1652-1678.
- Bigoni, M., C. Le Coq, S. Fridolfsson and G. Spagnolo, (2012). "Fines, Leniency and Rewards in Antitrust," *RAND Journal of Economics*, 43(2), 368-390.
- Bigoni, Maria, Sven-Olof Fridolfsson, Chloé Le Coq and Giancarlo Spagnolo, (2015). "Trust, Leniency and Deterrence", *Journal of Law, Economics and Organization* 31(4), 663-689.
- Branas-Garza, Pablo (2007). "Promoting helping behavior with framing in dictator games," *Journal of Economic Psychology*, Elsevier, vol. 28(4), pages 477-486.
- Breuer, L. (2013). "Tax Compliance and Whistleblowing: The Role of Incentives," *The Bonn Journal of Economics*, Vol 2, No. 2, pp. 7-44.
- Carson, T. Verdu, M. Wokutch, R. (2008). "Whistle-Blowing for Profit: An Ethical Analysis of the Federal False Claims Act", *Journal of Business Ethics*, Vol. 77, No. 3, pp. 361-376.
- Cookson, Richard, (2000). "Framing effects in public goods experiments." *Experimental Economics* 3.1: 55-79.
- Cotten S. and Santore R. (2016). "Whistleblowers, Amnesty, and Managerial Fraud: An Experimental Investigation." Unpublished manuscript.
- Dufwenberg, Martin, Simon Gächter, and Heike Hennig-Schmidt (2011). "The framing of games and the psychology of play." *Games and Economic Behavior* 73.2 (2011): 459-478.
- Dworkin, T.M. and J.P. Near (1997). "A Better Statutory Approach to Whistleblowing," *Business Ethics Quarterly* , 7: 1-16.
- Dyck, Alexander, Adair Morse, and Luigi Zingales (2010). "Who Blows the Whistle on Corporate Fraud?" *The Journal of Finance*, 65, 2213-53.
- Dyck, I. J., Adair Morse, and Luigi Zingales (2013). "How pervasive is corporate fraud?" *Rotman School of Management Working Paper* 2222608.
- Engstrom, David F. (2012). "Harnessing the Private Attorney General: Evidence from Qui Tam Litigation" *Columbia Law Review*, 112, 1244-1325.
- Engstrom, David F. (2013). "Public Regulation of Private Enforcement: Empirical Analysis of DOJ Oversight of Qui Tam Litigation Under the False Claims Act," *North-western University Law Review*, 107, 1689-1756.

- Engstrom, David F. (2014). "Private Enforcement's Pathways: Lessons from Qui Tam Litigation." *Columbia Law Review*, 115, 1913-2006.
- Engstrom, David F. (2014). "Whither Whistleblowing? Bounty Regimes, Regulatory Context, and the Challenge of Optimal Design." *Theoretical Inquiries L*.15 605
- Engstrom, David F. (2016). "Bounty Regimes." in *Research Handbook on Corporate Criminal Enforcement and Financial Misdealing* (Jennifer Arlen ed., Edward Elgar Press)
- Eckel, Catherine C., and Philip J. Grossman (1996). "Altruism in anonymous dictator games." *Games and Economic Behavior* 16: 181.
- Frey, Bruno S., and Felix Oberholzer-Gee (1997). "The cost of price incentives: An empirical analysis of motivation crowding-out." *The American Economic Review* 87.4 (1997): 746-755.
- Fehr, E., & U. Fischbacher (2004). "Third-party punishment and social norms." *Evolution and human behavior*, 25(2), 63-87.
- Fehr, Ernst, and Simon Gächter (2002). "Do Incentive Contracts Undermine Voluntary Cooperation?" Working Paper 34, Institute for Empirical Research in Economics, University of Zurich.
- Fehr, E. and B. Rockenbach (2003). "Detrimental effects of sanctions on human altruism." *Nature* 422, 137-140.
- Fehr, Ernst, and John A. List (2004). "The Hidden Costs and Returns of Incentives—Trust and Trustworthiness among CEOs." *Journal of the European Economic Association*, 2(5): 743-71.
- Feldman, Y., Lobel., O. (2010). "The Incentives Matrix: The Comparative Effectiveness of Rewards, Liabilities, Duties and Protections for Reporting Illegality", *Texas Law Review*, Vol 87
- Felli, L. and R., Hortala-Vallve (2016). "Collusion, Blackmail and Whistle-Blowing," *Quarterly Journal of Political Science*, 11(3): 279-312.
- Friebel, Guido and Sergei Guriev (2012). "Whistle-Blowing and Incentives in Firms," *Journal of Economics & Management Strategy*, 21(4), 1007-1027.
- Fischbacher, Urs (2007). "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10.2 (2007): 171-178.
- GAO-11-619 (2011). U.S. Gov't Accountability Office Report 11-619, Criminal Cartel Enforcement: Stakeholder Views on Impact of 2004 Antitrust Reform Are Mixed, but Support Whistleblower Protection, available at <http://www.gao.gov/new.items/d11619.pdf>.
- Gerber, A. S., Green, D. P., & Larimer, C. W., (2008). "Social pressure and voter turnout: Evidence from a large-scale field experiment." *American Political Science Review*, 102(01), 33-48.
- Gneezy, Uri, Stephan Meier, and Pedro Rey-Biel (2011). "When and why incentives (don't) work to modify behavior." *The Journal of Economic Perspectives* 25.4 (2011): 191-209.

