

# Motivating Whistleblowers\*

Jeffrey V. Butler<sup>+</sup>

Danila Serra<sup>♦</sup>

Giancarlo Spagnolo<sup>§</sup>

February 2017

## Abstract

We experimentally investigate employees' decisions to blow the whistle on a manager whose law-breaking benefits the firm but harms society. We investigate the effects of both financial rewards and non-monetary incentives, in the form of public scrutiny, on whistleblowing as well as their interaction with the visibility of harm, i.e., whether the harm to society stemming from the manager's malfeasance is known to the general public. Our results suggest that: i) financial rewards substantially increase the likelihood of whistleblowing; ii) public scrutiny and social judgment increase (decrease) whistleblowing when the negative externalities generated by fraud are visible (invisible) to the public. Ancillary results suggest an intriguing relationship between political orientation and responsiveness to public scrutiny.

**JEL Codes:** K42, C92, D04.

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

---

\*We thank seminar participants at LUISS Guido Carli University, Southern Methodist University and the University of California, Merced as well as conference participants at the Bay Area Behavioral and Experimental Workshop, the Texas Experimental Symposium, the North-American ESA 2016 conference, the ASSA 2017 annual meeting and the 2017 Bank of England/CEPR conference on Competition and Regulation in Banking for useful comments and suggestions. We are grateful to the Wallander and Hedelius foundation for financial support.

<sup>+</sup> Louisiana State University, Department of Economics. Email: jeffreybutler@lsu.edu.

<sup>♦</sup> Southern Methodist University, Department of Economics. Email: dserra@smu.edu.

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

## Introduction

Corporate fraud is widespread around the world. A recent survey of over 6000 organizations across 115 countries (2016 Global Crime Survey)<sup>1</sup> 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.<sup>2</sup> Dyck, Morse and Zingales (2013) estimated that between 1996 and 2004 about 15% of large<sup>3</sup> publicly traded US corporations 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, by blowing the whistle employees could potentially play a crucial role in uncovering illegal behavior and initiating internal or external investigations. However, while particular cases of whistleblowing have garnered the attention of the popular press in recent years, from the Enron scandal to the Snowden and Wikileaks-related cases, whistleblowing by employees is actually 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. Given the high costs associated with blowing the whistle, ranging from coworkers' disapproval and ostracism to lack of career advancement, job loss and outright harassment (e.g., Miceli and Near, 1994; Rothschild and Miethe, 1999), this rarity is unsurprising.<sup>4</sup> On top of this, psychological costs caused by conflicting moral norms – loyalty toward the firm on the one hand and fairness or justice concerns on the other – may also make employees reluctant to report wrongdoing taking place within their organization (Waytz, Dungan and Young, 2013). Finally, fear of media scrutiny or public disapproval might further reduce employees' willingness to blow the whistle. Alternatively, if the expectation is of public *approval*, media or public scrutiny might actually increase whistleblowing, a possibility we discuss below.

Despite the increasing attention on existing whistleblowers and widespread awareness of the benefits (and costs) of whistleblowing, little is known about how to motivate individuals to report illegal activities taking place within an organization. In this paper, we experimentally investigate the effectiveness of different policies which might produce additional whistleblowing. We focus on both monetary and non-monetary

---

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

<sup>2</sup> 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.

<sup>3</sup> “Large” is defined by having assets exceeding 750 million dollars.

<sup>4</sup> Many of these forms of retaliation are sufficiently opaque to escape whistleblower protection laws (e.g., lack of promotion) and the Ethics Resource Center (2014) reports a steady increase across time in the percentage of whistleblowers facing retaliation.

incentives. In particular, we assess the effects of financial rewards and public scrutiny on employees' willingness to blow the whistle. Moreover, we ask whether different sectors or different kinds of fraud require different policies, depending on the extent to which the social costs generated by fraud are visible and salient to the public – consider the context of insider trading versus Medicare fraud – as suggested by recent legal theory (e.g. Engstrom 2014).

The question of whether whistleblowers should be financially rewarded is not new. In fact it is central to the regulatory debate following 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).<sup>5</sup> US agencies administering these rewards schemes consider them a great success.<sup>6</sup> The limited empirical research available suggests that whistleblower reward schemes like the ones implemented in the US are indeed effective motivators of whistleblowing. In their well-known study, Dyck, Morse, and Zingales (2010) calculated that in sectors where the False Claim Act allows does not allow employees to obtain a financial reward, corporate fraud is unveiled by employees in 14% of the cases, while this percentage more than doubles (41%) when the False Claim Act can be applied, a highly significant difference.<sup>7</sup>

On the other hand, across the Atlantic regulatory agencies remain strongly opposed to financially rewarding whistleblowers even though the new EU Directive on financial fraud introduced the possibility.<sup>8</sup> One primary concern, expressed by many critics, is that whistleblower rewards would not be effective in

---

<sup>5</sup> 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 an FCA lawsuit on behalf of the United States Government and, if successful, to obtain up to 30% of recoveries plus fines. Another early whistleblower reward scheme against tax evasion is the one run by the IRS, which was substantially strengthened in 2006.

<sup>6</sup> The 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 since 2007, with \$343 million brought back in 2013 and \$310 million in 2014 (IRS Report 2015).

<sup>7</sup> 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.

<sup>8</sup> 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 the possibility of rewarding whistleblowers, arguing 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." Even though a study by Dyck, Morse and Zingales contradicting this statement (and further described below) circulated since 2007 and was published in *The Journal of Finance* in 2010.

eliciting additional valuable information because they could crowd out non-monetary incentives to “do the right thing,” while at the same time being costly to implement. Whether financial rewards will generally “crowd out” ethically-based whistleblowing may hinge on judgments about the morality of whistleblowing which, in turn, might be based partially upon the personal returns attached to the reporting act.<sup>9</sup> Starting at least with Titmuss (1970), both theoretical and experimental research has shown that in some circumstances financial incentives may indeed crowd out other types of motivations and lead to perverse outcomes (e.g., Bursztyjn and Jensen, *forthcoming*; Fehr et al., 2001; Fehr and Gächter, 2001; Fehr and List, 2004; Frey, 1997; Gneezy and Rustichini, 2000a, 2000b). However, 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 non-monetary incentives associated with blowing the whistle.

The class of non-monetary incentives we focus on, in particular, is public scrutiny and the concomitant specter of social approval or disapproval. We aim to shed light on whether and under what circumstances whistleblower anonymity, i.e. protection from public scrutiny, stimulates or hinders reporting. 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 may depend on how whistleblowers expect to be perceived by the public: will they be seen as snitches or as heroes? The answer to this question may lie in how salient the social costs of manager malfeasance are to the public. Consequently, a second contextual feature we vary is whether the public is aware of the costs imposed on them by law-breaking. This distinction may be important both theoretically and empirically. For instance, in 1971 economist Daniel Ellsberg leaked the Pentagon papers concerning American involvement in Vietnam. He is widely viewed as a hero, which may be in part due to the salience of the (literal, physical) public harm associated with this controversial war. Much more divided is the public opinion on Edward Snowden, seen by few as a hero and by many as a traitor. Perhaps not coincidentally, the public harm revealed by Snowden is diffuse, distant and difficult to quantify. In the current study we ask whether expectations of public praise or condemnation depend on the visibility and salience of public harm and whether these expectations, in turn, interact with variation in anonymity to substantially affect whistleblowing. This is an open and policy-

---

<sup>9</sup> See e.g. Carson, Verdu and Wokutch (2008) and references therein.

relevant question: 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 requiring them to first 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 managers within a firm. In our basic set-up, managers have the chance 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. Employees, who are not victims but rather beneficiaries of the manager's illicit behavior, are given the option of blowing the whistle on their manager. Whistleblowing leads to the automatic imposition of a monetary penalty on the manager. Across treatments we manipulate the presence of both financial rewards for, and public scrutiny of, 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, such as an aversion to social disapproval, in some treatments participants assigned the role of member of the public are allowed to send costless judgmental messages to 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 whistleblower rewards: in most of the situations we study, employees are significantly more likely to blow the whistle when doing so entails a personal financial gain. Our findings on how public scrutiny affects whistleblowing confirm 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. Together, these patterns are consistent with whistleblowers having an aversion to social disapproval and with the idea that the visibility of the social costs of fraud affects whistleblowers' expectations of how they expect to be judged by the public. Contrary to the crowding-out hypothesis, when whistleblowers are subject to public scrutiny we do not find that financial rewards are generally 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. As an

interesting ancillary result, we find that political orientation significantly affects employees' responsiveness to monetary and non-monetary incentives: while both right-leaning and left-leaning subjects respond to financial incentives, only left-leaning participants seem to be concerned about social approval or disapproval.

Overall, our investigation provides novel and important insights about the design and implementation of whistleblowing policies. Our findings suggest that financially rewarding whistleblowers is broadly effective and therefore generally desirable; that crowding out of non-pecuniary motivations for blowing the whistle in most situations should not be a reason for concern; and that protecting the identity of whistleblowers may be desirable in industries where the social costs of corporate misbehavior are less transparent or salient to the general public.

The remainder of the paper is organized as follows. Section 2 reviews the literature on whistleblowing. Section 3 describes the experiment and presents our hypotheses. Section 4 reports our empirical findings and Section 5 concludes.

## **1. Literature Review**

The economic analysis of whistleblowing has taken off in recent years, partially because the definition of whistleblowing is broad and encompasses a large set of illegal behaviors and reporting mechanisms. Any individual that exposes unlawful, dishonest or simply unethical behavior within an organization, public or private, is considered a whistleblower. To put our study in context, it is important to make three distinctions. The first is between “watchdog” and “traitorous” whistleblowers (Spagnolo, 2008; Breuer, 2013), where the former are innocent bystanders, while the latter are accomplices that betray their partners in crime. A further distinction is between “internal” and “external” whistleblowers, 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, e.g., 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 blowing the whistle on a public official. In this paper, we focus on watchdog whistleblowing by subordinate employees on the illegal activity of a team leader, a manager or a CEO, to some oversight authority that will punish the CEO for the unveiled wrongdoing with certainty. Our setting is therefore closer to the case of external whistleblowing, although it also encompasses internal whistleblowing in organizations with credible internal audit functions that are not captured by the management. We focus on the effectiveness of and interaction between financial and non-financial incentives, in the form of social approval or disapproval, on reporting decisions.

Theoretical economic analyses of (traitorous) whistleblower rewards programs begin with Spagnolo (2004), who constructs a dynamic model of collusion capturing the strategic features of illegal relationships with hold-up problems within the criminal team. He shows that offering rewards to the first whistleblower can lead to a first-best outcome: full deterrence with zero inspection probability. Aubert et al. (2006) focuses on within-firm employee whistleblowing, accounting for additional deterrence but also for several inefficiencies whistleblowing may generate. Friebe and Guriev (2012) study rewards for watchdog whistleblowers and show that rewards may negatively impact firm efficiency.<sup>10</sup> Felli et al. (2016) show how rewarding whistleblowing may prevent opportunistic behavior (collusion or blackmail) within a hierarchical structure. Finally, Heyes and Kapur (2008) derive optimal penalties for wrongdoing and optimal governmental responsiveness to whistleblowing under different assumptions about the intrinsic motivations of potential whistleblowers.

Empirical studies are rare and typically suffer from fundamental measurement and identification challenges: only illegal behavior that has been uncovered and only whistles that have been blown can be observed. Consequently, existing studies focus on either the infringements that have been discovered (e.g., Dyck et al., 2010) or on scenario-based survey data (e.g., Feldman and Lobel, 2010). Measurement and identification issues have led to a recent surge of experimental studies on whistleblowing. Laboratory experiments are particularly valuable to study deterrence of crimes, as they allow researchers to directly observe both wrongdoing and whistleblowing, and to measure responsiveness to changes in incentives in a controlled environment.

The experimental study most closely related to ours is by Schmolke and Utikal (2016), who 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 restore initial conditions. The authors study the effects of rewards for, 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. They find that even modest monetary rewards increase the probability of whistleblowing.<sup>11</sup> 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

---

<sup>10</sup> 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.

<sup>11</sup> 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.

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 sanctioned than when groups are randomly assigned. 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.<sup>12</sup> 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 over loyalty. Using survey questions, the authors find evidence of a significant loyalty/fairness tradeoff in the decision to blow the whistle. Differently from all of these studies, we employ a strongly framed laboratory experiment to investigate how financial and non-pecuniary incentives affect an employee's decision to blow the whistle on the wrongdoing of a manager in cases of wrongdoing that generate costs to the public while benefiting firm, including the employee.

More tangentially related to our study is the well-developed literature on (traitorous) whistleblowing in the context of illegal cartel formation among firms. Apesteguia, Dufwenberg and Selten (2007) were the first to study leniency and rewards to whistleblowers in an experiment on illegal cartel formation in the context of static Bertrand competition. Their results suggest that rewarding whistleblowers increases the likelihood of whistleblowing without reducing market prices. In a repeated game version of an analogous experiment, Bigoni et al. (2012) find that offering a monetary reward to the first whistleblower leads to high reporting rates that strongly deter cartel formation as predicted by theory (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), Hinloopen and Soetevent (2008), Bigoni et al. (2015), and Cotten and Santore (2016), among others.

Somewhat less directly related to our study is another growing strand of experimental literature that investigates whistleblowing in the context of corrupt transactions between public officials and citizens/firms. For instance, Abbink and Wu (2017) simulate both one-shot and repeated transactions between firms and public officials where firms can get illegal services through the payment of a bribe and find that whistleblower amnesty and monetary rewards strongly deter illegal transactions in a one-shot

---

<sup>12</sup> 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).



setting but that deterrence is limited in repeated relationships. Abbink et al. (2014), Buckenmaier et al. (2017), Schikora (2011) and Serra (2012) find similar results with amnesty alone. Breuer (2013) studies the effects of financial rewards for whistleblowers in a laboratory experiment on tax evasion and 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.

In sum, the existing literature, whether it investigates “watchdog” 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, or on the deterrence effects of whistleblowing on wrongdoing. Our contribution to the literature is threefold. First and foremost, we examine how social image concerns, operating through 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 how image motivations affect 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; see also the overview provided by Bursztyn and Jensen, *forthcoming*), there are no studies, to the best of our knowledge, investigating the relationship between whistleblowing and public scrutiny. This is an important relationship as the results of our analysis have the potential to inform policy about whether and in what contexts whistleblower anonymity is desirable. Second, 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 involving social costs that are less transparent or salient to the public, and consider whether the effects of financial and non-financial incentives differ across these contexts. Finally, our study sheds light on whether financial rewards may crowd-out employees' non-monetary motivations to blow the whistle. This is a first-order question since the assumption of crowding-out of intrinsic motives<sup>13</sup> is one of the primary reasons why financial rewards are opposed in the international policy debate on whistleblowing.

## **2. 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. In order to recreate in the lab the standard case where the

---

<sup>13</sup> See Fehr et al. (2001), Fehr and Gächter (2001), Fehr and List (2004), Frey (1997), Gneezy and Rustichini (2000a, 2000b) for examples of studies showing evidence of crowding-out of non-monetary motivations. For a survey of the crowding-out literature, see Gneezy, Meier and Rey-Biel (2011) and Bursztyn and Jensen (*forthcoming*).