- Gneezy, Uri, and Aldo Rustichini. (2000a). "Pay Enough or Don't Pay At All." *Quarterly Journal of Economics*, 115(3): 791–810.
- Gneezy, Uri, and Aldo Rustichini. (2000b). "A Fine Is a Price." *Journal of Legal Studies*, 29(1): 1–18.
- Greiner, Ben (2015). "Subject pool recruitment procedures: organizing experiments with ORSEE." *Journal of the Economic Science Association* 1.1 (2015): 114-125.
- Hamaguchi, Yasuyo, Toshiji Kawagoe and Aiko Shibata, (2009). "Group Size Effects on Cartel Formation and the Enforcement Power of Leniency Programs", *International Journal of Industrial Organization* 27(2), 145-165.
- Howse, R., & Daniels, R. J. (1995). "Rewarding Whistleblowers: The Costs and Benefits of an Incentive-Based Compliance Strategy," 525-549. Retrieved from http://repository.upenn.edu/law_series/4.
- Heyes, Anthony and Sadeep Kapur, (2009). "An Economic Model of Whistle-Blower Policy," *Journal of Law, Economics, & Organization*, 25(1), 157-182.
- Hinloopen, Jeroen and Soetevent, Adriaan (2008b). "Laboratory Evidence on the effectiveness of corporate leniency programs", *RAND Journal of Economics*, 39(2), 607-616.
- Linardi, S., & McConnell, M. A., (2011). "No excuses for good behavior: Volunteering and the social environment." *Journal of Public Economics*, 95(5), 445-454.
- Miceli, M. P. and Janet P. Near (1992). "Blowing the Whistle: The Organizational and Legal Implications for Companies and Employees," Lexington Books.
- Miceli, M. P. Janet P. Near, and Terry Morehead Dworkin, (2008). "Whistle-blowing in organizations," New York: Routledge.
- Rege, Mari, and Kjetil Telle (2004). "The impact of social approval and framing on cooperation in public good situations." *Journal of Public Economics* 88.7 (2004): 1625-1644.
- Reuben, Ernesto, and Matt Stephenson (2013). "Nobody likes a rat: On the willingness to report lies and the consequences thereof." *Journal of Economic Behavior & Organization* 93 (2013): 384-391.
- Schmolke, Klaus Ulrich and Utikal, Verena, (2016), "Whistleblowing: Incentives and Situational Determinants," FAU - Discussion Papers in Economics, No. 09/2016. Available at SSRN: <https://ssrn.com/abstract=2820475> or <http://dx.doi.org/10.2139/ssrn.2820475>.
- Serra, D. (2012). "Combining top-down and bottom-up accountability: evidence from a bribery experiment." *Journal of Law, Economics, and Organization*, 28(3), 569-587.
- Serra, D. and L. Wantchekon (2012). *New advances in experimental research on corruption*. Emerald, United Kingdom.
- Shikora, J. (2011). "Bringing good and bad Whistle-blowers to the Lab", Munich Discussion Paper No. 2011- 4, online: <http://epub.ub.uni-muenchen.de/12161/>.