“society” that is negatively affected by fraud is larger than the firm engaging in it, the public is made of six participants, i.e., double the number of each firm’s members. The experiment consists of four stages, only one of which is randomly chosen for payment at the end of the experimental session. Figure 1 displays the experimental stages.

The purpose of stage one is to create social ties between the members of a firm, generate a shared firm identity and, ideally, create a sense of loyalty between team members. Loyalty to the firm and to one’s manager is indeed an important feature of work within organizations and a crucial potential obstacle to employees’ decision to report wrongdoing (see Waytz et al., 2013). Therefore, when designing the experiment, we thought it was important to generate feelings of shared identity and social cohesion among members of a firm before they participate in the whistleblowing game. We accomplish this by having the three members of each firm engage in a series of team-building tasks with interdependent payoffs to create a sense of “shared fate,” a feature which has been shown to induce a sense of shared identity (Ashforth and Mael, 1989). The first task is the Kandinsky and Klee painting elicitation module first developed in Tajfel et al. (1971) in which subjects view a series of paintings and guess whether it is a Klee or a Kandinsky. Each member of the firm gets credit if at least one member of the firm guesses correctly. The second task consists of a series of addition problems. As before, each member of the firm earns money for each problem that at least one member of the firm solves correctly. The third task involves a series of multiplication problems, each of which involves multiplying two 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 the within firm cohesion ideally resulting from stage one. 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 choose a level of effort between 110 and 170, with their payoffs being determined by the difference between the minimum effort chosen in the group and their own effort multiplied by 0.75. Subjects 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 each firm, one participant is randomly chosen to be the “manager”

and the remaining two participants are assigned the role of “employees”.<sup>14</sup> By having two employees of identical status and a manager we aim to simulate most organizational set-ups where multiple individuals have the same tasks and respond to the same high-ranked supervisor or manager.<sup>15</sup> The employees engage in a real-effort task consisting of adding two-digit numbers, as in task two of stage one 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. There are a total of 12 problems per employee, resulting in maximum private earnings of 24 ECU per employee and maximum firm fund of, also, 24 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 (multiplying two-digit numbers, as in task three of stage one of the experiment) and answering at least eight of the twelve problems correctly. Alternatively, the manager can augment the fund by “breaking 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. Our payoff configuration implies that, as in real organizations, the manager always makes more money than the employees,<sup>16</sup> and his or her performance, either through legal or illegal practices, may add significant value to the firm and, therefore, benefit the employees.

As before, members of the public are only involved in individual decision-making. They have an initial endowment of 14 ECU and, like the employees, they engage in a real effort task consisting of adding two-digit numbers. The task generates 2 ECU for each correct answer. However, the final earnings of each member of the public also depend on the rule-breaking choice of the managers of the firms in the session, since a manager’s wrongdoing causes a loss of 2 ECU to each member of the public. Note that the decision to break the law would be socially efficient only if the manager were not able to augment the fund by successfully completing the multiplication task and if the firm fund were larger than 12 ECU. A high ability manager’s decision to break the law is thus always socially inefficient.<sup>17</sup> In order to keep the manager’s

---

<sup>14</sup> We chose the role allocation to be done randomly rather than based on individual performance in stage one as we wanted to have enough variation in CEO’s decision to break the law. Since such decision is likely to be correlated with CEO’s ability, role allocation by merit would have likely resulted in low frequency of law breaking.

<sup>15</sup> We also aimed to reduce each employee’s competitive feelings and inequality aversion toward the manager. Moreover, having two rather than one employee per firm doubles the number of (potential) whistleblower observations per session.

<sup>16</sup> Note that the manager’s wage equals the private earnings of the employee if the employee is highly productive, i.e. he or she solves all the 12 problems correctly. Even in this case, the manager ends up with higher earnings, since he or she receives half of the firm fund versus the one-fourth received by the employee.

<sup>17</sup> If the employees jointly generate a firm fund of 12 ECU, law-breaking behavior by a low ability manager would generate a firm surplus of 12 ECU while also generating a societal loss of 12 ECU. A high ability manager would always be able to solve the multiplication task correctly, therefore generating the firm surplus without any negative externalities on society.

decision to break the law comparable across firms and independent from efficiency concerns, we do not reveal the size of the firm fund to the manager before eliciting his or her decision to break the law.

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 on a law-breaking manager. Our use of the strategy method allows us to record each employee's willingness to report manager wrongdoing irrespective of whether the manager breaks the law. Had reporting been directly elicited, each employee's whistleblowing decision would have been conditional on the actual occurrence of law-breaking, compromising comparability across employees and resulting in fewer data points.<sup>18</sup> 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. This way, each employee knows that their decision will determine the earnings of a law-breaking manager with a 50 percent chance, no matter the whistleblowing decision of the other employee in the firm. With this design choice, we purposely abstract from the potential presence of collective action problems in the decision to blow the whistle and from the need to control for subject behavior and expectations in such a strategic situation. These aspects have been analyzed in other contexts (see e.g. Bigoni et al. 2012, 2015) and would have increased complexity and noise in the measuring of the effects we are interested in here.

Stage four concludes the remunerated portion of the experiment with a minimum-effort coordination game identical to the game subjects played in stage two. The purpose of this game is to identify the effects of the decisions made in the whistleblowing game – i.e., the manager's law-breaking decision and the employee's reporting decision – on firm cohesion.

After participating in the experiment, subjects fill out a survey. 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.<sup>19</sup> We chose these cases because the visibility of negative externalities varies substantially across the

---

<sup>18</sup> Whether and to what extent, the strategy elicitation affects observed behavior is the subject of an ongoing debate. While the evidence is mixed, a recent survey of the experimental literature by Brandts and Charness (2011) found no cases of treatment effects generated when using the strategy method and not observed when employing the direct-response method.

<sup>19</sup> For information on the Snowden case, see: [https://www.nytimes.com/2014/01/02/opinion/edward-snowden-whistle-blower.html?\\_r=0](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>.

cases as do financial incentives for the whistleblowers involved. Through the survey we also measure 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 *No Reward* condition, whistleblowing employees bear a cost of 5 ECU, while in the *Reward* condition an employee that blows the whistle against his or her manager also receives a financial reward of 10 ECU, i.e., whistleblowing results in a net financial gain for the employee of 5 ECU;
2. *Public Scrutiny vs. No Public Scrutiny*: 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 at all 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. By contrast, in the *No Public Scrutiny* treatments the public is informed of whistleblowing but cannot send messages of any kind to 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. On the other hand, 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 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 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.

In referring to the subject roles, experimental environment and available actions, we used the same contextual labels we used in Section 2.1 when describing the game. We chose to implement a framed experiment because, as recently discussed in Alekseev, Charness and Gneezy (2016), psychological and social factors may play a significant role in individuals' decisions to engage in and report on unlawful behavior and, in such situations, framing may help subjects more fully understand the decision-making context.<sup>20</sup>

The experiment consisted of four stages plus a questionnaire. Subjects were presented with the instructions for each stage on their computer screen immediately before that stage begun. Only one randomly selected stage of the experiment was used for actual payments. Experimental earnings 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 during the experiment did we ask subjects to reveal their names and, although actual names were used during the payment process for accounting purposes, we informed subjects that we would not register their names and therefore 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.

### **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.

---

<sup>20</sup>Framing 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). Alekseev, Charness and Gneezy (2016) provide a recent review of experiments employing either abstract or meaningful frames to present the decision-making setting to the experimental subjects. The general finding is that “evocative language either does not affect behavior or affects it in a desirable way by evoking the desired emotional response”.

**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.

A slightly weaker assumption is that while individuals still care only about monetary incentives, they care about both their own and others' earnings. In these purely *distributional* social preferences models (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002), predictions will vary by model when considering actions or contexts that change any individual's earnings. However, for contextual changes that leave all individuals' earnings unchanged, all of these models predict no change in behavior. Since the earnings consequences of whistleblowing for all parties are the same irrespective of whether negative externalities are visible to the public or whether whistleblowers are subject to public scrutiny, we have a second prediction.

**Prediction 2:** If individuals exhibit purely distributional social preferences, conditional on the presence or absence of rewards, whistleblowing propensity will be the same in the *No Public Scrutiny* treatments as in the *Public Scrutiny* treatments and the same in the *Visible Externalities* as in the *Invisible Externalities* treatments.

The act of whistleblowing itself does have distributional consequences in our experiment—it reduces the most highly remunerated individual's earnings (the manager) while either increasing (*Rewards*) or decreasing (*No Rewards*) the whistleblowing employee's earnings. We are unable to make clear ex ante predictions about how these distributional consequences will affect the prevalence of whistleblowing, however, even in the case of simple inequality aversion (Fehr and Schmidt, 1999) because our payoff parameters allow whistleblowing to either increase or decrease inequality depending on employee performance and the presence of whistleblower rewards.<sup>21</sup>

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

---

<sup>21</sup> To see this, consider the situation where both employees perform as well as possible and suppose distributional social preferences incorporate only the individual employee's and the manager's earnings. Before whistleblowing, the Employee's earnings would be  $(12 \text{ problems}) \times (2 \text{ ECU}) + (1/4) \times (24 \text{ ECU}) = 30 \text{ ECU}$ , while the manager's earnings would be  $24 \text{ ECU} + (1/2) \times (24 \text{ ECU}) = 36 \text{ ECU}$ , so inequality would be 6 ECU in favor of the manager. In the *Rewards* treatment, if the manager breaks the law an employee who blows the whistle would end up with earnings of 35 ECU while the manager would earn  $36 \text{ ECU} - 14 \text{ ECU} = 22 \text{ ECU}$  post-penalty. Consequently, whistleblowing would *increase* inequality from 6 ECU in favor of the manager to 12 ECU in favor of the employee. In the *No Rewards* treatment, whistleblowing puts the employee ahead by 3 ECU (25 ECU vs. 22 ECU) instead of behind by 6 ECU. Since rewards would therefore increase the employee's earnings and inequality conditional on whistleblowing, commonly stipulated distributional preferences such as inequality aversion would tend to reduce the effectiveness of financial rewards.

(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 experimental treatments manipulate factors which we believe will affect how whistleblowing is perceived and use this variation to make predictions. In particular, 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 the widespread moral norm of group loyalty,<sup>22</sup> an anti-social act, leading to social disapproval. In other words, the visibility of the negative externalities to the public is likely to affect whistleblowers' beliefs about 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 third prediction.

**Prediction 3:** If individuals have a preference for social approval or an aversion to social disapproval, i.e., their preferences exhibit social image motivations, allowing for public scrutiny and social judgment will increase whistleblowing in our *Visible Externalities* treatments relative to our *Invisible Externalities* treatments.

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 his or her actions. Intuitively, without financial rewards only pro-social types behave pro-socially, so behavior strongly signals type; however, if pro-social behavior is financially rewarded then some purely money-motivated types may also behavior pro-socially so that the relationship between behavior and type is less clear.

---

<sup>22</sup> 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.



If social image incentives, e.g., a desire to be viewed as a pro-social type, factored heavily into an individual's decision to take a particular action, then adding extrinsic incentives may backfire and “crowd out” that action. In our context, adding financial rewards may affect the public's perception of the motives behind whistleblowing act and, consequently, change the way whistleblowers are judged by the public. Either 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) social image incentives—so we have no prediction overall. However, one might expect the magnitude of the reduction in social image 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 leads to our last prediction.

**Prediction 4:** If individuals are social-image motivated and negative externalities are visible to the public, then 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 (Demel et al., 2016, and Fisman et al., 2016), may be more likely to be whistleblowers 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 subjects, by being more concerned with rule of law and law-breaking (Skitka and Tetlock, 1993; Graham et al., 2012) would be more likely to blow the whistle in order to punish such violations.

### 3. Results

We start by assessing the extent to which we were able to create social ties between members of the same firm in the stage one's tasks that preceded the whistleblowing game. As a measure of the resulting within-firm cohesion, we use the minimum effort chosen by members of a firm in the coordination game in stage two that followed our team building tasks. A comparison of the average minimum effort chosen by members of a firm and the average minimum effort chosen by members of the public, who did not engage in team building tasks,<sup>23</sup> reveals strong evidence of induced firm cohesion. The minimum effort chosen within firms is significantly higher than the minimum effort chosen by members of the public (123.94 vs. 119.21; two-

---

<sup>23</sup> As explained in Section 3, during the team-building stage of the experiment (stage one) members of the public engaged in the same tasks as the members of a firm, but their payoffs were determined solely by their individual performance in such tasks.

sided t-test p-value of 0.0003).<sup>24</sup> This finding suggests that we were successful in generating social cohesion and, possibly, in-group loyalty among members of a firm.

In what follows, we present and discuss the core results of the paper: 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 public's approval or disapproval of whistleblowers under the different treatments (Section 4.2). We conclude by describing managers' law-breaking behavior across treatments (Section 4.3).

#### 4.1.1 The decision to blow the whistle

Overall, about 33% of employees decided to blow the whistle against their law-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 Externalities* and the *Invisible Externalities* treatments simulate different types of illegal actions or different industries where the damages generated by fraud to the public are either 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 generally and substantially increase the prevalence of whistleblowing. This holds both when whistleblowers are subject to public scrutiny and judgment and when they are not. The lone exception, which we return to towards the end of this section, is that financial rewards are ineffective when the externalities caused by fraud are visible to the public and whistleblowers are shielded from social approval or disapproval. Also, in all cases except one the percentage of employees who blow the whistle is substantially and significantly larger than zero. These observations lead to our first result.

**Result 1:** We can reject the notion that employees are purely money maximizers, as whistleblowing is prevalent even when financially costly and it varies substantially with contextual variables having no direct earnings consequences.

Our first result implies that we (and European lawmakers) are justified in our desire to take into account non-pecuniary motivations when setting policies with regard to whistleblowing. The simplest widely used models of such motivations among economists are purely distributional preferences models. Unfortunately, this simple class of models apparently cannot capture important features of whistleblowers' motivations as, contradicting our second prediction, there is clear and substantial variation in behavior across treatments

---

<sup>24</sup> In the game, each member of a three-person group had to choose an effort level in the [110, 170] range, with payoffs being determined by: [minimum effort in the group – 0.75\*(own effort)].

where whistleblowing has identical distributional consequences (e.g., comparing treatment (*No Scrutiny, Reward, Invisible*) to (*Public Scrutiny, Reward, Invisible*)). This leads to our second result.

**Result 2:** We can also reject the idea that whistleblowers have purely distributional social preferences as there is significant variation in whistleblowing propensity across treatments in which whistleblowing yields identical earnings distributions.

Considering non-monetary, non-distributional, motivations, a plausible conjecture is that social image motivations factor heavily into pro-social behaviors like whistleblowing. If individuals care about social approval, then as explained above in Prediction 3, we would expect public scrutiny to have a different effect on whistleblowing depending on the visibility to the public of the costs imposed on them by manager malfeasance. In particular, when negative externalities are visible to the public, the possibility of public scrutiny, through expectations of social approval, should generally increase employees' willingness to blow the whistle while, when negative externalities are not visible to the public, public scrutiny should generally decrease whistleblowing, possibly because whistleblowers expect social disapproval. Behavior is mostly consistent with this pattern, as is discernible from Figure 2 and Table 2 by considering all pairwise comparisons of the form ( --, --, No Scrutiny) vs. (--, --, Public Scrutiny). The lone exception occurs when negative externalities are visible to the public but there are no whistleblower rewards.