- Spagnolo, Giancarlo, (2004). "Divide et Impera: Optimal Leniency Programs," CEPR Discussion Papers 4840.
- Spagnolo, Giancarlo, (2008). "Leniency and Whistleblowers in Antitrust," Ch.12 of P.Buccirossi (Ed.) Handbook of Antitrust Economics, 2008, M.I.T. Press.
- Titmuss, Richard M. (1970) "The gift relationship." *London* 19 (1970): 70.
- Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). "Social categorization and intergroup behaviour." *European Journal of Social Psychology*, 1(2), 149-178.
- Waytz, A., Dungan, J., and Young, L. (2013). "The whistleblower's dilemma and the fairness-loyalty tradeoff." *Journal of Experimental Social Psychology*, Vol. 49, pp. 1027-1033.
- Xiao, E., & D. Houser (2005). "Emotion expression in human punishment behavior." *Proceedings of the National Academy of Sciences of the United States of America*, 102(20), 7398-7401.
- Xiao, E., & D. Houser (2011). "Punish in public." *Journal of Public Economics*, 95(7), 1006-1017.
- Zingales, L. (2004) "Want to Stop Corporate Fraud? Pay Those Whistle-Blowers," *Washington Post*, January 18.

Appendix

Experiment Instructions

General Instructions

Thank you all for coming today. You are here to participate in an experiment. In addition to a \$10 participation fee, you may earn substantially more money from today's experiment. You will be paid privately and anonymously in cash at the end of your experimental session today.

Today's experiment consists of multiple stages. Separate instructions for each stage will appear on your computer screen at the beginning of each stage. You will have the chance to earn money in each stage of the experiment except the last stage, which will be a questionnaire. Earnings during the experiment will be denominated in Experimental Currency Units, or **ECU**. At the end of the session one of the remunerated stages of the experiment, i.e., not the questionnaire, will be randomly selected for payment.

Your earnings in the randomly selected stage will be converted to dollars at the exchange rate of:

$$\underline{2 \text{ ECU} = \$1}$$

After everybody has completed the final questionnaire, you will be paid the money you earned from the selected stage of the experiment **plus your participation fee of \$10**.

If you have any questions during the experiment, please raise your hand and wait for an experimenter to come to you. Please do not talk, exclaim or try to communicate with other participants during the experiment. Participants intentionally violating these rules may be asked to leave the experiment and may not be paid.

Please read and sign the Consent Form that you have been provided. Please raise your hand if you have any questions about any of the information on the Consent Form. We will proceed with the experiment once we have collected all signed consent forms.

Below we attach screenshots from Stage Three of the experiment, the Whistleblowing Game.

PUBLIC SCRUTINY and VISIBLE EXTERNALITIES

Member of the Public

For the next stage of today's experiment you will still be a member of the Public. All 6 members of the public from the previous stages will keep their roles. As before, there will be 2 firms, each made up of the same three people as in previous stages.

You will have the opportunity to engage in a task that will generate earnings for yourself. You will receive additional information as this stage of the experiment continues.

Click OK when you are ready to start this stage of today's experiment.

OK

Member of a Firm

For the next stage of today's experiment you will still be a member of Firm Green, as will the other two participants who were members of your firm in previous stages. Membership in each other firm will remain the same as well. Participants who were members of the public in previous stages will continue to be members of the public.

Within each firm, each firm member will be randomly assigned either the role of **manager** or the role of **employee**.

In each firm, one firm member will be randomly assigned the role of **manager**, while the other two members will be assigned the role of **employee**.

You will then have the opportunity to engage in a task that will generate earnings for yourself and for the other members of Firm Green.

You will receive additional information as this stage of the experiment continues.

Click OK when you are ready to start this stage of today's experiment.

OK

Member of the Public

You are a member of the Public. You have an initial endowment of 14 ECU. You will now engage in a number-addition task. Each correct answer will generate private earnings for you.

The participants that have been assigned the role of firm member will engage in a similar task. In each of the 2 firms, two members have been randomly assigned the role of employee and will engage in an addition task, while the remaining member has been assigned the role of manager. The manager will be given a fixed wage and will be given the opportunity to engage in a multiplication task.

Press OK when you are ready to begin your task.

Manager

You have been randomly assigned the role of **manager** of Firm Red




The other two members of Firm Red have been assigned the role of **employees**. Each employee will now engage in a number-addition task. Each correct answer will generate private earnings to the employee and will also generate a **firm surplus** that will later be redistributed among all the members of Firm Red. In the meantime, each member of the public will also engage in an addition task that will only generate private earnings for themselves.

As the manager of Firm Red, you will get a **fixed payment of 24 ECU**. You will then engage in a number-multiplication task. If you **answer more than 7 multiplication problems correctly**, you will augment the firm surplus generated by the two employees of your firm by **100%** of its original value. For example, if the workers in your firm create 5 ECU in surplus, you could increase this to 2×5 . **Once finalized, 1/2 of the firm surplus will be distributed back to you, 1/4 will be distributed back to one of your employees, and 1/4 will be distributed back to the other employee.**

An alternative way that **You may augment the firm surplus is by breaking the law**. Breaking the law will automatically augment the firm surplus by 100%. However, it will also generate a **loss of 2 ECU to the earnings of each of the 6 members of the public**. The employees and managers of all the other firms will face an identical situation to the one you face.