**Result 3:** Our data is largely consistent with Prediction 3. When the public is made aware of the costs imposed on them by manager malfeasance, public scrutiny tends to increase whistleblowing while, when these negative externalities are not visible, public scrutiny tends to decrease whistleblowing.

In other words, Result 3 suggests that individuals in our experiment directly value social (dis)approval and expect social approval to be more likely when the public realizes manager malfeasance directly harms them and social *dis*approval to be more likely when whistleblowing is more likely to be interpreted as an anti-social act (disloyalty toward the firm).

Moving beyond simple pairwise comparisons, in Table 3 we report marginal effects from a probit model estimate where the dependent variable is a dummy equal to 1 if the employee is willing to blow the whistle and 0 otherwise. In the first two columns, we split our data by the visibility of negative externalities for clarity, as behavior was substantially different across this dimension. In column 3, we pool our data across all treatments and include interaction terms between the *Reward* and *Public Scrutiny* treatment dummies and the *Visible Externality* dummy. In column 4, we additionally include a set of control variables that include demographics (age and gender), whether the subject is an economics major and the number of firms in the session. In order to proxy for employees' loyalty to the firm, our set of controls also includes the ratio

between firm performance and own performance in stage one of the experiment and the effort level chosen in the minimum effort game of stage two. 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, while the latter variable is a measure of firm cohesion, plausibly capturing 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. This dummy is equal to 1 for 56% of our participants.<sup>25</sup> In the final column, for completeness we include a triple interaction term involving all of our treatment dummies.

The second pattern that becomes more apparent in Table 3 is that rewards have a substantial and statistically significant main effect. In all estimates except for column 2, the marginal effect of financial rewards is to increase the prevalence of whistleblowing by about 30 percentage points when the negative externalities of fraud are not visible to the public. Even when the externalities are visible to the public (column 2), the estimated marginal effect is positive and large in magnitude albeit non-significant (p-value=0.11).

**Result 4:** Financial rewards generally substantially increase whistleblowing.

From Table 3 we can also re-confirm our impression of how the visibility of public harm interacts with public scrutiny to affect whistleblowing. In particular, either by considering *Visible* and *Invisible* treatments separately (columns 1 and 2) or by pooling the data and inspecting the estimated interactions between treatments, we can see that public scrutiny substantially and significantly decreases whistleblowing when the public is unaware of the costs imposed on them by manager malfeasance. When these externalities are clear to the public, on the other hand, public scrutiny increases whistleblowing substantially and, typically, significantly. All together, Table 3 (re-)confirms that our data support Prediction 3.

Next, we consider our fourth prediction, that rewards weaken social image motivations so that, overall, there should be a weaker relationship between rewards and whistleblowing when the act is subject to public scrutiny compared to when it is not. Evidence in support of Prediction 4 would be a negative and significant interaction between the treatment dummies *Public* and *Reward* as this would indicate that public scrutiny reduces the effectiveness of financial rewards. Since this estimated interaction (column 5) is positive,

---

<sup>25</sup> We employ the dummy rather than the continuous political orientation variable to permit comparability with analysis in a later section where we split our sample by political orientation and analyze whistleblowing by right- and left-leaning participants separately.

substantial in magnitude, but non-significant, our data provide little support to this effect. Moreover, the positive and significant coefficient of the triple interaction between *Public*, *Reward* and *Externalities* suggests that, contrary to Prediction 4, crowding out of image motivations does not occur when the externalities are visible to the public either; on the contrary, financial rewards seem to be even more effective when the public is aware of the costs generated by firm fraud, and whistleblowers are subject to public judgment.

**Result 5:** Public scrutiny does not weaken the effect of financial rewards, i.e., we find no evidence for the most common type of “crowding out.”

Interestingly, however, we do find evidence for a different type of crowding out. The negative and (sometimes marginally) significant interaction between *Visible* and *Reward* suggests that, absent public scrutiny, rewards are less effective in industries or cases of fraud where the public feels directly affected by managers’ law-breaking behavior. This pattern is also apparent in Figure 2, when comparing the first two bars in the left panel to the same two bars in the right panel: rewards strongly increase whistleblowing when whistleblowing is not subject to public scrutiny in the invisible externalities case, but have little effect when externalities are visible. Thus, even without true social image concerns (the public does not judge the whistleblowing in these cases), externality visibility alters the effect of financial rewards. Since we did not design our experiment to focus on this type of crowding out, we can only speculate about the underlying mechanism. One possibility is that individuals’ intrinsic motivations associated with whistleblowing are higher when the externalities are visible to the public; in this case, the introduction of financial rewards, absent public scrutiny, crowds out these motivations, resulting in an overall null effect of rewards. Another possibility is that the moral environment is more complex than we have been assuming and that, for example, whistleblowers learn about their own motivations through their actions – they self-signal in the terminology of BT. In this setting, when the whistleblower knows that the public is not aware of the costs imposed on them, blowing the whistle simply expresses a preference for justice or fairness – punishing the manager for bad behavior. When the whistleblower knows the public is aware of the harm imposed on them, motivations become more difficult to disentangle and, in particular, the “choosing sides” aspect, empathizing more with the public than with the in-group (firm), becomes more salient. Abstaining from whistleblowing would then become a self-signal about loyalty to the firm, made stronger in the BT sense by forgoing financial rewards, so that we would expect the patterns observed in the data.

#### **4.1.2 Firm Cohesion and the Interaction between Political Orientation and Public Scrutiny**

To conclude our analysis of whistleblowing, we now assess the effects that manager law-breaking and employee whistleblowing may have on firm cohesion, and we explore the relationship between propensity

to blow the whistle, public scrutiny and political orientation. To pursue the first objective, we assess changes in within-firm behavior in the minimum effort games played before and after the whistleblowing game. The before-after comparison shows a significant decline in the minimum effort observed within firms (123.94 vs. 121.72, with a two-sided t test p-value of 0.0075). In contrast, no significant change is observed among members of the public (119.21 vs. 119.11, with p-value equal to 0.9132). A closer look at the data shows that rule-breaking managers significantly increase their chosen effort in the stage 4 coordination game while whistleblowers tend to reduce their effort but not significantly so. However, the low rate of actual occurrence of whistleblowing<sup>26</sup> makes it impossible to investigate the consequences of whistleblowing on the relationships between members of a firm.<sup>27</sup>

As discussed in Section 2.2, in our post-experiment survey, we also 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 except, of course, the left-leaning dummy.

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 to generally disapprove of whistleblowing when it is unaware of the negative externalities associated with law-breaking behavior, and approve of whistleblowing when such externalities are known.

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 subjects in our student sample. We therefore searched for validation using a different sample of students characterized by predominantly right-leaning political views. We conducted a

---

<sup>26</sup> Recall that whistleblowing and punishment of the manager occur only if the manager breaks the law and if the employee randomly chosen (with a .5 chance) to determine payoffs had stated his or her willingness to blow the whistle. In practice, this occurred only in 3 out of 72 cases/firms.

<sup>27</sup> Note that firm members were not informed about co-workers' *willingness* to blow the whistle. Like in actual real life organizations, information about whistleblowing was available only if whistleblowing actually occurred.

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.

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 public scrutiny is absent 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 sixth result:

**Result 6:** Political orientation significantly impacts the effect of public scrutiny on whistleblowing: right-leaning subjects respond only to monetary incentives, while left-leaning subjects respond also to the possibility of social (dis)approval.

#### **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 managerial 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.

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 blowing the whistle in 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 (national security) and the Tenet (health care) case, less visible in the UBS (tax evasion) case and even less visible in the Enron (earnings management) 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, they 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 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.

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.<sup>28</sup> 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, in column 2 we add our political orientation dummy and in column 3 we employ the same set of controls as in Tables 3, 4 and 6. The estimates confirm that the presence of financial rewards to the whistleblower increases the probability that the public will approve of the whistleblower. Contrary to our expectations, the visibility of the social cost of fraud does not significantly affect the social approval of whistleblowers. We also do not find any systematic and significant relationship between political orientation and judgment of whistleblowers.<sup>29</sup>

**Result 7:** Financial rewards increase the social approval of the whistleblower.

### 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 managers is quite small, with a total of 72 observations. Overall, about 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 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 and that whistleblower rewards may have significant preventive/deterrent effects on corporate crime. 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 analysis<sup>30</sup> provides evidence of the impact of the manager's skills on the probability of breaking the law. In particular, the better the

---

<sup>28</sup> The table reports the UCSB data only. At SMU, a total of 24 members of the public participated in the Public Scrutiny treatment under visible or invisible externalities, always in the presence of financial rewards to the whistleblower. 12.50% did not send a message, 8% sent a message of disapproval, 37.5% sent a neutral message and 42% sent an approval message. The frequency of happy messages is higher under visible externalities (50% versus 33%) but not significantly so.

<sup>29</sup> The null effect of the visibility of the externalities caused by fraud is confirmed in the SMU sample, even though the small number of SMU observations leads us to interpret these messaging results with caution. Another noteworthy finding generated by the SMU sample is the higher likelihood of left-leaning subjects to approve of whistleblowing. About 80% of left-leaning subjects approve of whistleblowers versus 32% of right-leaning students (Fisher exact test p-value equal to 0.075).

<sup>30</sup> The corresponding table is not reported here. It is available from the authors upon request.

manager's performance in the stage one multiplication task, the lower the probability that the manager will decide to cheat to augment the firm fund. This finding seems in line with Baloria et al. (2015) who document that the companies that lobbied against whistleblower rewards provision in the Dodd-Frank Act are precisely those that are less well run and have weaker compliance programs and poorer governance structures (e.g. less separation between Chairman and CEO). These are also the firms for which whistleblower rewards are perceived by the market to be more needed and more likely to have positive effects in terms of improving management/governance and protecting shareholders.

#### **4. 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 being splashed across the covers of popular journals in recent years, the occurrence of whistleblowing is rare and the vast majority of white-collar crime remains undetected and unpunished (Dyck et al., 2013). 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 social image 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, as discussed in the legal debate. 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 while 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 evidence of the crowding out of social image motivations to blow the whistle, as financial rewards are equally 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 act as either an incentive for or a deterrent to blowing the whistle. Social judgment 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 deterrent when the opposite holds. This suggests that, in order to maximize whistleblowing, industries and corresponding cases of fraud should be classified based

on the perceived 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 provide novel insights about 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. Another interesting extension would be to incorporate collective action problems in the decision to blow the whistle. While we believe that whistleblowing on corporate fraud does not typically have the features of a social dilemma, whistleblowing on other, more visible, crimes may be a collective action problem. Indeed, if there is widespread awareness of the law-breaking actions or practices taking place within a firm, individuals' beliefs about fellow employees' reporting decisions may significantly affect willingness to blow the whistle. Finally, an unexpected yet interesting finding of our study is the interaction between political orientation and responsiveness to public scrutiny. In particular, our results suggest that, when deciding whether to report wrongdoing, right-leaning individuals are unaffected by the possibility of social judgment, while left-leaning subjects are highly responsive to it. Future research should assess the robustness of this novel finding to changes in the context and the decision set.

## 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, K. & K. Wu. (2017). Reward Self-Reporting to Deter Corruption: An Experiment on Mitigating Collusive Bribery. *Journal of Economic Behavior and Organization*, 133: 256-272.
- Abbink, K., U. Dasgupta, L. Ghanghadaran & T. Jain. (2014). Letting the Briber Go Free: An Experiment on Mitigating Harassment Bribes. *Journal of Public Economics* 111: 17-28.
- Alekseev, Aleksandr, Charness, Gary, Gneezy, Uri (2016). "Experimental Methods: When and Why Contextual Instructions are Important." *Journal of Economic Behavior and Organization* <http://dx.doi.org/10.1016/j.jebo.2016.12.005>
- 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, J. (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.
- Ashforth, B.E., and F. Mael (1989). "Social identity theory and the organization." *Academy of management review*, 14(1), 20-39.
- 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.
- Bénabou, R. and Tirole, J. (2004). "Willpower and Personal Rules," *Journal of Political Economy* 112(4): 848-886.
- Bénabou, R., and Tirole, J.(2006). "Incentives and prosocial behavior," *The American Economic Review*, 96(5), 1652-1678.
- Bigoni, M., Fridolfsson, S., Le Coq, C. and Spagnolo, G. (2012). "Fines, Leniency and Rewards in Antitrust," *RAND Journal of Economics*, 43(2), 368-390.
- Bigoni, M., Fridolfsson, S., LeCoq, C., and Spagnolo, G. (2015). "Trust, Leniency and Deterrence", *Journal of Law, Economics and Organization* 31(4), 663-689.
- Bolton, G.E., and A. Ockenfels. (2000). "ERC: A theory of equity, reciprocity, and competition." *American Economic Review*, 90(1): 166-193.
- Branas-Garza, P. (2007). "Promoting helping behavior with framing in dictator games," *Journal of Economic Psychology*, Elsevier, vol. 28(4), pages 477-486.
- Brandts, J. and G. Charness, (2011). "The strategy versus the direct-response method: A first survey of experimental comparisons." *Experimental Economics* 2011, 14(3): 375-398.
- Breuer, L. (2013). "Tax Compliance and Whistleblowing: The Role of Incentives," *The Bonn Journal of Economics*, Vol 2, No. 2, pp. 7-44.
- Buckenmaier, J., Dimant E. and L. Mittone (2017). "Experimental Evidence on Tax Evasion, Corruption and Incentives to Blow the Whistle", working paper.
- Burszty, L., and Jensen, R. (*forthcoming*). "Social Image and Economic Behavior in the Field: Identifying, Understanding, and Shaping Social Pressure." *Annual Review of Economics*.
- Carson, T., Verdu, M., and 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.
- Charness, G., and Rabin, M. (2002). "Understanding social preferences with simple tests." *The Quarterly Journal of Economics*, 117(3): 817-869.
- Cookson, R. (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.
- Demel, S., Barr, A., Miller, L. and Ubeda, P.(2016). "Labour market participation, political ideology and distributive preferences" CEDEX working paper.
- Dufwenberg, M., Gächter, S and Hennig-Schmidt, H. (2011). "The framing of games and the psychology of play." *Games and Economic Behavior* 73.2 (2011): 459-478.