Note that:

- One of the two employees will later be randomly chosen to learn of your choice and will be given the opportunity to **blow the whistle** if you decided to break the law.
- Blowing the whistle will cost the worker **5 ECU**, will generate a penalty of **14 ECU** to you and will earn the worker a reward of **10 ECU**.
- If you decide to **break the law**, the members of the public will be informed about the earnings they lost because of your law-breaking decision.

   NO MESSAGE

Please decide whether you would like to engage in the multiplication task or whether you would prefer to break the law.

Multiplication Task Break the law

Employee

You have randomly been assigned the role of **employee** in Firm Green. Another member of Firm Green has also been assigned the role of a employee, while the third member of your firm has been assigned the role of manager.

You and the other employee in Firm Green will now engage in a task that will generate earnings for yourself and all the other members of Firm Green. The task will be to add two numbers together. Each correct answer will generate private earnings to you and will also generate a **firm surplus** that will later be redistributed among you and the other members of Firm Green. In the meantime, each member of the public will also engage in an addition task that will only generate private earnings for himself or herself.

The **manager** of Firm Green will be given a fixed wage and will engage in a different task: a number multiplication task. If the manager answers more than 7 problems correctly, he or she will augment the **firm surplus** generated by you and the other employee by **100%** of its original value. For example, if the employees in your firm create S ECU in surplus, the manager could increase this to $2 * S$. Once finalized, 1/4 of the firm surplus will be distributed back to you, 1/4 will be distributed to the other employee, and 1/2 will be distributed to the manager.

Press OK when you are ready to begin your task.

Remaining Time 0

Please insert your answer into the space provided.

What is $23 + 73$?

Manager

Remaining Time 37

You chose to break the law.

While waiting for the participants to complete their tasks, you can still solve multiplication problems. However, the problems you solve will not generate earnings for you or the other members of your firm.

Please insert your answer into the space provided.

What is 35×20 ?

OK




Member of the Public

You are a member of the Public. While you engaged in your addition task, the **managers** of each of the 2 firms were given the chance to either engage in a multiplication task to augment the earnings of each member of their firm (including themselves), or to **break the law**.

Each manager knew that breaking the law would automatically augment the earnings of each member of the firm, but would also reduce your earnings as well as the earnings of every other member of the public by 2 ECU. You will soon be informed about the managers' decisions about whether to break the law.

Within each firm, both employees learned whether the manager broke the law. One employee was randomly selected to be given the chance to blow the whistle on the manager if the manager did indeed break the law. Blowing the whistle cost this randomly-selected worker 5 ECU, but also earned this employee a reward of 10 ECU. If the manager broke the law and the employee blew the whistle, the manager must pay a financial penalty of 14 ECU.

You are now given the chance to send a message to **the employees** that decided to blow the whistle on their manager if the manager broke the law. You can send one of the three messages below, or you can send no message at all. Please make your choice below.

   NO MESSAGE

Happy face Indifferent face Sad face No message

OK




Employee

You are an employee of Firm Green. You and the other employee generated a firm surplus of 0 ECU.

While you engaged in the addition task, your manager was given the chance to augment the firm surplus by 100% by engaging in a multiplication task and solving more than 7 problems correctly. Alternatively, instead of solving multiplication problems, the manager could have also augmented the firm surplus by breaking the law. If the manager chose to break the law, the manager knew that he or she would automatically augment the firm surplus, but that the earnings of each member of the public would be automatically reduced by 2 ECU. Note that the employees and managers of faced an identical situation.

You will now be given the chance to **blow the whistle** on your manager if he or she broke the law. Soon, either you or your co-employee will be randomly selected and, if your manager broke the law, the corresponding whistleblowing decision will be implemented.

Note that:
"Blowing the whistle will cost you 5 ECU and will generate a penalty of 14 ECU to your manager. Moreover, it will earn you a reward of 10 ECU.
"If your manager decided to break the law, the members of the public will be informed about the earnings they lost because of your manager' s law-breaking decision.
"If your manager decides to break the law and you decide to blow the whistle, each member of the public will be informed about your whistleblowing decision and will be given the chance to send you either no message or one of the messages below.

   NO MESSAGE

Please make your choice below.

If the manager of my firm broke the law:

I would like to blow the whistle I would like NOT to blow the whistle