- Dworkin, T.M. and Near, J.P. (1997). "A Better Statutory Approach to Whistleblowing," *Business Ethics Quarterly*, 7: 1-16.
- Dyck, A., Morse, A. and Zingales, L. (2010). "Who Blows the Whistle on Corporate Fraud?" *The Journal of Finance*, 65, 2213-53.
- Dyck, A., Morse, A. and Zingales, L. (2013). "How pervasive is corporate fraud?" Rotman School of Management Working Paper 2222608.
- Eckel, C. and Grossman, P.J. (1996). "Altruism in anonymous dictator games." *Games and Economic Behavior* 16: 181.
- Engstrom, D.F. (2012). "Harnessing the Private Attorney General: Evidence from Qui Tam Litigation" *Columbia Law Review*, 112, 1244-1325.
- Engstrom, D.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, D.F. (2014). "Private Enforcement's Pathways: Lessons from Qui Tam Litigation." *Columbia Law Review*, 115, 1913-2006.
- Engstrom, D.F. (2014). "Whither Whistleblowing? Bounty Regimes, Regulatory Context, and the Challenge of Optimal Design." *Theoretical Inquiries L*.15 605
- Engstrom, D.F. (2016). "Bounty Regimes." in *Research Handbook on Corporate Criminal Enforcement and Financial Misdealing* (Jennifer Arlen ed., Edward Elgar Press)
- Ethics Resource Center, 2014. "National Business Ethics Survey of the U.S. Workforce", Available at: <http://www.ethics.org/ecihome/research/nbes>, accessed 31/10/2016.
- Fehr, Ernst, and Klaus M. Schmidt (1999). "A theory of fairness, competition, and cooperation." *The quarterly journal of economics*, 114(3): 817-868.
- 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.
- Fischbacher, Urs (2007). "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10.2 (2007): 171-178.
- Fisman, Raymond, Pamela Jakiela, Shachar Kariv (2016). "Distributional Preferences and Political Behavior", working paper.
- 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.
- Friebel, Guido and Sergei Guriev (2012). "Whistle-Blowing and Incentives in Firms," *Journal of Economics & Management Strategy*, 21(4), 1007-1027.
- 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.
- Graham J, Nosek BA, Haidt J (2012). "The Moral Stereotypes of Liberals and Conservatives: Exaggeration of Differences across the Political Spectrum." *PLoS ONE* 7(12): e50092. doi:10.1371/journal.pone.0050092
- 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.
- Heyes, Anthony and Sadeep Kapur, (2009). "An Economic Model of Whistle-Blower Policy," *Journal of Law, Economics, & Organization*, 25(1), 157-182.
- Hinlopen, Jeroen and Soetevent, Adriaan (2008b). "Laboratory Evidence on the effectiveness of corporate leniency programs", *RAND Journal of Economics*, 39(2), 607-616.
- 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](http://repository.upenn.edu/law_series/4).

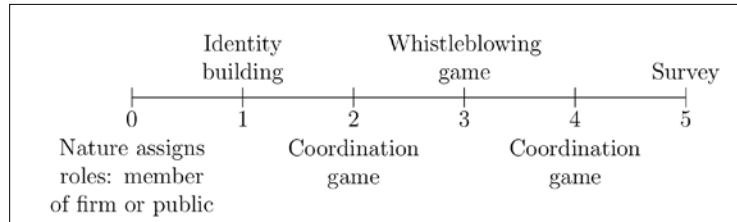
- Linardi, S., & McConnell, M. A., (2011). "No excuses for good behavior: Volunteering and the social environment." *Journal of Public Economics*, 95(5), 445-454.
- Miceli, M. P. and Janet P. Near (1992). "Blowing the Whistle: The Organizational and Legal Implications for Companies and Employees," Lexington Books.
- Miceli, M. P. and Janet P. Near (1994). "Relationship among Value Congruence, Perceived Victimization, and Retaliation against Whistle-blowers," *Journal of Management* 20(4), 773-794., 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.
- Rothschild, J., Miethe, T. (1999). "Whistle-Blower Disclosures and Management Retaliation", *Work and Occupations*, 26(1), 107-128.
- 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, Danila (2012). "Combining top-down and bottom-up accountability: evidence from a bribery experiment." *Journal of Law, Economics, and Organization*, 28(3), 569-587.
- Serra, Danila and Leonard 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.
- Skitka, Linda J. and Philip E. Tetlock (1993). "Providing Public Assistance: Cognitive and Motivational Processes Underlying Liberal and Conservative Policy Preferences," *Journal of Personality and Social Psychology* 65: 1205-23.
- Sunstein, Cass R. (1996). "On the Expressive Function of Law," *University of Pennsylvania Law Review* 144: 2021-2053.
- 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.



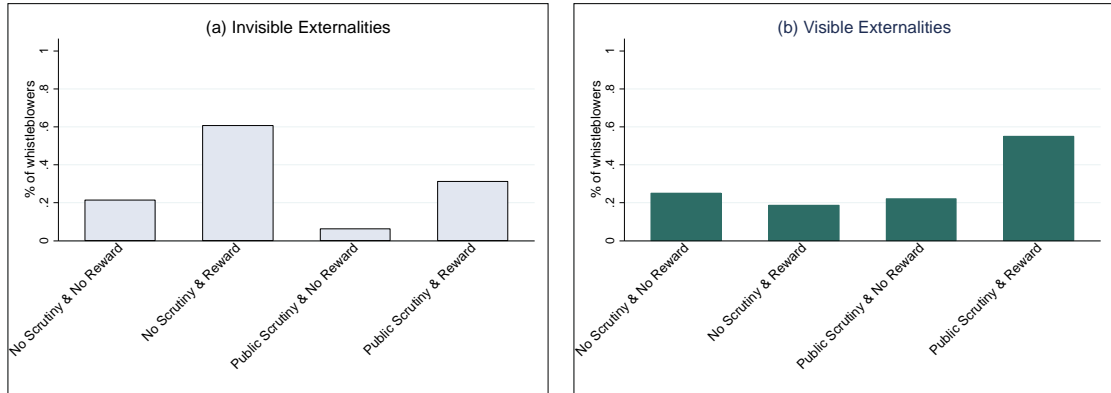
- Titmuss, Richard M. (1970). *The Gift Relationship: From Human Blood to Social Policy* (1970). Reprinted by the New Press, ISBN 1-56584-403-3 (reissued with new chapters 1997, John Ashton & Ann Oakley, LSE Books)
- Waytz, A., Dungan, J., and Young, L. (2013). "The whistleblower's dilemma and the fairness-loyalty 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.

## Figures and Tables

**Figure 1**  
**Stages of the Experiment**

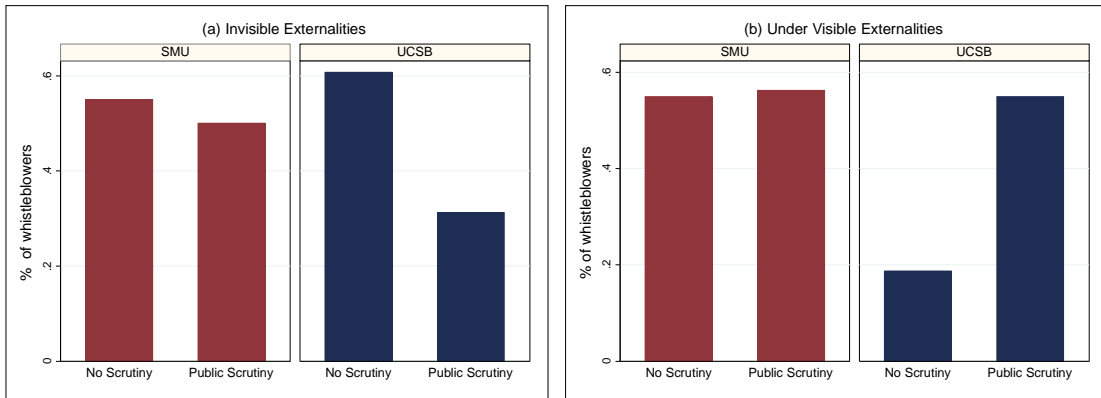


**Figure 2**  
**The effect of rewards and public scrutiny on whistleblowing**

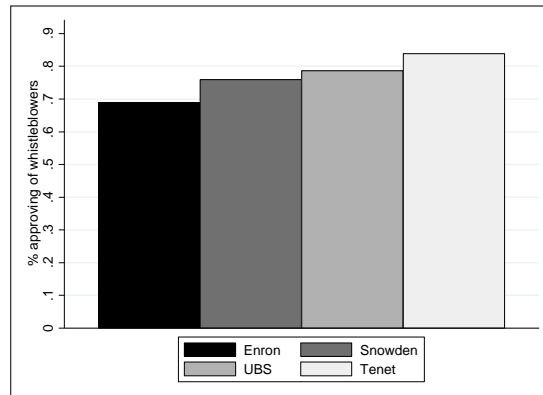


**Figure 3**

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



**Figure 4**  
**Social judgment of four whistleblowing cases (survey)**



**Table 1**  
**Sessions and treatments**

<b>Treatments</b>	<b>Invisible Externalities</b>		<b>Visible Externalities</b>		<b>Total</b>	
	Sessions	Subjects	Sessions	Subjects	Sessions	Subjects
<b>No Rewards &amp; No Public Scrutiny</b>	2	33	2	36	4	69
<b>No Rewards &amp; Public Scrutiny</b>	2	36	2	39	4	75
<b>Rewards &amp; No Public Scrutiny</b>	3	60	2	36	5	96
<b>Rewards &amp; Public Scrutiny</b>	2	36	3	48	5	84
<b>Total</b>	9	165	9	159	18	324

**Table 2**  
**Whistleblowing under different treatments**

	<b>No Rewards &amp; No Public Scrutiny</b>	<b>No Rewards &amp; Public Scrutiny</b>	<b>Rewards &amp; No Public Scrutiny</b>	<b>Rewards &amp; Public Scrutiny</b>
<b>Invisible Externalities</b>	<b>21.43%</b>	<b>6.25%</b>	<b>60.71%</b>	<b>31.25%</b>
Ho: No Scrutiny = Scrutiny	p-value = 0.222 (0.249) if Rewards=0		p-value = 0.060 (0.058) if Rewards=1	
Ho: Rewards = No Rewards	p-value = 0.016 (0.018) if No Scrutiny=1		p-value = 0.070 (0.086) if No Scrutiny=0	
<b>Visible Externalities</b>	<b>25.00%</b>	<b>22.22%</b>	<b>18.75%</b>	<b>55.00%</b>
Ho: No Scrutiny = Scrutiny	p-value = 0.849 (0.583) if Rewards=0		p-value = 0.027 (0.029) if Rewards=1	
Ho: Rewards = No Rewards	p-value = 0.669 (0.550) if No Scrutiny=1		p-value = 0.039 (0.041) if No Scrutiny=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.

**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	0.34*** (0.066)	0.15 (0.091)	0.35*** (0.072)	0.34*** (0.065)	0.29*** (0.084)
Public Scrutiny	-0.26*** (0.066)	0.17* (0.090)	-0.26*** (0.066)	-0.23** (0.092)	-0.29* (0.153)
Visible Externalities			-0.07 (0.128)	-0.02 (0.129)	0.08 (0.082)
Visible x Reward			-0.20** (0.099)	-0.20* (0.101)	-0.34*** (0.066)
Visible x Public Scrutiny			0.46*** (0.112)	0.41*** (0.134)	0.22 (0.185)
Public x Reward					0.11 (0.185)
Visible x Reward x Public					0.39* (0.213)
Controls	No	No	No	Yes	Yes
Observations	74	70	144	144	144

Note: We report marginal effects. 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. In column 5, the number of firms in the session and being an economics major are significant at the 90 percent confident level, with a positive and a negative sign, respectively. Robust standard errors, clustered at the session level, in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.



**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	0.26 (0.208)	0.36*** (0.101)	0.28* (0.153)	0.28** (0.113)
Public Scrutiny	-0.24 (0.243)	-0.93*** (0.033)	-0.22 (0.209)	-0.93*** (0.036)
Visible Externalities	0.09 (0.269)	0.03 (0.057)	0.40 (0.341)	0.09 (0.073)
Visible*Reward	-0.40** (0.196)	-0.27*** (0.066)	-0.48*** (0.124)	-0.23*** (0.073)
Visible*Public	0.41 (0.307)	0.98*** (0.018)	-0.01 (0.422)	0.98*** (0.024)
Public*Reward	0.16 (0.287)	0.93*** (0.003)	0.23 (0.233)	0.98*** (0.003)
Visible*Public*Reward	0.07 (0.445)	-0.28*** (0.050)	0.40 (0.468)	-0.21*** (0.050)
Controls	No	Yes	No	Yes
Observations	63	63	81	81

Note: We report marginal effects. 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. In column 4, the number of firms in the session and being an economics major are significant at the 90 percent confident level, with a positive and a negative sign, respectively. Robust standard errors, clustered at the session level, in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table 5**  
**SMU Sessions and Treatments**

<b>Treatments</b>	<b>Invisible Externalities</b>		<b>Visible Externalities</b>		<b>Total</b>	
	Sessions	Subjects	Sessions	Subjects	Sessions	Subjects
<b>Rewards &amp; No Public Scrutiny</b>	2	42	2	42	4	84
<b>Rewards &amp; Public Scrutiny</b>	2	33	2	36	4	69
<b>Total</b>	4	75	4	78	8	153

**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	-0.05 (0.175)	-0.09 (0.208)	-0.27 (0.290)
Visible Externalities	0.00 (0.159)	-0.14 (0.191)	-0.21 (0.212)
Public x Visible	0.06 (0.239)	0.00 (0.296)	0.17 (0.350)
Left-leaning		-0.29 (0.209)	-0.27 (0.232)
Left x Public Scrutiny		-0.75*** (0.059)	-0.81*** (0.053)
Left x Visible Externalities		0.39 (0.285)	0.51*** (0.188)
Left x Public x Visible		0.73*** (0.053)	0.73*** (0.055)
Controls	No	No	Yes
Observations	70	70	69

Note: The table reports marginal effects. 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. In column 3, the ratio between firm and own performance is statistically significant (p-value<0.01) with a positive sign. Robust standard errors in parentheses; \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

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

	No Rewards	Rewards
<b>Invisible Externalities</b>	58.33%	83.33%
<b>Visible Externalities</b>	33.33%	72.22%
H <sub>0</sub> : Rewards = No Rewards if Visible==0	p-value = 0.178 (0.185)	
H <sub>0</sub> : Rewards = No Rewards if Visible==1	p-value = 0.035 (0.042)	
H <sub>0</sub> : Visible = Invisible Ext. if Rewards=0	p-value = 0.219 (0.207)	
H <sub>0</sub> : 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.

**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		
	UCSB	UCSB	UCSB
Rewards	0.33** (0.131)	0.34** (0.133)	0.42*** (0.154)
Visible Externalities	-0.19 (0.135)	-0.18 (0.136)	-0.11 (0.155)
Left-leaning		-0.06 (0.142)	-0.18 (0.130)
Controls	No	No	Yes
Observations	54	54	54

Note: We report marginal effects. Controls are: age, gender, economics major, number of firms in the session, performance in the real effort task. None of the controls is consistently significant across specifications. Robust standard errors in parentheses: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table 9**  
**Percentages of managers breaking the law**

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

## 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: **2 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.

OK

## 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 5 ECU in surplus, the manager could increase this to 10 ECU. 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.

OK

*Employee*

Remaining Time 0

Please insert your answer into the space provided.

What is  $23 + 73$ ?

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

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.



Happy face     Indifferent face     Sad face     No message

OK

## 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  $S$  ECU in surplus, you could increase this to  $2 * S$ . **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

OK

## 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  $35 \times 20$ ?

OK